JMIR Cancer

Impact Factor (2024): 3.3
Volume 11 (2025) ISSN 2369-1999 Editor in Chief: Naomi Cahill, PhD, RD

Contents

Reviews

Assisted Reproductive Technology and Risk of Childhood Cancer Among the Offspring of Parents With Infertility: Systematic Review and Meta-Analysis (e65820)
Gao Song, Cai-qiong Zhang, Zhong-ping Bai, Rong Li, Meng-qun Cheng. . . . 8

Large Language Model Applications for Health Information Extraction in Oncology: Scoping Review (e65984)
David Chen, Saif Alnassar, Kate Avison, Ryan Huang, Srinivas Raman. . . . 22

The Efficacy of Digital Interventions on Adherence to Oral Systemic Anticancer Therapy Among Patients With Cancer: Systematic Review and Meta-Analysis (e64208)
Wan-Chuen Liao, Fiona Angus, Jane Conley, Li-Chia Chen. . . . 35

Application of Artificial Intelligence in Cardio-Oncology Imaging for Cancer Therapy–Related Cardiovascular Toxicity: Systematic Review (e63964)
Hayat Mushcab, Mohammed Al Ramis, Abdulrahman AlRujaib, Rawan Eskandarani, Tamara Sunbul, Anwar AlOtaibi, Mohammed Obaidan, Reman Al Harbi, Duaa Aljabri. . . . 50

Internet-Based Cognitive Behavioral Therapy Interventions for Caregivers of Patients With Cancer: Scoping Review (e67131)
Chun Shen, Jian Shi, Feng Liu, Xiao Lu. . . . 67

Family Experiences, Needs, and Perceptions in Home-Based Hospice Care for Patients With Terminal Cancer: Meta-Synthesis and Systematic Review (e71596)
Xin Deng, Kanokwan Hounsri, Violeta Lopez, Wilson Tam. . . . 84

Clinical Prediction Models Incorporating Blood Test Trend for Cancer Detection: Systematic Review, Meta-Analysis, and Critical Appraisal (e70275)
Pradeep Virdee, Kiana Collins, Claire Smith, Xin Yang, Sufen Zhu, Nia Roberts, Jason Oke, Clare Bankhead, Rafael Perera, FD Hobbs, Brian Nicholson. . . . 101

Benefits of Remote-Based Mindfulness on Physical Symptom Outcomes in Cancer Survivors: Systematic Review and Meta-Analysis (e54154)
Maria Komariah, Sidik Maulana, Shakira Amirah, Hesti Platini, Laili Rahayuwati, Ah Yusuf, Mohd Firdaus. . . . 119

The Effect of Nutritional Mobile Apps on Populations With Cancer: Systematic Review (e50662)
Krystal Ng, Murallitharan Munisamy, Joanne Lim, Mustafa Alshagga. . . . 133

Investigating the Complexity of Multidimensional Symptom Experiences in Patients With Cancer: Systematic Review of the Network Analysis Approach (e66087)
Vincent Richard, Allison Gilbert, Emanuela Pizzolla, Giovanni Briganti. . . . 152

Viewpoints

Chatbot for the Return of Positive Genetic Screening Results for Hereditary Cancer Syndromes: Prompt Engineering Project (e65848)
Emma Coen, Guilherme Del Fiol, Kimberly Kaphingst, Emerson Borsato, Jackilen Shannon, Hadley Smith, Aaron Masino, Caitlin Allen. . . . 172

Leveraging Artificial Intelligence for Digital Symptom Management in Oncology: The Development of CRCWeb (e68516)
Darren Liu, Yufen Lin, Runze Yan, Zhiyuan Wang, Delgersuren Bold, Xiao Hu. . . . 185

Process Re-Engineering and Data Integration Using Fast Healthcare Interoperability Resources for the Multidisciplinary Treatment of Lung Cancer (e53887)
Ching-Hsiung Lin, Bing-Yen Wang, Sheng-Hao Lin, Pei Shih, Chin-Jing Lee, Yung Huang, Shih Chen, Mei-Lien Pan. . . . 198

Adapting a Self-Guided eHealth Intervention Into a Tailored Therapist-Guided eHealth Intervention for Survivors of Colorectal Cancer (e63486)
Johanne Lyhne, Allan Smith, Tina Carstensen, Lisa Beatty, Adeola Bamgboje-Ayodele, Britt Klein, Lars Jensen, Lisbeth Frostholm. . . . 211

Developing Effective Frameworks for Large Language Model–Based Medical Chatbots: Insights From Radiotherapy Education With ChatGPT (e66633)
James Chow, Kay Li. . . . 225

Leveraging Patient-Reported Outcome Measures for Optimal Dose Selection in Early Phase Cancer Trials (e64611)
Bill Byrom, Anthony Everhart, Paul Cordero, Chris Garratt, Tim Meyer. . . . 1075

Original Papers

Ethical Design of Data-Driven Decision Support Tools for Improving Cancer Care: Embedded Ethics Review of the 4D PICTURE Project (e65566)
Marieke Bak, Laura Hartman, Charlotte Graafland, Ida Korfage, Alena Buyx, Maartje Schermer, 4D PICTURE Consortium. . . . 244

Usability, Acceptability, and Barriers to Implementation of a Collaborative Agenda-Setting Intervention (CASI) to Promote Person-Centered Ovarian Cancer Care: Development Study (e66801)
Rachel Pozzar, James Tulsky, Donna Berry, Jeidy Batista, Paige Barwick, Charlotta Lindvall, Patricia Dykes, Michael Manni, Ursula Matulonis, Nadine McCleary, Alexi Wright. . . . 262

A Digital Home-Based Health Care Center for Remote Monitoring of Side Effects During Breast Cancer Therapy: Prospective, Single-Arm, Monocentric Feasibility Study (e64083)
Hanna Huebner, Lena Wurmthaler, Chloë Goossens, Mathias Ernst, Alexander Mocker, Annika Krückel, Maximilian Kallert, Jürgen Geck, Milena Limpert, Katharina Seitz, Matthias Ruebner, Philipp Kreis, Felix Heindl, Manuel Hörner, Bernhard Volz, Eduard Roth, Carolin Hack, Matthias Beckmann, Sabrina Uhrig, Peter Fasching. . . . 277

Co-Designing a User-Centered Digital Health Tool for Supportive Care Needs of Patients With Brain Tumors and Their Caregivers: Interview Analysis (e53690)
Mahima Kalla, Ashleigh Bradford, Verena Schadewaldt, Kara Burns, Sarah Bray, Sarah Cain, Heidi McAlpine, Rana Dhillon, Wendy Chapman, James Whittle, Katharine J Drummond, Meinir Krishnasamy. . . . 296

Development and Implementation of a Personal Virtual Assistant for Patient Engagement and Communication in Postsurgical Cancer Care: Feasibility Cohort Study (e64145)
Miguel Bargas-Ochoa, Alejandro Zulbaran-Rojas, M Finco, Anthony Costales, Areli Flores-Camargo, Rasha Bara, Manuel Pacheco, Tina Phan, Aleena Khichi, Bijan Najafi. . . . 308

Evaluation of Large Language Models in Tailoring Educational Content for Cancer Survivors and Their Caregivers: Quality Analysis (e67914)
Darren Liu, Xiao Hu, Canhua Xiao, Jinbing Bai, Zahra Barandouzi, Stephanie Lee, Caitlin Webster, La-Urshalar Brock, Lindsay Lee, Delgersuren Bold, Yufen Lin. . . . 323

Identifying Complex Scheduling Patterns Among Patients With Cancer With Transportation and Housing Needs: Feasibility Pilot Study (e57715)
Allan Fong, Christian Boxley, Laura Schubel, Christopher Gallagher, Katarina AuBuchon, Hannah Arem. . . . 336

Patient Voices: Multimethod Study on the Feasibility of Implementing Electronic Patient-Reported Outcome Measures in a Comprehensive Cancer Center (e56625)
Cinzia Brunelli, Sara Alfieri, Emanuela Zito, Marco Spelta, Laura Arba, Linda Lombi, Luana Caselli, Augusto Caraceni, Claudia Borreani, Anna Roli, Rosalba Miceli, Gabriele Tine', Ernesto Zecca, Marco Platania, Giuseppe Procopio, Nicola Nicolai, Luigi Battaglia, Laura Lozza, Morena Shkodra, Giacomo Massa, Daniele Loiacono, Giovanni Apolone. . . . 350

Barriers and Facilitators to the Preadoption of a Computer-Aided Diagnosis Tool for Cervical Cancer: Qualitative Study on Health Care Providers' Perspectives in Western Cameroon (e50124)
Magali Jonnalagedda-Cattin, Alida Moukam Datchoua, Virginie Yakam, Bruno Kenfack, Patrick Petignat, Jean-Philippe Thiran, Klaus Schönenberger, Nicole Schmidt. . . . 366

Association Between Risk Factors and Major Cancers: Explainable Machine Learning Approach (e62833)
Xiayuan Huang, Shushun Ren, Xinyue Mao, Sirui Chen, Elle Chen, Yuqi He, Yun Jiang. . . . 378

AI-Based Identification Method for Cervical Transformation Zone Within Digital Colposcopy: Development and Multicenter Validation Study (e69672)
Tong Wu, Yuting Wang, Xiaoli Cui, Peng Xue, Youlin Qiao. . . . 394

Using ChatGPT to Improve the Presentation of Plain Language Summaries of Cochrane Systematic Reviews About Oncology Interventions: Cross-Sectional Study (e63347)
Jelena Šuto Pavi i , Ana Maruši , Ivan Buljan. . . . 409

Assessing the Quality and Reliability of ChatGPT's Responses to Radiotherapy-Related Patient Queries: Comparative Study With GPT-3.5 and GPT-4 (e63677)
Ana Grilo, Catarina Marques, Maria Corte-Real, Elisabete Carolino, Marco Caetano. . . . 421

Developing and Assessing a Scalable Digital Health Tool for Pretest Genetic Education in Patients With Early-Onset Colorectal Cancer: Mixed Methods Design (e59464)
Jessica Rivera Rivera, Moran Snir, Emilie Simmons, Tara Schmidlen, Misha Sholeh, Melinda Maconi, Carley Geiss, Hayden Fulton, Laura Barton, Brian Gonzalez, Jennifer Permuth, Susan Vadaparampil. . . . 440

Perception and Counseling for Cardiac Health in Breast Cancer Survivors Using the Health Belief Model: Qualitative Analysis (e71062)
Sarah Marrison, Nicholas Shungu, Vanessa Diaz. . . . 459

Usability and Implementation Considerations of Fitbit and App Intervention for Diverse Cancer Survivors: Mixed Methods Study (e60034)
Zakery Dabbagh, Reem Najjar, Ariana Kamberi, Ben Gerber, Aditi Singh, Apurv Soni, Sarah Cutrona, David McManus, Jamie Faro. . . . 470

Exploring Older Adult Cancer Survivors' Digital Information Needs: Qualitative Pilot Study (e59391)
Lorelei Newton, Helen Monkman, Claire Fullerton. . . . 483

Analysis of the Relationship Between Rural-Urban Status and Use of Digital Health Technology Among Older Cancer Survivors Based on the Health Information National Trends Survey: Cross-Sectional Analysis (e66636)
Samantha Werts-Pelter, Zhao Chen, Jennifer Bea, Amanda Sokan, Cynthia Thomson. 489

Exploring the Impact of the Multimodal CAPABLE eHealth Intervention on Health-Related Quality of Life in Patients With Melanoma Undergoing Immune-Checkpoint Inhibition: Prospective Pilot Study (e58938)
Itske Fraterman, Lucia Sacchi, Henk Mallo, Valentina Tibollo, Savannah Glaser, Stephanie Medlock, Ronald Cornet, Matteo Gabetta, Vitali Hisko, Vadzim Khadakou, Ella Barkan, Laura Del Campo, David Glasspool, Alexandra Kogan, Giordano Lanzola, Roy Leizer, Manuel Ottaviano, Mor Peleg, Konrad niatała, Aneta Lisowska, Szymon Wilk, Enea Parimbelli, Silvana Quaglini, Mimma Rizzo, Laura Locati, Annelies Boekhout, Lonneke van de Poll-Franse, Sofie Wilgenhof. 502

A Machine Learning Approach Using Topic Modeling to Identify and Assess Experiences of Patients With Colorectal Cancer: Explorative Study (e58834)
Kelly Voigt, Yingtao Sun, Ayush Patandin, Johanna Hendriks, Richard Goossens, Cornelis Verhoef, Olga Husson, Dirk Grünhagen, Jiwon Jung. 518

Exploring Motives Behind Ideal Melanoma Survivorship Care Plans With Multiple Stakeholders: A Cocreation Study (e55746)
Nadia Kamminga, Marjolein Lugtenberg, Julia Van den Broek, Tamar Nijsten, Marlies Wakkee, Kasia Tabeau. 530

Co-Designing Priority Components of an mHealth Intervention to Enhance Follow-Up Care in Young Adult Survivors of Childhood Cancer and Health Care Providers: Qualitative Descriptive Study (e57834)
Sharon Hou, Brianna Henry, Rachelle Drummond, Caitlin Forbes, Kyle Mendonça, Holly Wright, Iqra Rahamatullah, Perri Tutelman, Hailey Zwicker, Mehak Stokoe, Jenny Duong, Emily Drake, Craig Erker, Michael Taccone, Liam Sutherland, Paul Nathan, Maria Spavor, Karen Goddard, Kathleen Reynolds, Fiona Schulte. 544

Predicting Overall Survival in Patients with Male Breast Cancer: Nomogram Development and External Validation Study (e54625)
Wen-Zhen Tang, Shu-Tian Mo, Yuan-Xi Xie, Tian-Fu Wei, Guo-Lian Chen, Yan-Juan Teng, Kui Jia. 565

Challenges of Cross-Sectoral Video Consultation in Cancer Care on Patients' Perceived Coordination: Randomized Controlled Trial (e60158)
Fereshteh Baygi, Theis Bitz Trabjerg, Lars Jensen, Maria Munch Storsveen, Sonja Wehberg, Jeffrey Sisler, Jens Søndergaard, Dorte Gilså Hansen. 582

Assessing Health Information Seeking Behaviors Among Targeted Social Media Users Using an Infotainment Video About a Cancer Clinical Trial: Population-Based Descriptive Study (e56098)
Jonathan Sommers, Don Dizon, Mark Lewis, Erik Stone, Richard Andreoli, Vida Henderson. 595

Evaluation of Douyin Short Videos on Mammography in China: Quality and Reliability Analysis (e59483)
Hongwu Yang, Chuangying Zhu, Chunyan Zhou, Ruibin Huang, Lipeng Huang, Peifen Chen, Shanshan Zhu, Huanpeng Wang, Chunmin Zhu.
6 0 3

Comparison of Electronic Surveillance With Routine Monitoring for Patients With Lymphoma at High Risk of Relapse: Prospective Randomized Controlled Phase 3 Trial (Sentinel Lymphoma) (e65960)
Katell Le Dû, Adrien Chauchet, Sophie Sadot-Lebouvier, Olivier Fitoussi, Bijou Fontanet, Arnaud Saint-Lezer, Frédéric Maloisel, Cédric Rossi, Sylvain Carras, Anne Parcelier, Magali Balavoine, Anne-Lise Septans. 616

Preliminary Effectiveness of a Telehealth-Delivered Exercise Program in Older Adults Living With and Beyond Cancer: Retrospective Study (e56718)
Emily Dunston, Sonal Oza, Yang Bai, Maria Newton, Leslie Podlog, Kish Larson, Darren Walker, Rebecca Zingg, Pamela Hansen, Adriana Coletta. 632

A Novel Telehealth Exercise Program Designed for Rural Survivors of Cancer With Cancer-Related Fatigue: Single-Arm Feasibility Trial (e59478)
Ryan Marker, Andrew Kittelson, Jared Scorsone, Ian Moran, John Quindry, Heather Leach. 644

Changes in Physical Activity Across Cancer Diagnosis and Treatment Based on Smartphone Step Count Data Linked to a Japanese Claims Database: Retrospective Cohort Study (e58093)
Yoshihide Inayama, Ken Yamaguchi, Kayoko Mizuno, Sachiko Tanaka-Mizuno, Ayami Koike, Nozomi Higashiyama, Mana Taki, Koji Yamanoi, Ryusuke Murakami, Junzo Hamanishi, Satomi Yoshida, Masaki Mandai, Koji Kawakami. 660

An App-Based Intervention With Behavioral Support to Promote Brisk Walking in People Diagnosed With Breast, Prostate, or Colorectal Cancer (APPROACH): Process Evaluation Study (e64747)
Fiona Kennedy, Susan Smith, Rebecca Beeken, Caroline Buck, Sarah Williams, Charlene Martin, Phillippa Lally, Abi Fisher. 668

Development of a Mobile App to Support Head and Neck Cancer Caregiving: Mixed Methods Study (e66471)
Katherine Sterba, Evan Graboyes, Jessica Burris, Megan Scallion, Hannah Kinder, Jama Olsen, Benjamin Toll, Kent Armeson, Terry Day, Bhishamjit Chera, Kenneth Ruggiero. 685

Supporting Medication Adherence in Pediatric Patients Undergoing Hematopoietic Stem Cell Transplant Using the BMT4me mHealth App: Mixed Methods Usability Study (e66847)
Mariam Kochashvili, Parishma Guttoo, Emre Sezgin, Ahna Pai, Rajinder Bajwa, Wendy Landier, Cynthia Gerhardt, Micah Skeens. 700

Digital Health Intervention to Reduce Malnutrition Among Individuals With Gastrointestinal Cancer Receiving Cytoreductive Surgery Combined With Hyperthermic Intraperitoneal Chemotherapy: Feasibility, Acceptability, and Usability Trial (e67108)
Yu Lin, Ryan Hagen, Benjamin Powers, Sean Dineen, Jeanine Milano, Emma Hume, Olivia Sprow, Sophia Diaz-Carraway, Jennifer Permuth, Jeremiah Deneve, Amir Alishahi Tabriz, Kea Turner. 718

Mobile Electronic Patient-Reported Outcomes and Interactive Support During Breast and Prostate Cancer Treatment: Health Economic Evaluation From Two Randomized Controlled Trials (e53539)
Marie-Therése Crafoord, Joakim Ekstrand, Kay Sundberg, Marie Nilsson, Maria Fjell, Ann Langius-Eklöf. 732

Analyzing Geospatial and Socioeconomic Disparities in Breast Cancer Screening Among Populations in the United States: Machine Learning Approach (e59882)
Soheil Hashtarkhani, Yiwang Zhou, Fekede Kumsa, Shelley White-Means, David Schwartz, Arash Shaban-Nejad. 753

Exploring the Social Media Discussion of Breast Cancer Treatment Choices: Quantitative Natural Language Processing Study (e52886)
Daphna Spiegel, Isabel Friesner, William Zhang, Travis Zack, Gianna Yan, Julia Willcox, Nicolas Prionas, Lisa Singer, Catherine Park, Julian Hong. 765

Breast Cancer Screening Participation and Internet Search Activity in a Japanese Population: Decade-Long Time-Series Study (e64020)
Noriaki Takahashi, Mutsuhiro Nakao, Tomio Nakayama, Tsutomu Yamazaki. 775

Assessing the Data Quality Dimensions of Partial and Complete Mastectomy Cohorts in the All of Us Research Program: Cross-Sectional Study (e59298)
Matthew Spotnitz, John Giannini, Yechiam Ostchega, Stephanie Goff, Lakshmi Anandan, Emily Clark, Tamara Litwin, Lew Berman. 787

Examining Demographic, Geographic, and Temporal Patterns of Melanoma Incidence in Texas From 2000 to 2018: Retrospective Study (e67902)
Kehe Zhang, Madison Taylor, Jocelyn Hunyadi, Hung Doan, Adewole Adamson, Paige Miller, Kelly Nelson, Cici Bauer. 806

Assessing Public Interest in Mammography, Computed Tomography Lung Cancer Screening, and Computed Tomography Colonography Screening Examinations Using Internet Search Data: Cross-Sectional Study (e53328)
Zachary Zippi, Isabel Cortopassi, Rolf Grage, Elizabeth Johnson, Matthew McCann, Patricia Mergo, Sushil Sonavane, Justin Stowell, Brent Little. 816

Spatiotemporal Correlation Analysis for the Incidence of Esophageal and Gastric Cancer From 2010 to 2019: Ecological Study (e66655)
Zixuan Cui, Chen Suo, Yidan Zhao, Shuo Wang, Ming Zhao, Ruilin Chen, Linyao Lu, Tiejun Zhang, Xingdong Chen. 826

Analyzing Online Search Trends for Kidney, Prostate, and Bladder Cancers in China: Infodemiology Study Using Baidu Search Data (2011-2023) (e57414)
Shuangquan Lin, Lingxing Duan, Xiangda Xu, Haichao Cao, Xiongbing Lu, Xi Wen, Shanzun Wei. 837

Identifying Adverse Events in Outpatients With Prostate Cancer Using Pharmaceutical Care Records in Community Pharmacies: Application of Named Entity Recognition (e69663)
Yuki Yanagisawa, Satoshi Watabe, Sakura Yokoyama, Kyoko Sayama, Hayato Kizaki, Masami Tsuchiya, Shungo Imai, Mitsuhiro Someya, Ryoo Taniguchi, Shuntaro Yada, Eiji Aramaki, Satoko Hori. 857

Implementation of a Quality Improvement and Clinical Decision Support Tool for Cancer Diagnosis in Primary Care: Process Evaluation (e65461)
Sophie Chima, Barbara Hunter, Javiera Martinez-Gutierrez, Natalie Lumsden, Craig Nelson, Dougie Boyle, Kaleswari Somasundaram, Jo-Anne Manski-Nankervis, Jon Emery. 868

The Role of Online Support, Caregiving, and Gender in Preventative Cancer Genetic Testing Participation: Cross-Sectional Study From a National Study (e67650)
Lavlin Agrawal, Richelle DaSouza, Pavankumar Mulgund, Pankaj Chaudhary. 881

User and Developer Views on Using AI Technologies to Facilitate the Early Detection of Skin Cancers in Primary Care Settings: Qualitative Semistructured Interview Study (e60653)
Owain Jones, Natalia Calanzani, Suzanne Scott, Rubeta Matin, Jon Emery, Fiona Walter. 897

Social Media as a Platform for Cancer Care Decision-Making Among Women: Internet Survey-Based Study on Trust, Engagement, and Preferences (e64724)
Anna Johnson, Grace Longfellow, Clara Lee, Benjamin Ormseth, Gary Skolnick, Mary Politi, Yonaira Rivera, Terence Myckatyn. 916

Examining How Technology Supports Shared Decision-Making in Oncology Consultations: Qualitative Thematic Analysis (e70827)
Alan Yung, Tim Shaw, Judy Kay, Anna Janssen. 936

Treatment Outcomes From Erlotinib and Gefitinib in Advanced Epidermal Growth Factor Receptor–Mutated Nonsquamous Non–Small Cell Lung Cancer in Aotearoa New Zealand From 2010 to 2020: Nationwide Whole-of-Patient-Population Retrospective Cohort Study (e65118)
Phyu Aye, Joanne Barnes, George Laking, Laird Cameron, Malcolm Anderson, Brendan Luey, Stephen Delany, Dean Harris, Blair McLaren, Elliott Brenman, Jayden Wong, Ross Lawrenson, Michael Arendse, Sandar Tin Tin, Mark Elwood, Philip Hope, Mark McKeage. 951

Predicting Early-Onset Colorectal Cancer in Individuals Below Screening Age Using Machine Learning and Real-World Data: Case Control Study (e64506)
Chengkun Sun, Erin Mobley, Michael Quillen, Max Parker, Meghan Daly, Rui Wang, Isabela Visintin, Ziad Awad, Jennifer Fishe, Alexander Parker, Thomas George, Jiang Bian, Jie Xu. 969

Large Language Model Approach for Zero-Shot Information Extraction and Clustering of Japanese Radiology Reports: Algorithm Development and Validation (e57275)
Yosuke Yamagishi, Yuta Nakamura, Shouhei Hanaoka, Osamu Abe. 984

Leveraging Digital Twins for Stratification of Patients with Breast Cancer and Treatment Optimization in Geriatric Oncology: Multivariate Clustering Analysis (e64000)
Pierre Heudel, Mashal Ahmed, Felix Renard, Arnaud Attye. 995

A Deep Learning–Enabled Workflow to Estimate Real-World Progression-Free Survival in Patients With Metastatic Breast Cancer: Study Using Deidentified Electronic Health Records (e64697)
Gowtham Varma, Rohit Yenukoti, Praveen Kumar M, Bandlamudi Ashrit, K Purushotham, C Subash, Sunil Ravi, Verghese Kurien, Avinash Aman, Mithun Manoharan, Shashank Jaiswal, Akash Anand, Rakesh Barve, Viswanathan Thiagarajan, Patrick Lenehan, Scott Soefje, Venky Soundararajan. 1010

Interpretable Machine Learning to Predict the Malignancy Risk of Follicular Thyroid Neoplasms in Extremely Unbalanced Data: Retrospective Cohort Study and Literature Review (e66269)
Rui Shan, Xin Li, Jing Chen, Zheng Chen, Yuan-Jia Cheng, Bo Han, Run-Ze Hu, Jiu-Ping Huang, Gui-Lan Kong, Hui Liu, Fang Mei, Shi-Bing Song, Bang-Kai Sun, Hui Tian, Yang Wang, Wu-Cai Xiao, Xiang-Yun Yao, Jing-Ming Ye, Bo Yu, Chun-Hui Yuan, Fan Zhang, Zheng Liu. 1027

Next-Generation Sequencing–Based Testing Among Patients With Advanced or Metastatic Nonsquamous Non–Small Cell Lung Cancer in the United States: Predictive Modeling Using Machine Learning Methods (e64399)
Alan Brnabic, Ilya Lipkovich, Zbigniew Kadziola, Dan He, Peter Krein, Lisa Hess. 1044

Evaluating an AI Chatbot “Prostate Cancer Info” for Providing Quality Prostate Cancer Screening Information: Cross-Sectional Study (e72522)
Otis Owens, Michael Leonard. 1094

Research Letter

Design and Use of Patient-Facing Electronic Patient-Reported Outcomes and Sensor Data Visualizations During Outpatient Chemotherapy (e62711)
Christianna Bartel, Leeann Chen, Weiyu Huang, Qichang Li, Qingyang Li, Jennifer Fedor, Krina Durica, Carissa Low. 1083

Short Paper

Monthly Variations in Colorectal Cancer Screening Tests Among Federally Qualified Health Center Patients in Missouri: Quality Improvement Project (e64809)
Jane McElroy, Jamie Smith, Kevin Everett. 1087

Corrigenda and Addenda

Correction: Benefits of Remote-Based Mindfulness on Physical Symptom Outcomes in Cancer Survivors: Systematic Review and Meta-Analysis (e71958)
Maria Komariah, Sidik Maulana, Shakira Amirah, Hesti Platini, Laili Rahayuwati, Ah Yusuf, Mohd Firdaus. 1106

Assisted Reproductive Technology and Risk of Childhood Cancer Among the Offspring of Parents With Infertility: Systematic Review and Meta-Analysis

Gao Song[1*], MM; Cai-qiong Zhang[1*], MM; Zhong-ping Bai[1*], MM; Rong Li[1*], MM; Meng-qun Cheng[2*], MM

[1]Department of Pharmacy, Puer People's Hospital, Puer, China

[2]Department of Reproductive Medicine, Puer People's Hospital, 44 Zhenxing Avenue, Puer, Yunnan, China

[*]all authors contributed equally

Corresponding Author:
Meng-qun Cheng, MM
Department of Reproductive Medicine, Puer People's Hospital, 44 Zhenxing Avenue, Puer, Yunnan, China

Abstract

Background: The relationship between assisted reproductive technology (ART) and childhood cancer risk has been widely debated. Previous meta-analyses did not adequately account for the impact of infertility, and this study addresses this gap.

Objective: Our primary objective was to assess the relative risk (RR) of childhood cancer in infertile populations using ART versus non-ART offspring, with a secondary focus on comparing frozen embryo transfer (FET) and fresh embryo transfer (fresh-ET).

Methods: A literature review was conducted through PubMed, Embase, Cochrane, and Web of Science, with a cutoff date of July 10, 2024. The study was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY 202470119). Inclusion criteria were based on the PICOS (Population, Intervention, Comparison, Outcomes, and Study Design) framework: infertile or subfertile couples (population), ART interventions (in vitro fertilization [IVF], intracytoplasmic sperm injection [ICSI], FET, and fresh-ET), non-ART comparison, and childhood cancer risk outcomes. Data abstraction focused on the primary exposures (ART vs non-ART and FET vs fresh-ET) and outcomes (childhood cancer risk). The risk of bias was assessed using the Newcastle-Ottawa Quality Assessment Scale, and the evidence quality was evaluated with the Grading of Recommendations Assessment, Development, and Evaluation (GRADE). Pooled estimates and 95% CIs were calculated using random effects models.

Results: A total of 18 studies were included, published between 2000 and 2024, consisting of 14 (78%) cohort studies and 4 (22%) case-control studies, all of which were of moderate to high quality. The cohort studies had follow-up periods ranging from 3 to 18 years. Compared with non-ART conception, ART conception was not significantly associated with an increased risk of childhood overall cancer (RR 0.95, 95% CI 0.71 - 1.27; GRADE quality: low to moderate). Subgroup analyses of IVF (RR 0.86, 95% CI 0.59 - 1.25), ICSI (RR 0.76, 95% CI 0.26 - 2.2), FET (RR 0.98, 95% CI 0.54 - 1.76), and fresh-ET (RR 0.75, 95% CI 0.49 - 1.15) showed similar findings. No significant differences were found for specific childhood cancers, including leukemia (RR 0.99, 95% CI 0.79 - 1.24), lymphoma (RR 1.22, 95% CI 0.64 - 2.34), brain cancer (RR 1.22, 95% CI 0.73 - 2.05), embryonal tumors (RR 1, 95% CI 0.63 - 1.58), retinoblastoma (RR 1.3, 95% CI 0.73 - 2.31), and neuroblastoma (RR 1.02, 95% CI 0.48 - 2.16). Additionally, no significant difference was observed in a head-to-head comparison of FET versus fresh-ET (RR 0.99, 95% CI 0.86 - 1.14; GRADE quality: moderate).

Conclusions: In conclusion, this study found no significant difference in the risk of childhood cancer between offspring conceived through ART and those conceived through non-ART treatments (such as fertility drugs or intrauterine insemination) in infertile populations. While infertility treatments may elevate baseline risks, our findings suggest that whether individuals with infertility conceive using ART or non-ART methods, their offspring do not face a significantly higher risk of childhood cancer. Further research, especially comparing infertile populations who conceive naturally, is needed to better understand potential long-term health outcomes.

Trial Registration: INPLASY 202470119; https://inplasy.com/?s=202470119

(JMIR Cancer 2025;11:e65820) doi:10.2196/65820

KEYWORDS

assisted reproductive technology; childhood cancer; infertility; subfertile; risks; systematic review

Introduction

Over the last century, global fertility rates have significantly declined, and it is projected that by 2060, fertility will fall below replacement levels [1,2]. This trend is closely linked to an increase in infertility, which can be caused by factors such as ovulation disorders, tubal abnormalities, uterine issues, and sperm abnormalities [3]. Assisted reproductive technology (ART) has helped many infertile couples achieve parenthood. Since ART's introduction in 1978, over 10 million children have been born using this technology [4], with approximately 1 million children conceived via ART each year. As ART usage increases, monitoring the long-term health risks associated with it, particularly childhood cancer, becomes crucial [5].

The relationship between ART and childhood cancer has been widely studied, but the results remain controversial due to inconsistent findings [6,7]. One of the key reasons for this inconsistency is the use of different reference groups. Few studies distinguish between children born to parents with infertility and those born to parents who conceived naturally [5,8]. It is essential to differentiate the effects of parental infertility from those of ART treatment, particularly given the challenge of small sample sizes in many studies. Furthermore, most studies are conducted within a single health care system or region, which limits their ability to fully assess cancer risk in offspring conceived through ART.

Previous reviews and meta-analyses have not adequately addressed infertility as a factor, possibly due to the limited availability of relevant studies [9-14]. However, recent large national cohort studies have compared offspring of parents with infertility with controls, and follow-up periods have extended beyond 10 years [15-22]. Given these advances, we conducted a systematic review and meta-analysis to assess the relative risk (RR) of childhood cancer in ART versus non-ART offspring in infertile populations and to compare frozen embryo transfer (FET) with fresh embryo transfer (fresh-ET). This study provides new insights into the relationship between ART modalities and pediatric cancer risk, which could help guide clinical ART fertility treatments.

Methods

Overview

This study was retrospectively registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY 202470119). The systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and included all published articles on ART exposure and childhood cancer risk in the offspring of parents with infertility [23].

Search Strategy and Inclusion Criteria

We conducted a systematic literature search with a deadline of July 10, 2024, using PubMed, Embase, Cochrane, and Web of Science. The electronic search strategy was initially developed by the author (GS) and subsequently reviewed by the author with extensive search experience (MQC). We first tested the search by adapting it for each database and validating it against previously published meta-analyses on relevant topics to ensure the comprehensiveness of our approach. The validated search strategy was implemented simultaneously across each database on July 10, 2024, using the search terms "ART," "children," "cancer," and "risk." The detailed Boolean expressions of the search strategy are presented in Multimedia Appendix 1.

The inclusion criteria were constructed using the PICOS (Population, Intervention, Comparison, Outcomes, and Study Design) framework:

- Population: infertile or subfertile couples.
- Intervention: ART, including in vitro fertilization (IVF), intracytoplasmic sperm injection (ICSI), FET, and fresh-ET.
- Comparison: non-ART, defined as infertile or subfertile couples who did not conceive through ART but may have conceived naturally or with induced ovulation induction (OI) or intrauterine insemination (IUI).
- Outcomes: risk of childhood cancer, including overall childhood cancers and specific types such as leukemia, lymphomas, brain cancer, embryonal tumors, retinoblastoma, and neuroblastoma.
- Study design: randomized controlled trials and observational studies (eg, cohort or case-control studies).

Studies lacking sufficient data to calculate RR estimates and their 95% CIs were excluded. Additionally, conference abstracts, reviews, non-English articles, duplicate data, and non–peer-reviewed publications were excluded.

Study Selection

On July 10, 2024, 2 researchers (CQZ and RL) conducted literature searches, reviewed the results, and imported them into Endnote X8 (Clarivate Analytics). CQZ was responsible for deduplication and the initial screening of studies, while RL reviewed CQZ's selections. Both researchers then independently performed further screening based on the predefined inclusion criteria. Any disagreements were resolved through discussion between CQZ and RL. If a consensus could not be reached, a third researcher, GS, was consulted.

Data Extraction

Data extraction was carried out by CQZ using a prespecified and tested form in Microsoft Excel. RL then reviewed the extracted data for accuracy. The information extracted included the first author, year of publication, age at follow-up, study design, study timeframe, country, data source, duration of follow-up, type of cancer reported, ART type (IVF, ICSI, FET, or fresh-ET), and case-control or exposure-nonexposure data. If any data were missing, the authors were contacted to obtain the necessary information.

Quality Assessment and Risk of Bias

The Newcastle-Ottawa Quality Assessment Scale (NOS) was used for the quality assessment of the included studies [24]. Two authors (CQZ and RL) independently conducted the NOS evaluation, and any disagreements were resolved through discussions with the corresponding author or GS. Studies were categorized into low (total score ≥7), moderate (total score 5 - 6), and high (total score ≤4) risk of bias.

Publication bias for the primary outcomes, such as ART versus non-ART and FET versus fresh-ET, was assessed using funnel plots and the Egger test. If the points on the funnel plot were symmetrically distributed, it indicated no or low bias; asymmetry suggested the presence of publication bias. The Egger test was performed to quantitatively assess publication bias, with a $P<.05$ indicating significant bias. Sensitivity analyses were conducted for the primary outcomes.

Data Analysis and Synthesis

The RR and 95% CIs were chosen to assess the association between ART and childhood cancer in infertile offspring. Outcomes were combined using the DerSimonian and Laird random effects model [25]. All analyses were visualized using Stata 17 statistical software, and in meta-analyses, $P<.05$. Heterogeneity was analyzed using the I^2 statistic. A high degree of heterogeneity was indicated if the I^2 value was greater than 50%. Subgroup analyses were conducted based on the following four criteria: (1) continents, (2) duration of follow-up, (3) reported cancer type, and (4) operational versus nonoperational. Subgroup differences were assessed using the Q test, and statistical significance was defined as a $P<.05$. Regardless of the level of heterogeneity, a random-effects model was consistently applied to ensure the robustness of the analysis across different study designs and populations. Sensitivity analyses were performed by excluding each study individually.

Quality of Evidence

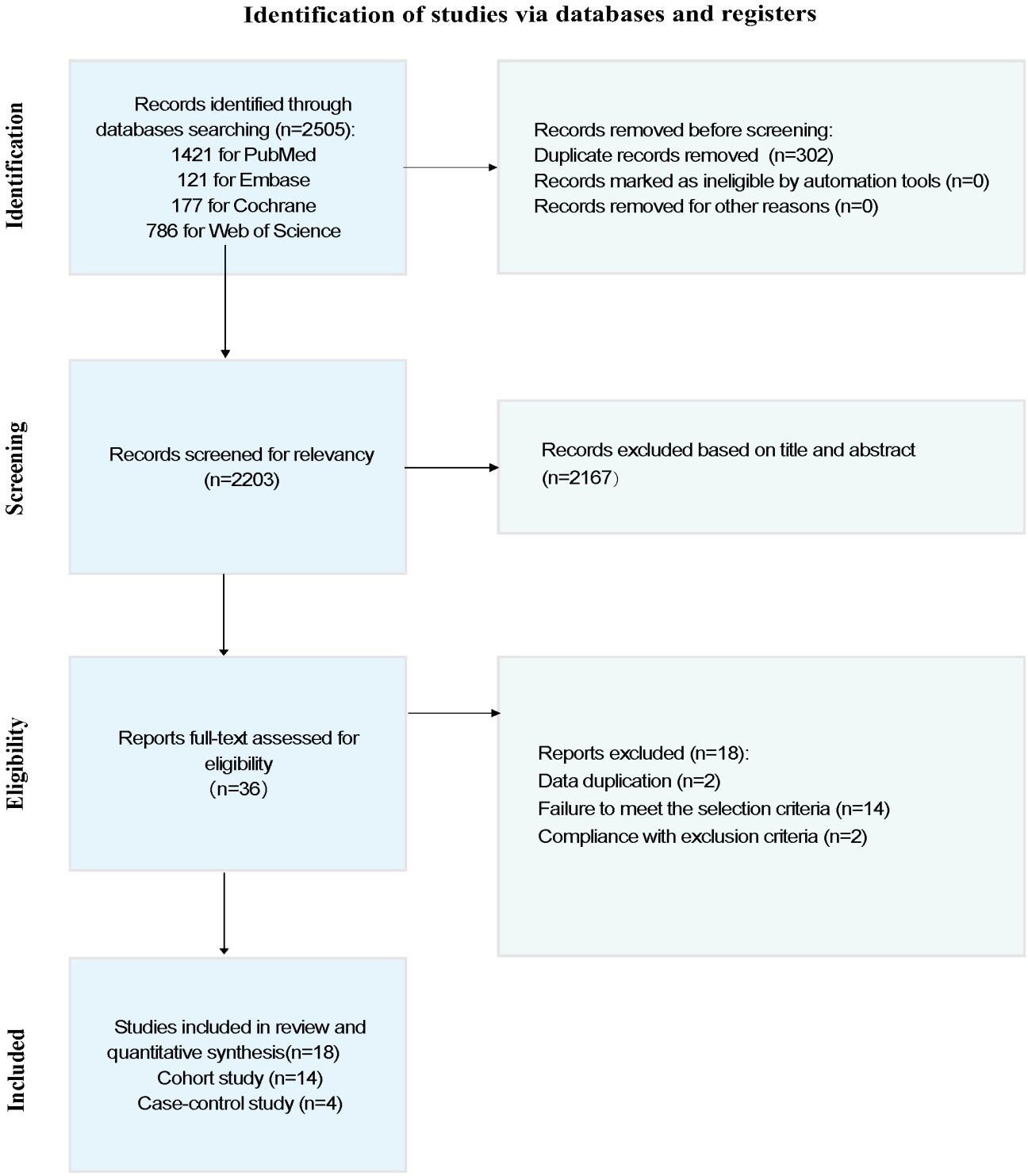
GRADE (Grading of Recommendations Assessment, Development, and Evaluation) is a systematic approach for evaluating the quality of evidence by assessing 5 domains: methodological limitations (eg, risk of bias), heterogeneity of results (eg, inconsistency), generalizability of findings (eg, indirectness), precision of estimates, and risk of publication bias [26]. The overall certainty of the evidence is categorized into 4 levels, ranging from high to very low.

Results

Search Results and Study Characteristics

A total of 2505 articles were obtained from the systematic search, of which 302 (12.06%) were duplicates. We screened the titles and abstracts to exclude 2167 (86.51%) articles that did not meet the eligibility criteria and subsequently removed them. The full manuscripts of the 36 articles were screened to exclude 18 (50%) articles that did not meet the eligibility criteria. These included different papers by the same authors with duplicate data. Data that did not involve subfertile offspring were excluded. A total of 18 studies [15-22,27-36] were thus included in this review. The NOS quality of the included studies was either moderate or high (Table S1 in Multimedia Appendix 2) [15-22,27-36]. The PRISMA flowchart depicts the article screening process (Figure 1).

Figure 1. Flowchart illustrating the article selection process.



Of the 18 included studies, 14 (77%) were cohort studies [15-22,27-32] and 4 (22%) were case-control studies [33-36]. All cohort studies reported overall cancer occurrence risk, while the 4 case-control studies focused only on specific types of cancer, including retinoblastoma, leukemia, and neuroblastoma. All studies were published in English and covered multiple countries and regions, including Australia, Israel, Denmark, the United States, France, Finland, Sweden, the Netherlands, Taiwan, the United Kingdom, and Norway. The studies were published across nearly 2 decades, with the earliest study published in 2000 [27] and the most recent study published in 2024 [16]. Most cohort studies had follow-up durations ranging from 3 to 18 years, with the shortest follow-up period being 3 years [30] and the longest extending to 18 years [17].

Of the 18 studies, 10 (56%) [15-20,33-36] compared ART with non-ART and 11 (61%) [16,17,20-22,27-32] compared FET with fresh-ET. Of the 10 studies comparing ART with non-ART, 6 (60%) were cohort studies [15-20] involving 480,852 ART patients and 716,144 non-ART patients. Four (4/10, 40%) [33-36] were case-control studies involving 563 ART patients and 1521 non-ART patients. Of the 18 studies, 11 (61%)

comparative studies of FET versus fresh-ET were cohort studies involving 176,800 FET patients and 723,327 fresh-ET patients.

Comparison of Childhood Overall Cancer Risk by ART Conception and Non-ART Conception

Of the 18 studies, 6 (33%) studies have compared the risk of childhood overall cancer in offspring of ART versus non-ART conceptions (Table S2 in Multimedia Appendix 2) [15-20]. The results showed that there was no significant increase in the risk of childhood overall cancer in ART-conceived offspring compared with non-ART conception (RR 0.95, 95% CI 0.71 - 1.27; I^2=82%) (Figure 2). A high degree of heterogeneity was observed. We performed subgroup analyses based on continent, follow-up duration, reported cancer types, and whether artificial insemination procedures were involved. No significant differences were observed within the subgroups (Table 1). However, when the non-ART control group was defined as nonoperational (ie, using only OI or fertility drugs), the RR for childhood overall cancer in the ART group was 1.23 (95% CI 0.98 - 1.54). Based on the GRADE evidence quality assessment, the quality of the comparison between ART and non-ART was rated as "low to moderate" due to serious risk of bias and inconsistency (Table S3 in Multimedia Appendix 2). The Egger test did not detect significant publication bias (*P*=.66), and the adjusted RR was 0.812 (95% CI 0.549 - 1.074), indicating robust results (Figure S1 in Multimedia Appendix 3).

Figure 2. Comparison of childhood overall cancer risk by ART conception and non-ART conception [15-20]. ART: assisted reproductive technology; RR: relative risk.

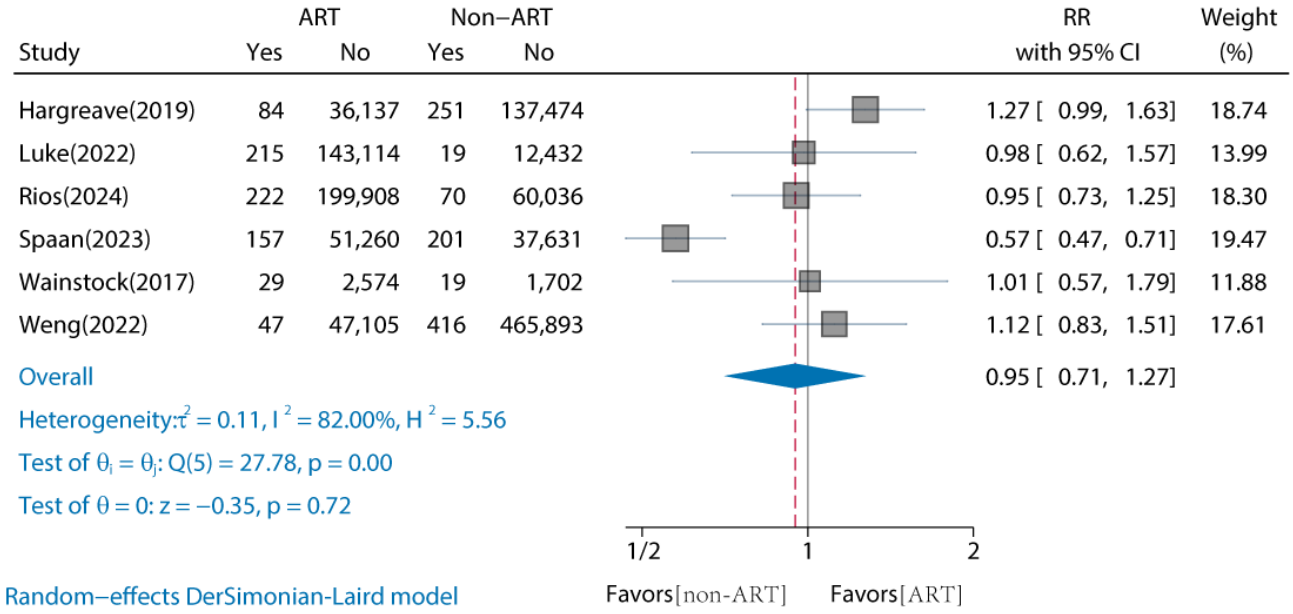


Table . Comparison of childhood overall cancer risk by ART[a] conception and non-ART conception by subgroup analysis.

		Studies, n	ART, n	Non-ART, n	RR[b] (95% CI)	I^2 (%)	P Heterogeneity	P Between groups
Overall		6	480,852	716,144	0.95 (0.71 - 1.27)	82	<.001	—[c]
Continents								.73
	Asian	2	49,755	468,030	1.09 (0.84 - 1.43)	0	.76	—
	Europe	3	287,768	235,663	0.88 (0.54 - 1.43)	91.83	<.001	—
	North America	1	143,114	12,451	0.98 (0.62 - 1.57)	—	—	—
Duration of follow-up (years)								.69
	≤10	3	390,611	538,866	1.02 (0.84 - 1.22)	0	.73	—
	>10	3	90,241	177,278	0.89 (0.49 - 1.62)	91.61	<.001	—
Reported cancers type								.84
	Neoplasm	1	2603	1721	1.01 (0.57 - 1.79)	—	—	—
	Overall cancer	5	478,249	714,423	0.94 (0.68 - 1.31)	85.5	<.001	—
Operational versus nonoperational[d]								.10
	Non-ART (nonoperational)	2	38,711	139,176	1.23 (0.98 - 1.54)	0	.47	—
	Non-ART (operational)	4	441,387	575,992	0.87 (0.61 - 1.22)	82.08	<.001	—

[a]ART: assisted reproductive technology.

[b]RR: relative risk.

[c]Not applicable.

[d]We set non-ARTs that only use fertility drugs or ovulation induction as a nonoperational factor, and those that involve artificial insemination or intrauterine insemination operations as an operational factor.

In addition, we compared IVF, ICSI, FET, and fresh-ET conceptions with non-ART conceptions separately. The results showed no significant differences between either (P≥.05). The corresponding RRs were for IVF (RR 0.86, 95% CI 0.59 - 1.25; I^2=70.18%), ICSI (RR 0.76, 95% CI 0.26 - 2.2; I^2=94.61%), FET (RR 0.98, 95% CI 0.54 - 1.76; I^2=83.18%), and fresh-ET (RR 0.75, 95% CI 0.49 - 1.15; I^2=81.85%) (Figures S1-S4 in Multimedia Appendix 4) [15-18,20].

Comparison of Childhood Overall Cancer Risk by FET Conception and Fresh-ET Conception

Of the 18 studies, 11 (61%) cohort studies compared the risk of childhood overall cancer in FET versus fresh-ET conceived offspring (Table S4 in Multimedia Appendix 2) [16,17,20-22,27-32]. The results showed no significant increase in the risk of childhood overall cancer for FET-conceived offspring compared to fresh-ET (RR 0.99, 95% CI 0.86 - 1.14; Figure 3). The interstudy heterogeneity was low (I^2=24.45%). Subgroup analyses by continent, follow-up duration, and cancer type revealed no significant differences (Table 2). Funnel plots and the Egger test indicated no publication bias (t=0.53, P=.61; adjusted RR 0.98, 95% CI 0.856 - 1.125; Figure S2 in Multimedia Appendix 3). Based on the GRADE assessment, the quality of the comparison between FET and fresh-ET was rated as “moderate” due to a serious risk of bias (Table S3 in Multimedia Appendix 2).

Figure 3. Comparison of childhood overall cancer risk by FET conception and fresh-ET conception [16,17,20-22,27-32]. FET: frozen embryo transfer; fresh-ET: fresh embryo transfer; RR: relative risk.

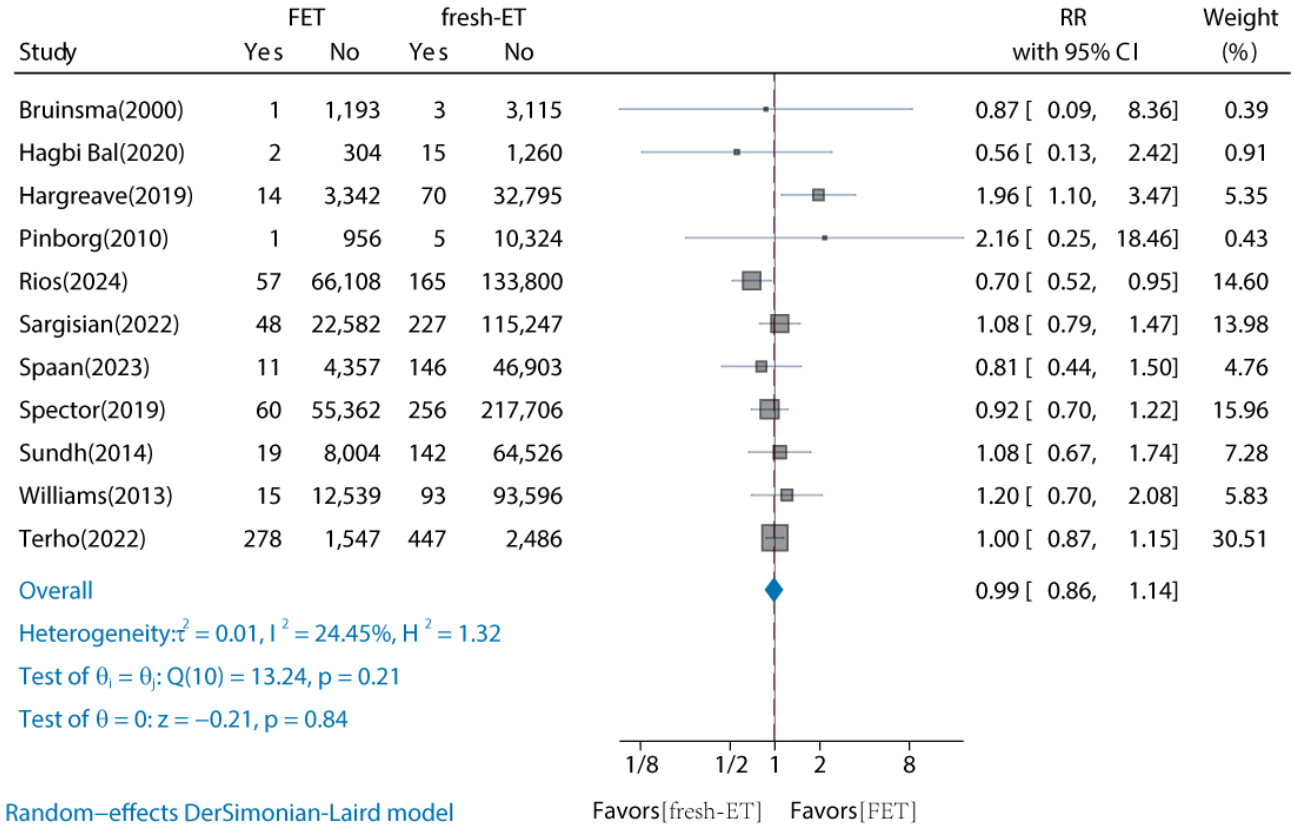


Table . Comparison of childhood overall cancer risk by FET[a] conception and fresh-ET[b] conception by subgroup analysis.

		Studies, n	RR[c] (95% CI)	I^2 (%)	P Heterogeneity	P Between groups
Overall		11	0.99 (0.86 - 1.14)	24.45	.21	—[d]
Continents						.21
	Asian	1	0.56 (0.13 - 2.42)	—	—	—
	Europe	8	1.02 (0.84 - 1.23)	43.58	.09	—
	North America	1	0.92 (0.7 - 1.22)	—	—	—
	Oceania	1	0.87 (0.09 - 8.36)	—	—	—
Duration of follow-up (years)						.37
	≤10	8	0.92 (0.79 - 1.08)	0	.47	—
	>10	3	1.13 (0.75 - 1.72)	64.27	.06	—
Reported cancers type						.94
	Neoplasm	2	0.99 (0.87 - 1.14)	0	.44	—
	Overall cancer	9	1 (0.82 - 1.23)	36.27	.13	—

[a]FET: frozen embryo transfer.

[b]fresh-ET: fresh embryo transfer.

[c]RR: relative risk.

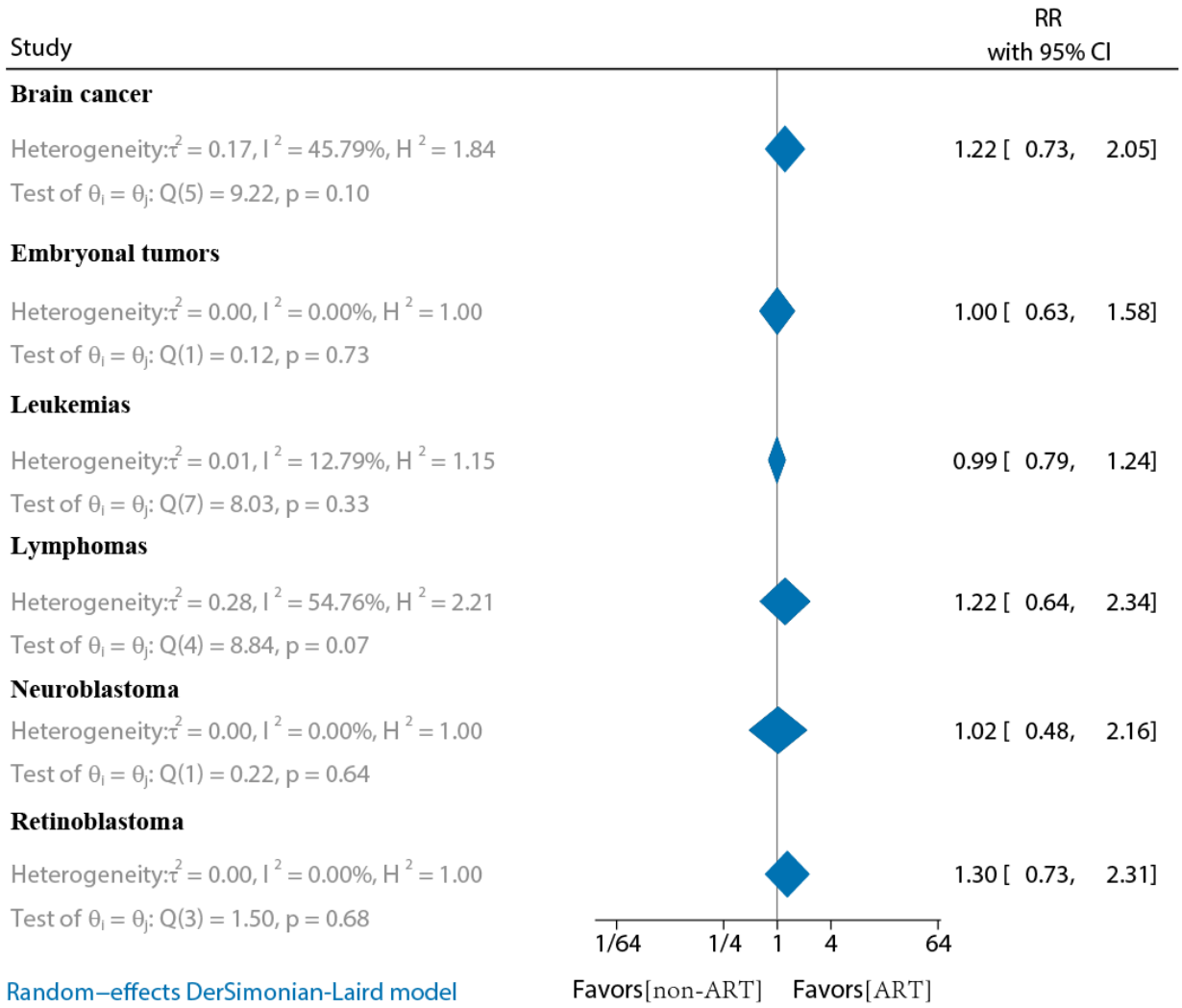
[d]Not applicable.

Comparison of Childhood-Specific Cancer Risk by ART Conception and Non-ART Conception

In total, 10 studies compared the risk of childhood-specific cancer in the offspring of ART versus non-ART conceptions (Tables S5 and S6 in Multimedia Appendix 2) [15-20,33-36]. The main studies included 6 cohort studies and 4 case-control studies. The results showed that none of the ART-conceived offspring had a significantly increased risk of childhood-specific cancer compared to non-ART conception (P≥.05). The main ones included leukemia (RR 0.99, 95% CI 0.79 - 1.24;

I^2=12.79%), lymphoma (RR 1.22, 95% CI 0.64 - 2.34; I^2=54.76%), brain cancer (RR 1.22, 95% CI 0.73 - 2.05; I^2=45.79%), embryonal tumors (RR 1, 95% CI 0.63 - 1.58; I^2=0%), retinoblastoma (RR 1.3, 95% CI 0.73 - 2.31; I^2=0%), and neuroblastoma (RR 1.02, 95% CI 0.48 - 2.16; I^2=0%) (Figure 4).

Figure 4. Comparison of childhood-specific cancer risk by ART conception and non-ART conception. ART: assisted reproductive technology; RR: relative risk.



Sensitivity Analysis

To explore the stability of the meta-analysis results, sensitivity analyses were performed by excluding each study individually. The results demonstrated the robustness of findings for both ART versus non-ART conception and FET versus fresh-ET conception regarding childhood cancer risk. The results remained consistent even after excluding the study by Spaan et al [17] (RR 1.1, 95% CI 0.95 - 1.26; Figure S3 in Multimedia Appendix 3) [15-22,27-32].

Discussion

Principal Findings

This study, to our knowledge, is the first systematic review and meta-analysis to use infertile or subfertile populations as the reference group. We found no significant increase in childhood overall cancer risk in ART-conceived offspring compared to non-ART. This result was consistent across different ART methods, including IVF, ICSI, FET, and fresh-ET. Furthermore, no significant differences were observed between FET and fresh-ET in terms of childhood cancer risk. Despite the heterogeneity of the studies included, the results were robust across sensitivity analyses, supporting the stability and reliability of our findings.

Comparison to Prior Work

To date, 6 meta-analyses have examined the association between ART fertility treatments and childhood cancer risk. The meta-analyses by Wang et al [12], Chiavarini et al [14], and Hargreave et al [10] found a significant correlation between ART and childhood cancer risk, while those by Raimondi et al [9], Gilboa et al [11], and Zhang et al [13] did not support such an association. Recent large-scale cohort studies have yet to reach a consensus on this issue. Some studies report a significant association between ART conception and increased childhood cancer risk [19,21,30,37,38]. For instance, a large Nordic study by Sargisian et al [21], which included data from Denmark, Norway, Sweden, and Finland, found a significantly increased risk of childhood cancer in ART-conceived offspring compared

to naturally conceived offspring (RR 1.13, 95% CI 1.01 - 1.26). However, other studies have not observed this association [16,17,20]. The inconsistencies in these findings may be due to differences in control group selection, sample size, and follow-up duration [5,6,8,39].

Recent evidence suggests that epigenetic changes may play a key role in causing infertility, rather than being simply a result of fertility treatments [40,41]. Couples experiencing infertility may already have a higher risk of epigenetic defects in their gametes, which fertility treatments have only helped to reveal [42]. The present meta-analysis provided new insights, and we selected an appropriate control group to eliminate the effects of infertile or subfertile. We included 6 large cohort studies involving 480,852 ART conceptions and 716,144 non-ART conceptions. In our analysis of the non-ART control group, we performed subgroup analysis by categorizing it into operational (IUI or artificial insemination) and nonoperational (OI or fertility drugs) groups. When the non-ART control group was defined as nonoperational, the RR for childhood overall cancer in the ART group increased to 1.23 (95% CI 0.98 - 1.54), approaching the statistical significance threshold. The review by Berntsen et al [5] suggested that a scientific control group should consist of children of low-fertility parents who conceived naturally. However, obtaining such controls is challenging because they are rarely included in registry data. Additionally, it was noted that children born through fertility measures, such as ovarian stimulation or IUI, could also serve as suitable controls. Due to limitations in current published studies, we focused on studies with the latter control group approach. Future research should aim to include offspring born to low-fertility parents who conceived naturally to better understand the long-term effects of both infertility and fertility treatments.

ICSI has become increasingly common worldwide, with approximately one-third of fresh ART cycles using conventional IVF and two-thirds using ICSI [43,44]. Despite the invasive nature of ICSI and ongoing concerns about the health of children born through this method, our meta-analysis, which included 2 eligible studies, showed that the risk of childhood overall cancer in ICSI-conceived offspring was not significantly higher compared to non-ART offspring (RR 0.76, 95% CI 0.26 - 2.2; I^2=94.61%). However, there was considerable heterogeneity between the studies, and further research is needed to confirm the long-term safety of ICSI regarding childhood cancer risk.

FET accounts for 32.6% of all ART treatment cycles in Europe, showing a clear increasing trend [45]. Large cohort studies and meta-analyses have provided short-term health data on FET, such as perinatal outcomes [46-50]. Compared to singletons born after fresh-ET, infants born after FET generally have higher birth weights and a higher risk of LGA (large for gestational age) in suprapregnant children but lower perinatal mortality. Singletons born after FET are at an increased risk of LGA and preterm labor compared to naturally conceived offspring [50]. However, data on the long-term health of FET offspring are limited. Studies comparing FET with naturally conceived offspring suggest an increased cancer risk in FET-conceived children [20,21].

In our meta-analysis, which included 4 large cohort studies with infertile populations as the comparison group, we found no significant increase in childhood overall cancer risk in FET-conceived offspring compared to non-ART offspring (RR 0.98, 95% CI 0.54 - 1.76; I^2=83.18%). Given the high heterogeneity, these results should be interpreted with caution. Additionally, to our knowledge, no meta-analyses have compared childhood cancer risk between FET and fresh-ET conceived offspring. Our analysis, which included 11 cohort studies with 176,800 FET-conceived and 723,327 fresh-ET-conceived individuals, found no significant difference in cancer risk between the 2 groups (RR 0.99, 95% CI 0.86 - 1.14), with low heterogeneity (I^2=24.45%). No significant bias was found, and sensitivity analyses confirmed the stability of the results.

Several studies have explored the association between ART conception and specific childhood cancers, including leukemia [16,51], lymphoma [52], hepatoblastoma [31,53], retinoblastoma [54], and central nervous system tumors [15]. A 2013 meta-analysis by Hargreave et al [10] reported an increased risk of cancers such as leukemia (RR 1.65), neuroblastoma (RR 4.04), and retinoblastoma (RR 1.62). A 2019 meta-analysis by Chiavarini et al [14] found that ART significantly increased the risk of hematological neoplasms (odds ratio [OR] 1.3, 95% CI 1.08 - 1.58) and neurological cancers (OR 1.21, 95% CI 1.01 - 1.46). Furthermore, a 2020 study by Zhang et al [13] showed a significantly increased risk of hematologic cancers (RR 1.39), other solid tumors (RR 1.57), and leukemia (RR 1.31). Leukemia is one of the most common childhood cancers and a leading cause of death in children, followed closely by lymphoma and central nervous system tumors [55]. Although several studies have suggested that ART is associated with an increased risk of childhood leukemia, most compared ART offspring with those conceived naturally [16,19,20,30,56]. In contrast, our analysis included 8 studies on leukemia, 3 of which had follow-up durations of more than 10 years, and 2 were case-control studies. The results showed no significant increase in leukemia risk in ART offspring compared to non-ART offspring (RR 0.99, 95% CI 0.79 - 1.24; I^2=12.79%). When cohort studies were analyzed separately, the results remained unchanged (RR 1.1, 95% CI 0.87 - 1.4; I^2=4.05%). Additionally, no significant differences were found in further analyses of other specific childhood cancers, including lymphoma (RR 1.22, 95% CI 0.64 - 2.34), brain cancer (RR 1.22, 95% CI 0.73 - 2.05), embryonal tumors (RR 1, 95% CI 0.63 - 1.58), retinoblastoma (RR 1.3, 95% CI 0.73 - 2.31), and neuroblastoma (RR 1.02, 95% CI 0.48 - 2.16).

Strengths and Limitations

Strengths

One strength of this study is the use of a more appropriate control group, that is, infertile or subfertile populations, which enhances the reliability of the comparisons and helps address the risk of epigenetic defects associated with infertility. Additionally, our estimates were not significantly affected by recall bias, which is common in case-control studies [57]. Parents of children with cancer may be more likely to recall

past events, potentially overestimating cancer risk. We combined 4 eligible case-control studies, mainly focusing on specific cancer types, and performed subgroup analyses, showing no significant differences across subgroups. Furthermore, large cohort studies with long-term and comprehensive data were recently included, reducing the risks of selection, attrition, and recall bias, while providing more opportunities to observe rare cancer exposures, thus enhancing the credibility of the findings.

Limitations

First, the sample size of infertile or subfertile populations was small. Despite a comprehensive search, the limited number of studies, especially on ICSI and FET offspring, may reduce confidence in the findings. Larger sample sizes are needed in future research for greater statistical power. Second, this meta-analysis did not classify the non-ART control group further. The lack of distinction between naturally conceived offspring from low-fertility parents and those conceived through ovarian stimulation or IUI may introduce confounding, affecting the cancer risk baseline. Future studies should differentiate these groups to better assess ART's impact on childhood cancer risk. Third, while some studies reported male infertility, gender-specific analyses were not performed, preventing separate calculations for male and female infertility. Future studies should address this to explore gender-specific effects on offspring health after ART. Fourth, the included studies used raw data without adjusting for factors like age, gender, birth order, socioeconomic status, and history of abortion. This lack of adjustment may affect result interpretation. Future studies should include adjusted data for more accurate conclusions.

Conclusions

This study found no significant difference in the risk of childhood cancer between offspring conceived through ART and those conceived through non-ART treatments (such as fertility drugs or OI/IUI) in infertile populations. While infertility treatments may elevate baseline risks, our findings suggest that whether individuals with infertility conceive using ART or non-ART methods, their offspring do not face a significantly higher risk of childhood cancer. Further research, especially comparing infertile populations who conceive naturally, is needed to better understand potential long-term health outcomes.

Acknowledgments

This study was funded by the Yunnan Province Science and Technology Department (no. 202201AY070001-294) and the Science and Technology Popularization Program (no. 202404AM350308).

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary files.

Authors' Contributions

Conceptualization: GS, MQC
Data curation: CQZ, RL, GS, ZPB
Formal analysis: CQZ, RL, GS
Writing – original draft: GS, MQC
Writing – review & editing: GS, MQC, CQZ, RL

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategy.
[DOC File, 54 KB - cancer_v11i1e65820_app1.doc]

Multimedia Appendix 2
Summary tables.
[DOC File, 160 KB - cancer_v11i1e65820_app2.doc]

Multimedia Appendix 3
Funnel charts and sensitivity analyses.
[DOC File, 138 KB - cancer_v11i1e65820_app3.doc]

Multimedia Appendix 4

Forest plot of cancer risk in offspring of different types of ART (IVF, ICSI, FET, and fresh-ET). ART: assisted reproductive technology; FET: frozen embryo transfer; fresh-ET: fresh embryo transfer; ICSI: intracytoplasmic sperm injection; IVF: in vitro fertilization.
[DOC File, 49 KB - cancer_v11i1e65820_app4.doc]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 25 KB - cancer_v11i1e65820_app5.docx]

References

1. Ono M, Kuji N, Ueno K, Kojima J, Nishi H. The long-term outcome of children conceived through assisted reproductive technology. Reprod Sci 2024 Mar;31(3):583-590. [doi: 10.1007/s43032-023-01339-0]
2. Ishihara O, Jwa SC, Kuwahara A, et al. Assisted reproductive technology in Japan: a summary report for 2018 by the Ethics Committee of the Japan Society of Obstetrics and Gynecology. Reprod Med Biol 2021 Jan;20(1):3-12. [doi: 10.1002/rmb2.12358] [Medline: 33488278]
3. Carson SA, Kallen AN. Diagnosis and management of infertility: a review. JAMA 2021 Jul 6;326(1):65-76. [doi: 10.1001/jama.2021.4788] [Medline: 34228062]
4. Pinborg A, Wennerholm UB, Bergh C. Long-term outcomes for children conceived by assisted reproductive technology. Fertil Steril 2023 Sep;120(3 Pt 1):449-456. [doi: 10.1016/j.fertnstert.2023.04.022] [Medline: 37086833]
5. Berntsen S, Söderström-Anttila V, Wennerholm UB, et al. The health of children conceived by ART: “the chicken or the egg?”. Hum Reprod Update 2019 Mar 1;25(2):137-158. [doi: 10.1093/humupd/dmz001] [Medline: 30753453]
6. Graham ME, Jelin A, Hoon AH Jr, Wilms Floet AM, Levey E, Graham EM. Assisted reproductive technology: short‐and long‐term outcomes. Develop Med Child Neuro 2023 Jan;65(1):38-49. [doi: 10.1111/dmcn.15332]
7. Stern JE, Farland LV, Hwang SS, et al. Assisted reproductive technology or infertility: what underlies adverse outcomes? Lessons from the Massachusetts Outcome Study of Assisted Reproductive Technology. F S Rev 2022 Oct;3(4):242-255. [doi: 10.1016/j.xfnr.2022.06.003] [Medline: 36505962]
8. Zhang S, Luo Q, Meng R, Yan J, Wu Y, Huang H. Long-term health risk of offspring born from assisted reproductive technologies. J Assist Reprod Genet 2024 Mar;41(3):527-550. [doi: 10.1007/s10815-023-02988-5] [Medline: 38146031]
9. Raimondi S, Pedotti P, Taioli E. Meta-analysis of cancer incidence in children born after assisted reproductive technologies. Br J Cancer 2005 Oct;93(9):1053-1056. [doi: 10.1038/sj.bjc.6602838]
10. Hargreave M, Jensen A, Toender A, Andersen KK, Kjaer SK. Fertility treatment and childhood cancer risk: a systematic meta-analysis. Fertil Steril 2013 Jul;100(1):150-161. [doi: 10.1016/j.fertnstert.2013.03.017] [Medline: 23562045]
11. Gilboa D, Koren G, Barer Y, et al. Assisted reproductive technology and the risk of pediatric cancer: a population based study and a systematic review and meta analysis. Cancer Epidemiol 2019 Dec;63:101613. [doi: 10.1016/j.canep.2019.101613] [Medline: 31606679]
12. Wang T, Chen L, Yang T, et al. Cancer risk among children conceived by fertility treatment. Int J Cancer 2019 Jun 15;144(12):3001-3013. [doi: 10.1002/ijc.32062] [Medline: 30548591]
13. Zhang Y, Gao R, Chen H, et al. The association between fertility treatments and the incidence of paediatric cancer: a systematic review and meta-analysis. Eur J Cancer 2020 Oct;138:133-148. [doi: 10.1016/j.ejca.2020.08.001]
14. Chiavarini M, Ostorero A, Naldini G, Fabiani R. Cancer risk in children and young adults (offspring) born after medically assisted reproduction: a systematic review and meta-analysis. J 2019;2(4):430-448. [doi: 10.3390/j2040028]
15. Luke B, Brown MB, Wantman E, et al. The risks of birth defects and childhood cancer with conception by assisted reproductive technology. Hum Reprod 2022 Oct 31;37(11):2672-2689. [doi: 10.1093/humrep/deac196] [Medline: 36112004]
16. Rios P, Herlemont P, Fauque P, et al. Medically assisted reproduction and risk of cancer among offspring. JAMA Netw Open 2024 May 1;7(5):e249429. [doi: 10.1001/jamanetworkopen.2024.9429] [Medline: 38696167]
17. Spaan M, Pontesilli M, van den Belt-Dusebout AW, et al. Cancer risk in children, adolescents, and young adults conceived by ART in 1983–2011. Human Reproduction Open 2023 Jun 1;2023(3):hoad027. [doi: 10.1093/hropen/hoad027]
18. Wainstock T, Walfisch A, Shoham-Vardi I, et al. Fertility treatments and pediatric neoplasms of the offspring: results of a population-based cohort with a median follow-up of 10 years. Am J Obstet Gynecol 2017 Mar;216(3):e1-e14. [doi: 10.1016/j.ajog.2017.01.015]
19. Weng SS, Huang YT, Huang YT, Li YP, Chien LY. Assisted reproductive technology and risk of childhood cancers. JAMA Netw Open 2022 Aug 1;5(8):e2230157. [doi: 10.1001/jamanetworkopen.2022.30157] [Medline: 36044210]
20. Hargreave M, Jensen A, Hansen MK, et al. Association between fertility treatment and cancer risk in children. JAMA 2019 Dec 10;322(22):2203-2210. [doi: 10.1001/jama.2019.18037] [Medline: 31821431]
21. Sargisian N, Lannering B, Petzold M, et al. Cancer in children born after frozen-thawed embryo transfer: a cohort study. PLoS Med 2022 Sep;19(9):e1004078. [doi: 10.1371/journal.pmed.1004078] [Medline: 36048761]
22. Terho AM, Tiitinen A, Martikainen H, Gissler M, Pelkonen S. Health of singletons born after frozen embryo transfer until early adulthood: a Finnish register study. Hum Reprod 2022 Nov 24;37(12):2899-2907. [doi: 10.1093/humrep/deac211]

23. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
24. Tang X, Han YP, Chai YH, et al. Association of kidney function and brain health: a systematic review and meta-analysis of cohort studies. Ageing Res Rev 2022 Dec;82:101762. [doi: 10.1016/j.arr.2022.101762] [Medline: 36374833]
25. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
26. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength. J Clin Epidemiol 2013 Jul;66(7):726-735. [doi: 10.1016/j.jclinepi.2013.02.003] [Medline: 23570745]
27. Bruinsma F, Venn A, Lancaster P, Speirs A, Healy D. Incidence of cancer in children born after in-vitro fertilization. Hum Reprod 2000 Mar;15(3):604-607. [doi: 10.1093/humrep/15.3.604]
28. Bal MH, Harlev A, Sergienko R, et al. Possible association between in vitro fertilization technologies and offspring neoplasm. Fertil Steril 2021 Jul;116(1):105-113. [doi: 10.1016/j.fertnstert.2020.12.013] [Medline: 33597091]
29. Pinborg A, Loft A, Aaris Henningsen AK, Rasmussen S, Andersen AN. Infant outcome of 957 singletons born after frozen embryo replacement: the Danish National Cohort Study 1995-2006. Fertil Steril 2010 Sep;94(4):1320-1327. [doi: 10.1016/j.fertnstert.2009.05.091]
30. Spector LG, Brown MB, Wantman E, et al. Association of in vitro fertilization with childhood cancer in the United States. JAMA Pediatr 2019 Jun 1;173(6):e190392. [doi: 10.1001/jamapediatrics.2019.0392] [Medline: 30933244]
31. Sundh KJ, Henningsen AKA, Källen K, et al. Cancer in children and young adults born after assisted reproductive technology: a Nordic cohort study from the Committee of Nordic ART and Safety (CoNARTaS). Hum Reprod 2014 Sep;29(9):2050-2057. [doi: 10.1093/humrep/deu143] [Medline: 24990274]
32. Williams CL, Bunch KJ, Stiller CA, et al. Cancer risk among children born after assisted conception. N Engl J Med 2013 Nov 7;369(19):1819-1827. [doi: 10.1056/NEJMoa1301675] [Medline: 24195549]
33. Foix-L'Hélias L, Aerts I, Marchand L, et al. Are children born after infertility treatment at increased risk of retinoblastoma? Hum Reprod 2012 Jul;27(7):2186-2192. [doi: 10.1093/humrep/des149] [Medline: 22587995]
34. Rudant J, Amigou A, Orsi L, et al. Fertility treatments, congenital malformations, fetal loss, and childhood acute leukemia: the ESCALE study (SFCE). Pediatr Blood Cancer 2013 Feb;60(2):301-308. [doi: 10.1002/pbc.24192] [Medline: 22610722]
35. Ajrouche R, Rudant J, Orsi L, et al. Maternal reproductive history, fertility treatments and folic acid supplementation in the risk of childhood acute leukemia: the ESTELLE study. Cancer Causes Control 2014 Oct;25(10):1283-1293. [doi: 10.1007/s10552-014-0429-8]
36. Munzer C, Menegaux F, Lacour B, et al. Birth-related characteristics, congenital malformation, maternal reproductive history and neuroblastoma: the ESCALE study (SFCE). Int J Cancer 2008 May 15;122(10):2315-2321. [doi: 10.1002/ijc.23301] [Medline: 18076072]
37. Borgmann-Staudt A, Michael S, Sommerhaeuser G, et al. The use of assisted reproductive technology by European childhood cancer survivors. Curr Oncol 2022 Aug 15;29(8):5748-5762. [doi: 10.3390/curroncol29080453] [Medline: 36005191]
38. Farland LV, Liu CL, Diop H, et al. Hospitalizations up to 8 years following delivery in assisted reproductive technology-treated and subfertile women. Fertil Steril 2022 Mar;117(3):593-602. [doi: 10.1016/j.fertnstert.2021.11.012] [Medline: 35058044]
39. Chen M, Heilbronn LK. The health outcomes of human offspring conceived by assisted reproductive technologies (ART). J Dev Orig Health Dis 2017 Aug;8(4):388-402. [doi: 10.1017/S2040174417000228]
40. Doornbos ME, Maas SM, McDonnell J, Vermeiden JPW, Hennekam RCM. Infertility, assisted reproduction technologies and imprinting disturbances: a Dutch study. Hum Reprod 2007 Sep;22(9):2476-2480. [doi: 10.1093/humrep/dem172] [Medline: 17586835]
41. Ludwig M, Katalinic A, Gross S, Sutcliffe A, Varon R, Horsthemke B. Increased prevalence of imprinting defects in patients with Angelman syndrome born to subfertile couples. J Med Genet 2005 Apr;42(4):289-291. [doi: 10.1136/jmg.2004.026930] [Medline: 15805153]
42. Niemitz EL, Feinberg AP. Epigenetics and assisted reproductive technology: a call for investigation. Am J Hum Genet 2004 Apr;74(4):599-609. [doi: 10.1086/382897] [Medline: 14991528]
43. Dyer S, Chambers GM, de Mouzon J, et al. International Committee for Monitoring Assisted Reproductive Technologies world report: assisted reproductive technology 2008, 2009 and 2010. Hum Reprod 2016 Jul;31(7):1588-1609. [doi: 10.1093/humrep/dew082] [Medline: 27207175]
44. Calhaz-Jorge C, De Geyter C, Kupka MS, et al. Assisted reproductive technology in Europe, 2013: results generated from European registers by ESHRE. Hum Reprod 2017 Oct 1;32(10):1957-1973. [doi: 10.1093/humrep/dex264]
45. European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE), Wyns C, Bergh C, et al. ART in Europe, 2016: results generated from European registries by ESHRE. Hum Reprod Open 2020;2020(3):hoaa032. [doi: 10.1093/hropen/hoaa032] [Medline: 32760812]
46. Pelkonen S, Koivunen R, Gissler M, et al. Perinatal outcome of children born after frozen and fresh embryo transfer: the Finnish cohort study 1995-2006. Hum Reprod 2010 Apr;25(4):914-923. [doi: 10.1093/humrep/dep477] [Medline: 20124395]

47. Maheshwari A, Pandey S, Amalraj Raja E, Shetty A, Hamilton M, Bhattacharya S. Is frozen embryo transfer better for mothers and babies? Can cumulative meta-analysis provide a definitive answer? Hum Reprod Update 2018 Jan 1;24(1):35-58. [doi: 10.1093/humupd/dmx031] [Medline: 29155965]
48. Sha T, Yin X, Cheng W, Massey IY. Pregnancy-related complications and perinatal outcomes resulting from transfer of cryopreserved versus fresh embryos in vitro fertilization: a meta-analysis. Fertil Steril 2018 Feb;109(2):330-342. [doi: 10.1016/j.fertnstert.2017.10.019]
49. Terho AM, Pelkonen S, Opdahl S, et al. High birth weight and large-for-gestational-age in singletons born after frozen compared to fresh embryo transfer, by gestational week: a Nordic register study from the CoNARTaS group. Hum Reprod 2021 Mar 18;36(4):1083-1092. [doi: 10.1093/humrep/deaa304]
50. Elias FTS, Weber-Adrian D, Pudwell J, et al. Neonatal outcomes in singleton pregnancies conceived by fresh or frozen embryo transfer compared to spontaneous conceptions: a systematic review and meta-analysis. Arch Gynecol Obstet 2020 Jul;302(1):31-45. [doi: 10.1007/s00404-020-05593-4] [Medline: 32445067]
51. Reigstad MM, Larsen IK, Myklebust T, et al. Risk of cancer in children conceived by assisted reproductive technology. Pediatrics 2016 Mar;137(3):e20152061. [doi: 10.1542/peds.2015-2061] [Medline: 26908669]
52. European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE), Wyns C, De Geyter C, et al. ART in Europe, 2017: results generated from European registries by ESHRE. Hum Reprod Open 2021;2021(3):hoab026. [doi: 10.1093/hropen/hoab026] [Medline: 34377841]
53. Marino JL, Moore VM, Willson KJ, et al. Perinatal outcomes by mode of assisted conception and sub-fertility in an Australian data linkage cohort. PLoS One 2014;9(1):e80398. [doi: 10.1371/journal.pone.0080398] [Medline: 24416127]
54. Lerner-Geva L, Boyko V, Ehrlich S, et al. Possible risk for cancer among children born following assisted reproductive technology in Israel. Pediatr Blood Cancer 2017 Apr;64(4):e26292. [doi: 10.1002/pbc.26292] [Medline: 27748017]
55. Wu Y, Deng Y, Wei B, et al. Global, regional, and national childhood cancer burden, 1990–2019: an analysis based on the Global Burden of Disease Study 2019. J Adv Res 2022 Sep;40:233-247. [doi: 10.1016/j.jare.2022.06.001]
56. Petridou ET, Sergentanis TN, Panagopoulou P, et al. In vitro fertilization and risk of childhood leukemia in Greece and Sweden. Pediatr Blood Cancer 2012 Jun;58(6):930-936. [doi: 10.1002/pbc.23194] [Medline: 21618418]
57. Schulz KF, Grimes DA. Case-control studies: research in reverse. Lancet 2002 Feb;359(9304):431-434. [doi: 10.1016/S0140-6736(02)07605-5]

Abbreviations

ART: assisted reproductive technology
FET: frozen embryo transfer
fresh-ET: fresh embryo transfer
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
ICSI: intracytoplasmic sperm injection
INPLASY: International Platform of Registered Systematic Review and Meta-Analysis Protocols
IUI: intrauterine insemination
IVF: in vitro fertilization
LGA: large for gestational age
NOS: Newcastle-Ottawa Quality Assessment Scale
OI: ovulation induction
OR: odds ratio
PICOS: Population, Intervention, Comparison, Outcomes, and Study Design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RR: relative risk

Edited by N Cahill; submitted 27.08.24; peer-reviewed by ACF De Farias Santos, A Palani, E Stern Judy; revised version received 26.01.25; accepted 27.01.25; published 12.03.25.

Please cite as:
Song G, Zhang CQ, Bai ZP, Li R, Cheng MQ
Assisted Reproductive Technology and Risk of Childhood Cancer Among the Offspring of Parents With Infertility: Systematic Review and Meta-Analysis
JMIR Cancer 2025;11:e65820
URL: https://cancer.jmir.org/2025/1/e65820
doi:10.2196/65820

Large Language Model Applications for Health Information Extraction in Oncology: Scoping Review

David Chen[1], BMSc; Saif Addeen Alnassar[2], BASc; Kate Elizabeth Avison[2], BASc; Ryan S Huang[1], MSc; Srinivas Raman[3], MD

[1]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[2]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

[3]Department of Radiation Oncology, BC Cancer Vancouver, 600 W 10th Ave, Vancouver, BC, Canada

Corresponding Author:
Srinivas Raman, MD
Department of Radiation Oncology, BC Cancer Vancouver, 600 W 10th Ave, Vancouver, BC, Canada

Abstract

Background: Natural language processing systems for data extraction from unstructured clinical text require expert-driven input for labeled annotations and model training. The natural language processing competency of large language models (LLM) can enable automated data extraction of important patient characteristics from electronic health records, which is useful for accelerating cancer clinical research and informing oncology care.

Objective: This scoping review aims to map the current landscape, including definitions, frameworks, and future directions of LLMs applied to data extraction from clinical text in oncology.

Methods: We queried Ovid MEDLINE for primary, peer-reviewed research studies published since 2000 on June 2, 2024, using oncology- and LLM-related keywords. This scoping review included studies that evaluated the performance of an LLM applied to data extraction from clinical text in oncology contexts. Study attributes and main outcomes were extracted to outline key trends of research in LLM-based data extraction.

Results: The literature search yielded 24 studies for inclusion. The majority of studies assessed original and fine-tuned variants of the BERT LLM (n=18, 75%) followed by the Chat-GPT conversational LLM (n=6, 25%). LLMs for data extraction were commonly applied in pan-cancer clinical settings (n=11, 46%), followed by breast (n=4, 17%), and lung (n=4, 17%) cancer contexts, and were evaluated using multi-institution datasets (n=18, 75%). Comparing the studies published in 2022 - 2024 versus 2019 - 2021, both the total number of studies (18 vs 6) and the proportion of studies using prompt engineering increased (5/18, 28% vs 0/6, 0%), while the proportion using fine-tuning decreased (8/18, 44.4% vs 6/6, 100%). Advantages of LLMs included positive data extraction performance and reduced manual workload.

Conclusions: LLMs applied to data extraction in oncology can serve as useful automated tools to reduce the administrative burden of reviewing patient health records and increase time for patient-facing care. Recent advances in prompt-engineering and fine-tuning methods, and multimodal data extraction present promising directions for future research. Further studies are needed to evaluate the performance of LLM-enabled data extraction in clinical domains beyond the training dataset and to assess the scope and integration of LLMs into real-world clinical environments.

(JMIR Cancer 2025;11:e65984) doi:10.2196/65984

KEYWORDS

artificial intelligence; chatbot; data extraction; AI; conversational agent; health information; oncology; scoping review; natural language processing; NLP; large language model; LLM; digital health; health technology; electronic health record

Introduction

The advent of electronic health records (EHR) has allowed clinicians to leverage their access to vast amounts of longitudinal, patient-level clinical text data that inform patient diagnoses, prognoses, and management [1]. However, the majority of useful clinical data are stored as unstructured free text that requires manual extraction into meaningful clinical features; therefore, clinicians spend more time on administrative work reviewing EHRs instead of practising patient-facing medicine [1]. To address this task of extracting key attributes from unstructured clinical text, natural language processing (NLP) methods have classically applied rule-based and machine-learning methods to identify important entities in text and categorize them based on categories of interest [2]. For instance, the extraction of cancer staging information from clinical text requires an NLP algorithm to recognize references to cancer staging in clinical texts and categorize these references

according to defined cancer staging nomenclature, such as the TNM classification of malignant tumors system.

Rule-based classification relies on domain expert-designed rules, heuristics, ontologies, and pattern-matching techniques to extract information from text. In contrast, machine learning-based approaches use statistical models trained on large-scale labeled text data to automatically learn patterns and generalize these learned competencies in data extraction to unlabeled testing data. The emergence of deep learning models, a subfield of machine learning that focuses on artificial neural network models with multiple processing layers, has been particularly effective at modeling the hierarchical structure of natural language and demonstrated superior performance across diverse NLP tasks, including but not limited to data extraction [3].

One particularly promising deep learning architecture, known as the transformer model, has gained worldwide attention for its generative language competency and strong performance in question answering, sentence completion, and sentence classification tasks compared to other deep learning models [4]. Deep learning–based transformer models may require less time and fewer resources needed to manually annotate training datasets compared to classical machine learning models and can better address nuanced edge cases in data extraction that may not be explicitly accounted for in rule-based data approaches [5,6]. However, these models are often limited by their need for large-scale computational resources and training data [7,8].

Modern LLMs are commonly built using adaptations of the transformer architecture and trained on large corpora of text to enable human-like natural language competency. Due to their extensive training dataset, LLMs such as BERT and GPT may have zero-shot capabilities, meaning they can perform tasks without prior task-specific training [9]. Emerging research on fine-tuning LLMs with custom datasets and prompt engineering for conversational LLMs has yielded promising performance improvements for specialized NLP tasks compared to baseline LLMs.

Given the longitudinal nature of cancer care, the vast amount of clinical text associated with cancer patient EHRs necessitates the development of automated methods for data extraction from these clinical records into structured data, which is useful for review by oncologists. The broad natural language competency of LLMs encourages the design of specialized LLM applications for data extraction from unstructured clinical text, reducing the oncologists' time and effort spent in manually reviewing patient EHRs to extract key information to inform their clinical decision-making.

The emergence of several recent pilot studies of LLM-enabled data extraction prompts the need for a scoping review to map the current landscape, including definitions, frameworks, and future directions for this novel tool in clinical data extraction. This review seeks to address this gap in the literature by characterizing primary research articles that evaluated an LLM tool applied to data extraction from unstructured clinical text into structured data.

Methods

We queried OVID Medline on June 2, 2024, using oncology ("neoplasms," "cancer," "onco," "tumor") and generative LLM ("natural language processing," "artificial intelligence," "generative," "large language model") keywords in consultation with a librarian. Non-English articles, nonprimary research articles, articles published before 2000, and articles published in nonpeer-reviewed settings were excluded. The full search strategy is detailed in Multimedia Appendix 1. Following the deduplication of articles (n=10) using the Covidence review management tool, the literature search yielded 817 articles for manual screening.

We conducted abstract screening followed by full-text screening of articles in duplicate (KA and SA), including primary research articles that tested a large language model, were applied in oncology contexts, and evaluated the performance of data extraction from text. The articles that evaluated an NLP-based algorithm that did not assess an LLM, were secondary research articles, applied in only nononcology settings, and did not evaluate or report the performance of data extraction from the clinical text were excluded. Screening conflicts were resolved through consensus discussion with a third reviewer (DC).

We extracted key study attributes from the included full-text papers in duplicate (KA and SA), including clinical domain, LLM attributes (eg, model, use of fine-tuning, use of prompt engineering), the dataset used for training and testing, primary study outcomes, model training methodology, and model evaluation processes. The LLMs were coded as baseline if they were applied "out of the box" without additional fine-tuning. LLMs were coded as (1) fine-tuned LLMs: the study described training the baseline LLM on a custom dataset intended to yield improved data extraction performance compared to the baseline LLM alone; (2) zero-shot LLMs: they were applied "out-of-the-box" without additional prompt engineering, (3) prompt engineered LLMs: the study described adaptations to prompting procedures, such as one-shot or few-shot prompting, designed to yield improved data extraction performance compared to the baseline LLM alone. Data extraction conflicts were resolved through consensus discussion with a third reviewer (DC).

The synthesis of extracted data involved grouping studies based on similarities in the evaluated specific model, clinical domain applied, and shared themes of strengths and limitations, based on outcomes reported by the studies. The appraisal process involved the completion of a standardized data extraction form to systematically code in duplicate (KA and SA) which articles commented on which themes of strengths and limitations, and the discrepancies were resolved through discussion (DC and SR). The risk of bias was assessed using ROBINS-I (Version 2) in duplicate (KA and SA), with conflicts resolved through consensus discussion with a third reviewer (DC). Cohen κ score was used to assess inter-rater concordance. This scoping review followed the PRISMA-ScR reporting guideline.

Results

The literature search yielded 817 papers, of which 24 papers met the inclusion criteria (Figure 1). Most included papers exhibited moderate (n=15, 62.5%) risk or low (n=9, 37.5%) risk of bias (Figure 2). The most common domains for moderate risk of bias included bias due to confounding (n=21, 87.5%) and bias in the selection of the reported result (n=21, 87.5%). No papers scored a high risk of bias in any domain. ROBINS-I risk of bias assessment exhibited moderate inter-rater concordance based on an κ score of 0.43.

Figure 1. Search and filtering strategy used to select large language model studies evaluating data extraction performance for inclusion in this review. LLM: large language model.

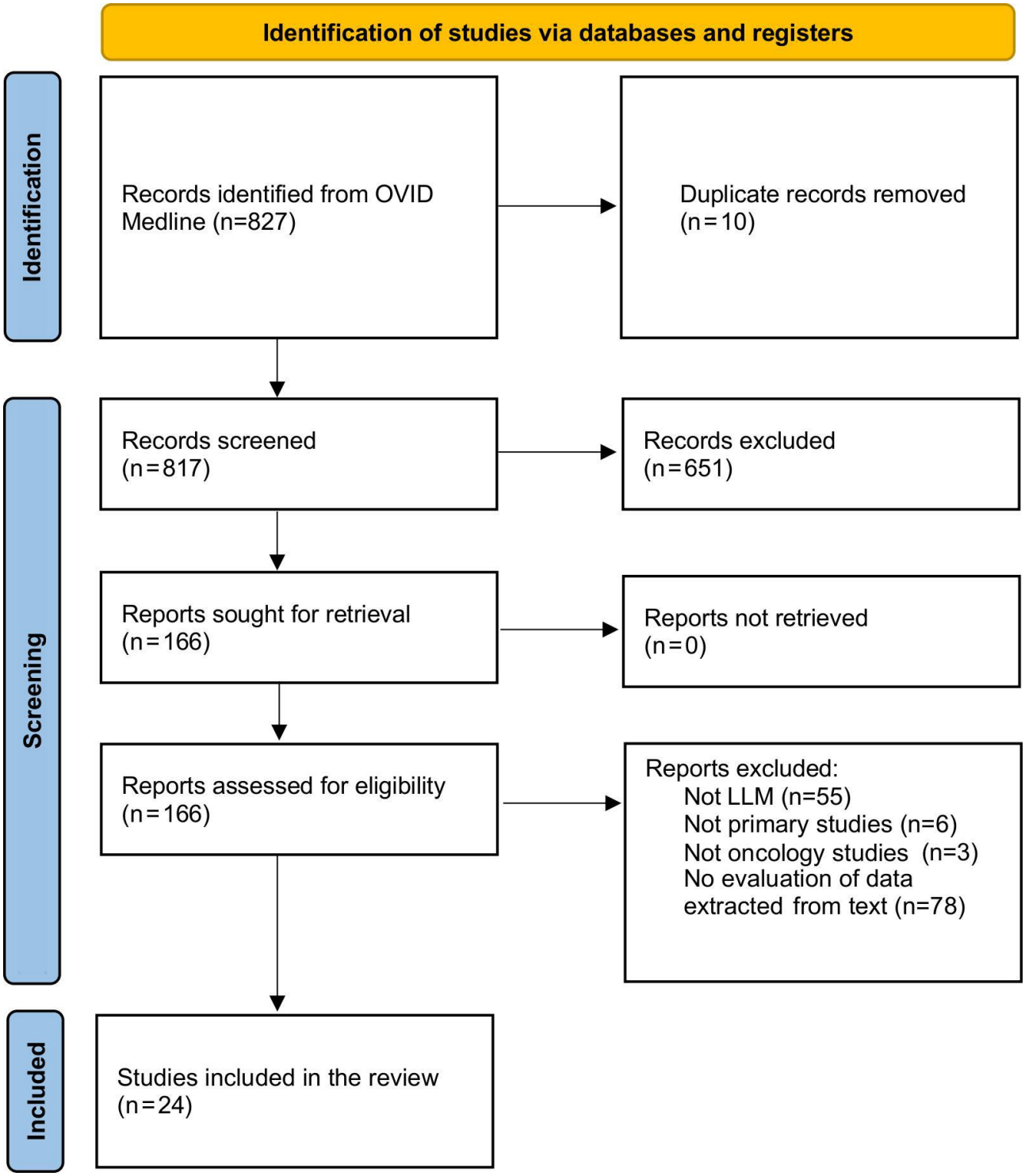


Figure 2. Risk of bias assessment using the ROBINS-I tool displayed as a traffic light plot for each included study [1,3,5,6,7,8-12,13,14-16,17,18,19,20-26].

Characteristics of the studies included in the study and published between 2019 - 2024 are shown in Table 1. The most common LLMs reported in these studies included BERT and its variants, as well as ChatGPT. Additional details related to methodology are reported in Multimedia Appendix 2.

Table . Characteristics of studies included in the review.

Study ID	Clinical domain	Baseline model	Baseline or fine-tuned LLM[a]	Zero-shot or prompt - engineered LLM	LLM main outcomes
Solarte-Pabon 2023[10]	Breast	BERT; RoBERTa	Fine-tuned	Zero-shot	F-scores: BETA: 0.9371; Multilingual BERT: 0.9463; RoBERTa Biomedical: 0.9501; RoBERTa BNE: 0.9454
Choi 2023 [11]	Breast	ChatGPT-3.5	Baseline	Prompt-engineered	Accuracy: 87.7%
Zhou 2022 [3]	Breast	BERT	Fine-tuned	Zero-shot	F1-score: 0.866 and 0.904 for exact and permissive matches respectively
Zhang 2019 [1]	Breast	BERT	Fine-tuned	Zero-shot	NER:[b] 93.53%; Relation extraction: 96.73% (best model, BERT+ Bi-LSTM-CRF)
Seong 2023 [5]	Colorectal	Bi-LSTM with a CRF layer; BioBERT	Fine-tuned	Zero-shot	Bi-LSTM-CRF:[c] Precision: 0.9844; F1-score:0.9848; Pre trained word embedding performed better than the one hot encoding pre-processing
Laios 2023 [12]	Gynecology	RoBERTa	Baseline	Zero-shot	AUROC:[d] 0.86; AUPRC:[e] 0.87; F1: 0.77; Accuracy: 0.81
Liu 2021 [13]	Liver	BERT	Fine-tuned	Zero-shot	APHE[f]: 98.40%; PDPH[g]: 90.67%
Fink 2023 [14]	Lung	ChatGPT-3.5; ChatGPT-4.0	Baseline	Prompt-engineered	Overall accuracy: GPT-4: 98.6%; GPT-3.5: 84% Metastatic ID accuracy: GPT-4: 98.1%; GPT-3.5: 90.3% Oncologic progression accuracy: GPT-4 F1: 0.96; GPT-3.5: 0.91 Oncologic reasoning correctness: GPT-4: 4.3; GPT-3.5: 3.9 accuracy: GPT-4: 4.4; GPT-3.5: 3.3
Chen 2023 [15]	Lung	BERT	Fine-tuned	Zero-shot	Macro F1-score: Task 1:0.92; Task 2: 0.82; Task 3: 0.74
Lyu 2023 [16]	Lung	ChatGPT-4.0	Baseline	Zero-shot	Translate: 4.27/5; Provided specific suggestions based on findings in 37% of all cases
Yu 2021 [7]	Lung	BERT; RoBERTa	Fine-tuned	Zero-shot	BERT Lenient: 0.8999 BERT Strict: 0.8791

Study ID	Clinical domain	Baseline model	Baseline or fine-tuned LLM[a]	Zero-shot or prompt - engineered LLM	LLM main outcomes
Martin-Noguerol 2024 [17]	Neurology	BERT	Fine-tuned	Zero-Shot	HGG: Precision: 79.17; Sensitivity: 76; F1:77.55; Metastasis: Precision: 73.91; Sensitivity: 77.27; F1: 75.56; AUC: 76.64
Fang 2022_1 [18]	Endocrine	BERT-BiLSTM-CRF	Fine-tuned	Zero-shot	Strict F1-score: 91.27%; Relaxed F1-score: 95.57%
Huang 2024 [19]	Pan-cancer	ChatGPT-3.5	Baseline	Prompt-engineered	Accuracy 0.89; F1 0.88; Kappa 0.80; Recall 0.89; Precision 0.89, Coverage 0.95
Arya 2024 [6]	Pan-cancer	BERT	Fine tuned	Zero-shot	Predict imaging scan site: Precision:99.4%; Recall:99.4%; F1-score: 99.3%; AUROC:99.4%; Accuracy:99.9%; Predict cancer presence: Precision:88.8%; Recall:89.2%; F1:88.8%; AUROC:97.6%; Accuracy:93.4%; Predict cancer status: Precision:85.6%; Recall:85.5%; F1-score: 85.5%; AUROC:97%; Accuracy:93.1%
Truhn 2024 [9]	Pan-cancer	ChatGPT-4.0	Baseline	Zero-shot	Experiment 1: Correct T-stage: 99%; Correct N-stage: 95; Correct M stage: 94; Lymph nodes; 99% Experiment 3: 100% accuracy
Hu 2024 [8]	Lung	ChatGPT-4.0	Baseline	Prompt-engineered	Prompt Base: Accuracy: 0.937; Precision: 0.860; Recall: 0.917; F1-score:0.882; Prior medical knowledge: Accuracy: 0.940; Precision:0.900; Recall: 0.864; F1:0.867; PMK-EN[h]: Accuracy: 0.896; Precision:0.871: Recall:0.776; F1: 0.786
Elmarakeby 2023 [20]	Pan-cancer	BERT	Fine-tuned	Zero-shot	AUC: ClinicalBERT: 0.93; DFCI-Imaging-BERT: 0.95 F1: ClinicalBERT: 0.72; DFCI-Imaging-BERT: 0.78

Study ID	Clinical domain	Baseline model	Baseline or fine-tuned LLM[a]	Zero-shot or prompt - engineered LLM	LLM main outcomes
Tan 2023 [21]	Pan-cancer	GatorTron; BERT; PubMedGPT	Fine-tuned	Prompt-engineered	Accuracy: GatorTron: 0.8916; BioMegatron:0.8861; BioBERT:0.8861; RoBERTa:0.8813; PubMedGPT:0.8762; DeBERTa:0.8746; BioClinicalBERT: 0.8746; BERT: 0.8708
Fang 2022_2 [22]	Pan-cancer	BERT	Baseline	Zero-shot	ROC:[i] 0.94
Mitchell 2022 [23]	Pan-cancer	BERT	Fine-tuned	Zero-shot	Group level site accuracy: 93.53%; Histology codes: 97.6%
Lu 2021 [24]	Pan-cancer	BERT	Fine-tuned	Zero-shot	Symptom domains: 0.931; problems with cognitive and social attributes on pain interference: 0.916; problems on fatigue: 0.929
Percha 2021 [25]	Breast	ALBERT; BART; ELECTRA; RoBERTa; XLNet	Fine-tuned	Zero-shot	ALBERT was the best-performing model in 22 out of the 43 fields
Gao 2021 [26]	Pan-cancer	BlueBERT	Fine-tuned	Zero-shot	BERT does not outperform baseline models–quantifiable measures not available

[a]LLM: large language model.

[b]NER: named entity recognition.

[c]Bi-LSTM-CRF: bidirectional-long short term memory-conditional random field.

[d]AUROC: area under the receiver operating characteristic.

[e]AUPRC: area under the precision-recall curve.

[f]APHE: hyperintense enhancement in the arterial phase.

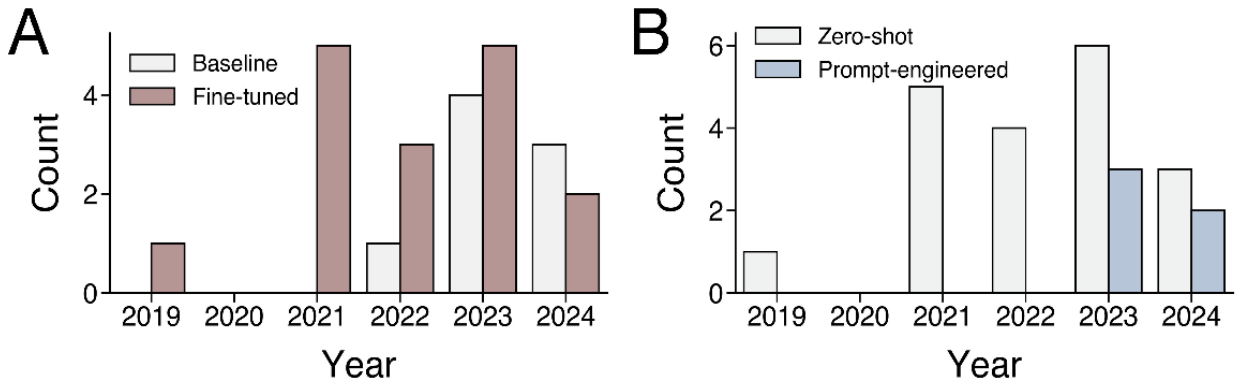
[g]PDPH: hypointense in the portal and delayed phases.

[h]PMK-EN: Prior Medical Knowledge-English Prompt

[i]ROC: receiver operating characteristic.

Most studies evaluated either the original or fine-tuned variants of the BERT LLM (n=18, 75%) in studies published between 2019 - 2024, followed by the Chat-GPT conversational LLM (n=6, 25%), upon application to data extraction from clinical texts in oncology, in studies published between 2023 - 2024. The LLMs for data extraction were commonly applied in pan-cancer clinical settings (n=11, 46%), followed by breast (n=4, 17%), lung (n=4, 17%), neurological (n=2, 8%), colorectal (n=1, 4%), gynecological (n=1, 4%), and liver (n=1, 4%) cancer contexts. The author teams of these studies belonged to institutions in the United States (n=11, 46%), China (n=4, 17%), Korea (n=3, 12%), Germany (n=2, 8%), Spain (n=2, 8%), the United Kingdom (n=1, 4%), and Singapore (n=1, 4%). Most studies were evaluated on datasets sourced from multiple institutional centers (n=18, 75%) compared to a single institutional center (n=6, 25%). Regarding the year of study publication, we observed a higher number of studies published between 2022 - 2024 (n=18, 75%) compared to 2019 - 2021 (n=6, 25%) (Figure 3). Notably, upon a comparison of studies published between 2022 - 2024 with studies between 2019 - 2021, the proportion of studies that reported the use of the fine-tuning method was lower (10/18, 55.6% vs 6/6, 100%) (Figure 3A), whereas the proportion of studies that reported the use of prompt engineering was higher (5/18, 28% vs 0/6, 0%) (Figure 3B).

Figure 3. Number of studies that evaluated (A) fine-tuning and (B) prompt engineering methodologies to optimize large language model data extraction performance.



Discussion

Principal Findings

Our scoping review of 24 studies highlights significant research interest in designing, evaluating, and deploying LLMs for data extraction from clinical text in oncology. The most commonly used LLMs for data extraction from clinical text in oncology include BERT and Chat-GPT, two of the most well-known LLMs in NLP research. These LLMs were most frequently applied in pan-cancer clinical contexts, reflecting their generalized natural language competency, regardless of clinical domain and context-specific terminologies and nomenclature. We observed a notable trend toward increasing utilization and refinement of LLM techniques over time, particularly in the areas of fine-tuning and prompt engineering. Given the common application of fine-tuning [26-28] and prompt-engineering [1,29,30] techniques in the design of deep learning- and LLM-based models in oncology, respectively, the emergence of optimized LLMs using these techniques represents a promising future direction for enhancing their data-processing capabilities. Despite these advancements, mixed reports of data extraction performance underscore the imperative for further assessment of these models across specific topics and use cases before their deployment as tools in cancer research and clinical care. Compared to historical statistical NLP and machine learning-based methods for data extraction in oncology, LLMs have been broadly evaluated for comparable applications, such as extracting tumor and cancer characteristics and patient-related demographic data [31].

The data processing competency of LLMs makes them a useful tool for automating repetitive, rule-based tasks, such as data extraction from clinical text on EHRs, to generate medical evidence about specific patients and patient populations that can inform patient care and population health guidelines respectively. Notably, LLMs have already shown competency in pilot studies of automated data extraction in biology [32], materials science materials science [33], and pharmacology [33], suggesting their generalized ability to extract relevant named entities from the clinical text that may be useful to synthesize medical knowledge. Across the included studies in this review, we found that LLMs offer several benefits for data extraction in clinical oncology, though further benchmarking against representative datasets and classical machine learning or statistical NLP approaches is required to determine their superior performance. In general, LLMs exhibited positive performance metrics compared to baseline human or statistical NLP approaches or were deemed feasible and acceptable in cross-sectional studies. These LLMs harbor the potential to balance accuracy and efficiency when processing large-scale, complex, unstructured text datasets found in EHRs [19]. Using LLM approaches for clinical data extraction as a supportive tool along with a human reviewer may reduce the potential for errors associated with human-led manual data extraction alone, thereby enhancing the reliability of clinical data analyses and interpretations [34].

Moreover, LLMs may curtail the resources required for data extraction, which is traditionally a labor- and time-intensive process [35]. For instance, our review highlighted the generalized performance of LLM-enabled data extraction across various text types in oncology, including histological and pathological classification [9,36], imaging report classification [8,14], and data extraction from postoperative surgery reports [5]. By automating the extraction and preliminary analysis of clinical text data, these models may free up valuable time for health care professionals, allowing them to focus more on patient-facing care and synthesis of medical knowledge from LLM-extracted information rather than the burden of administrative data management [10,12,37]. This shift not only improves clinical efficiency and cost-effectiveness but also reduces the serious risks of burnout among clinical staff by mitigating some of the repetitive administrative tasks associated with data handling [11,38].

Additionally, the versatility of LLMs across different clinical text contexts is notable. Whether dealing with structured data formats or the myriad forms of unstructured data present in EHRs, such as physician's notes and diagnostic reports, the general human-like natural language competencies of LLMs enable these "out-of-the-box" solutions to automatically adapt to and extract relevant information from varied data sources. This adaptability is crucial in precision oncology, where data from multiple data formats—such as imaging reports, next-generation sequencing results, and laboratory results—must be integrated and analyzed to generate personalized patient profiles and treatment strategies [39]. Our review highlighted

that current state-of-the-art evaluations of LLMs for data extraction in oncology have primarily focused on clinical text as input. However, we also highlight the recent emergence of multimodal LLMs capable of processing both image- and text-based inputs, serving as a new frontier for clinical decision support [40]. Taken together, future research to optimize data extraction for specific text formats in oncology—each with their own nuances—may improve extraction accuracy, enhance reliability, and produce results that can be trusted by clinicians and readily inform clinical decision-making [41].

The distribution of studies included in our scoping review reflects a predominant application of LLMs in pan-cancer clinical domains, accounting for nearly half of all research studies. This suggests that researchers leverage the versatility of LLMs to address broad oncological challenges across multiple cancer types, likely due to the generalizable nature of these models for various cancer data [42]. Breast and lung cancer also constituted a large portion of the studies, which can likely be attributed to their high prevalence and extensive clinical data availability, providing a rich dataset for deploying and testing the efficacy of LLMs [43]. The focus on these specific cancers indicates a targeted approach, where models are fine-tuned to address unique data extraction challenges, such as cancer type-specific nomenclature and lexicons. This underscores the potential of LLMs to be customized for specialized medical fields while also highlighting their broad "out-of-the-box" utility in general oncology. For instance, Gao et al [44] reported that BlueBERT did not outperform baseline nonLLM models in pan-cancer contexts, while Fang et al [22] and Mitchell et al (2022) [23] reported that the data extraction performance of BERT exceeded 90% accuracy in pan-cancer contexts. The mixed performance reported by different pilot studies of data extraction performance within the same clinical domain may be confounded by study-specific factors, including the prompting methodology, benchmark dataset, and definitions of performance metrics. These findings align with similar reports of mixed performance across different tasks and clinical text datasets within cancer type-specific domains [45-47], highlighting the need for systematic benchmarks to assess LLM data extraction reliability and domain-specific limitations. Standardizing performance metrics and defining critical thresholds for acceptable performance of data extraction accuracy remain open research questions to be addressed.

Our analysis reveals an increasing trend in the use of fine-tuning and prompt-engineering techniques in studies on LLMs, with 16 (67%) studies incorporating fine-tuning and 5 (21%) using prompt engineering. This progression suggests a maturation in the application of LLMs in clinical settings, where research has transitioned from developing baseline models for simple data extraction to the optimization of existing models using novel model adaptations and prompting methodologies tailored to the intricacies of medical data extraction. Fine-tuning allows models to adapt to the unique linguistic and contextual challenges presented by medical texts, potentially improving the accuracy and relevance of extracted information [29]. In comparison, prompt engineering enables the creation of more effective queries that align closely with the specific information needs of specialty fields such as oncology, steering LLMs toward more precise data retrieval [48]. For instance, Huang et al [19] demonstrated that providing LLMs with example outputs for few-shot learning and chain-of-thought reasoning methods for prompting yielded higher classification performance compared to baseline zero-shot applications of LLMs for data extraction. The careful design of prompting methodologies personalized to specific tasks and clinical domains within oncology may yield more accurate and efficient data extraction performance [49].

Despite the promising applications of LLMs in clinical oncology, our review also highlights notable disadvantages, particularly in cases of poor data extraction accuracy and performance [8,9]. Among the 24 reviewed studies, 9 (38%) cited accuracy as a limitation of LLMs for data extraction. These shortcomings underscore the critical need for cautious integration of LLMs into clinical workflows. The variability in performance can be attributed to the complex and diverse nature of clinical data, which may include nuanced medical terminologies and varied presentation styles across different documents [50]. These challenges emphasize the necessity for ongoing refinement and testing of these models under real-world conditions. Another minor disadvantage is the token limit of many LLMs, including both BERT and ChatGPT [20,42,44]. This limitation may complicate the extraction process, requiring models to be adapted to longer texts and resulting in reduced performance of these models [51]. Future research directions, as indicated by the reviewed studies, should involve performance benchmarks against existing statistical and machine learning–based methods and the extension of LLM tool validation to external, hold-out cohorts from additional clinical domains beyond those used in initial training datasets [7,16,24]. This would help ensure that the models are robust and reliable across various medical specialties and global oncology patient populations. While LLMs hold significant potential to revolutionize data management in oncology, their integration into clinical practice must be approached with careful planning and systematic evaluation to truly harness their capabilities without compromising patient care quality and privacy. The interpretation of both advantages and disadvantages of LLMs requires individualized consideration of each study, on a case-by-case basis given the heterogeneity in benchmark datasets, study designs, and reported outcomes.

Limitations

We acknowledge the limitations inherent in our scoping review. First, the rapid evolution of LLM technologies means that newer advancements may not have been fully represented in the reviewed studies due to the delays in publication cycles, leading to the omission of recent models. Second, the heterogeneity in study designs, datasets, and methodologies across included articles may affect the generalizability of findings in external contexts not evaluated in the same conditions as the original studies. Third, the majority of included studies originated from high-resource settings, primarily the United States, which may limit the applicability of results to lower-resource or structurally different health care systems. Fourth, while the risk of publication bias was not formally evaluated in our review, the tendency to publish studies with positive results may overrepresent the strengths of these LLMs without an

understanding and consideration of their limitations and nonpublished, negative results. Fifth, more recent journals that publish artificial intelligence research may not be indexed in the search databases yet, limiting the completeness of the search results in this scoping review. Sixth, this scoping review searched only one literature database, which may have resulted in the omission of relevant studies from other sources and limited the comprehensiveness of the findings.

Conclusion

In conclusion, the application of LLMs in oncology represents a forward leap in the digital transformation of health care data management. The potential to enhance data extraction processes and improve clinical decision-making is significant yet tempered by the current technological and methodological limitations. Ongoing research and development will be vital in harnessing the full potential of these models, ultimately leading to their more widespread adoption in clinical practice.

Authors' Contributions

Conceptualization: DC, SR
Data curation: KA, RH, SA
Formal analysis: DC, KA, RH, SA
Funding acquisition: SR
Investigation: DC, KA, RH, SA
Methodology: DC, SR
Project administration: DC, SR
Visualization: DC
Supervision: SR
Writing – original draft: DC, KA, RH, SA
Writing – review & editing: SR

Conflicts of Interest

None declared.

Multimedia Appendix 1
Scoping review full search strategy for MEDLINE.
[DOCX File, 7 KB - cancer_v11i1e65984_app1.docx]

Multimedia Appendix 2
Methodology characteristics of included studies.
[XLSX File, 12 KB - cancer_v11i1e65984_app2.xlsx]

Checklist 1
PRISMA-ScR reporting guideline.
[PDF File, 677 KB - cancer_v11i1e65984_app3.pdf]

References

1. Zhang X, Zhang Y, Zhang Q, et al. Extracting comprehensive clinical information for breast cancer using deep learning methods. Int J Med Inform 2019 Dec;132:103985. [doi: 10.1016/j.ijmedinf.2019.103985]
2. Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018 Jan;77:34-49. [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
3. Zhou S, Wang N, Wang L, Liu H, Zhang R. CancerBERT: a cancer domain-specific language model for extracting breast cancer phenotypes from electronic health records. J Am Med Inform Assoc 2022 Jun 14;29(7):1208-1216. [doi: 10.1093/jamia/ocac040]
4. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online on May 24, 2019 URL: http://arxiv.org/abs/1810.04805
5. Seong D, Choi YH, Shin SY, Yi BK. Deep learning approach to detection of colonoscopic information from unstructured reports. BMC Med Inform Decis Mak 2023 Feb 7;23(1):28. [doi: 10.1186/s12911-023-02121-7] [Medline: 36750932]
6. Arya A, Niederhausern A, Bahadur N, et al. Artificial intelligence-assisted cancer status detection in radiology reports. Cancer Res Commun 2024 Apr 9;4(4):1041-1049. [doi: 10.1158/2767-9764.CRC-24-0064] [Medline: 38592452]
7. Yu Z, Yang X, Dang C, et al. A study of social and behavioral determinants of health in lung cancer patients using transformers-based natural language processing models. AMIA Annu Symp Proc 2021;2021:1225-1233. [Medline: 35309014]
8. Hu D, Liu B, Zhu X, Lu X, Wu N. Zero-shot information extraction from radiological reports using ChatGPT. Int J Med Inform 2024 Mar;183:105321. [doi: 10.1016/j.ijmedinf.2023.105321]

9. Truhn D, Loeffler CM, Müller-Franzes G, et al. Extracting structured information from unstructured histopathology reports using generative pre-trained transformer 4 (GPT-4). J Pathol 2024 Mar;262(3):310-319. [doi: 10.1002/path.6232] [Medline: 38098169]
10. Solarte-Pabón O, Montenegro O, García-Barragán A, et al. Transformers for extracting breast cancer information from Spanish clinical narratives. Artif Intell Med 2023 Sep;143:102625. [doi: 10.1016/j.artmed.2023.102625] [Medline: 37673566]
11. Choi HS, Song JY, Shin KH, Chang JH, Jang BS. Developing prompts from large language model for extracting clinical information from pathology and ultrasound reports in breast cancer. Radiat Oncol J 2023 Sep;41(3):209-216. [doi: 10.3857/roj.2023.00633] [Medline: 37793630]
12. Laios A, Kalampokis E, Mamalis ME, et al. RoBERTa-assisted outcome prediction in ovarian cancer cytoreductive surgery using operative notes. Cancer Control 2023;30:10732748231209892. [doi: 10.1177/10732748231209892] [Medline: 37915208]
13. Liu H, Zhang Z, Xu Y, et al. Use of BERT (Bidirectional Encoder Representations from Transformers)-based deep learning method for extracting evidences in Chinese radiology reports: development of a computer-aided liver cancer diagnosis framework. J Med Internet Res 2021 Jan 12;23(1):e19689. [doi: 10.2196/19689] [Medline: 33433395]
14. Fink MA, Bischoff A, Fink CA, et al. Potential of ChatGPT and GPT-4 for data mining of free-text CT reports on lung cancer. Radiology 2023 Sep;308(3):e231362. [doi: 10.1148/radiol.231362] [Medline: 37724963]
15. Chen S, Guevara M, Ramirez N, et al. Natural language processing to automatically extract the presence and severity of esophagitis in notes of patients undergoing radiotherapy. JCO Clin Cancer Inform 2023 Jul;7(7):e2300048. [doi: 10.1200/CCI.23.00048] [Medline: 37506330]
16. Lyu Q, Tan J, Zapadka ME, et al. Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. Vis Comput Ind Biomed Art 2023 May 18;6(1):9. [doi: 10.1186/s42492-023-00136-5] [Medline: 37198498]
17. Martín-Noguerol T, López-Úbeda P, Pons-Escoda A, Luna A. Natural language processing deep learning models for the differential between high-grade gliomas and metastasis: what if the key is how we report them? Eur Radiol 2024 Mar;34(3):2113-2120. [doi: 10.1007/s00330-023-10202-4] [Medline: 37665389]
18. Fang A, Hu J, Zhao W, et al. Extracting clinical named entity for pituitary adenomas from Chinese electronic medical records. BMC Med Inform Decis Mak 2022 Mar 23;22(1):72. [doi: 10.1186/s12911-022-01810-z] [Medline: 35321705]
19. Huang J, Yang DM, Rong R, et al. A critical assessment of using ChatGPT for extracting structured data from clinical notes. NPJ Digit Med 2024 May 1;7(1):106. [doi: 10.1038/s41746-024-01079-8] [Medline: 38693429]
20. Elmarakeby HA, Trukhanov PS, Arroyo VM, et al. Empirical evaluation of language modeling to ascertain cancer outcomes from clinical text reports. BMC Bioinformatics 2023 Sep 2;24(1):328. [doi: 10.1186/s12859-023-05439-1] [Medline: 37658330]
21. Tan RSYC, Lin Q, Low GH, et al. Inferring cancer disease response from radiology reports using large language models with data augmentation and prompting. J Am Med Inform Assoc 2023 Sep 25;30(10):1657-1664. [doi: 10.1093/jamia/ocad133] [Medline: 37451682]
22. Fang C, Markuzon N, Patel N, Rueda JD. Natural language processing for automated classification of qualitative data from interviews of patients with cancer. Value Health 2022 Dec;25(12):1995-2002. [doi: 10.1016/j.jval.2022.06.004] [Medline: 35840523]
23. Mitchell JR, Szepietowski P, Howard R, et al. A question-and-answer system to extract data from free-text oncological pathology reports (CancerBERT Network): development study. J Med Internet Res 2022 Mar 23;24(3):e27210. [doi: 10.2196/27210]
24. Lu Z, Sim JA, Wang JX, et al. Natural language processing and machine learning methods to characterize unstructured patient-reported outcomes: validation study. J Med Internet Res 2021 Nov 3;23(11):e26777. [doi: 10.2196/26777] [Medline: 34730546]
25. Percha B, Pisapati K, Gao C, Schmidt H. Natural language inference for curation of structured clinical registries from unstructured text. J Am Med Inform Assoc 2021 Dec 28;29(1):97-108. [doi: 10.1093/jamia/ocab243] [Medline: 34791282]
26. Roslidar R, Saddami K, Arnia F, Syukri M, Munadi K. A study of fine-tuning CNN models based on thermal imaging for breast cancer classification. Presented at: 2019 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom); Aug 22-24, 2019; Banda Aceh, Indonesia p. 77-81. [doi: 10.1109/CYBERNETICSCOM.2019.8875661]
27. Chougrad H, Zouaki H, Alheyane O. Deep convolutional neural networks for breast cancer screening. Comput Methods Programs Biomed 2018 Apr;157:19-30. [doi: 10.1016/j.cmpb.2018.01.011] [Medline: 29477427]
28. Nasir MU, Ghazal TM, Khan MA, et al. Breast cancer prediction empowered with fine-tuning. Comput Intell Neurosci 2022;2022:5918686. [doi: 10.1155/2022/5918686] [Medline: 35720929]
29. Nguyen D, Swanson D, Newbury A, Kim YH. Evaluation of ChatGPT and Google Bard using prompt engineering in cancer screening algorithms. Acad Radiol 2024 May;31(5):1799-1804. [doi: 10.1016/j.acra.2023.11.002] [Medline: 38103973]
30. Khanmohammadi R, Ghanem AI, Verdecchia K, et al. Iterative prompt refinement for radiation oncology symptom extraction using teacher-student large language models. arXiv. Preprint posted online on 2024 URL: https://arxiv.org/abs/2402.04075

31. Savova GK, Danciu I, Alamudun F, et al. Use of natural language processing to extract clinical cancer phenotypes from electronic medical records. Cancer Res 2019 Nov 1;79(21):5463-5470. [doi: 10.1158/0008-5472.CAN-19-0579] [Medline: 31395609]
32. Jung SJ, Kim H, Jang KS. LLM based biological named entity recognition from scientific literature. Presented at: 2024 IEEE International Conference on Big Data and Smart Computing (BigComp); Feb 18-21, 2024; Bangkok, Thailand p. 433-435. [doi: 10.1109/BigComp60711.2024.00095]
33. Schilling-Wilhelmi M, Ríos-García M, Shabih S, et al. From text to insight: large language models for materials science data extraction. arXiv. Preprint posted online on Dec 2, 2024 URL: http://arxiv.org/abs/2407.16867
34. Hong MKH, Yao HHI, Pedersen JS, et al. Error rates in a clinical data repository: lessons from the transition to electronic data transfer--a descriptive study. BMJ Open 2013 May 28;3(5):e002406. [doi: 10.1136/bmjopen-2012-002406] [Medline: 23793682]
35. Meddeb A, Ebert P, Bressem KK, et al. Evaluating local open-source large language models for data extraction from unstructured reports on mechanical thrombectomy in patients with ischemic stroke. J NeuroIntervent Surg 2024 Aug 2:jnis-2024-022078. [doi: 10.1136/jnis-2024-022078]
36. Huang H, Lim FXY, Gu GT, et al. Natural language processing in urology: Automated extraction of clinical information from histopathology reports of uro-oncology procedures. Heliyon 2023 Apr;9(4):e14793. [doi: 10.1016/j.heliyon.2023.e14793]
37. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98. [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
38. De Hert S. Burnout in healthcare workers: prevalence, impact and preventative strategies. Local Reg Anesth 2020;13:171-183. [doi: 10.2147/LRA.S240564] [Medline: 33149664]
39. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. Nat Rev Cancer 2022 Feb;22(2):114-126. [doi: 10.1038/s41568-021-00408-3] [Medline: 34663944]
40. Chen D, Huang RS, Jomy J, et al. Performance of multimodal artificial intelligence chatbots evaluated on clinical oncology cases. JAMA Netw Open 2024 Oct 1;7(10):e2437711. [doi: 10.1001/jamanetworkopen.2024.37711] [Medline: 39441598]
41. Belyaeva A, Cosentino J, Hormozdiari F, Eswaran K, Shetty S, Corrado G, et al. Multimodal llms for health grounded in individual-specific data. In: Maier AK, Schnabel JA, Tiwari P, Stegle O, editors. Machine Learning for Multimodal Healthcare Data: Springer Nature Switzerland; 2024, Vol. 14315:86-102 URL: https://link.springer.com/10.1007/978-3-031-47679-2_7 [doi: 10.1007/978-3-031-47679-2_7]
42. Truhn D, Eckardt JN, Ferber D, Kather JN. Large language models and multimodal foundation models for precision oncology. NPJ Precis Oncol 2024 Mar 22;8(1):72. [doi: 10.1038/s41698-024-00573-2] [Medline: 38519519]
43. Hwang KT. Clinical databases for breast cancer research. In: Noh DY, Han W, Toi M, editors. Translational Research in Breast Cancer: Springer Singapore; 2021:493-509. [doi: 10.1007/978-981-32-9620-6_26]
44. Gao S, Alawad M, Young MT, et al. Limitations of transformers on clinical text classification. IEEE J Biomed Health Inform 2021 Sep;25(9):3596-3607. [doi: 10.1109/JBHI.2021.3062322] [Medline: 33635801]
45. Chen D, Parsa R, Hope A, et al. Physician and artificial intelligence chatbot responses to cancer questions from social media. JAMA Oncol 2024 Jul 1;10(7):956-960. [doi: 10.1001/jamaoncol.2024.0836] [Medline: 38753317]
46. Samaan JS, Yeo YH, Rajeev N, et al. Assessing the accuracy of responses by the language model ChatGPT to questions regarding bariatric surgery. Obes Surg 2023 Jun;33(6):1790-1796. [doi: 10.1007/s11695-023-06603-5] [Medline: 37106269]
47. Haver HL, Ambinder EB, Bahl M, Oluyemi ET, Jeudy J, Yi PH. Appropriateness of breast cancer prevention and screening recommendations provided by ChatGPT. Radiology 2023 May;307(4):e230424. [doi: 10.1148/radiol.230424] [Medline: 37014239]
48. Marvin G, Hellen N, Jjingo D, Nakatumba-Nabende J. Prompt engineering in large language models. In: Jacob IJ, Piramuthu S, Falkowski-Gilski P, editors. Data Intelligence and Cognitive Informatics: Springer Nature Singapore; 2024:387-402 URL: https://link.springer.com/10.1007/978-981-99-7962-2_30
49. Wang J, Shi E, Yu S, Wu Z, Ma C, Dai H, et al. Prompt engineering for healthcare: methodologies and applications. arXiv. Preprint posted online on Mar 23, 2024 URL: http://arxiv.org/abs/2304.14670
50. Ullah E, Parwani A, Baig MM, Singh R. Challenges and barriers of using large language models (LLM) such as ChatGPT for diagnostic medicine with a focus on digital pathology - a recent scoping review. Diagn Pathol 2024 Feb 27;19(1):43. [doi: 10.1186/s13000-024-01464-7] [Medline: 38414074]
51. Liu NF, Lin K, Hewitt J, et al. Lost in the middle: how language models use long contexts. Trans Assoc Comput Linguist 2024 Feb 23;12:157-173. [doi: 10.1162/tacl_a_00638]

Abbreviations

EHR: electronic health record
LLM: large language model
NLP: natural language processing

Edited by N Cahill; submitted 30.08.24; peer-reviewed by D Hu, K You, K Patel; revised version received 23.01.25; accepted 27.01.25; published 28.03.25.

Please cite as:
Chen D, Alnassar SA, Avison KE, Huang RS, Raman S
Large Language Model Applications for Health Information Extraction in Oncology: Scoping Review
JMIR Cancer 2025;11:e65984
URL: https://cancer.jmir.org/2025/1/e65984
doi:10.2196/65984

The Efficacy of Digital Interventions on Adherence to Oral Systemic Anticancer Therapy Among Patients With Cancer: Systematic Review and Meta-Analysis

Wan-Chuen Liao[1,2], MDS; Fiona Angus[1,3], MRes; Jane Conley[1], MPharm; Li-Chia Chen[1], PhD

[1]Centre for Pharmacoepidemiology and Drug Safety, Division of Pharmacy and Optometry, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester Academic Health Science Centre, Oxford Road, Manchester, United Kingdom

[2]School of Dentistry, College of Medicine, National Taiwan University, Taipei, Taiwan

[3]Department of Pharmacy, The Christie NHS Foundation Trust, Manchester, United Kingdom

Corresponding Author:
Fiona Angus, MRes
Centre for Pharmacoepidemiology and Drug Safety, Division of Pharmacy and Optometry, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester Academic Health Science Centre, Oxford Road, Manchester, United Kingdom

Abstract

Background: Digital interventions have been increasingly applied in multidisciplinary care plans to improve medication adherence to oral systemic anticancer therapy (SACT), the crucial lifesaving treatments for many cancers. However, there is still a lack of consensus on the efficacy of those digital interventions.

Objectives: This systematic review and meta-analysis aimed to investigate the efficacy of digital interventions in improving adherence to oral SACTs in patients with cancer.

Methods: This systematic review and meta-analysis followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement guidelines. The protocol has been registered at PROSPERO (no. CRD42024550203). Fully published, randomized controlled trials (RCTs) in English on adults with cancer assessing digital interventions for improving adherence to oral SACTs were retrieved from MEDLINE, Embase, APA PsycINFO, and CINAHL Plus up to May 31, 2024. Adherence measures compared between digital intervention users and nonusers were extracted. The proportions of poor adherence were synthesized using a random-effects model. The pooled results were reported as the odds ratio and 95% CI. The heterogeneity was assessed with the I^2 test (%). The mean difference and 95% CI were calculated from the mean adherence score and SD. A risk of bias assessment was conducted using version 2 of the Cochrane Risk of Bias Assessment Tool (RoB 2) for RCTs, which ensured that a quality assessment of all included studies was conducted as recommended by the Cochrane Collaboration.

Results: This study included 13 RCTs on digital interventions for improving adherence to oral SACTs in patients with cancer. The 13 RCTs, published between 2016 and 2024, were conducted in the United States, South Korea, France, Egypt, Finland, Australia, Colombia, Singapore, and Turkey. The technologies used were mobile apps (n=4), reminder systems (n=4), telephone follow-ups (n=3), and interactive multimedia platforms (n=2). Adherence was measured by surveys (n=8), relative dose intensity (n=2), pill count (n=1), self-reported missed doses (n=1), a smart pill bottle (n=1), and urine aromatase inhibitor metabolite assays (n=1). Concerns regarding risk of bias primarily involved randomization, missing outcome data, and outcome measurement, including nonblinded randomization, subjective patient-reported data, and difficulties in distinguishing between missed appointments and actual medication nonadherence. Pooled results from 11 trials showed that digital technology users had significantly lower risk of poor adherence (odds ratio 0.60, 95% CI 0.47 - 0.77). Two studies reported positive mean differences in adherence scores comparing digital intervention users and nonusers. However, due to considerable heterogeneity (I^2=73.1%), it is difficult to make a definitive conclusion from the pooled results about the effect of digital interventions upon adherence to oral anticancer therapy.

Conclusions: Digital intervention users exhibited significantly lower risk of poor oral SACTs adherence than nonusers. Acknowledging individual variation and tailoring digital technologies to prioritize patient needs is essential.

Trial Registration: PROSPERO CRD42024550203; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024550203

(JMIR Cancer 2025;11:e64208) doi:10.2196/64208

KEYWORDS

efficacy; digital interventions; oral systemic anticancer therapy; medication adherence; cancer; oral; patients with cancer; therapy; systematic review; meta-analysis; care plans; medication; treatments; mobile app; mobile applications; mHealth; multimedia platforms; digital technology; self-reported; mobile phone

Introduction

Medication adherence is a major public health concern, and nonadherence is responsible for 8% of global health expenditure and imposes a substantial economic burden on health care systems [1]. The advance in innovative treatments has led to an increasing number of cancers being classified as a long-term condition [2]. There is an increasing amount of research on measuring adherence [3], quantifying adherence rates in various drugs and cancer [4,5], investigating how to improve drug adherence [6], and identifying predictors of nonadherence [7].

Oral systemic anticancer therapy (SACT) has become increasingly accessible over the past 10 years, comprising 25% of oncology prescriptions globally [8] due to the advantages of being noninvasive, less intrusive, and more convenient [9]. However, they are prone to nonadherence as patients take medicines away from the medical setting. Many patients struggle to adhere to daily oral SACTs, with an adherence rate varying from 16% to 100% based on the settings and types of medicine [10].

Adherence is crucial to aiding successful patient outcomes of oral SACTs, while nonadherence can lead to disease progression, increased hospitalizations, and higher health care costs [11]. Factors such as complicated regimens, insufficient monitoring, poor communication, a lack of community support, mental health concerns, drug efficacy views, adverse effects, and financial load might contribute to nonadherence to oral SACT [6]. Clinicians may also neglect to mention the need for adherence and possible adverse effects, and patients may not have an adequate support system or understand the necessity of the medication [12]. Meanwhile, it has been asserted that interventions, including patient education and counseling, can improve treatment adherence [13].

Educational resources and various forms of communication have been used to build educational programs for patients in health care [14]. It is suggested that there is a link between continuous patient education and optimal adherence after a study showed that almost 50% of patients forgot their doctors' instructions immediately after being told them [15]. Patient-centered care and individualized interventions incorporating digital strategies have emerged as promising directions for research and development [16].

Innovative digital approaches include telemedicine, which refers to the provision of clinical services remotely using communication tools such as video or telephone. It encompasses activities such as diagnosis, monitoring, advice, reminders, education, interventions, and remote admissions, offering benefits such as reduced travel costs and time [17]. Smart home technology is another app that integrates computing solutions into living spaces to provide various services, including health care. Using telecommunication and web technologies can involve remote monitoring systems that enable patients to receive support while remaining in their homes [18].

Recent evidence suggests that digital interventions improve medication adherence in patients with chronic conditions. A meta-analysis involving 11 studies across various diseases demonstrated that reminder-based interventions, including text messages, phone calls, and video calls, significantly improved adherence, with 65.94% of prescribed doses taken in the reminder groups compared with 54.71% in control groups (P=.04) [19].

In oncology, digital tools such as apps [20], text messages [21], mobile games [22], phone calls [23], and multimedia interactive information technologies [14] have been used to increase medical adherence. Specific benefits of the digital approach include aiding in treatment recall, promoting healthy lifestyle habits, and suggesting that patient-focused educational initiatives could enhance treatment adherence and quality of life [14,24]. According to Karaaslan-Eşer and Ayaz-Alkaya [25], digital apps are easy to use, safe, provide access to medical professionals, offer guidance on managing symptoms with real-time feedback, and send timely notifications to enhance treatment adherence.

However, previous publications on the digital approach to increasing adherence have been limited to targeted oral SACT [26], specific digital tools (such as mobile [27], app-based design [20], text message [28], or telemedicine [23]), and specific diseases [29,30], with previous reviews lacking synthesized results from a meta-analysis [31,32]. Furthermore, medications for cancer treatment differ from those for other chronic conditions, as dosing is often less stable. SACTs are often adjusted by clinicians in response to treatment-related side effects and disease progression, leading to fluctuating dosages that complicate patient adherence [33].

Given these unique challenges, further investigation is warranted to evaluate the efficacy of digital interventions on adherence, specifically for patients with cancer taking oral SACT. This knowledge gap can be explored by undertaking this systematic review and meta-analysis examining their efficacy.

Methods

Protocol Registration

This systematic review and meta-analysis followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement guidelines (Multimedia Appendix 1) [34]. The protocol has been registered at PROSPERO (no. CRD42024550203). There were no deviations from the registered protocol.

Selection Criteria

The inclusion and exclusion criteria of this study are summarized as follows (Table 1).

Table . Inclusion and exclusion criteria of this study.

	Inclusion criteria	Exclusion criteria
Population and conditions	• Patients with cancer aged 18 years and older. • Patients diagnosed with cancer. • Patients with cancer taking oral SACTs[a].	• Patients with cancer including pediatrics, children, adolescents, neonates, or infants. • Studies that include mixed age groups of participants with cancer. • Patients with cancer taking nonoral SACTs[a]. • Patients with cancer exclusively receiving injectable SACTs[a].
Intervention and comparator	• The use of digital interventions such as: 1. Mobile apps 2. Web-based platforms 3. Wearable devices 4. Telemedicine interventions 5. Reminder systems (eg, text message reminders) 6. Virtual support groups or web-based communities • Comparator: standard or usual care without digital interventions.	• Studies that use nondigital interventions to improve adherence. • Studies with no suitable or appropriate comparator.
Outcome	• Adherence measures such as: 1. Medication possession ratio 2. Proportion of days covered 3. Self-reported adherence measures (eg, questionnaires and surveys) 4. Pharmacy refill data 5. Medication event monitoring systems (eg, smart pill bottles and electronic pill caps) 6. Biological markers	• The study does not contain outcome measures related to adherence. • Adherence measures are based solely on subjective reporting (unless validated self-reported measures were used).
Study type	• Human studies	• Animal or in vitro studies
Language	• English	• Non-English language
Publication	• Randomized controlled trials and clinical trials (comparative interventional trials)	• Review papers, systematic reviews, meta-analyses, cross-sectional studies, case-control studies, pilot studies, feasibility studies, editorials, commentaries, letters, opinion pieces, conference abstracts, gray literature, and non–peer-reviewed sources.

[a]SACTs: systemic anticancer therapies.

Types of Studies

Randomized controlled trials (RCTs) and clinical trials (nonrandomized, comparative interventional trials) were included. Review papers, systematic reviews, meta-analyses, cross-sectional studies, case-control studies, pilot studies, feasibility studies, editorials, commentaries, letters, opinion pieces, conference abstracts, gray literature, and non–peer-reviewed sources were excluded.

Types of Participants

This study included participants who met the following criteria: (1) patients aged 18 years and older, (2) patients diagnosed with cancer, and (3) patients taking oral SACTs. Patients younger than 18 years, studies that included mixed-age groups of participants, patients with cancer taking nonoral SACTs, and patients with cancer exclusively receiving injectable SACTs were all excluded.

Types of Interventions

The digital interventions were categorized according to the existing literature and the Cochrane Effective Practice and Organisation of Care (EPOC) taxonomy of health system interventions. EPOC outlined 4 categories of information and communication technology that health care organizations use for managing and delivering health care: health information systems, the application of information and communication technology, smart home technologies, and telemedicine [35].

To improve their adherence to oral SACTs, patients with cancer who used digital interventions, such as mobile apps, web-based platforms, wearable devices, telemedicine interventions, reminder systems (eg, text message reminders), virtual support

groups, or web-based communities, were included. Studies using nondigital interventions to enhance adherence were excluded.

Types of Outcome Measures

As there is no gold standard for measuring adherence and its associated outcomes, studies that reported adherence to oral SACTs, measured by various methods including self-reported adherence measures (such as the Morisky Medication Adherence Scale Score [36]), pharmacy refill data, medication event monitoring systems (including smart pill bottles and electronic pill caps), and biological markers, and presented as continuous or dichotomous data, such as the medication possession ratio [37], the proportion of days covered [37], or the proportion of adherence or nonadherence, were included in this review. Any studies that did not contain outcome measures related to adherence and studies that used adherence measures based solely on subjective reporting (unless validated self-reported measures were used) were excluded.

Data Sources and Search Strategies

A comprehensive electronic database search was conducted on MEDLINE, Embase, APA PsycINFO, and CINAHL Plus from their inception to May 31, 2024, as this review began in June 2024. MEDLINE and Embase are widely recommended for studying health care interventions [38], while APA PsycINFO and CINAHL Plus, although narrower in scope, are also well suited for this field. These databases focus on subject-specific rather than population-based information. Although there is no established guideline for the number of databases to include in a search, the combination of 2 broad and 2 focused databases is considered appropriate for the subject area of this review. Various structured search strategies were used, using controlled vocabulary and keywords based on the study's inclusion and exclusion criteria (Table 1) (Multimedia Appendix 2).

Study Selection

The title and abstract of papers retrieved from the electronic databases search were first screened by 2 reviewers (FA and WCL) independently according to the selection criteria (Table 1) using the predesigned electronic screening form. Each paper was rated as "included," "further check," or "excluded." The intraclass correlation coefficient (2-way mixed-effects model with absolute agreement [39]) and 95% CI were calculated for the consistency between 2 reviewers (FA and WCL) in record screening. Any discrepancy was resolved by discussing between reviewers and, if necessary, with a third reviewer (LCC) to reach a consensus. The full texts of potentially eligible papers were further reviewed independently by 2 reviewers (FA and WCL) to conclude the selection of studies.

Data Extraction and Management

The data for each study were independently extracted by 2 reviewers (FA and WCL) using the standardized and piloted electronic data extraction sheet. Disagreements were adjudicated by a third reviewer (LCC). Study information (study title, lead author, country, and year of publication), study design, setting, targeted population (cancer and oral SACT), intervention (digital apps), comparison, outcome measures, and follow-up period were extracted. Study results, including continuous data (such as mean adherence scale score and SD) and dichotomous data (such as the proportion of adherent or nonadherent patients), were retrieved. If raw data are unavailable, risk ratio, hazard ratio, mean (SD), median (range) of adherence duration, or any other results that can be converted into raw data were extracted. Duplicates were identified using EndNote 20 (Clarivate Analytics) through its default 1-step auto-deduplication process, which applies the matching criteria of "author," "year," and "title." This process was used to aid in screening the studies.

Risk of Bias Assessment

Controlling the risk of bias in a systematic review is crucial, as bias can distort the true effect of interventions [40]. Quality assessment of all included studies was conducted using version 2 of the Cochrane Risk of Bias Assessment Tool (RoB 2) for RCTs as recommended by the Cochrane Collaboration [41]. By assessing bias across 5 critical methodological aspects of each RCT, namely, the randomization process, deviations from the intended intervention, missing outcome data, outcome measurement, and selection of reported results [41], the included studies were categorized into "'low risk of bias," "some concerns," or "high risk of bias" using the RoB 2 tool. The results were subsequently tabulated. Risk of bias assessment was conducted independently and in duplicate by the 2 reviewers (FA and WCL).

Data Analysis

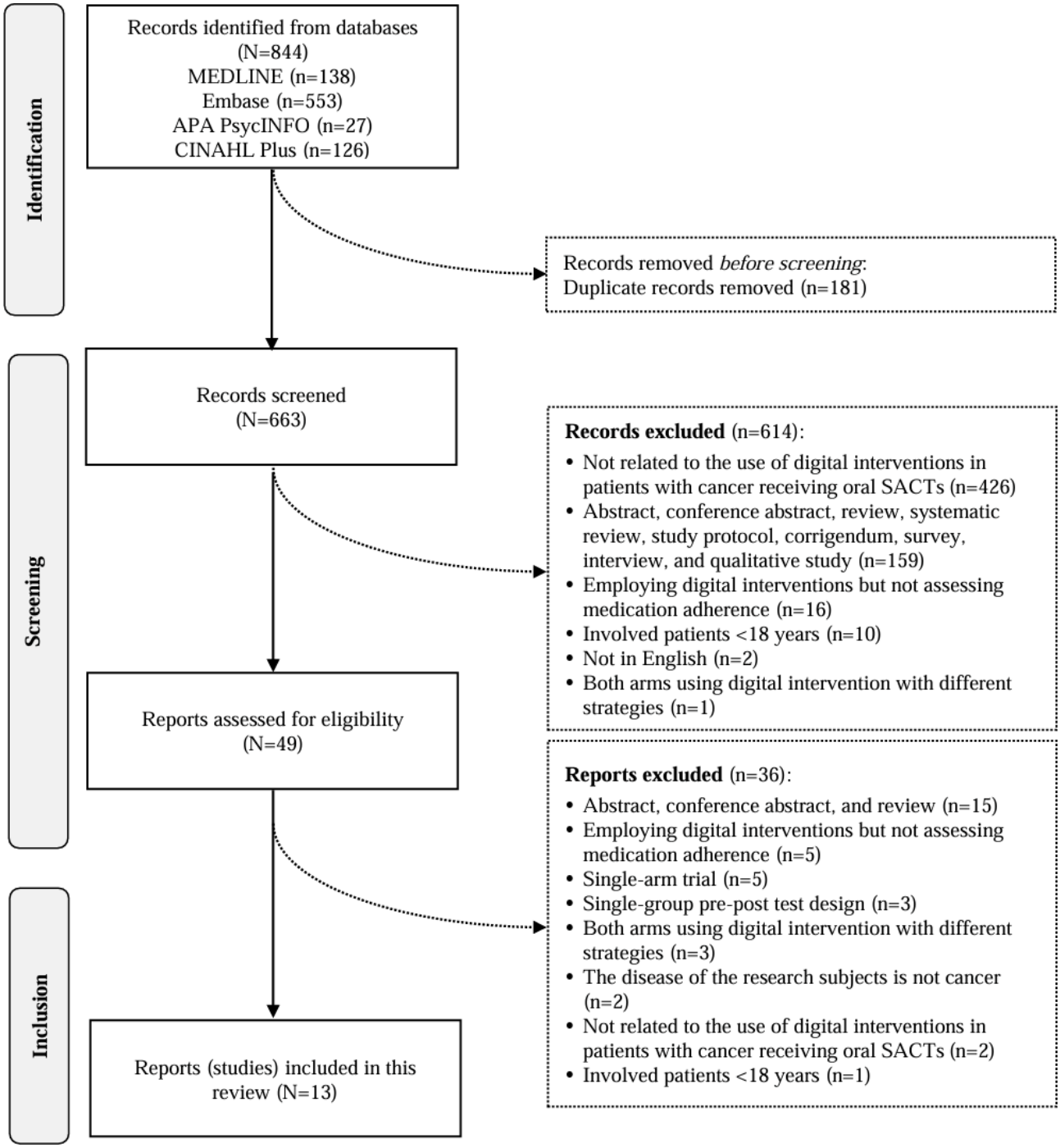
All outcomes were compared between the exposed group (digital intervention users) and the nonexposed group (those receiving standard care). The proportions of poor adherence were synthesized using a random-effects model (Der-Simonian and Laird method [42]). The pooled results were reported as odds ratio and 95% CI. The heterogeneity was assessed with the I^2 test (%). If appropriate, the mean difference and 95% CI of the adherence scale scores between the exposed and nonexposed groups were calculated and synthesized. The meta-analysis was conducted in STATA (Release 14; StataCorp LLC).

Results

Selection of Study

Of the 844 records identified from the electronic databases search, 181 duplicates were deleted. After screening titles and abstracts, 614 records were removed due to the irrelevance to digital interventions in patients with cancer receiving oral SACTs (n=426), being not fully published original interventional papers (n=159), not assessing medication adherence (n=16), involving patients younger than 18 years (n=10), not being in English (n=2), and both arms using digital interventions (n=1). After the full-text screening of the remaining 49 studies, 36 were excluded, leaving 13 studies (2611 participants) for inclusion in this review (Figure 1). The intraclass correlation coefficient between the 2 reviewers (WCL and FA) is 0.886 (95% CI 0.868-0.902), indicating good consistency. Since both authors demonstrated consistency and agreement at the full-text screening stage, the intraclass correlation coefficient was calculated solely for the abstract screening.

Figure 1. Selection of studies. APA: American Psychological Association; SACTs: systemic anticancer therapies.



Characteristics of Study

The 13 included RCTs, published from 2016 to 2024, were conducted in various countries: the United States (n=3) [21,33,43], South Korea (n=2) [22,24], France (n=2) [44,45], Egypt (n=1) [23], Finland (n=1) [46], Australia (n=1) [47], Colombia (n=1) [14], Singapore (n=1) [48], and Turkey (n=1) [25]. The studies involved patients with breast cancer (n=5) [21,22,24,47,48], various types of cancer (n=5) [25,33,43-45], chronic myeloid leukemia (n=1) [46], colorectal or gastric cancer (n=1) [23], and multiple myeloma (n=1) [14]. Digital interventions included mobile apps (n=4) [22,24,25,43], reminder systems (n=4) [21,33,47,48], telephone follow-ups (n=3) [23,44,45] and interactive multimedia platforms (n=2) [14,46]. According to the EPOC taxonomy [35], 7 RCTs used smart-home technologies [22,24,25,33,43,47,48], 4 used telemedicine [23,33,44,45], and 2 used information and communication technology [14,46] (Table 2). There were 1305 patients in the digital intervention group and 1306 patients in the control group.

Table . Characteristics of included studies.

Author, year, country	Cancer type, age of patients (years)	Digital intervention Tools or technology and intensity of intervention	EPOC[a]	Control	Adherence measure
Kekale et al (2016), Finland [46]	Chronic myeloid leukemia, median (range): 60 (25-83).	30-minute face-to-face counseling and multimedia interactive information technologies comprising a 5-minute video and daily text messages for 9 months.	Information and communication technology	Standard treatment	MMAS[b]
Kim et al (2018), South Korea [22]	Metastatic breast cancer, mean (SD): 50.9 (7.0)	Mobile game. Play the game for >30 minutes, 3 times weekly, for 3 weeks.	Smart-home technologies	Routine care	K-MARS[c]
Sikorskii et al (2018), United States [33]	Various types of cancer[d], mean (SD): 61 (12).	Reminder phone calls consisting of daily adherence reminder calls.	Telemedicine	Standard care	RDI[e]
Eldeib et al (2019), Egypt [23]	Metastatic colorectal or gastric cancer, mean (SD): intervention group: 49.98 (10.7); control group: 44.8 (12.65)	Follow-up phone calls involving weekly phone calls for the 11 cycles of treatment.	Telemedicine	Standard care	Pill count method
Greer et al (2020), United States [43]	Various types of cancer[f], mean (SD): 53.30 (12.91)	Mobile app with patients using the app for 12 weeks.	Smart-home technologies	Standard care	MMAS[b]
Hershman et al (2020), United States [21]	Early-stage breast cancer, median (range): 60.9 (30.7 - 82.4)	Text message twice a week for 3 years.	Smart-home technologies	No text messaging	Urine test
Tan et al (2020), Singapore [48]	Breast cancer, median (range): 61 (32-80)	Text message weekly for 1 year.	Smart-home technologies	Standard care	SMAQ[g]
Bouleftour et al (2021), France [44]	Various types of cancer[h], median (Q1-Q3): 70 (62-78)	Follow-up phone calls with calls at baseline, 3rd, 6th, 12th, and 24th weeks.	Telemedicine	Routine care	MMAS[b]
Karaaslan-Eser and Ayaz-Alkaya (2021), Turkey [25]	Various types of cancer[i], mean (SD): intervention group: 60.33 (9.31); control group: 62.14 (9.97)	Mobile app, which was a weekly record of symptoms and severity for 6 months.	Smart-home technologies	Standard care	OCAS[j]
Mir et al (2022), France [45]	Various advanced or metastatic cancer[k], median (range): 62 (20-92)	Follow-up by phone or internet (web portal) weekly for first month, biweekly from second to fourth month, and then 3 weekly from the fifth month onward.	Telemedicine	Usual care	RDI[e] and questionnaire
Park et al (2022), South Korea [24]	Breast cancer, mean (SD): 53.33 (8.71)	Mobile app and smart pill bottle reminder with smart pill bottle reminder daily for 4 weeks.	Smart-home technologies	Usual care	Automatic smartphone records
Singleton et al (2023), Australia [47]	Breast cancer, mean (SD): 55.1 (11.1)	Text messages comprising 4 text messages weekly for 6 months.	Smart-home technologies	Usual care	Self-reported missed doses within the last 7 days

Author, year, country	Cancer type, age of patients (years)	Digital intervention Tools or technology and intensity of intervention	EPOC[a]	Control	Adherence measure
Guio et al (2024), Colombia [14]	Multiple myeloma, mean (SD): intervention group: 65.19 (10.45); control group: 62.25 (11.89)	Multimedia interactive information technologies. Contents are presented to patients and caregivers at the start of each 4-month cycle.	Information and communication technology	Conventional educational approach	MAQ[l]

[a]EPOC: Effective Practice and Organisation of Care.

[b]MMAS: Morisky Medical Adherence Scale.

[c]K-MARS: Korean version of the Medication Adherence Rating Scale.

[d]Breast, colorectal, gastrointestinal, leukemia, liver, lung, lymphoma, melanoma, myeloma, pancreatic, prostate, renal, sarcoma, brain, esophageal, and other cancer.

[e]RDI: relative dose intensity (defined as the ratio of the dose delivered over time to the prescribed dose intensity).

[f]Hematologic, non–small cell lung, breast, high-grade glioma, sarcoma, gastrointestinal, genitourinary, melanoma, and nongastrointestinal stromal tumor sarcoma.

[g]SMAQ: Simplified Medication Adherence Questionnaire.

[h]Hematologic, breast, prostate, pulmonary, kidney, colon, cerebral, rectum, sarcoma, and other cancers.

[i]Colorectal cancer, gastrointestinal stromal tumor, lung cancer, renal cell carcinoma, hepatocellular carcinoma, cholangiocarcinoma, breast cancer, pancreatic cancer, and glioblastoma.

[j]OCAS: Oral Chemotherapy Adherence Scale.

[k]Endocrine, breast, digestive, renal, central nervous system, sarcoma, gynecological, lung, hematological, melanoma, and other.

[l]MAQ: Medication Adherence Questionnaire.

Quality Assessment

The 13 included RCTs raised concerns primarily related to the randomization process, missing outcome data, and outcome measurement; there were no high risks identified in any of the 5 areas of bias. The randomization was conducted by the principal investigator (KM) in one study [46] and lacked blinding in another [23]. In several studies, adherence outcomes were derived subjectively from patient-reported data via self-completed questionnaires [14,22,25,43-48]. In addition, challenges in differentiating missed appointments from actual medication nonadherence [21] and the possibility of smart pill bottles being opened without medication intake [24] further compounded measurement bias (Multimedia Appendix 3).

The challenges in recording outcome measures were found in 2 studies [21,24]. The authors of these RCTs made assumptions about the absence of urine samples as an indicator of nonadherence and the correlation between opening smart bottles and actual medication intake. While both studies used a sampling check or additional survey to support their assumptions, these diverse approaches contributed to increased heterogeneity and potential biases in this meta-analysis.

Characteristics of the Interventions

Four studies used mobile apps to integrate educational materials into their platforms [22,24,25,43]. Although the app (ILOVEBREAST) by Kim et al [22] functioned as a game, it still served as an educational tool for patients. Standard features of these mobile apps include side effects and symptom management [22,25,43], lifestyle guidance [43], and addressing adherence concerns [24,43]. Two of these studies incorporated additional digital technologies into their mobile apps, such as smart pill bottle reminders [24] and integrated Fitbit for monitoring physical activity [43] (Multimedia Appendix 4).

Moreover, standard features across mobile apps and other digital technologies included disease management and patient education about specific cancer types. Three studies directly targeted adherence through their digital technologies, either by questioning patients about their adherence [23,44] or by measuring it [24]. The remaining studies indirectly addressed adherence by focusing on related features. Some text messages covered a variety of content related to not only medication adherence but also physical activity, healthy diet, well-being, side effects management, physician recommendations, and providing support [21,47]. In addition, 3 studies used digital interventions to identify problems, particularly symptoms and toxicities [25,44,45]. In 1 study, health care professionals were able to access patient data and communicate with nurse navigators via a web portal [45] (Multimedia Appendix 4).

The delivery mode of digital technologies in the 13 RCTs varied. Mobile apps involve self-administration by patients, constituting a passive delivery method, although 2 studies personalized the app experience with features such as customized medication dosing timetables and symptom recording [25,43]. Reminder systems, either via text message or phone call, were passively delivered through telecommunication companies [48] or an interactive voice response system [33], with reminders predominantly generic. Telephone follow-ups were tailored to individual patients and proactively delivered by trained nurses [44,45] or a single principal investigator [23]. Interactive multimedia platforms, although passively delivered, provided bespoke content. One study combined multimedia interactive

platforms with face-to-face counseling sessions delivered by trained nurses [46] (Multimedia Appendix 4).

The duration of digital interventions in the 13 RCTs ranged from 3 weeks [22] to 3 years [21], with 1 study comprising 44 months in 11 undefined-length cycles [23]. Reminder systems were predominantly weekly, except for some studies conducted daily [33] or biweekly reminders [21]. Several studies used reminder systems to enhance adherence to oral SACTs. These systems varied, with some studies using smartphone messages [25,46,48], smart pill boxes [24], or telephone calls [33] to remind patients about their medication. Mobile apps were recommended for daily [22,24] or weekly use [25], except 1 study with unspecified frequency [33]. Telephone follow-ups varied from weekly [23] to less regular pattern [44,45]. One study combined follow-up phone calls with a web portal for web-based communication and patient information sharing [45]. Multimedia interactive platform engagement varied from monthly [14] to unspecified frequencies [46], with text messages being sent daily in 1 study [46] (Table 2).

Adherence Measurement

Adherence was the primary outcome in 11 RCTs, while 2 studies assessed it as a secondary outcome [45,47]. Various subjective measures, including surveys [14,22,25,43-46,48], relative dose intensity (RDI) [33,45], pill count [23], self-reported missed doses [47], and a smart pill bottle [24], were used across the 13 RCTs. One study used a more objective measure of adherence using time-to-adherence failure, defined by urine aromatase inhibitor metabolite assay results [21] (Table 2).

Adherence Rate

The pooled result from 11 studies [14,21,23-25,33,43,45-48] showed that users of digital technology had a significantly lower risk of poor adherence to oral SACTs than nonusers (odds ratio 0.60, 95% CI 0.47-0.77; I^2=73.1%) (Table 3). A trend was observed where smaller studies favored the digital intervention group [14,25,46], while larger studies favored the control group or showed no significant difference [21,33,43,45,48]. However, definitive conclusions cannot be drawn due to substantial heterogeneity (I^2=73.1%) [40]. In 1 study, only the proportion of medium adherence was reported, with no significant difference observed between the intervention (92/183, 77.2%) and control (91/183, 81.3%) groups [44].

Table . Proportion of patients with poor adherence in the included studies.

Study	Type of digital technology	Follow-up	Event rate[a]	Odds ratio (95% CI)
Kekäle et al (2016) [46]	• Face-to-face counseling • Interactive multimedia platforms	9 months	1/35 vs 9/33	0.08 (0.01 - 0.66)
Sikorskii et al (2018) [33]	• Reminder phone calls • Disease self-management tool kits	12 weeks	0/106 vs 0/108	1.02 (0.02 - 51.82)
Eldeib et al (2019) [23]	• Follow-up phone calls	11 cycles	0/44 vs 3/38	0.13 (0.01 - 2.73)
Greer et al (2020) [43]	• Mobile app	12 weeks	11/80 vs 20/86	0.53 (0.23 - 1.18)
Hershman et al (2020) [21]	• Text message	3 years	238/290 vs 268/313	0.77 (0.50 - 1.19)
Tan et al (2020) [48]	• Text message	1 year	59/123 vs 55/121	1.11 (0.67 - 1.83)
Karaaslan-Eser and Ayaz-Alkaya (2021) [25]	• Text message	6 months	16/38 vs 28/39	0.29 (0.11 - 0.74)
Mir et al (2022) [45	• Follow-up by phone or internet (web portal)	6 months	15/255 vs 26/265	0.57 (0.30 - 1.11)
Park et al (2022) [24]	• Mobile app integrated with a smart pill bottle reminder	4 weeks	1/30 vs 3/27	0.28 (0.03 - 2.83)
Singleton et al (2023) [47]	• Text message	6 months	3/42 vs 8/47	0.38 (0.09 - 1.52)
Guio et al (2024) [14]	• Interactive multimedia platforms	At least 100 days following transplantation or 3 months after maintenance	1/16 vs 13/16	0.02 (0.01 - 0.17)
Overall	N/A[b]	N/A	345/1059 vs 433/1093	0.60 (0.47 - 0.77); I^2=73.1%

[a]Event rate refers to the proportion of poor adherence in each study, measured by the specific method used in the study. Digital intervention users versus nonusers. Some event rate values have been converged based on the adherence data provided by studies.

[b]N/A: not applicable

Adherence Scale Score and RDI

Two studies reported adherence scale scores [22,44]. Although the results were not pooled, the mean difference was calculated (Table 4). These 2 studies generated positive mean differences, indicating that digital technology users experienced an increase or improvement in oral SACT adherence compared with nonusers. The mean (SD) of the RDI for the intervention group and the control group were 0.89 (0.03) (n=122) and 0.92 (0.03) (n=117) in one study [33], and 0.84 (0.26) (n=255) and 0.80 (0.21) (n=265) in another study [45]. A value of RDI<0.8 indicated underadherence, as reported in 1 study [33].

Table . Adherence scale score and mean difference of the included studies.

Study	Digital technology	Follow-up	Adherence scale	Mean (SD) score[a]	Mean difference[b] (95% CI)
Kim et al (2018) [22]	Mobile game	3 weeks	Korean version of the medication adherence rating scale	7.6 (0.7) (n=34) vs 6.5 (0.5) (n=38)	1.10 (0.82-1.38)[c]
Karaaslan-Eser and Ayaz-Alkaya (2021) [25]	Text message	6 months	Oral chemotherapy adherence scale	81.22 (8.05) (n=38) vs 73.36 (10.44) (n=39)	7.86 (3.81-11.91)[c]

[a]Digital intervention users versus nonusers.

[b]Mean difference represents the adherence score difference between digital intervention users and standard care patients, with higher scores indicating better adherence.

[c]P<.01.

Discussion

Principal Results

This study investigated the efficacy of digital interventions in improving adherence to oral SACTs and found that digital intervention users had a significantly lower risk of poor adherence to oral SACTs than nonusers. In addition, digital technology users demonstrated improved or increased adherence scores compared with nonusers.

Interactive and patient-focused digital supports have revolutionized the possibilities for improving medication adherence [16]. An overview of reviews indicates that incorporating digital technologies with direct clinician contact is likely to increase adherence [31]. A systematic review confirmed the efficacy of digital interventions in improving short-term treatment adherence among patients with cancer receiving oral chemotherapy [32]. Our pooled meta-analysis results also support this, as they showed a significantly reduced risk of poor adherence to oral SACTs among users of digital tools.

The efficacy of digital tools in achieving success can be attributed to various factors, for example, providing instructional resources, dosage aids, engagement with health care providers, digital medicine, self-monitoring, and quickly implementable technical methods [16]. Patient awareness of their drug regimen and the goals, benefits, and potential adverse events is critical for optimal adherence [49]. Digital can offer medication information and instructional help as educational resources [22,24,25,43]. Digital-based interventions such as personalized dosing schedules help patients organize and improve drug adherence [43]. Face-to-face counseling, proposed as a single consultation experience, was also included in our review for its potential to enhance patient adherence [46,50].

Implications

Medication adherence is crucial in oncology therapy, yet low adherence rates, as low as 14% for some cancer regimens, significantly impact patient health outcomes and strain health care systems and budgets [51]. This indicates that personalized interventions may improve adherence [51,52]. With more than 4.57 billion web users globally, 91% are accessing it via mobile devices, and smartphone usage—projected to increase by 8% annually [53], as well as digital health tools including phones and wearable devices—offer promising avenues for enhancing health care outcomes, cost-effectiveness, and patient acceptance [27].

Telemedicine offers greater flexibility than in-person interventions, allowing for addressing nonadherence wherever and whenever it occurs, such as between appointments or outside of clinic settings [54]. Telemedicine for reminder and follow-up phone calls was also a method of implementation used in several studies examined [23,33,44]. Digital medicine involves tools such as electronic pill bottles and wearable electronic devices. These devices enhance adherence and can track when containers are opened, although this does not verify intake [55]. Moreover, digital treatments may have drawbacks, including the cost and time needed for transferring or connecting with electronic equipment [16].

One study investigated whether using 1 or 2 digital tools improved adherence [56]. Both groups received weekly automated voice responses over 8 weeks, with the intervention group receiving additional daily text messages for 21 - 28 days. Results suggested that the extra text messages improved adherence and symptom management in patients taking oral anticancer agents. Another similar study showed that additional text messages could positively impact patients by promoting behavior change and improving self-care [28]. This highlights the potential for diverse clinical outcomes with varying types and quantities of digital tools.

Furthermore, social inequality is often correlated with the reduced use of digital technology in health care, contributing to a digital health divide [57]. For instance, older adults are less likely to use the web [58] or smartphones [59], and individuals with lower incomes face greater barriers to web access [60]. This inequality results in disparities in access to digital tools and hampers the implementation of digital interventions in health care [61]. To enhance accessibility, patients and health care professionals need to be involved in the development of these interventions, ensuring that they meet the needs of diverse patient populations. In addition, educational campaigns should aim to raise awareness and provide training on digital tools while also challenging stereotypes about older adults' technological capabilities and reinforcing patients' confidence in maintaining their privacy when using such interventions [61].

Strengths and Limitations

This review focuses on managing medication adherence at home for patients with cancer who are prescribed oral SACTs. All studies included are RCTs, considered the gold standard for measuring intervention efficacy [62]. We excluded single-group pre-post test designs to ensure randomization and aimed to cover various contemporary digital tools to assess their efficacy on medication adherence. One study had a 3-year follow-up, offering valuable insights into long-term impact [21]. The pooled meta-analysis results provide an integrated understanding of digital tools' efficacy in supporting medication adherence among patients with cancer.

While digital interventions hold promise, we acknowledge several limitations in this study, including various cancer types and oral SACT classes introducing disease uniqueness and drug response variability, potentially impacting medication adherence and intervention efficacy.

Despite including only RCTs, these studies exhibited considerable variability in research design, data collection methods, outcome measures, and the digital interventions used, as well as diversity in the cancer types investigated. The inability to conduct a patient-blinded experiment due to patient expectations of additional digital support is recognized [23,25]. Follow-up phone calls by different health care professionals may introduce bias [44,45]. Furthermore, reliance on subjective self-monitoring or self-reporting for medication adherence evaluation poses potential errors [24,33,46]. Small sample sizes

in some trials may limit statistical power and significance between intervention and control groups.

This heterogeneity is inherent to the subject matter [63]. Methodological heterogeneity was notable (I^2=73.1%), but it was accounted for by using a random-effects model in the meta-analysis, which assumes a normal distribution of underlying effects [40]. Also, due to the significant heterogeneity, the publication bias assessment test was not conducted to avoid presenting potentially misleading results. Acknowledging these limitations is crucial for interpreting the research results and allows readers to evaluate the significance and scope of the study more comprehensively. Another limitation of the study was that subgroup analyses were not conducted due to lack of data. This could have been used to investigate heterogeneous results or ask specific questions about a cancer type or intervention type.

This review included a variety of adherence and outcome measures due to the lack of consensus on these metrics. While self-reported adherence may be less robust due to recall bias and social desirability effects [64], only those studies using validated tools widely accepted in adherence research were included. Although these tools facilitate low-burden data collection, self-reported adherence may not always accurately reflect actual behavior, necessitating cautious interpretation of results. This diversity in outcome measures provides a comprehensive view of adherence-related consequences, which is crucial for understanding the broader context of digital interventions but may also complicate the ability to draw definitive conclusions.

Cancer populations encompass low-, middle-, and high-income regions globally, each with varying access to digital technologies and health care systems. Most studies have been conducted in high-income regions, which limits the generalizability of the results to low- and middle-income areas. In addition, the limited and diverse regional patient inclusion across these studies may further restrict the applicability of the findings to broader conditions [23-25,43].

Recommendations

Future interventions should be developed that focus on patient-centered, motivation-driven, and culturally adapted digital tools and be tailored for individuals with different types of cancer or oral SACTs. Efforts should focus on minimizing the threshold and difficulties associated with using digital tools and ensuring accessibility and ease of implementation for patients of all ages. Investigating patients' preferences for digital interventions could also increase usage rates. Monitoring health care professionals' responses and perspectives on digital interventions, alongside tracking patients' medication adherence, would provide valuable insights. To prevent alert fatigue [21], future research could explore optimal timing and frequency for implementing digital interventions. Qualitative studies could be conducted to delve deeper into the experiences of digital intervention users in real-world therapeutic settings, complementing quantitative findings.

Conclusions

Considering the growing use of oral SACTs and their higher patient acceptance over intravenous therapy, addressing medication adherence is vital in clinical oncology. Digital interventions offer effective support, enhancing adherence to oral SACTs and improving treatment outcomes while providing convenience for patients. This study highlights the significant benefits of digital technology in promoting adherence. Future research should focus on refining and personalizing digital tools to better meet individual patients' needs.

Acknowledgments

The authors thank Mr Tony Wei for data extraction and technical support and Dr Tristan P Martin for his guidance on conducting journal surveys. Dr Wan-Chuen Liao acknowledges the National Science and Technology Council, Taiwan, for supporting her Overseas Project for Postgraduate Research (112-2917-I-002-025) at The University of Manchester from 2023 to 2024. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other information content in this manuscript.

Authors' Contributions

The study was conceptualized by LCC and FA with data curated by WCL and FA. Formal analysis was conducted by WCL, FA, and JC. There was no funding acquisition for the study. The investigation was conducted by WCL, FA, and JC, and the methodology was developed by LCC, WCL, FA, and JC. Project administration was primarily by LCC and with support from WCL and FA. Resources and software were provided by The University of Manchester, and supervision was done by LCC. Validation was done by WCL and FA, and visualization was prepared by WCL and FA. The original draft was written by WCL and FA, and the review and editing of the writing were done by WCL, FA, and LCC. This statement has been written according to the CRediT taxonomy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 checklist.
[PDF File, 147 KB - cancer_v11i1e64208_app1.pdf]

Multimedia Appendix 2
Electronic search strategy.
[DOCX File, 19 KB - cancer_v11i1e64208_app2.docx]

Multimedia Appendix 3
The risk-of-bias assessment for the randomized controlled trials.
[PDF File, 21 KB - cancer_v11i1e64208_app3.pdf]

Multimedia Appendix 4
Characteristics of interventions.
[DOCX File, 25 KB - cancer_v11i1e64208_app4.docx]

References

1. Bosworth HB, Granger BB, Mendys P, et al. Medication adherence: a call for action. Am Heart J 2011 Sep;162(3):412-424. [doi: 10.1016/j.ahj.2011.06.007] [Medline: 21884856]
2. Pituskin E. Cancer as a new chronic disease: oncology nursing in the 21st century. Can Oncol Nurs J 2022;32(1):87-92. [Medline: 35280062]
3. Huang WC, Chen CY, Lin SJ, Chang CS. Medication adherence to oral anticancer drugs: systematic review. Expert Rev Anticancer Ther 2016;16(4):423-432. [doi: 10.1586/14737140.2016.1159515] [Medline: 26935964]
4. Angus F, Wang Y, Rigg A, Chen LC. Investigating adherence to tyrosine kinase inhibitors in renal cancer. J Oncol Pharm Pract 2023 Jun 5;29. [doi: 10.1177/10781552241259354]
5. Hall AE, Paul C, Bryant J, et al. To adhere or not to adhere: rates and reasons of medication adherence in hematological cancer patients. Crit Rev Oncol Hematol 2016 Jan;97:247-262. [doi: 10.1016/j.critrevonc.2015.08.025] [Medline: 26412718]
6. Tokdemir G, Kav S. The effect of structured education to patients receiving oral agents for cancer treatment on medication adherence and self-efficacy. Asia Pac J Oncol Nurs 2017;4(4):290-298. [doi: 10.4103/apjon.apjon_35_17] [Medline: 28966956]
7. Moon Z, Moss-Morris R, Hunter MS, Carlisle S, Hughes LD. Barriers and facilitators of adjuvant hormone therapy adherence and persistence in women with breast cancer: a systematic review. Patient Prefer Adherence 2017;11:305-322. [doi: 10.2147/PPA.S126651] [Medline: 28260867]
8. Lin C, Clark R, Tu P, Bosworth HB, Zullig LL. Breast cancer oral anti-cancer medication adherence: a systematic review of psychosocial motivators and barriers. Breast Cancer Res Treat 2017 Sep;165(2):247-260. [doi: 10.1007/s10549-017-4317-2] [Medline: 28573448]
9. O'Mahony C, Murphy KD, Byrne S. A mixed methods analysis of the monitoring of oral anti-cancer therapies. Eur J Oncol Nurs 2021 Oct;54:102026. [doi: 10.1016/j.ejon.2021.102026] [Medline: 34487968]
10. Foulon V, Schöffski P, Wolter P. Patient adherence to oral anticancer drugs: an emerging issue in modern oncology. Acta Clin Belg 2011;66(2):85-96. [doi: 10.2143/ACB.66.2.2062525] [Medline: 21630604]
11. Matsui D. Medication adherence issues in patients: focus on cost. CA Cancer J Clin 2013;2013:33-42. [doi: 10.2147/CA.S30125]
12. McCue DA, Lohr LK, Pick AM. Improving adherence to oral cancer therapy in clinical practice. Pharmacotherapy 2014 May;34(5):481-494. [doi: 10.1002/phar.1399] [Medline: 24877187]
13. Wilhelmsen NC, Eriksson T. Medication adherence interventions and outcomes: an overview of systematic reviews. Eur J Hosp Pharm 2019 Jul;26(4):187-192. [doi: 10.1136/ejhpharm-2018-001725] [Medline: 31338165]
14. Guio J, Melo A, Saldarriaga MM, et al. A novel digital educational strategy improves treatment adherence and quality of life in patients with multiple myeloma. J Cancer Educ 2024 Feb;39(1):50-57. [doi: 10.1007/s13187-023-02374-w] [Medline: 37875743]
15. Ley P. Memory for medical information. Br J Soc Clin Psychol 1979 Jun;18(2):245-255. [doi: 10.1111/j.2044-8260.1979.tb00333.x] [Medline: 454984]
16. Haga SB. Toward digital-based interventions for medication adherence and safety. Expert Opin Drug Saf 2020 Jun;19(6):735-746. [doi: 10.1080/14740338.2020.1764935] [Medline: 32370574]
17. Gogia S. Chapter 1—overview. In: Gogia S, editor. Fundamentals of Telemedicine and Telehealth: Academic Press; 2020:3-10.
18. Alam MR, Reaz MBI, Ali MAM. A review of smart homes—past, present, and future. IEEE Trans Syst, Man, Cybern C 2012;42(6):1190-1203. [doi: 10.1109/TSMCC.2012.2189204]
19. Fenerty SD, West C, Davis SA, Kaplan SG, Feldman SR. The effect of reminder systems on patients' adherence to treatment. Patient Prefer Adherence 2012;6:127-135. [doi: 10.2147/PPA.S26314] [Medline: 22379363]
20. Ali EE, Leow JL, Chew L, Yap KYL. Patients' perception of app-based educational and behavioural interventions for enhancing oral anticancer medication adherence. J Cancer Educ 2018 Dec;33(6):1306-1313. [doi: 10.1007/s13187-017-1248-x] [Medline: 28707206]

21. Hershman DL, Unger JM, Hillyer GC, et al. Randomized trial of text messaging to reduce early discontinuation of adjuvant aromatase inhibitor therapy in women with early-stage breast cancer: SWOG S1105. J Clin Oncol 2020 Jul 1;38(19):2122-2129. [doi: 10.1200/JCO.19.02699] [Medline: 32369401]
22. Kim HJ, Kim SM, Shin H, Jang JS, Kim YI, Han DH. A mobile game for patients with breast cancer for chemotherapy self-management and quality-of-life improvement: randomized controlled trial. J Med Internet Res 2018 Oct 29;20(10):e273. [doi: 10.2196/jmir.9559] [Medline: 30578205]
23. Eldeib HK, Abbassi MM, Hussein MM, Salem SE, Sabry NA. The effect of telephone-based follow-up on adherence, efficacy, and toxicity of oral capecitabine-based chemotherapy. Telemed J E Health 2019 Jun;25(6):462-470. [doi: 10.1089/tmj.2018.0077] [Medline: 30048216]
24. Park HR, Kang HS, Kim SH, Singh-Carlson S. Effect of a smart pill bottle reminder intervention on medication adherence, self-efficacy, and depression in breast cancer survivors. Cancer Nurs 2022;45(6):E874-E882. [doi: 10.1097/NCC.0000000000001030] [Medline: 34661562]
25. Karaaslan-Eşer A, Ayaz-Alkaya S. The effect of a mobile application on treatment adherence and symptom management in patients using oral anticancer agents: a randomized controlled trial. Eur J Oncol Nurs 2021 Jun;52:101969. [doi: 10.1016/j.ejon.2021.101969] [Medline: 33991868]
26. Geynisman DM, Wickersham KE. Adherence to targeted oral anticancer medications. Discov Med 2013 Apr;15(83):231-241. [Medline: 23636140]
27. Gambalunga F, Iacorossi L, Notarnicola I, Serra V, Piredda M, De Marinis MG. Mobile health in adherence to oral anticancer drugs. Comput Inform Nurs 2021 Jun 19;39(1):17-23. [doi: 10.1097/CIN.0000000000000643]
28. Spoelstra SL, Given CW, Sikorskii A, et al. Proof of concept of a mobile health short message service text message intervention that promotes adherence to oral anticancer agent medications: a randomized controlled trial. Telemed J E Health 2016 Jun;22(6):497-506. [doi: 10.1089/tmj.2015.0126] [Medline: 26716365]
29. Pereira-Salgado A, Westwood JA, Russell L, et al. Mobile health intervention to increase oral cancer therapy adherence in patients with chronic myeloid leukemia (The REMIND System): clinical feasibility and acceptability assessment. JMIR Mhealth Uhealth 2017 Dec 6;5(12):e184. [doi: 10.2196/mhealth.8349] [Medline: 29212628]
30. Hurtado-de-Mendoza A, Cabling ML, Dilawari A, et al. Providers' perspectives on adherence to hormonal therapy in breast cancer survivors. Is there a role for the digital health feedback system? Health Technol 2019 Mar;9(2):175-184. [doi: 10.1007/s12553-018-0267-x]
31. Dang TH, Forkan ARM, Wickramasinghe N, et al. Investigation of intervention solutions to enhance adherence to oral anticancer medicines in adults: overview of reviews. JMIR Cancer 2022 Apr 27;8(2):e34833. [doi: 10.2196/34833] [Medline: 35475978]
32. Puspitasari AW, Kristina SA, Prabandari YS. Digital intervention to improve oral chemotherapy adherence: systematic review. IJPR 2021 Feb 2;13(2):1310-1319 [FREE Full text] [doi: 10.31838/ijpr/2021.13.02.185]
33. Sikorskii A, Given CW, Given BA, et al. An automated intervention did not improve adherence to oral oncolytic agents while managing symptoms: results from a two-arm randomized controlled trial. J Pain Symptom Manage 2018 Nov;56(5):727-735. [doi: 10.1016/j.jpainsymman.2018.07.021] [Medline: 30096441]
34. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med 2009 Aug 18;151(4):W65-W94. [doi: 10.7326/0003-4819-151-4-200908180-00136] [Medline: 19622512]
35. EPOC taxonomy. Effective Practice and Organisation of Care (EPOC). 2015. URL: https://epoc.cochrane.org/epoc-taxonomy [accessed 2024-05-09]
36. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986 Jan;24(1):67-74. [doi: 10.1097/00005650-198601000-00007] [Medline: 3945130]
37. Andrade SE, Kahler KH, Frech F, Chan KA. Methods for evaluation of medication adherence and persistence using automated databases. Pharmacoepidemiol Drug Saf 2006 Aug;15(8):565-574. [doi: 10.1002/pds.1230] [Medline: 16514590]
38. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. Centre for Reviews and Dissemination.: CRD, University of York; 2009. URL: https://www.york.ac.uk/media/crd/Systematic_Reviews [accessed 2024-05-30]
39. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
40. Deeks JJ, Higgins JPT, Altman DG. Chapter 10: analyzing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 64: Cochrane; 2024. URL: https://training.cochrane.org/handbook [accessed 2025-04-09]
41. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919. [doi: 10.1136/bmj.i4919]
42. Kontopantelis E, Reeves D. Metaan: random-effects meta-analysis. The Stata Journal: Promoting communications on statistics and Stata 2010 Sep;10(3):395-407. [doi: 10.1177/1536867X1001000307]
43. Greer JA, Jacobs JM, Pensak N, et al. Randomized trial of a smartphone mobile app to improve symptoms and adherence to oral therapy for cancer. J Natl Compr Canc Netw 2020 Feb;18(2):133-141. [doi: 10.6004/jnccn.2019.7354] [Medline: 32023526]

44. Bouleftour W, Muron T, Guillot A, et al. Effectiveness of a nurse-led telephone follow-up in the therapeutic management of patients receiving oral antineoplastic agents: a randomized, multicenter controlled trial (ETICCO study). Support Care Cancer 2021 Aug;29(8):4257-4267. [doi: 10.1007/s00520-020-05955-3]
45. Mir O, Ferrua M, Fourcade A, et al. Digital remote monitoring plus usual care versus usual care in patients treated with oral anticancer agents: the randomized phase 3 CAPRI trial. Nat Med 2022 Jun;28(6):1224-1231. [doi: 10.1038/s41591-022-01788-1] [Medline: 35469070]
46. Kekäle M, Söderlund T, Koskenvesa P, Talvensaari K, Airaksinen M. Impact of tailored patient education on adherence of patients with chronic myeloid leukaemia to tyrosine kinase inhibitors: a randomized multicentre intervention study. J Adv Nurs 2016 Sep;72(9):2196-2206. [doi: 10.1111/jan.12978] [Medline: 27113362]
47. Singleton AC, Raeside R, Partridge SR, et al. Supporting women's health outcomes after breast cancer treatment comparing a text message intervention to usual care: the EMPOWER-SMS randomised clinical trial. J Cancer Surviv 2023 Dec;17(6):1533-1545. [doi: 10.1007/s11764-022-01209-9]
48. Tan EH, Wong ALA, Tan CC, et al. Improving medication adherence with adjuvant aromatase inhibitor in women with breast cancer: a randomised controlled trial to evaluate the effect of short message service (SMS) reminder. Breast 2020 Oct;53:77-84. [doi: 10.1016/j.breast.2020.06.012] [Medline: 32652462]
49. Martin LR, Williams SL, Haskard KB, Dimatteo MR. The challenge of patient adherence. Ther Clin Risk Manag 2005 Sep;1(3):189-199. [Medline: 18360559]
50. Tiao DK, Chan W, Jeganathan J, et al. Inflammatory bowel disease pharmacist adherence counseling improves medication adherence in Crohn's disease and ulcerative colitis. Inflamm Bowel Dis 2017 Aug;23(8):1257-1261. [doi: 10.1097/MIB.0000000000001194] [Medline: 28719539]
51. Dang TH, Wickramasinghe N, Forkan ARM, et al. Co-design, development, and evaluation of a mobile solution to improve medication adherence in cancer: design science research approach. JMIR Cancer 2024 Apr 3;10(1):e46979. [doi: 10.2196/46979] [Medline: 38569178]
52. Peh KQE, Kwan YH, Goh H, et al. An adaptable framework for factors contributing to medication adherence: results from a systematic review of 102 conceptual frameworks. J Gen Intern Med 2021 Sep;36(9):2784-2795. [doi: 10.1007/s11606-021-06648-1] [Medline: 33660211]
53. Kaye R, Rosen-Zvi M, Ron R. Digitally-enabled remote care for cancer patients: here to stay. Semin Oncol Nurs 2020 Dec;36(6):151091. [doi: 10.1016/j.soncn.2020.151091] [Medline: 33223410]
54. Basit SA, Mathews N, Kunik ME. Telemedicine interventions for medication adherence in mental illness: a systematic review. Gen Hosp Psychiatry 2020;62:28-36. [doi: 10.1016/j.genhosppsych.2019.11.004] [Medline: 31775066]
55. Aldeer M, Javanmard M, Martin RP. A review of medication adherence monitoring technologies. ASI 2018 May 6;1(2):14. [doi: 10.3390/asi1020014]
56. Spoelstra SL, Given CW, Sikorskii A, et al. Feasibility of a text messaging intervention to promote self-management for patients prescribed oral anticancer agents. Oncol Nurs Forum 2015 Nov;42(6):647-657. [doi: 10.1188/15.ONF.647-657] [Medline: 26488833]
57. Jahnel T, Dassow HH, Gerhardus A, Schüz B. The digital rainbow: digital determinants of health inequities. Digit Health 2022;8:20552076221129093. [doi: 10.1177/20552076221129093] [Medline: 36204706]
58. Exploring the UK's digital divide. Office for National Statistics. 2019. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/articles/exploringtheuksdigitaldivide/2019-03-04 [accessed 2024-05-09]
59. Mobile fact sheet. Center PR. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2025-04-09]
60. Measuring digital development. Facts and figures 2021. Union. IT. 2021. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2021.pdf [accessed 2024-12-21]
61. König LM, Krukowski RA, Kuntsche E, et al. Reducing intervention- and research-induced inequalities to tackle the digital divide in health promotion. Int J Equity Health 2023 Dec 4;22(1):249. [doi: 10.1186/s12939-023-02055-6] [Medline: 38049789]
62. Hariton E, Locascio JJ. Randomised controlled trials—the gold standard for effectiveness research. BJOG 2018 Dec;125(13):1716-1716 [FREE Full text] [doi: 10.1111/1471-0528.15199]
63. Ioannidis JPA, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. BMJ 2008 Jun 21;336(7658):1413-1415. [doi: 10.1136/bmj.a117] [Medline: 18566080]
64. Stirratt MJ, Dunbar-Jacob J, Crane HM, et al. Self-report measures of medication adherence behavior: recommendations on optimal use. Transl Behav Med 2015 Dec;5(4):470-482. [doi: 10.1007/s13142-015-0315-2] [Medline: 26622919]

Abbreviations

EPOC: Effective Practice and Organisation of Care
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RDI: relative dose intensity

RoB 2: Risk of Bias Assessment Tool
SACT: systemic anticancer therapy

Edited by N Cahill; submitted 11.07.24; peer-reviewed by K Matthew, K Sonnex, T Muller, X Wang; revised version received 14.01.25; accepted 15.01.25; published 16.04.25.

Please cite as:
Liao WC, Angus F, Conley J, Chen LC
The Efficacy of Digital Interventions on Adherence to Oral Systemic Anticancer Therapy Among Patients With Cancer: Systematic Review and Meta-Analysis
JMIR Cancer 2025;11:e64208
URL: https://cancer.jmir.org/2025/1/e64208
doi:10.2196/64208

Application of Artificial Intelligence in Cardio-Oncology Imaging for Cancer Therapy–Related Cardiovascular Toxicity: Systematic Review

Hayat Mushcab[1], PhD; Mohammed Al Ramis[2]; Abdulrahman AlRujaib[3]; Rawan Eskandarani[4], MD; Tamara Sunbul[5], MD; Anwar AlOtaibi[1], BSc; Mohammed Obaidan[2]; Reman Al Harbi[6]; Duaa Aljabri[7], PhD

[1]Research Office, Johns Hopkins Aramco Healthcare, Medical Access Road 1, Dhahran, Saudi Arabia

[2]College of Medicine, University College Dublin, Dublin, Ireland

[3]College of Medicine, Royal College of Surgeons in Ireland-Bahrain, Muharraq, Bahrain

[4]Emergency Department, King Fahad Medical City, Riyadh, Saudi Arabia

[5]Health Informatics Department, Johns Hopkins Aramco Healthcare, Dhahran, Saudi Arabia

[6]College of Medicine, University of Jeddah, Jeddah, Saudi Arabia

[7]College of Public Health, Imam Abdulrahman bin Faisal University, Dammam, Saudi Arabia

Corresponding Author:
Hayat Mushcab, PhD
Research Office, Johns Hopkins Aramco Healthcare, Medical Access Road 1, Dhahran, Saudi Arabia

Abstract

Background: Artificial intelligence (AI) is a revolutionary tool yet to be fully integrated into several health care sectors, including medical imaging. AI can transform how medical imaging is conducted and interpreted, especially in cardio-oncology.

Objective: This study aims to systematically review the available literature on the use of AI in cardio-oncology imaging to predict cardiotoxicity and describe the possible improvement of different imaging modalities that can be achieved if AI is successfully deployed to routine practice.

Methods: We conducted a database search in PubMed, Ovid MEDLINE, Cochrane Library, CINAHL, and Google Scholar from inception to 2023 using the AI research assistant tool (Elicit) to search for original studies reporting AI outcomes in adult patients diagnosed with any cancer and undergoing cardiotoxicity assessment. Outcomes included incidence of cardiotoxicity, left ventricular ejection fraction, risk factors associated with cardiotoxicity, heart failure, myocardial dysfunction, signs of cancer therapy–related cardiovascular toxicity, echocardiography, and cardiac magnetic resonance imaging. Descriptive information about each study was recorded, including imaging technique, AI model, outcomes, and limitations.

Results: The systematic search resulted in 7 studies conducted between 2018 and 2023, which are included in this review. Most of these studies were conducted in the United States (71%), included patients with breast cancer (86%), and used magnetic resonance imaging as the imaging modality (57%). The quality assessment of the studies had an average of 86% compliance in all of the tool's sections. In conclusion, this systematic review demonstrates the potential of AI to enhance cardio-oncology imaging for predicting cardiotoxicity in patients with cancer.

Conclusions: Our findings suggest that AI can enhance the accuracy and efficiency of cardiotoxicity assessments. However, further research through larger, multicenter trials is needed to validate these applications and refine AI technologies for routine use, paving the way for improved patient outcomes in cancer survivors at risk of cardiotoxicity.

Trial Registration: PROSPERO CRD42023446135; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023446135

(JMIR Cancer 2025;11:e63964) doi:10.2196/63964

KEYWORDS

artificial intelligence; cardiology; oncology; cancer therapy–induced; cardiotoxicity; cardiovascular toxicity; machine learning; imaging; radiology

Introduction

The World Cancer Research Fund International reported 18.1 million cancer cases in the year 2020, with breast and lung cancer being at the top of the list, representing 12.5% and 12.2% of all cases, respectively [1]. Breast cancer is the most commonly diagnosed type of cancer globally [2]. In 2020, the International Agency for Research on Cancer reported 27,885 new cancer cases, with nearly 47% of these cases ending with death [3].

In the United States, there are currently 17 million cancer survivors; by 2030, that number is predicted to rise to 22 million. For many cancer survivors, cardiovascular disease (CVD) is the leading cause of noncancer morbidity and mortality. Studies show that compared to the general population, patients with cancer have a 2 - 6 times higher chance of dying from CVD. Considering the progress made in cancer therapies and the decrease in cancer-related fatalities, comprehensive cardiovascular care is essential to improving these patients' overall results [4].

In recent years, there has been a notable advancement in the fight against cancer. However, a new problem has come to light: the potential for lifesaving cancer treatments to cause unintended damage to the heart. This is where cardio-oncology, a rapidly developing field, comes into play. It focuses on the crucial relationship between cancer treatment and heart health, focusing on controlling and preventing cardiovascular toxicity [5].

Cardiovascular toxicity, commonly known as cardiotoxicity, defined by the 2022 European Society of Cardiology Cardio-Oncology guidelines, is the term used to describe the harm inflicted upon the heart muscle or cardiovascular system due to different cancer treatments. Although chemotherapy and radiation therapy are essential tools in the fight against cancer, they can have negative side effects on the heart. These adverse effects can include anything from mild alterations in cardiac function to potentially fatal issues, including heart failure [6].

According to the American Society of Clinical Oncology (ASCO), the survivorship of cancer in the United States is approximately 67% and 18% for 5 and 20 years or more after diagnosis, respectively [7], especially if diagnosed early [8]. However, patients receiving cancer treatments such as chemotherapy, radiotherapy, and targeted agents have a 20% chance of developing myocardial dysfunction, with up to 7% to 10% having cardiomyopathy or heart failure [9]—in other words, therapy-induced cardiotoxicity [10,11]. Therapy-induced cardiotoxicity depends on the type of treatment, such as mediastinal and left-sided radiotherapy, anthracycline-based chemotherapy, and trastuzumab (targeted therapy), and other risk factors such as age, stage of diagnosis, ethnicity, and pre-existing CVDs [12].

Trastuzumab is a targeted therapy that uses drugs and other substances to precisely identify and attack specific types of cancer cells [13]. It is a humanized immunoglobulin G1 monoclonal antibody that is used to treat HER2+ (human epidermal growth factor receptor 2) breast cancer. Recently, it has also been approved to treat HER2+ advanced gastric cancer. The use of trastuzumab on patients with HER2+ breast cancer, which constitutes 20% of breast cancer cases, has demonstrated a significant reduction in recurrence risk, morbidity, and mortality. However, not all patients with HER2+ breast cancer respond to trastuzumab treatment due to resistance [14]. Recently, targeted therapy has been increasingly used in treating cancer, which has resulted in a significant improvement in the overall survival of patients with cancer. However, it can cause systemic toxicity, particularly cardiovascular toxicity [15].

Moreover, one of the most effective chemotherapy agents for several cancer types is anthracycline-based chemotherapy [16]. The American National Cancer Institute defines anthracycline as a type of antibiotic extracted from certain types of *Streptomyces* bacteria; it kills cancer cells by causing damage to their DNA and interfering with their reproduction [17,18]. The anthracycline chemotherapy agents include doxorubicin, epirubicin, daunorubicin, idarubicin, mitoxantrone, and valrubicin [18]. Although anthracyclines have been proven effective in treating various types of cancer, they do not come without adverse effects, which can limit their therapeutic potential [16]. These adverse effects range from mild and short-term to severe and long-term side effects [19]. Thus, early detection of cardiac dysfunction or cardiotoxicity allows the administration of the appropriate cardiac care, improving the overall outcome [20].

Long-term, dose-dependent risks of cardiotoxicity with anthracyclines are well-established [19]. Therefore, the recommended current practice by ASCO is a comprehensive assessment before initiating the treatment that includes a history and physical examination, screening for CVD risk factors, and an echocardiogram [21]. ASCO also recommends that clinicians manage modifiable cardiovascular risk factors (smoking, hypertension, diabetes, and obesity); the clinicians may incorporate several strategies, such as the use of dexrazoxane for cardioprotection, continuous infusion, or liposomal formulation of doxorubicin during the administration of anthracycline therapy [21]. In addition to cardiac imaging during the routine clinical assessment before therapy initiation (echocardiogram and cardiac magnetic resonance imaging [MRI]), ASCO recommends routine surveillance for cardiac function in patients considered to be at increased risk of developing cardiac dysfunction or heart failure [21,22].

The current method for cardiac function surveillance is "echocardiography" [14] to assess the left ventricular ejection fraction (LVEF) and the global longitudinal strain (GLS) [23]. Echocardiography has many advantages, making it the first modality of choice to monitor cardiotoxicity. These advantages include its ability to provide real-time imaging; availability and accessibility; noninvasiveness; and low cost [23]. However, echocardiography has limitations that hinder the detection of early signs of cardiotoxicity. Some of these limitations include the fact that echocardiography is entirely user-dependent, subjectivity in results interpretation, and variability in the image quality [23]. These limitations can result in the inability to detect subclinical cardiotoxicity and the early signs of cardiac dysfunction, which are crucial for personalized treatment plans that aim to improve the patient's prognosis [23]. Moreover, other CVD manifestations, such as myocardial perfusion and mitochondrial dysfunction, may precede a myocardial injury detected by echocardiography; this can only be recognized by a higher level of imaging modalities, which use targeted radiotracers such as cardiac magnetic resonance imaging (CMR) and nuclear imaging to provide information on specific mechanisms of cardiotoxicity [24].

With the recent emergence of artificial intelligence (AI) and machine learning (ML), their applications have meritoriously contributed to many advancements, with a promising potential for more across different areas, including imaging in the medical field [23,25]. One of the potential advancements is the rise of

stable diffusion, a generative model; it is anticipated that it might fill the gap in low-quality medical images by generating data on the missing details of the pathology with pattern recognition [25,26]. AI can generate this data by processing large amounts of readily available imaging data through artificial neural networks inspired by the connectionism of the biological neural network in the brain [25]. Tasks executed by AI algorithms in medical image processing include image acquisition, analysis, segmentation, feature extraction, visualization, registration, and classification [25]. Using AI-augmented imaging in the assessment of cardiotoxicity can help in recognizing subclinical cardiotoxicity caused by anthracyclines in addition to being able to reproduce the images more accurately by enhancing the imaging quality produced by the echocardiograph, which eventually will allow better monitoring and earlier detection of cardiac dysfunction [27]. For more detailed definitions of cancer treatments, cardiovascular toxicity, imaging modalities, and the application of AI in healthcare, please refer to Multimedia Appendix 1.

Despite its potential, the evidence base of AI imaging solutions for cardiovascular care in general and predicting cardiotoxicity in particular has been limited to date. Therefore, further research about AI's usefulness and effectiveness in the routine practice of cardio-oncology care is necessary. This systematic review aims to review the available literature on the use of AI in cardio-oncology imaging to predict cardiotoxicity and describe the possible improvement of each modality for cardio-oncology imaging when deploying AI to routine practice.

Methods

This systematic review was conducted according to the PRISMA (Preferred Reporting Items for Reviews and Meta-Analyses) guidelines from July 1 to August 1, 2023. The review is registered in PROSPERO (International Prospective Register of Systematic Reviews; CRD42023446135).

Search Strategy

The literature search for this review was performed using PubMed, MEDLINE, Cochrane Library, CINAHL, and Google Scholar for relevant studies from inception until June 2023. An AI research assistant (Elicit) was also used to search for relevant papers using the same terminology. In addition, PROSPERO was searched for ongoing similar systematic reviews. The first and senior authors are experienced in systematically reviewing the literature and have published several reviews. In addition, the authors have consulted experts using Editage services to achieve a high level of reliability. Please see Multimedia Appendix 2 for a detailed search strategy.

Terminology

In order to achieve the objective of this review, the databases were searched using keywords and their Medical Subject Headings (MeSH) terms connected by the Boolean operators "AND," "OR," and "*."

The search used the following terms and their MeSH terms: artificial intelligence, AI, deep learning, machine learning, cardio-oncology, cardiotoxicity, cardiac toxicity, cancer treatment, cancer therapy, "artificial intelligence," "machine learning," "AI augmentation," "deep learning," OR "AI" AND "Cardio-oncology," "Cardiotoxicity," "Cardiovascular toxicity," OR "Cardiac toxicity" AND "chemotherapy," "anthracycline," "cytotoxic regimens," "immunotherapy," "Cancer treatment," OR "therapy-induced" AND "Imaging," "Echocardiogram," "Echo*," "Cardiac magnetic resonance imaging," "CMR," "Multigated acquisition," "MUGA," "Cardiac computed tomography," OR "CCT."

Eligibility Criteria and Study Selection

Original studies reporting AI outcomes in adult patients diagnosed with any type of cancer and undergoing cardiotoxicity assessment were included. Outcomes included incidence of cardiotoxicity, LVEF, risk factors associated with cardiotoxicity, heart failure, myocardial dysfunction, signs of cancer therapy–related cardiovascular toxicity (CTR-CVT), echocardiography, and CMR. Non-English studies, case reports, literature reviews, studies on children, and studies that did not include CTR-CVT were excluded from this review.

Quality Assessment

The first, second, and third authors (MR, HM, and AR) independently assessed the included articles according to the 42-item Checklist for Artificial Intelligence in Medical Imaging (CLAIM) [28]. CLAIM is modeled after the Standards for Reporting of Diagnostic Accuracy Studies guideline. It addresses the application of AI in medical imaging, including classification, image reconstruction, text analysis, and workflow optimization [28]. Subsequently, the first, second, and third authors cross-checked each other's articles, and conflicts were resolved through group discussion.

Risk of Bias

Finally, both HM and AO independently assessed the risk of bias for each study across ROBINS-I's (risk of bias in nonrandomized studies - of interventions) 7 domains: confounding, selection of participants, classification of exposures, deviation from intended exposure, missing data, measurement of outcomes, and selection of reported results. Each domain was rated as low, moderate, serious, or critical based on the each domain's algorithm, with the most severe rating across all domains determining the overall assessment for each study. Any disagreements in the assessments were discussed until a consensus was reached, with one reviewer (DJ) ensuring consistent application of judgments. Additionally, we engaged a fifth-year medical student experienced in systematic reviews and various research projects to independently evaluate the risk of bias for all included studies using the same tool.

Statistical Extraction and Analysis

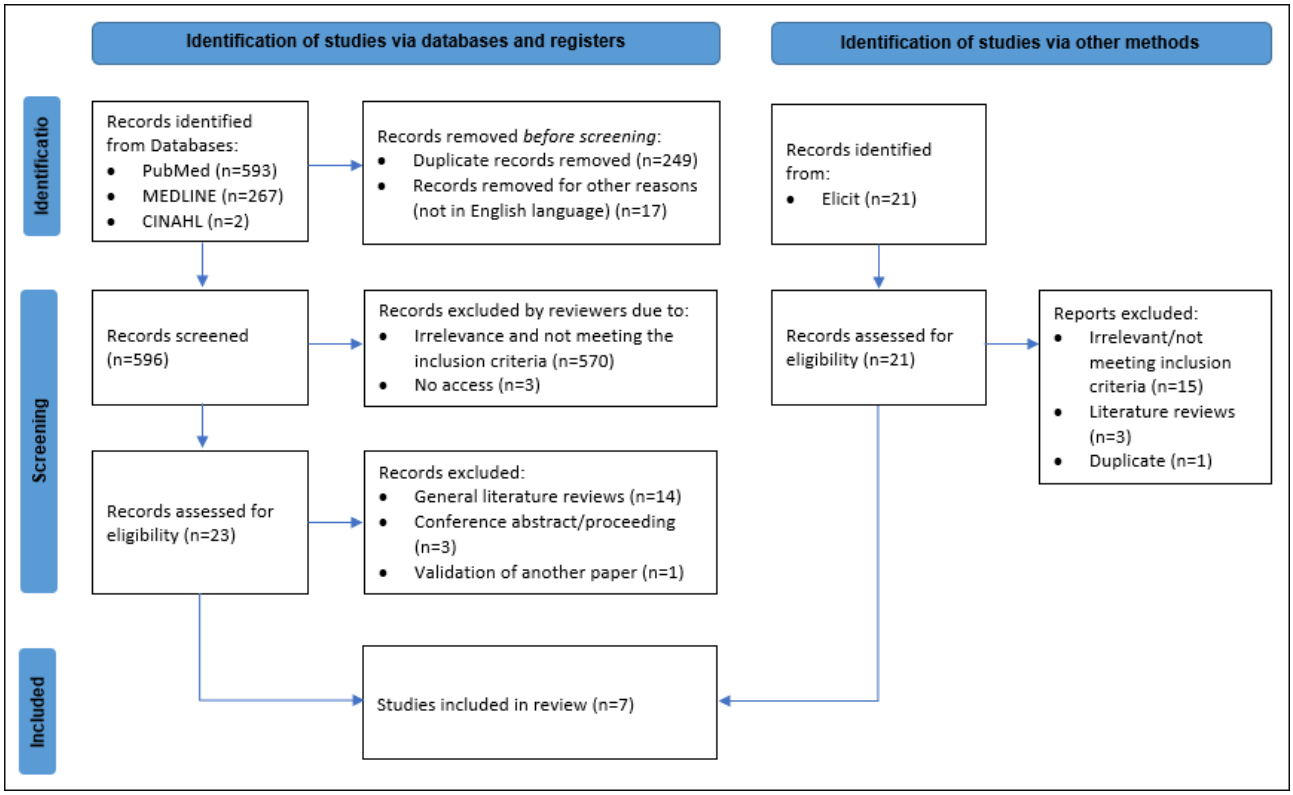
According to the CLAIM checklist, the first, second, and third authors (MR, HM, and AR) extracted data from the included studies. All discrepancies were resolved after a discussion, with HM acting as an arbitrator. Descriptive information about each study was recorded, including publication details (author, year, and country), sample size, cancer type, imaging technique, AI model, outcomes, and limitations. AO performed analysis, and figures were generated using RStudio (version 2023.06.0; Posit PBC).

Results

Study Selection

A total of 883 articles were identified in the database search, comprising 593 articles from PubMed, 267 from MEDLINE, 2 from CINAHL, and 21 from Elicit. After eliminating duplicate titles and articles in non-English languages, 617 articles remained. Then, the title and abstract of the 617 articles were screened independently by the first and second authors (MR and HM), and 44 remained. The authors reviewed full texts and 7 articles met the inclusion criteria (Figure 1).

Figure 1. PRISMA flow diagram of the systematic search of the databases for artificial intelligence in cardio-oncology imaging. PRISMA: Preferred Reporting Items for Reviews and Meta-Analyses.



Quality Assessment of Included Studies

The quality assessment used the 42-item CLAIM. The distribution and percentages of different sections and items of CLAIM compliance are depicted in Figures 2 and 3. These sections include title/abstract, introduction, methods, results, discussion, and other information. Each section is categorized into “No” and “Yes” groups, indicating whether it is reported in the selected articles.

Figure 2. CLAIM sections compliance. CLAIM: Checklist for Artificial Intelligence in Medical Imaging.

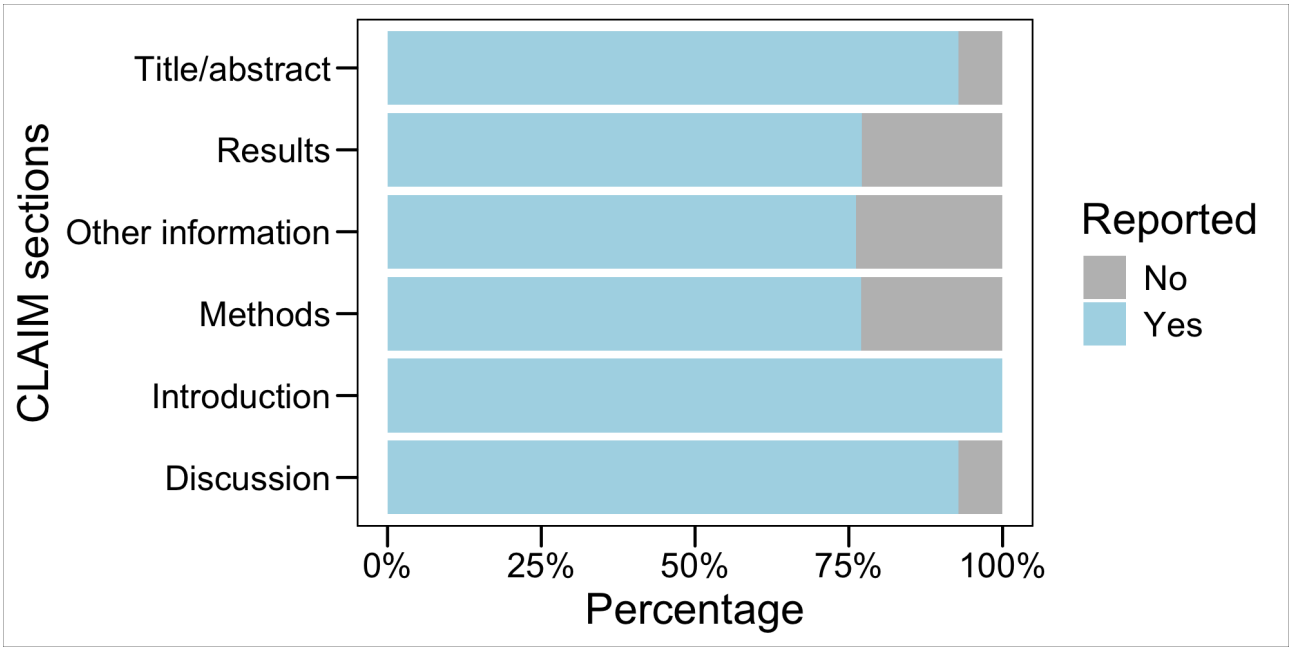


Figure 3. CLAIM items compliance. CLAIM: Checklist for Artificial Intelligence in Medical Imaging.

Based on the data, the title/abstract section was compliant in 93% of the articles (1 article was compliant with the title but not the abstract, which was considered as half compliant). An introduction section was included in all 7 articles, representing 100% compliance. Methods had 77% compliance, results represented 77% compliance, and there was a discussion in 93% of the articles, while other information was 76% compliant. Items 5 and 6 of the checklist—specific to the study methodology and design—were met as follows: 5 studies were conducted prospectively, while the remaining 2 were conducted retrospectively. Moreover, 4 studies were reported as feasibility studies, 2 were exploratory studies, and 1 was a model creation study. Finally, items 10 and 27 of the CLAIM criteria did not apply to the 7 studies.

Characteristics of the Included Studies

In total, 7 studies conducted between 2018 and 2023 were included, with 5 from the United States (Kar et al [29-31], Zhang et al [32], Edalati et al [33]), 1 from China (Shen et al [34]), and 1 from Taiwan (Chang et al [35]). Of these, 6 studies involved patients with breast cancer with additional cancers (eg, sarcoma, lymphoma, leukemia) in some cohorts. Imaging modalities included MRI (4 studies: 3 displacement encoding with stimulated echoes [DENSE] MRI, 1 CMR), echocardiography (n=2), and nongated, noncontrast chest computed tomography (CT) (n=1). AI approaches varied: 4 studies (57%) used convolutional neural networks (CNNs), 1 (14%) used ML, and 6 (86%) implemented image segmentation. Table 1 provides demographic and descriptive data and Table 2 provides details of the AI components of the included studies.

Table . Summary of the studies included in this review: demographic and descriptive data.

Author, year, country	Study design	Sample size	Gender	Age (years)	Treatment	Cancer type	Imaging technique
Shen et al, 2023, China [34]	Retrospective, multicenter	N=1468	Male: n=785, female: n=683	>60: n=617, <60: n=851	Anthracycline	Diffuse large B-cell lymphoma	Nongated and noncontrast chest computed tomography for coronary artery calcium scoring echocardiography for cancer therapy–related cardiac dysfunction and major adverse cardiovascular event
Chang et al, 2022, Taiwan [35]	Prospective, single center, with 3 years of follow-up	N=211	n=211	55.8 (SD 10.28)	Anthracycline, trastuzumab	Breast cancer; stage I: n=50; stage II: n=101; stage III: n=52; stage IV: n=8	Echocardiography
Kar et al, 2023, United States [31]	Prospective	N=32	Female n=32	Baseline: 59.4 (SD 9.7); 3 months: 59.6 (SD 9.7); 6 months: 59.6 (SD 9.7)	Anthracycline, trastuzumab, radiotherapy	Breast cancer	• DENSE[a] • Magnetic resonance imaging • Transesophageal echocardiogram
Kar et al, 2022, United States [30]	Prospective	IG[b]: n=30; CG[c]: n=30	IG female: n=30; CG female: n=30	IG: 54 (SD 9), CG: 50 (SD 13)	Anthracycline, trastuzumab	Breast cancer	• DENSE • Magnetic resonance imaging
Kar et al, 2021, United States [29]	Prospective	N=42	Female: n=42	55.5 (SD 8.6)	Anthracycline, trastuzumab	Breast cancer	• DENSE • Magnetic resonance imaging
Zhang et al, 2018, United States [32]	Retrospective, 10 years	Hypertrophy cardiomyopathy: n=260; echo: n=14,035; amyloidosis: n=81; CIC[d]: n=152; pulmonary arterial hypertension: n=27	CIC female: n=152	CIC: 55	Trastuzumab, pertuzumab	Breast cancer	Echocardiography
Edalati et al, 2022, United States [33]	Prospective	CG: n=10, IG: n=10	CG male: n=5; CG female: n=5: IG male: n=5; IG female: n=5	CG: 52.6 (SD 21.2); IG: 47.6 (SD 13.6)	Not applicable	Breast cancer: n=4, sarcoma: n=3, lymphoma: n=1, leukemia: n=1, myeloma: n=1	Cardiac magnetic resonance imaging

[a]DENSE: displacement encoding with stimulated echoes.

[b]IG: intervention group.

[c]CG: control group.

[d]CIC: chemotherapy-induced cardiotoxicity.

Table . Summary of the studies included in this review: details of the artificial intelligence components in the included studies.

Author, year, country	Artificial intelligence solution	Main outcomes	Limitation
Shen et al, 2023, China [34]	Artificial intelligence coronary artery calcium scoring: • Deep learning algorithm • Image segmentation • Bound the range of the heart area • Detect and segment the calcified lesions in coronary arteries • Calculate coronary artery calcium score	• Cancer therapy–related cardiac dysfunction • Major adverse cardiovascular events	• A larger sample is needed to validate the model's accuracy • The study was limited to Chinese patients
Chang et al, 2022, Taiwan [35]	Machine learning: • Multilayer perceptron • A tree-based estimator was used to compute essential features, and 15 features were included in our multilayer perceptron model based on experts' judgments.	• Cancer therapy–related cardiac dysfunction • Symptomatic heart failure with reduced ejection fraction	• A relatively small number of included patients
Kar et al, 2023, United States [31]	Validated advanced artificial intelligence methodologies (DeepLabV3+) with fully convolutional networks: • Segmenting the DENSE[a] magnitude images for chamber quantification • Segmenting the DENSE phase images for phase-unwrapping and 3D strain analysis	• Global longitudinal strain • Cancer therapy–related cardiac dysfunction • Adverse cardiac events	• Single-center study without external validation • No integration between cancer therapy–related cardiac dysfunction risk analysis by combining circulating troponin levels with global longitudinal strain measurements for a practical bivariable prognostic approach
Kar et al, 2022, United States [30]	An FCN[b]-based solution adapted from the DeepLabV3+ network: • Phase-unwrapping FCN. • Compared with conventional unwrapping techniques, validation via phantom setup with known displacements and 3D strain analysis in healthy patients. • Left ventricular volume was estimated with previously validated DeepLabV3+. • Computation of 3D myocardial strains with the meshfree Radial Point Interpolation Method	Global longitudinal strain	• Comparing the performance of phase unwrapping with DeepLabV3+ to another FCN such as PhaseNet. • The relationship between the wrapped phase and wrap count can be leveraged with more arbitrary shapes rather than round and ellipsoidal shapes only.
Kar et al, 2021, United States [29]	An automated left ventricular chamber quantification tool (deep learning): • DCNN[c] and DeepLabV3+ with ResNet-50 backbone • Some layers of the original ResNet-50 to tailor DCNN for cardiac image segmentation • DENSE-based results were validated by corresponding steady-state free precession data in the same patients who were trained using an identical DeepLabV3+ DCNN. • Chamber quantification and strain analysis were done after the image-based reconstruction of the full 3D left ventricle.	• Left ventricular end diastolic diameter • Left ventricular ejection fraction • Myocardial strains analyzed with the radial point interpolation method	• Backbone networks such as Xception, Inception, ResNet-101, U Net, and others were not tested for left ventricular segmentation.

Author, year, country	Artificial intelligence solution	Main outcomes	Limitation
Zhang et al, 2018, United States [32]	A computer vision pipeline for automated 2D echocardiogram interpretation: • Convolutional neural network for view classification • Image segmentation • Measurements of cardiac structure and function disease detection	• Automated identification of 23 viewpoints segmentation of cardiac chambers across 5 common views • Quantification of structure and function • Detection of hypertrophic cardiomyopathy • Detection of cardiac amyloid • Detection of pulmonary arterial hypertension	• Problems with segmentation • Forced normalization to the lower strain value because of the lack of electrocardiogram information, which can result in biases in measurements, estimate of strain • Lack of distinguished diagnosis of hypertrophy cardiomyopathy, amyloid, or any hypertrophic disease • Lack of comparison of deep learning models to onesbuilt using hand-selected features (left atrial mass or septal thickness)
Edalati et al, 2022, United States [33]	EasyScan: • Otsu method: segment heart region • Trained regression network: distance map calculation	• Scan time difference • Accuracy of cardiac plane prescriptions • Signal to noise ratio • Contrast to noise ratio • Overall image quality (sharpness and magnetic resonance image degradation) • Ejection fraction • Absolute wall thickening	N/A[d]

[a]DENSE: displacement encoding with stimulated echoes.

[b]FCN: fully convolutional network.

[c]DCNN: deep convolutional neural network.

[d]N/A: not applicable.

The included studies revealed significant clinical heterogeneity across the studies. Study designs ranged from retrospective (eg, Shen et al [34]: n=1468; Zhang et al [32]: n=260) to prospective (eg, Chang et al [35]: n=211; Kar et al [29-31]: n=32 - 42), impacting sample size and follow-up duration (eg, 3 years in Chang et al [35] vs 10 years in Zhang et al [32]). Imaging modalities differed in application: echocardiography (Chang et al [35], Zhang et al [32]) assessed LVEF and GLS; DENSE MRI (Kar et al [29-31]) focused on strain analysis; CT (Shen et al [34]) targeted coronary artery calcium scoring (CACS); and CMR (Edalati et al [33]) evaluated image quality and efficiency. AI techniques showed varied sophistication—CNNs (eg, DeepLabV3+ in Kar et al [29-31], CNN pipeline in Zhang et al [32]) and deep learning (Shen et al [34]) enhanced segmentation and classification, while ML with multilayer perceptron (Chang et al [35]) predicted outcomes like heart failure with reduced ejection fraction.

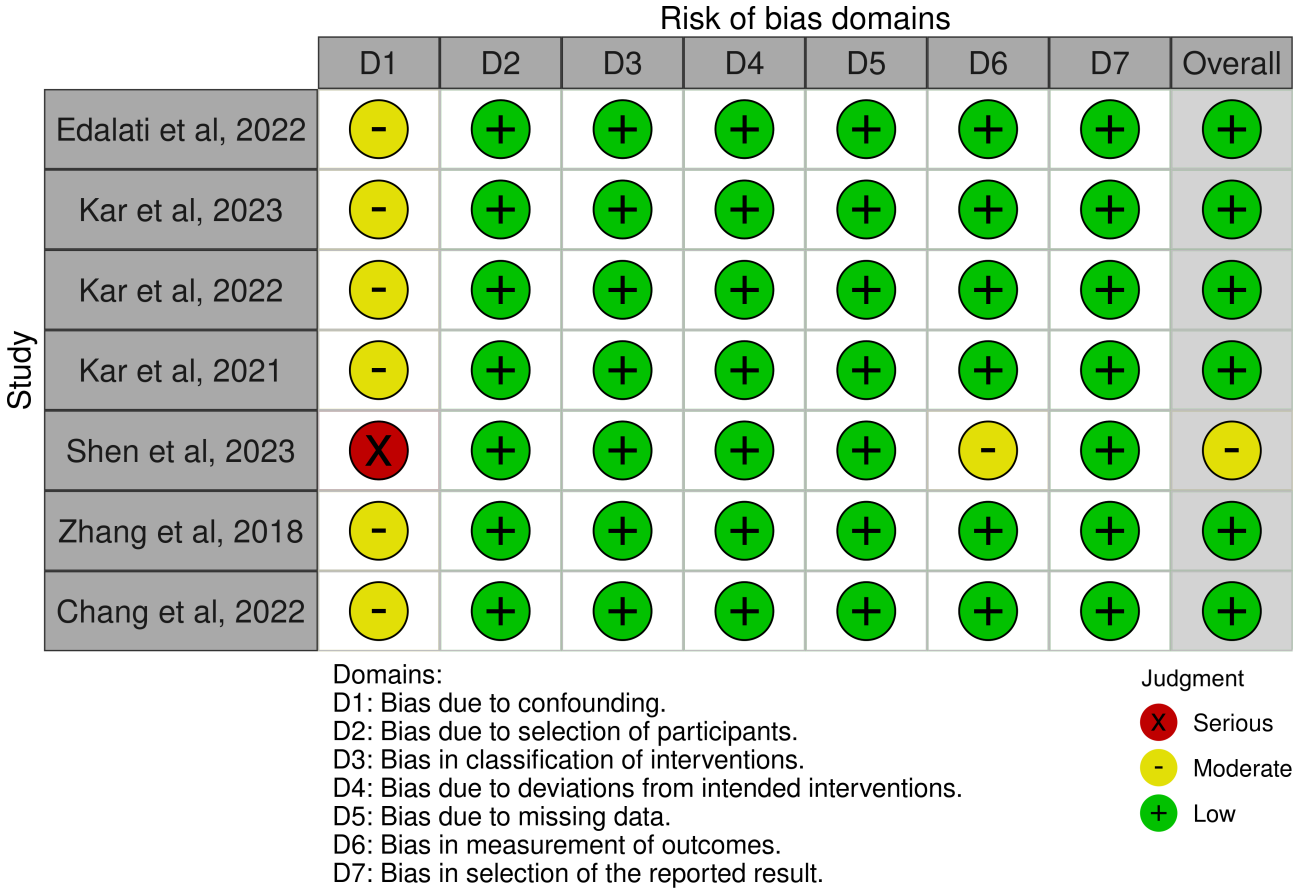
Outcomes centered on CTR-CVT, with cancer therapy–related cardiac dysfunction assessed in 5 studies (Shen et al [34], Chang et al [35], Kar et al [31], Zhang et al [32], Edalati et al [33]), GLS in 3 (Kar et al [29-31]), and LVEF in 3 (Kar et al [29], Edalati et al [33], Zhang et al [32]). Shen et al [34] uniquely linked CACS to major adverse cardiovascular events (MACE). At the same time, Edalati et al [33] emphasized scan time and signal-to-noise ratio. AI improved detection accuracy (eg, automated CACS in Shen et al [34], GLS computation in Kar et al [30,31]) and efficiency (eg, EasyScan in Edalati et al [33]) compared to manual methods. However, direct comparisons across studies were limited by outcome diversity.

Common limitations included small sample sizes (eg, Chang et al [35], Edalati et al [33]), single-center designs (eg, Kar et al [31], Chang et al [35]), and lack of external validation (eg, Kar et al [31]). Geographic restriction (Shen et al [34], Chinese patients) and technical challenges (eg, segmentation issues in Zhang et al [32]) further constrained generalizability.

Risk of Bias Assessment

The risk of bias assessment began with general considerations for all studies, which included establishing a minimal set of confounders identified by the reviewers as likely to introduce bias in the observed associations. Next, each study was described individually within the framework of an ideal target trial. The consensus results from the evaluations of the 7 nonrandomized studies are depicted in the “traffic light” plot shown in Figure 4.

Figure 4. Traffic light plot of risk of bias assessment.



Discussion

Summary of Included Studies

In 2018, Zhang et al [32] published their work on automating echocardiographic cardiac images using 14,035 echocardiograms collected retrospectively spanning 10 years. Their study included 152 patients diagnosed with CTR-CVT and other patients with other heart conditions such as hypertrophic cardiomyopathy (n=260), amyloidosis (n=81), and pulmonary arterial hypertension [32]. Zhang developed a model for view classification in just a few steps. First, they taught the machine to recognize individual echocardiographic views, where models were trained using manual labels assigned to individual images. Then, they used deep learning architecture for view classification, designed to mimic how the visual system works [32]. This process refers to multiple layers of neurons, processing nodes tuned to recognize features within an image. Afterward, they trained a 13-layer CNN and assessed the accuracy using 5-fold cross-validation. Finally, they used t-distributed stochastic neighbor embedding (an algorithm for visualizing high-dimensional data) to cluster the output of the top layer to visualize the output of their view classification network [32]. By training the CNNs, Zhang could perform image segmentation to locate cardiac chambers that derived cardiac structure and function measurements to develop disease classification models [32]. Zhang's approach is intended to enable data mining and knowledge extraction from the enormous number of archived echocardiograms, which will have a significant clinical impact by introducing relatively low-cost quantitative metrics into clinical practice and enabling causal insights that require systematic longitudinal tracking of patients [32]. The study results favored using AI-automated measurements over manual measurements across 11 internal consistency metrics. One of these is the correlation between left atrial and left ventricular volumes. This work is argued to have laid the basis for using automated interpretation to support serial patient tracking. Limitations to the study are the length of the analysis period and room for bias. Moreover, the study did not include the number of males or females involved, which may affect the results.

Using a different imaging modality, Edalati et al [33] developed EasyScan, which is automated cardiac planning, by developing, training, and validating 2 deep neural networks on preacquired cardiac MRI datasets (also known as cardiovascular magnetic resonance). EasyScan is implemented with the CMR scanner for automatic slice planning and shimming. The trial included 10 healthy individuals (5 males and 5 females) and 10 cardio-oncology patients (5 males and 5 females) undergoing 2 identical CMR protocols (manual cardiac planning versus AI-based EasyScan) to assess the time difference and accuracy of the cardiac plane. Moreover, Cine images were obtained for the study participants with standard cardiac volume shim and AI-shim to assess the signal-to-noise ratio, contrast-to-noise ratio, overall IQ (sharpness and magnetic resonance image degradation), LVEF, and absolute wall thickening [33]. EasyScan demonstrated accelerated cardiac exams compared to standard manual cardiac planning and achieved an improved and more uniform B0 magnetic field homogeneity using the

AI-shim technique compared to volume shimming [33]. Eldalati argued that his results suggest many potential positive outcomes of implementing AI, including a more straightforward and faster workflow chain by minimizing technique complexity. However, a significant limitation of this study is the cohort size, as it is considered small compared to other papers in this field.

Kar et al [29-31] used AI, deep learning, segmentation, and fully convolutional networks (FCN) on the DENSE MRI sequence imaging modality in their 3 studies. In the study published in 2021, Kar et al [29] investigated the automation of measuring left-ventricular strain with a quantification tool via segmentation with a supervised deep convolutional neural network (DCNN) before strain analysis with DENSE images [29]. Kar and her team were able to introduce a novel and automated DCNN architecture–based chamber quantification methodology for detecting the extent of left-ventricular myocardium in single-scan DENSE MRI for patients with breast cancer susceptible to cardiotoxicity. Kar et al identified accurate segmentation, chamber quantification, and subsequent strain analysis in the myocardium as the main critical requirements for engineering and developing this solution. After validation, Kar et al emphasized that their DCNN-based segmentation can provide accurate estimates of the left-ventricular chamber quantification required in strain analysis.

Kar et al argued that their model can perform fast and inexpensive automated measurements of cardiac strain as the model can detect altered material properties. However, the thresholds that define cardiac dysfunction caused by cancer therapy are still an area that needs to be further studied [29].

In 2022, Kar and her team continued their work using DENSE in developing another direct MRI-based, FCN-based, deep-learning semantic segmentation approach for computing GLS for patients with breast cancer [30]. This time, they computed myocardial strains directly from the unwrapped phases with the radial point interpolation method. They compared the results of 30 patients with 30 healthy individuals, and the difference in GLS results between the participants demonstrated that the FCN is sensitive to unwrapping left ventricular data in a heterogeneous cohort [30]. Moving forward with their work on GLS computation, Kar and her team investigated early alterations in prognostic factors such as GLS with standard Cox proportional hazards regression for estimating the risk of CTR-CVT incidents in patients with breast cancer undergoing cancer treatment using their previously developed AI-FCN.

Moving forward, Kar and her team carried out a trial using their tool to estimate the risk of developing cardiotoxicity in patients with breast cancer using data from their previous studies [31]. The trial proved their hypothesis that GLS computation can be used for early detection of CTR-CVT as an independent prognostic method of left ventricular dysfunction [31]. The advantage Kar et al had in their studies was that they were able to validate their solution internally within their center. However, their trials did not come without limitations. The solutions were not validated externally with other centers, and there was a greater sample for better accuracy measures [29-31]. In addition, the phase unwrapping approach for GLS measures was not compared to phase wrapping with another FCN, such as PhaseNet, which is considered a significant limitation in their conclusion [30].

Concurrently, in 2022, Chang et al [35] conducted another single-center prospective study and included a larger sample size of 211 patients diagnosed with breast cancer at different stages [35]. Chang et al [35] aimed to establish an AI-based predictive model for CTR-CVT using a cardio-oncology program. They prospectively collected clinical information and echocardiographic images from patients with breast cancer over 1 year. In their study, 2 echo technicians performed an echocardiogram independently to measure the LVEF at baseline, 3 months, 6 months, and 1 year after patients received their treatment. A cardiologist with a validated reliability and reproducibility interpreted the images. Moving forward with the AI solution, data were validated using a data mart for further analysis. Then, we compared the accuracy, precision, sensitivity, specificity, and area under the curve of the random forest, logistic regression, support vector clustering, LightGBM, K-nearest neighbour, and multilayer perceptron models. This process yielded the best accuracy in predicting CTR-CVT [35]. Moreover, the multilayer perceptron showed the best results in predicting heart failure with a reduced ejection fraction as an early sign of myocardial dysfunction after the occurrence of CTR-CVT [35].

Shen et al [34] conducted the most recent study in China in 2023. The study aimed to evaluate whether the pretreatment CACS can stratify the risk of CTR-CVT and MACEs in patients with diffuse large B-cell lymphoma (DLBCL). They retrospectively collected nongated and noncontrast chest CT scans of 1468 patients from 4 health centers in China, then used a deep-learning–based algorithm software (CACScoreDoc) to calculate the automatic CACS. CACScoreDoc automatically calculated the CACS and transmitted the results to the doctors after uploading the CT images to the software. The study showed that automating CACS derived from chest CT scans done before receiving the treatment is potentially helpful in identifying patients at risk of developing CTR-CVT and MACEs in patients with DLBCL receiving anthracycline chemotherapy, which can guide clinicians to implement cardiovascular protective strategies and minimize CTR-CVT in DLBCL patients [34].

Although cardiovascular events that are caused by cancer medications vary in prevalence from one type of cancer and its medication to another, they are still the second most common cause of mortality in cancer survivors. To accurately predict the risk of cardiotoxicity among individuals receiving cancer treatment is still a great challenge in the cardio-oncology field due to high cost, limited access to care, and inadequate compliance with screening protocols. Therefore, noninvasive, low-cost, accessible, innovative approaches to predict high-risk individuals and detect cardiotoxicity early among patients with cancer are critically needed to enable optimal screening, early diagnosis, and timely interventions [36].

Current Versus Future AI Practice

The current tool used to investigate signs of cardiotoxicity is medical imaging, with the 2 most used imaging modalities for this purpose being the echocardiograph and CMR. However, although these modalities have helped the medical field to

achieve significant improvement in prognosis in this area, some drawbacks hold them back from being optimal methods of investigation. The echocardiograph is entirely user-dependent in image reproducibility and results interpretation, leaving ample room for bias and inconsistency. On the other hand, the CMR is not always available due to its high cost. Therefore, more robust, cost-effective methods and imaging protocols are needed in this cardio-oncology area to optimize patient care [36].

Many health care disciplines have moved toward advancing artificial intelligence and developing better ML algorithms as they continue to improve patient care quality significantly. With the availability of enormous volumes of patient data and accessibility of proper hardware, AI and ML can accelerate the pace of change in health care. These technologies can sift through the data and analyze it much faster than humans, leading to increased efficiency. ML is used to predict clinical risk factors by feeding it with an enormous volume of data retrieved from patient medical records or national datasets and registries or detect cardiotoxicity via deep learning of patients' cardiovascular images. In this review, the authors focused their assessment on using AI and ML in cardiovascular imaging to increase the diagnostic strength and accuracy in detecting CTR-CVT.

This review included 7 studies that intended to assess the implementation of AI in cardiovascular imaging among patients with cancer. These studies examine the use of AI on MRI, echocardiogram, and CT imaging modalities with different AI technologies such as ML, CNNs, and image segmentation.

The future of imaging AI in cardio-oncology holds substantial promise. This convergence of cutting-edge technologies, encompassing molecular imaging, wearable devices, multiomics data, and predictive modeling, is poised to transform cardiotoxicity management in patients with cancer. These advancements enable early detection and personalized risk assessment and promise targeted interventions, ultimately enhancing patient outcomes and survivorship. This future trajectory in imaging AI aligns with the significant advancements witnessed from ML to deep learning in AI, revolutionizing robotics and autonomous systems' capabilities and enabling them to perceive, learn, and adapt with increased efficiency and accuracy in complex environments. These models, leveraging AI algorithms trained on diverse patient cohorts and multimodal imaging data, could assist clinicians in formulating proactive strategies for long-term cardiac care in cancer survivors, thereby enhancing overall cardiovascular health and quality of life.

Challenges of AI in Health Care

As promising as AI and ML sound to the advancement of imaging in health care and the prediction of the risk of developing cardiotoxicity among patients receiving cancer treatment specifically, there are methodological and practical limitations preventing these technologies from reaching their full potential. The evidence base needs more prospective validation of the technology and current workflow, including evidence on the length of analysis required for validation and the interoperator and interobserver variability to eliminate manufactured variations that limit reproducibility [23]. Moreover, their usefulness in health care depends on incorporating the AI tool in clinical decision-making as part of the clinical practice routine, and that concern needs further investigation [37]. Another inadequacy of AI applications in health care is the systematic biases affecting patient demographics, such as gender imbalance [38]. It is worth mentioning that AI requires training on all kinds of populations with different demographics to guarantee equal performance from one population to another. It is recommended that multiple massive datasets be combined either retrospectively or prospectively to improve the generalizability of the ML process and the training of AI models, which was not achieved by all the included studies in this review [39].

Review Limitations

The first limitation we had while conducting this review was the limited published evidence in the literature about the application of imaging AI in cardio-oncology to predict CTR-CVT. Therefore, we could not specify the cancer type or treatment under investigation. Second, even though there is significant literature on AI and imaging with different modalities, when we narrowed it down to our criteria, which was patients with cancer who are undergoing cardiotoxicity assessment, the literature search resulted in 3 different imaging modalities rather than studying AI with one specific imaging technique at a time. This resulted in different outcomes that prevented us from proceeding with a meta-analysis.

The use of AI in the medical field is a relatively new research area. This review could be used to stimulate further research. It can be used as groundwork for lab work to improve AI models or inspire new ones. In addition, this review highlights the positive outcomes of different studies in this area and their limitations. It may encourage experts to improve the AI and ML models and eventually implement them into medical imaging, possibly leading to the advancement of the field. However, given this field's rapidly evolving nature, additional studies may have been published since the initial search process for this paper.

Conclusions

In conclusion, this systematic review highlights the promising potential of AI in enhancing cardio-oncology imaging for predicting cardiotoxicity in patients with cancer. Through analyzing 7 studies conducted between 2018 and 2023, it became evident that AI methodologies, including ML and deep learning, can significantly improve the accuracy and efficiency of cardiotoxicity assessments across various imaging modalities, such as echocardiography and CMR.

The review underscores that AI-driven tools have demonstrated improved clinical outcomes by enabling earlier detection of cardiovascular complications associated with cancer therapies. However, while the findings are encouraging, the limited number of studies and their varying methodologies indicate a need for further research. This includes conducting larger, multicenter trials to validate AI applications in diverse patient populations and refine these technologies for routine clinical use.

In light of these insights, collaboration among data scientists, health care professionals, and researchers is essential to advancing AI's integration in cardio-oncology. This collaboration will pave the way for personalized medicine approaches, ultimately enhancing patient care and improving the quality of life for cancer survivors at risk of cardiotoxicity.

Acknowledgments

The authors would like to thank Ms Amnah Hazimi, a fifth-year medical student, for her time and effort in reviewing the articles for the risk of bias assessment. No funding was acquired for this review.

Data Availability

Data are available upon request.

Authors' Contributions

HM was responsible for the conception of the work, developing the search strategy, conducting a systematic database search, writing the manuscript, and supervising all coauthors throughout the project. MR contributed by performing a literature review, conducting a parallel systematic database search, writing sections of the manuscript, and conducting the CLAIM assessment. AR also engaged in the literature review and participated in the CLAIM assessment. As the toxicologist, RE critically reviewed the materials and contributed to manuscript revisions and edits. TS focused on reviewing material related to artificial intelligence and medicine, in addition to contributing to manuscript review and edits. AO, the biostatistician, conducted the analysis for the CLAIM assessment and created the traffic plot for risk of bias assessment. Medical students MO and RAH assisted with the literature review. Finally, DJ supported the development of the search strategy, contributed to the systematic database search, and contributed to reviewing and editing the manuscript. The only artificial intelligence tool used in writing this paper was Grammarly for proper language editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions.
[DOCX File, 58 KB - cancer_v11i1e63964_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File, 40 KB - cancer_v11i1e63964_app2.docx]

Checklist 1
PRISMA checklist.
[PDF File, 274 KB - cancer_v11i1e63964_app3.pdf]

References

1. Worldwide cancer data. World Cancer Research Fund International. URL: https://www.wcrf.org/preventing-cancer/cancer-statistics/worldwide-cancer-data/ [accessed 2023-07-17]
2. Breast cancer facts and statistics. BreastCancer.org. URL: https://www.breastcancer.org/facts-statistics [accessed 2023-07-10]
3. Saudi Arabia. International Agency for Research on Cancer. URL: https://gco.iarc.fr/today/en/fact-sheets-populations#countries [accessed 2023-07-10]
4. Sadler D, Okwuosa T, Teske AJ, et al. Cardio oncology: digital innovations, precision medicine and health equity. Front Cardiovasc Med 2022;9:951551. [doi: 10.3389/fcvm.2022.951551] [Medline: 36407451]
5. Koutsoukis A, Ntalianis A, Repasos E, Kastritis E, Dimopoulos MA, Paraskevaidis I. Cardio-oncology: a focus on cardiotoxicity. Eur Cardiol 2018 Aug;13(1):64-69. [doi: 10.15420/ecr.2017:17:2] [Medline: 30310475]
6. Lyon AR, López-Fernández T, Couch LS, et al. 2022 ESC Guidelines on cardio-oncology developed in collaboration with the European Hematology Association (EHA), the European Society for Therapeutic Radiology and Oncology (ESTRO) and the International Cardio-Oncology Society (IC-OS). Eur Heart J 2022 Nov 1;43(41):4229-4361. [doi: 10.1093/eurheartj/ehac244] [Medline: 36017568]
7. American Society of Clinical Oncology. Survivorship: during and after treatment. URL: https://www.cancer.org/cancer/survivorship.html [accessed 2025-04-16]

8. Survival rates for breast cancer. Cancer.org. URL: https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html [accessed 2025-04-19]
9. Cardiotoxicity: cancer treatment and the heart. Cleveland Clinic. URL: https://my.clevelandclinic.org/health/diseases/16858-chemotherapy--the-heart-cardiotoxicity [accessed 2023-07-18]
10. Florescu DR, Nistor DE. Therapy-induced cardiotoxicity in breast cancer patients: a well-known yet unresolved problem. Discoveries (Craiova) 2019 Mar 31;7(1):e89. [doi: 10.15190/d.2019.2] [Medline: 32309607]
11. Bikiewicz A, Banach M, von Haehling S, Maciejewski M, Bielecka-Dabrowa A. Adjuvant breast cancer treatments cardiotoxicity and modern methods of detection and prevention of cardiac complications. ESC Heart Fail 2021 Aug;8(4):2397-2418. [doi: 10.1002/ehf2.13365] [Medline: 33955207]
12. Gernaat SAM, Ho PJ, Rijnberg N, et al. Risk of death from cardiovascular disease following breast cancer: a systematic review. Breast Cancer Res Treat 2017 Aug;164(3):537-555. [doi: 10.1007/s10549-017-4282-9] [Medline: 28503723]
13. Targeted therapy. American Cancer Society. URL: https://www.cancer.org/cancer/managing-cancer/treatment-types/targeted-therapy.html [accessed 2023-07-17]
14. Boekhout AH, Beijnen JH, Schellens JHM. Trastuzumab. Oncologist 2011;16(6):800-810. [doi: 10.1634/theoncologist.2010-0035] [Medline: 21632460]
15. Hahn VS, Zhang KW, Sun L, Narayan V, Lenihan DJ, Ky B. Heart failure with targeted cancer therapies: mechanisms and cardioprotection. Circ Res 2021 May 14;128(10):1576-1593. [doi: 10.1161/CIRCRESAHA.121.318223] [Medline: 33983833]
16. Venkatesh P, Kasi A. Anthracyclines. StatPearls.: Treasure Island URL: https://www.ncbi.nlm.nih.gov/books/NBK538187/ [accessed 2025-04-19]
17. Anthracycline. National Cancer Institute. URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/anthracycline [accessed 2023-07-11]
18. Anthracycline chemo regimen seems better than regimen without anthracycline for early-stage, HER2-negative, high-risk disease. BreastCancer.org. URL: https://www.breastcancer.org/research-news/chemo-w-anthracycline-better-for-some [accessed 2023-07-11]
19. Braybrooke J, Bradley R, Gray R. Anthracycline-containing and taxane-containing chemotherapy for early-stage operable breast cancer: a patient-level meta-analysis of 100 000 women from 86 randomised trials. Lancet 2023 Apr 15;401(10384):1277-1292. [doi: 10.1016/S0140-6736(23)00285-4] [Medline: 37061269]
20. Awadalla M, Hassan MZO, Alvi RM, Neilan TG. Advanced imaging modalities to detect cardiotoxicity. Curr Probl Cancer 2018 Jul;42(4):386-396. [doi: 10.1016/j.currproblcancer.2018.05.005] [Medline: 30297038]
21. Armenian SH, Lacchetti C, Barac A, et al. Prevention and monitoring of cardiac dysfunction in survivors of adult cancers: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol 2017 Mar 10;35(8):893-911. [doi: 10.1200/JCO.2016.70.5400] [Medline: 27918725]
22. Alkofide H, Alnaim L, Alorf N, Alessa W, Bawazeer G. Cardiotoxicity and cardiac monitoring among anthracycline-treated cancer patients: a retrospective cohort study. Cancer Manag Res 2021;13:5149-5159. [doi: 10.2147/CMAR.S313874] [Medline: 34234558]
23. Madan N, Lucas J, Akhter N, et al. Artificial intelligence and imaging: opportunities in cardio-oncology. Am Heart J Plus 2022 Mar;15:100126. [doi: 10.1016/j.ahjo.2022.100126] [Medline: 35693323]
24. Cannizzaro MT, Inserra MC, Passaniti G, et al. Role of advanced cardiovascular imaging in chemotherapy-induced cardiotoxicity. Heliyon 2023 Apr;9(4):e15226. [doi: 10.1016/j.heliyon.2023.e15226] [Medline: 37095987]
25. Navin M. Medical imaging being transformed with GAN: MRI to CT scan and many others. Medium. URL: https://medium.com/analytics-vidhya/medical-imaging-being-transformed-with-gan-mri-to-ct-scan-and-many-others-18a307ef528 [accessed 2023-09-17]
26. Rumsfeld JS, Mortazavi BJ, Krumholz HM. Artificial intelligence and cardiovascular care. In: Fuster and Hurst's The Heart, 15th edition: McGraw Hill; 2022. URL: https://accessmedicine.mhmedical.com/content.aspx?bookid=3134§ionid=264102768 [accessed 2025-04-19]
27. Yagi R, Goto S, MacRae CA, Deo RC. Expanded adaptation of an artificial intelligence model for predicting chemotherapy-induced cardiotoxicity using baseline electrocardiograms. Eur Heart J 2022 Oct 3;43(Supplement_2). [doi: 10.1093/eurheartj/ehac544.2577]
28. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020 Mar;2(2):e200029. [doi: 10.1148/ryai.2020200029] [Medline: 33937821]
29. Kar BJ, Cohen MV, McQuiston SP, Malozzi CM. A deep-learning semantic segmentation approach to fully automated MRI-based left-ventricular deformation analysis in cardiotoxicity. Magn Reson Imaging 2021 May;78:127-139. [doi: 10.1016/j.mri.2021.01.005] [Medline: 33571634]
30. Kar J, Cohen MV, McQuiston SA, Poorsala T, Malozzi CM. Direct left-ventricular global longitudinal strain (GLS) computation with a fully convolutional network. J Biomech 2022 Jan;130:110878. [doi: 10.1016/j.jbiomech.2021.110878] [Medline: 34871894]
31. Kar J, Cohen MV, McQuiston SA, Malozzi CM. Can global longitudinal strain (GLS) with magnetic resonance prognosticate early cancer therapy-related cardiac dysfunction (CTRCD) in breast cancer patients, a prospective study? Magn Reson Imaging 2023 Apr;97:68-81. [doi: 10.1016/j.mri.2022.12.015] [Medline: 36581216]

32. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice. Circulation 2018 Oct 16;138(16):1623-1635. [doi: 10.1161/CIRCULATIONAHA.118.034338] [Medline: 30354459]
33. Edalati M, Zheng Y, Watkins MP, et al. Implementation and prospective clinical validation of AI-based planning and shimming techniques in cardiac MRI. Med Phys 2022 Jan;49(1):129-143. [doi: 10.1002/mp.15327] [Medline: 34748660]
34. Shen H, Lian Y, Yin J, et al. Cardiovascular risk stratification by automatic coronary artery calcium scoring on pretreatment chest computed tomography in diffuse large B-cell lymphoma receiving anthracycline-based chemotherapy: a multicenter study. Circ Cardiovasc Imaging 2023 Feb;16(2):e014829. [doi: 10.1161/CIRCIMAGING.122.014829] [Medline: 36748459]
35. Chang WT, Liu CF, Feng YH, et al. An artificial intelligence approach for predicting cardiotoxicity in breast cancer patients receiving anthracycline. Arch Toxicol 2022 Oct;96(10):2731-2737. [doi: 10.1007/s00204-022-03341-y] [Medline: 35876889]
36. Martinez DSL, Noseworthy PA, Akbilgic O, et al. Artificial intelligence opportunities in cardio-oncology: overview with spotlight on electrocardiography. Am Heart J Plus 2022 Mar;15:100129. [doi: 10.1016/j.ahjo.2022.100129] [Medline: 35721662]
37. Chen H, Ouyang D, Baykaner T, Jamal F, Cheng P, Rhee JW. Artificial intelligence applications in cardio-oncology: leveraging high dimensional cardiovascular data. Front Cardiovasc Med 2022;9:941148. [doi: 10.3389/fcvm.2022.941148] [Medline: 35958422]
38. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. Proc Natl Acad Sci U S A 2020 Jun 9;117(23):12592-12594. [doi: 10.1073/pnas.1919012117] [Medline: 32457147]
39. Alabed S, Maiter A, Salehi M, et al. Quality of reporting in AI cardiac MRI segmentation studies - a systematic review and recommendations for future studies. Front Cardiovasc Med 2022;9:956811. [doi: 10.3389/fcvm.2022.956811] [Medline: 35911553]

Abbreviations

AI: artificial intelligence
ASCO: American Society of Clinical Oncology
CACS: coronary artery calcium score
CLAIM: Checklist for Artificial Intelligence in Medical Imaging
CMR: cardiac magnetic resonance
CT: computed tomography
CTR-CVT: cancer therapy–related cardiovascular toxicity
CVD: cardiovascular disease
DCNN: deep convolutional neural network
DENSE: displacement encoding with stimulated echoes
DLBCL: diffuse large B-cell lymphoma
FCN: fully convolutional networks
GLS: global longitudinal strain
HER2: human epidermal growth factor receptor 2
LVEF: left ventricular ejection fraction
MACE: major adverse cardiovascular event
MeSH: Medical Subject Headings
ML: machine learning
MRI: magnetic resonance imaging
PRISMA: Preferred Reporting Items for Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
ROBINS-I: risk of bias in nonrandomized studies - of interventions

Edited by N Cahill; submitted 04.07.24; peer-reviewed by AA Alzaydi, MA Washmi, V Vinay; revised version received 02.03.25; accepted 03.03.25; published 09.05.25.

Please cite as:
Mushcab H, Al Ramis M, AlRujaib A, Eskandarani R, Sunbul T, AlOtaibi A, Obaidan M, Al Harbi R, Aljabri D
Application of Artificial Intelligence in Cardio-Oncology Imaging for Cancer Therapy–Related Cardiovascular Toxicity: Systematic Review
JMIR Cancer 2025;11:e63964
URL: https://cancer.jmir.org/2025/1/e63964
doi:10.2196/63964

Internet-Based Cognitive Behavioral Therapy Interventions for Caregivers of Patients With Cancer: Scoping Review

Chun Tong Shen[1*], BSN; Jian Shi[1*], PhD; Feng Xia Liu[2*], MSN; Xiao Meng Lu[3*], BSN

[1]Department of Oncology, The Fourth Hospital of Hebei Medical University, No. 12 Health Road, Shi Jiazhuang, China

[2]Department of Nursing, The Fourth Hospital of Hebei Medical University, Shi Jiazhuang, China

[3]Department of Radiotherapy, The Fourth Hospital of Hebei Medical University, Shi Jiazhuang, China

[*]all authors contributed equally

Corresponding Author:
Jian Shi, PhD
Department of Oncology, The Fourth Hospital of Hebei Medical University, No. 12 Health Road, Shi Jiazhuang, China

Abstract

Background: Cancer imposes significant physical and emotional distress not only on patients, but also on their caregivers. In recent years, there has been a growing focus on the mental and physical well-being of caregivers. Among various psychological interventions, cognitive behavioral therapy (CBT) is widely recognized as one of the most effective approaches. However, traditional CBT is often limited by time and geographical constraints, resulting in delayed or inefficient support for caregivers. Internet-based cognitive behavioral therapy (ICBT) presents a valuable alternative for alleviating the caregiving burden and the negative emotions experienced by caregivers.

Objectives: This study aimed to provide a scoping review of ICBT interventions for caregivers of patients with cancer, examining intervention content, outcome measures, and effectiveness and to offer insights and references for the development and clinical applications of ICBT programs tailored to caregivers of patients with cancer in China.

Methods: Relevant literature was systematically searched in PubMed, Web of Science, Cochrane Library, CINAHL, Embase, China National Knowledge Infrastructure (CNKI), Wanfang Data, and VIP Chinese Journal Database. The search timeframe was from database inception to June 6, 2024. Inclusion criteria encompassed intervention studies that implemented cognitive behavioral therapy for caregivers of patients with cancer via the internet, WeChat (Tencent), or mobile electronic devices. This category includes both randomized and nonrandomized controlled trials.

Results: A total of 12 studies met the criteria and were included in the review. The intervention content included the following components: treatment initiation and brief introduction (5/12, 41%), cognitive education and restructuring (7/12, 58%), emotional expression and coping (6/12, 50%), cognitive restructuring and reinforcement (4/12, 33%), behavioral training and activation (9/12, 75%), problem-solving techniques (4/12, 33%), communication (5/12, 41%), and completion of treatment with follow-up consolidation (3/12, 25%). The intervention duration typically ranged from 6 to 8 weeks. Outcome indicators encompassed feasibility and acceptability, anxiety, depression, caregiver burden, and quality of life. ICBT demonstrated positive effects for caregivers of patients with cancer. Most intervention programs were feasible and acceptable, with 2 out of 5 feasibility studies reporting recruitment rates below 50%. Attrition rates across studies ranged from 3% to 16%, and caregivers expressed satisfaction with the information, quality, and skills provided. ICBT exhibits a moderate effect in diminishing negative emotions among caregivers and alleviating caregiver stress. However, its impact on improving quality of life is not statistically significant, underscoring the need for long-term follow-up.

Conclusions: The implementation of ICBT for caregivers of patients with cancer has demonstrated beneficial outcomes, attributed to its practicality and flexibility, which contribute to its greater acceptance among caregivers. Nevertheless, there is significant heterogeneity in intervention format, duration, and outcome indicators. It is necessary to develop optimal intervention strategies and secure online platforms based on the cultural background in China to improve the quality of life of caregivers.

(JMIR Cancer 2025;11:e67131) doi:10.2196/67131

KEYWORDS

cancer; oncology; caregivers; informal caregivers; internet; scoping review; cognitive behavioral therapy

Introduction

The incidence and mortality rates of cancer are rapidly increasing globally. According to the International Agency for Research on Cancer, there were 19.29 million new cancer cases and 9.96 million cancer deaths worldwide in 2020, and the rise in cancer incidence and mortality rates has resulted in a significant disease burden on people [1]. The diagnosis and long-term treatment of cancer not only cause adversity for patients but also impose psychological stress and burdens on caregivers [2]. Caregivers of patients with cancer refer to informal caregivers, including family members, partners, or friends. They provide unpaid social, emotional, and economic support to a family member with cancer requiring care and are involved throughout the patient's symptom management and nursing [3,4]. Caregivers attend to the daily needs of patients and fulfill family responsibilities; they also serve as the patient's primary emotional support. Due to complex treatment environments, a lack of disease-related knowledge, and significant economic burdens, caregivers often experience negative emotions such as anxiety and depression [5]. Zhou et al [6] found that 60.7% of caregivers of patients with cancer experience sleep disturbances. Yang et al [7] conducted a survey involving 116 caregivers of terminally ill patients with cancer receiving home care, revealing that 83.62% of the caregivers reported experiencing moderate to severe fatigue, primarily characterized by physical fatigue. Geng et al [8] reported that the prevalence of anxiety and depression among caregivers of patients with cancer was 46.55% and 42.30%, respectively, with 62% of caregivers bearing a heavy burden that negatively affected their daily lives. Therefore, attention should be given to the physical and mental health of caregivers, along with the provision of appropriate supportive care.

Current interventions for caregivers of patients with cancer include psychosocial support, education, and informational support [9]. Cognitive behavioral therapy (CBT) has received considerable attention owing to its robust theoretical foundation, brief treatment duration, and well-defined structural approach. However, traditional CBT is often influenced by economics, time, and spatial factors, preventing some caregivers from accessing effective help and support [10]. In recent years, with the rise of the "Internet+Healthcare" service model, internet-based cognitive behavioral therapy (ICBT) has emerged. ICBT is an internet-based treatment approach that uses tools such as computers and mobile devices to deliver the core content and skills of CBT through text, video, images, and audio [11].

ICBT addresses the limitations of CBT in its application. Some caregivers concentrate on caregiving behaviors, frequently suppressing their own emotions, which may lead to distress stemming from a deficiency in caregiving skills. ICBT provides caregivers with a discreet online platform that allows them to access relevant information at any time through simple and user-friendly self-service methods, facilitating timely communication with health care professionals and enhancing their cognitive abilities. In addition, techniques such as emotional guidance and relaxation training are used to alleviate caregiver stress and improve their quality of life (QoL) [12,13]. Existing studies have shown that ICBT can mitigate the anticipatory grief experienced by caregivers of patients with cancer, decrease caregiving burden, and improve their self-efficacy [14].

Currently, research on the application of ICBT for caregivers of patients with cancer is steadily growing. However, there is significant heterogeneity in the forms of online interventions, intervention content, and outcome indicators. To gain a comprehensive understanding of the current research status of ICBT, this study uses a scoping review to systematically analyze pertinent studies from both domestic and international contexts. Our goal is to provide references to promote the use and dissemination of ICBT among caregivers of patients with cancer in China.

Methods

Study Design and Framework

This scoping review adhered to the methodological framework developed by the Joanna Briggs Institute (JBI; 2019) [15]. The reporting follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Checklist 1).

Research Questions

The review addressed three key questions: (1) what are the intervention components of ICBT for caregivers of patients with cancer?; (2) what are the intervention forms, duration, and evaluation time points for ICBT?; and (3) what are the outcome indicators and effects of ICBT interventions?

Search Strategy

A systematic search was conducted across 9 databases, including CNKI, Wanfang Database, China Biomedical Literature Database, VIP, PubMed, Web of Science, Embase, CINAHL, and Cochrane Library. The search timeframe extended from the establishment of the databases to June 6, 2024. A combination of subject headings and free-text terms was used. The search strategy was formulated with the guidance of a librarian. The English search terms were "neoplas*, carcinoma*, tumor, oncology, cancer*;" "Cognitive Behavio*, Behavio* Therap*, Cognitive Therap*, ICBT, cognitive behavioural therapy, CCBT;" "online, network, Internet, smartphone, telephone, computer;" and "caregiver*, spouse, family, informal caregiver, couple*." The search strategy for each database was documented in the Multimedia Appendix 1.

Study Selection

The eligibility criteria is presented in Textbox 1.

Textbox 1. Inclusion and exclusion criteria for article selection.

Inclusion criteria:

- Study participants: caregivers of confirmed (by pathology or imaging) patients with cancer, including offspring, parents, and spouses, aged 18 years or older.
- The intervention emphasizes the implementation of cognitive behavioral therapy via the internet, WeChat, mobile devices, or other applications.
- Literature type: original research, including randomized controlled trials or quasiexperimental studies.
- Published literature in both Chinese and English.

Exclusion criteria:

- Literature for which the full text could not be obtained.
- Duplicated publications.
- Conference abstracts.
- Research protocols, reviews, and case studies.

Data Extraction and Quality Assessment

The literature search results were imported into EndNote X9 for duplicate removal. Two independent reviewers (CTS) and (XML) screened titles, abstracts, and full texts against the predefined inclusion and exclusion criteria. Discrepancies were resolved through discussion with a third researcher to reach consensus. One reviewer (CTS) extracted study data using a standardized Microsoft Excel form, capturing authors, publication year, country, design, population characteristics, sample size, interventions, and outcomes, with a second reviewer (XML) independently verifying the accuracy and completeness of all extracted data. For included randomized controlled trials, we conducted quality assessments using the Cochrane Risk of Bias Tool (version 5.1.0) [16], categorizing studies as grade A (low risk), B (moderate risk), or C (high risk), with any discrepancies resolved through consultation with a third researcher to reach consensus.

Results

Study Selection and Characteristics

The systematic search identified 1005 records, with 12 studies meeting inclusion criteria after screening (Figure 1) [14,16-26]. The studies (2013 - 2023) represented diverse geographic regions, such as the United States (4/12, 33%), Australia (3/12, 25%), China (3/12, 25%), Lithuania (1/12, 8%), and Germany (1/12, 8%). Study designs included randomized controlled trials (9/12, 75%), quasi-experimental (1/12, 8%), mixed-methods (1/12, 8%), and feasibility studies (1/12, 8%). Quality assessment of randomized controlled trials (RCTs) indicated moderate methodological rigor (B-level). Table 1 details the basic characteristics of the included studies.

Figure 1. Literature screening process diagram.

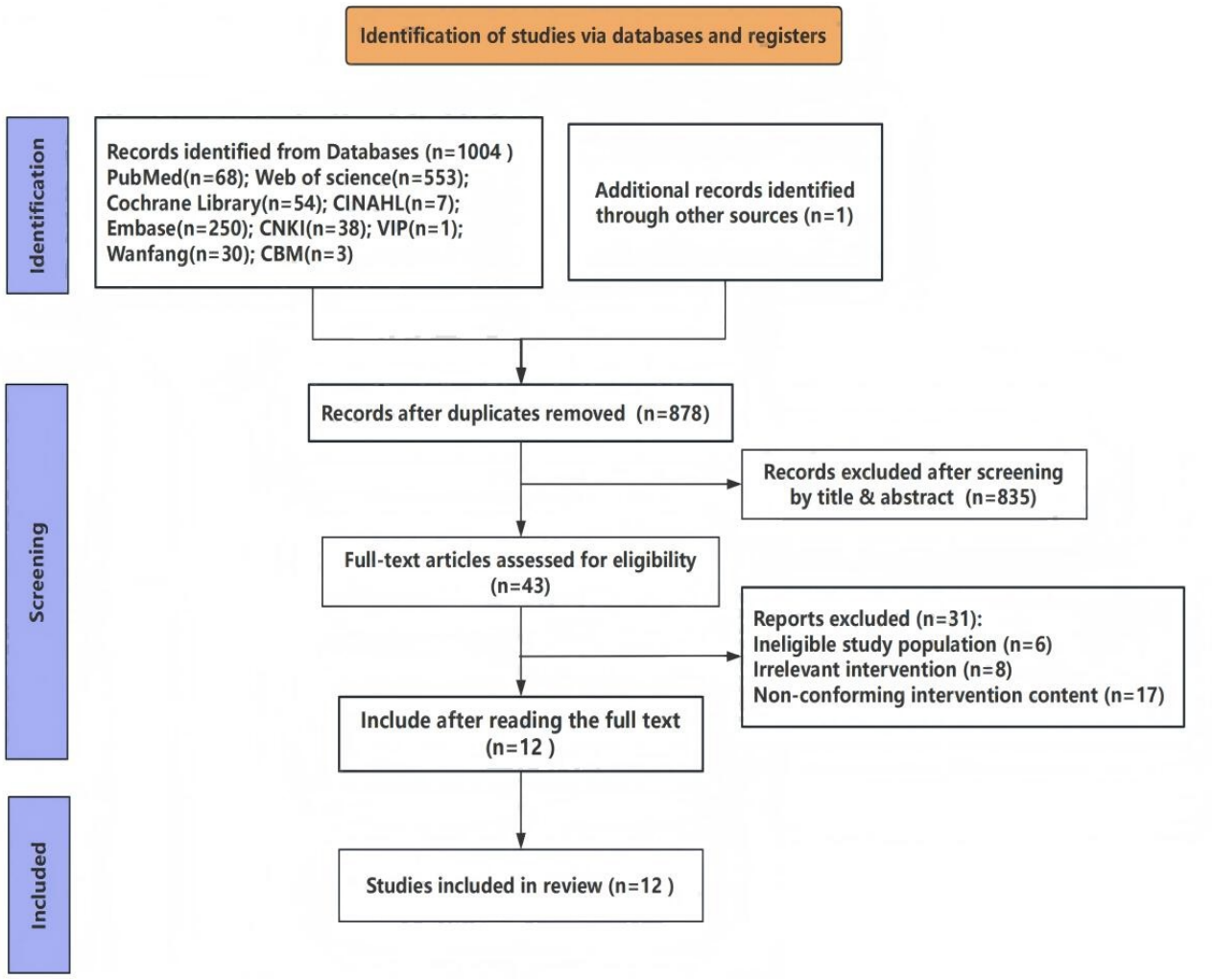


Table . Basic characteristics of included literature.

Reference	Study type	Interference objects	Sample (example, T/C)	Intervention and groups	Outcome indicators; evaluation time	Results
Scott et al [17]	Feasibility of the intervention	Carers of adult patients with cancer receiving curative treatment	13	• Therapist-administered 6-week ICBT[a] program via website and email. • CCO[b] components included: starting treatment; coping with physical symptoms and side effects; coping with emotional distress; body image, identity and sexuality; family and friends; completing treatment.	Primary outcomes: feasibility and acceptability; secondary outcomes: negative affect, distress, QoL; 2-time point evaluation (pre-post)	Engagement was modest; CCO resulted in large overall reductions in negative affect (Cohen *d*=0.88) and small reductions in cancer-specific distress (Cohen *d*=0.37), small to moderate increases in QoL.
Chambers et al [18]	RCT[d]	Patients with cancer and caregivers	345/345	• A 5-session psychologist cognitive-behavioral intervention delivered by telephone (psychoeducation; coping and stress management skills; problem solving; cognitive therapy; enhancing support networks). • A single session of nurse-led self-management intervention.	Primary outcomes: psychological; cancer-specific distress and posttraumatic growth; 4-time point evaluation (pre-3-6-12 months).	The psychologist-led intervention demonstrated reductions both psychological distress (Cohen *d*=0.2, *P*<.001) and cancer-specific distress (Cohen *d*=0.77, *P*<.001), while also enhancing positive adjustment (Cohen *d*=0.82, *P*<.001) from baseline to 12 months.

Reference	Study type	Interference objects	Sample (example, T/C)	Intervention and groups	Outcome indicators; evaluation time	Results
Mosher et al [19]	RCT	Patients with lung cancer and their family caregivers	51 pairs/55 pairs	• Psychologists and clinical social workers delivered the TSM[e] intervention to participants via telephone, with 4 weekly sessions. • TSM components included: relaxation; cognitive restructuring; problem-solving; self-soothing/emotion-focused approach; pleasant activities; activity pacing; communication; plan for continued skills practice. • Education or support	Primary outcomes: depression and anxiety; secondary outcomes: self-efficacy and caregiver burden; 3-time point evaluation (baseline-2 - 6 weeks postintervention).	Small effects in favor of TSM were found regarding caregiver self-efficacy for managing their own emotions and perceived social constraints from the patient.
Kubo et al [20]	2-arm RCT	Patients undergoing cancer chemotherapy and caregivers	Patients: 54/43; caregivers: 17/14	• Psychologists, psychosocial workers, and nurses implemented the 8-week Headspace program through website or mobile apps. • Headspace: encourage participants to first complete a 30-day mindfulness meditation foundation course; they can also choose 10 to 30 days of related symptom or meditation courses. • UC[f]	Primary outcomes: mindfulness and quality of life; secondary outcomes: distress; posttraumatic growth; fatigue; sleep quality; 2-time point evaluation (pre-post).	Headspace significantly improved mindfulness (P=.03) with borderline significant effects on PTGI[g] new possibilities (P=.06) versus controls
Biliunaite et al [21]	2-arm RCT	Caregivers of individuals with dementia, cancer, or other illnesses	31/32			

Reference	Study type	Interference objects	Sample (example, T/C)	Intervention and groups	Outcome indicators; evaluation time	Results
				• Therapists implemented an 8-week ICBT program through the Slaugau Artima website. • ICBT: introduction; thoughts; stress and relaxation; problem solving; communication; anxiety; behavioral activation; and maintenance. • UC	Primary outcome: CBI[h]; secondary outcomes: depression, anxiety, stress, and QoL; 2-time point evaluation (pre-post).	ICBT showed large effects on burden reduction (P<.001) and stress (P<.001), moderate effects on anxiety (P=.004) and depression (P=.01), and significant QoL enhancement (P=.001).
Luo et al [22]	RCT	Parents of children diagnosed with cancer	52/51	• Psychologists, doctors, and nurses implemented the 8-week a mobile device–based resilience training program. • Resilience training program: understanding the purpose of intervention, relaxation technique training, problem-solving skills, cognitive restructuring, promoting good relationships, and cultivating positive performance and beliefs, etc. • UC	Primary outcome: resilience; secondary outcomes: depressive symptoms and QoL; 3-time point evaluation (pretreatment, 2 and 6 months after the intervention began).	At the 6-month follow-up, the intervention demonstrated statistically significant improvements in resilience (P=.01) and depressive symptoms (P=.04), but failed to show significant QoL enhancement (P=.38), although the experimental group showed numerically higher QoL scores than controls.
Wakefield et al [23]	3-arm RCT	Parents or caregivers of children	19/18/19		Feasibility; acceptability; safety; efficacy (QoL; psychological outcomes); 4-time point evaluation (baseline, 2‐4 weeks post-intervention; 2‐4 weeks post-booster and 6 months postintervention).	Most Cascade parents were satisfied and reported experiencing benefits from the program. However, Cascade did not improve their main outcomes, including parents' quality of life, depression and anxiety.

Reference	Study type	Interference objects	Sample (example, T/C)	Intervention and groups	Outcome indicators; evaluation time	Results
				• Psychologist-delivered online intervention. • Cascade: introduction and behavioral activation; identifying and challenging unhelpful thoughts; mindfulness and disengagement; skills for fostering relationships and living a rich life after cancer; booster session. • Peer-support or waitlist.		
Trevino et al [24]	RCT	Older adults with cancer and their caregivers	14 pairs/14 pairs	• Social workers provide 7-session over the telephone. • MAC[i]: this includes providing information on treatment methods for the elderly, addressing the widespread shame associated with psychological services among the elderly, and integrating strategies to address cancer-specific stressors during the intervention process. • UC	Feasibility, acceptability, participant adherence, anxiety, depression, and QoL; 2-time point evaluation (pre-post).	85.7% participants completed all 7 sessions, over 80% of caregivers rated MAC as "moderately" to "very" helpful. MAC dyads experienced a greater reduction in anxiety than dyads in usual care with smaller changes in depression and quality of life.
Kaiser et al [25]	RCT	Caregivers of cancer bereavement	44/43			

Reference	Study type	Interference objects	Sample (example, T/C)	Intervention and groups	Outcome indicators; evaluation time	Results
				• Caregivers completed 2 self-scheduled 45-minute writing sessions weekly via a website. • ICBT: 10 structured writing tasks, 3 modules (self-confrontation; cognitive reappraisal; and social sharing). • Waitlist	Primary outcome: prolonged grief; secondary outcomes: depression, anxiety, posttraumatic stress, posttraumatic growth, somatization, sleep quality, and mental and physical health; 5-time point evaluation (baseline, posttreatment, and 3、6-12 months).	ICBT reduced symptoms of prolonged grief (Cohen *d*=0.80; *P*<.001) to a clinically significant extent. It had favorable effects on depression, anxiety, posttraumatic stress, posttraumatic growth, and overall mental health but not on somatization, sleep quality, or physical health.
Yang et al [26]	RCT	Patients undergoing cervical cancer chemotherapy and their spouses	53 pairs/53 pairs	• Psychological counselors and nurses delivered an 8-week ICBT program through WeChat (Tencent) public platform and offline group sessions. • ICBT: psychological diagnosis; cognitive education; behavior training; emotional expression; consolidate follow-up. • UC	The patients' distress and QoL; 2-time point evaluation (pre-1 month after intervention).	ICBT significantly reduced the detection rate of psychological distress (*P*<.05) and improved quality of life (*P*<.01) among patients.
Carr et al [27]	Mixed research	Caregivers of a phase 1 oncology trial patient	23 quantitative cases; 5 qualitative cases		Primary outcome: acceptability and feasibility; secondary outcomes: caregiver distress; caregiver burden; PAC[k]; caregiver grief; anxiety and depression; 2-time point evaluation (pre-post).	Although the P1CaLL[l] pilot achieved limited recruitment feasibility (45.3% enrollment), its high acceptability (84% completion) and preliminary efficacy signals across multiple caregiver outcomes (stress reduction, isolation mitigation, and self-control improvement) support further development.

Reference	Study type	Interference objects	Sample (example, T/C)	Intervention and groups	Outcome indicators; evaluation time	Results
				• Clinician-administered telephone intervention comprised 4 CBMS[j]+4 randomized CBT sessions across 9 weeks. • CBSM: mind-body connection; coping skills; communication; and social support; metta meditation. • CBT: intro to CBT-tracking automatic thoughts; identifying distorted thoughts; challenging distorted thoughts; core beliefs or relapse prevention		
Wang [14]	Quasiexperimental study	Caregivers of patients with cancer	38/38	• Physicians and nurses administered ICBT through a website and WeChat groups, including twice-weekly digital content and monthly peer support video sessions. • ICBT: basic knowledge; symptom education; common knowledge of home care; relaxation training; social support • UC	Care burden, anticipatory grief, and self-efficacy; 3-time point evaluation (pretreatment, posttreatment 6-12 weeks).	ICBT demonstrated significant efficacy in reducing caregivers' anticipatory grief levels (P<.001), alleviating caregiving burden (P<.001), and enhancing self-efficacy (P<.001).

[a]ICBT: internet-based cognitive behavioral therapy.

[b]CCO: cancer coping online.

[c]QoL: quality of life.

[d]RCT: randomized controlled trial.

[e]TSM: telephone-based symptom management.

[f]UC: usual care.

[g]PTGI: posttraumatic growth inventory.

[h]CBI: caregiver burden inventory.

[i]MAC: managing anxiety from cancer.

[j]CBSM: cognitive behavioral stress-management.

[k]PAC: positive aspects of caregiving.

[l]P1CaLL: Phase 1 Caregiver LifeLine.

Intervention Content

The included studies incorporated the following key components in their ICBT interventions for caregivers (Textbox 2).

Textbox 2. Key components of internet-based cognitive behavioral therapy interventions for caregivers.

- Treatment introduction: a total of 5 studies [14,17,21,23,26] initiated interventions with an introductory phase, wherein researchers presented the medical center and care team to establish caregiver-provider collaboration. Before intervention delivery, caregivers underwent baseline assessments, received psychoeducation on cognitive-behavioral therapy principles, and were guided on study protocols to optimize adherence. In addition, they were encouraged to set personalized goals to enhance engagement and motivation.
- Cognitive education and restructuring: a total of 7 studies [14,18,19,22,24-26] incorporated structured modules to help caregivers reframe maladaptive cognitions. Through perspective-taking exercises and peer-sharing sessions, participants were educated on disease-related knowledge to foster accurate perceptions of cancer. This component also emphasized symptom recognition training to improve timely and appropriate caregiving responses.
- Emotional expression and coping: a total of 6 studies [17-19,21,26,27] assessed caregivers' emotional states using standardized questionnaires or individual interviews, followed by discussions on psychosocial impacts. To mitigate distress, interventions introduced techniques such as cognitive restructuring diaries, mindfulness meditation, and progressive relaxation training.
- Cognitive restructuring and reinforcement: a total of 4 studies [22,23,25,27] focused on identifying and modifying maladaptive thought patterns. Caregivers were taught to recognize automatic thoughts and common cognitive distortions. Through guided exercises, they practiced challenging unhelpful beliefs to cultivate healthier cognitive frameworks.
- Behavioral training and activation: the most frequently implemented component (9 studies [14,17,19-24,26]) involved skill-building through evidence-based techniques such as progressive muscle relaxation, diaphragmatic breathing, music therapy, and guided imagery. Caregivers selected preferred modalities based on individual capacity and preferences. Protocols also emphasized self-care strategies, including scheduled personal time and reward systems.
- Problem-solving techniques: a total of 4 studies [18,19,21,22] trained caregivers in structured problem-solving. Participants learned to deconstruct challenges, identify barriers, evaluate coping strategies, and implement solutions. Postintervention, they monitored outcomes and adjusted approaches through reflective practice to enhance long-term adaptive skills.
- Communication: a total of 5 studies [17,19,21,23,27] addressed communication dynamics, exploring how caregiving roles influenced interpersonal interactions. Caregivers practiced maintaining or improving intimacy through verbal and nonverbal techniques, sustaining social connections, and fostering supportive relationships.
- Completion of treatment and consolidation follow-up: a total of 3 studies [17,21,26] concluded with a review phase, summarizing key concepts and reinforcing long-term skill retention. Caregivers were guided in self-directed practice, goal reflection, and future planning to sustain intervention benefits long-term skill retention. Caregivers were guided in self-directed practice, goal reflection, and future planning to sustain intervention benefits.

The included studies typically incorporated 3‐5 core intervention modules. Following each intervention session, participants were required to complete structured homework assignments, which included documenting caregiving-related emotional experiences, evaluating automatic thoughts using thought records, emotional expression through writing exercises [21], completion of structured writing tasks [25], and home practice of acquired skills [19,20].

The reviewed literature revealed 2 distinct ICBT delivery models for caregivers. Seven studies [14,17,21-23,25,27] implemented caregiver-specific interventions focusing on cognitive restructuring, emotional regulation training, and evidence-based relaxation techniques to enhance multidimensional wellbeing. Alternatively, 5 studies [18-20,24,26] used dyadic approaches that simultaneously engaged both patients and caregivers through adapted protocols delivered in either parallel or joint therapeutic sessions. Both models demonstrated effectiveness in addressing the psychological needs of caregivers while accounting for different caregiving contexts.

Intervention Elements

The ICBT interventions examined in this study comprised several key components, such as delivery modality, provider qualifications, intervention duration, and evaluation timelines. The primary delivery modalities included web-based platforms (4/7, 57%) studies with real-time psychologist support or email feedback [14,17,20-22,25,26], telephone sessions (45‐60 min) [18,19,24,27], and video conferencing [23]. Interventions were predominantly delivered by psychotherapists, with 1 study [22] using a multidisciplinary team (psychologists, physicians, and nurses). Intervention duration varied (most commonly 6‐8 weeks [17,20-22,26]) depending on content and format. Telephone sessions typically lasted 45‐60 minutes [19,24], though some flexible protocols permitted completion within 1 week [14,17,21,26]. Effectiveness was evaluated at 3 time

points: (1) baseline (preintervention, confirming group comparability [*P*>.05]), (2) postintervention, and (3) follow-up (1‐12 months) in 4 studies [18,22,23,25]. Qualitative components (semistructured interviews) were included in 2 studies [17,27] to assess participant experiences.

Outcome Indicators and Effectiveness Evaluation of Interventions

The ICBT interventions evaluated 7 primary outcome measures: feasibility (n=4 studies [17,23,24,27]), acceptability (n=4 studies [17,23,24,27]), QoL (n=8 studies [17,20-26]), caregiver burden (n=4 studies [14,19,21,27]), psychological distress (n=11 studies [14,17-25,27]), posttraumatic growth (n=3 studies [18,20,25]), and self-efficacy (n=3 studies [14,19,23]).

Feasibility

Feasibility was assessed based on recruitment, retention, and completion rates. In total, 2 studies [17,27] reported recruitment rates below 50%, primarily attributed to participants' aversion to online support, lack of interest, demanding work schedules, and substantial caregiving commitments. Although most participants completed all intervention modules, attrition occurred due to heightened psychological distress, time constraints, or deterioration of the care recipient's health [20]. Attrition rates generally ranged from 3% to 16%, with one exception reaching 31% [17]. Qualitative analysis suggests this discontinuation pattern may result from both the lack of personalized engagement in digital interventions and deliberate withdrawal after achieving therapeutic objectives.

Acceptability

Acceptability was assessed through participant-reported satisfaction with both intervention content and engagement modalities. Many caregivers expressed positive feedback regarding the intervention, stating that "the online intervention is convenient, time-saving, and practical," "they were satisfied with the information and quality provided," "the skills learned were relevant to cancer treatment," and "the intervention courses helped alleviate stress." However, 1 study [17] indicated that 33% of participants felt the intervention did not adequately address the needs of caregivers, while another study revealed disagreement about whether patients and caregivers should be treated together. Specifically, 45% of caregivers preferred individual interventions, 36.4% favored some combined treatment, and 18.2% preferred fully integrated interventions.

Psychological Outcomes of Caregiver Intervention

Cognitive-behavioral interventions demonstrated measurable benefits for caregiver mental health. The stress management program by Carr et al [27] significantly improved caregivers' stress coping abilities (effect size *r*=0.39), while 2 other trials [14,21] reported statistically significant reductions in caregiver burden (*P*<.05). Caregivers often experience negative emotions such as anxiety, depression, or sadness due to prolonged caregiving and stress [28]. ICBT interventions showed moderate efficacy in alleviating anxiety and depression [17,21], with the structured writing intervention ("Online-Trauertherapie") by Kaiser et al [25] producing effect sizes ranging from 0.29 to 0.84 across multiple psychological domains. Notably, Trevino et al [24] found patient-caregiver anxiety changes were positively correlated, though other psychological outcomes showed nonsignificant associations. Some studies [23,29] reported limited effects on anxiety or depression, potentially due to low baseline distress levels or requiring specific intervention components (eg, guided imagery) to achieve psychological benefits.

QoL

The included studies reported mixed effects on caregiver's QoL. A total of 4 trials [17,20,21,25] demonstrated statistically significant QoL improvements following intervention. Notably, one spouse-focused intervention [26] showed significant patient QoL benefits, suggesting potential secondary effects. However, 3 studies [22-24] found no significant QoL changes for caregivers, potentially due to shorter intervention durations or differing outcome measures.

Discussion

Principal Findings and Comparison With Previous Work

This study examines the existing research on ICBT interventions for caregivers of patients with cancer. A systematic review of 12 eligible studies revealed that ICBT is a feasible and acceptable approach for this population. The findings suggest that ICBT may alleviate anxiety and depressive symptoms among caregivers, with some studies additionally reporting reduced caregiver burden and enhanced self-efficacy.

These findings are consistent with previous studies, demonstrating the acceptability and feasibility of ICBT when applied to caregivers. For instance, Meichsner et al [30] observed high satisfaction and enhanced well-being among dementia caregivers following ICBT. Similarly, Tur et al [31] documented strong participant satisfaction in an ICBT program for prolonged grief disorder, with 75% of participants achieving clinically significant reductions in depressive symptoms and 50% demonstrating meaningful improvements in grief-related cognitions. Further corroborating these results, Titov et al [32] reported that 63% of individuals with generalized anxiety disorder experienced significant anxiety reduction after ICBT, with concurrent improvements in comorbid depression.

However, the effects of ICBT on caregivers' QoL were inconsistent across studies. While some improvement in QoL was observed, the overall effect was not statistically significant, possibly due to the relatively short intervention duration. Although follow-up assessments were conducted in some studies, no significant differences in QoL were detected. Li et al [33] suggested that improvements in QoL generally require a longer time to manifest compared with behavioral or mental health changes. In addition, multiple factors, such as caregivers' socioeconomic status, age, caregiving duration, coping strategies, the patient's clinical condition, and available social support may influence intervention outcomes [34]. Therefore, future studies should implement multidimensional assessments of caregiver well-being (addressing physical, psychological, emotional, and social domains) to inform personalized interventions, while extending intervention durations to better evaluate long-term QoL outcomes.

This scoping review synthesizes existing literature on ICBT interventions for caregivers of patients with cancer, offering key insights into the current evidence base. Our findings highlight the feasibility and acceptability of most programs; however, recruitment challenges were evident, with a substantial proportion of eligible caregivers declining participation. Furthermore, enrolled participants exhibited high attrition rates. To enhance intervention engagement and retention, future studies should focus on precise population targeting, refined inclusion criteria, and proactive baseline assessments to identify and support at-risk caregivers.

Population Specificity and Methodological Variability

Among the included studies, 3 investigations [21,24,25] specifically targeted subgroups with distinct characteristics, such as caregivers exhibiting elevated anxiety levels or heightened caregiving burdens. Another 3 studies [19,26,27] enhanced research specificity by focusing on particular cancer types. Several studies treated patient-caregiver dyads as intervention units [24,26], demonstrating that caregiver support could reduce patients' psychological distress and improve their QoL. This reciprocal relationship reflects how caregivers' comprehensive support impacts patients' recovery and emotional state, while patients' conditions similarly affect caregivers' wellbeing. Li et al [35] found dyadic collaboration improved both QoL and coping outcomes, though some caregivers preferred individual interventions. Future research should systematically compare the efficacy of caregiver-only interventions versus dyadic intervention approaches.

In total, 2 studies [17,27] included qualitative postintervention evaluations. Carr et al [27] reported that caregivers benefited from recognizing automatic thought patterns, which improved their awareness of irrational thinking and overall wellbeing. However, participants suggested improvements, including initial in-person contact with facilitators, more personalized resources, and better scheduling to accommodate caregiving duties. While ICBT applications for caregivers remain exploratory, mixed-methods approaches combining quantitative outcome measures with qualitative insights can strengthen the evidence base for optimizing interventions [36].

Exploring Intervention Content and Delivery Formats

Effective intervention design must prioritize both engagement and usability through personalized support programs that address caregivers' specific needs. A critical yet frequently overlooked component is the implementation of preparatory phases, as evidenced by our review finding that only 5 studies incorporated preintervention assessments. These preliminary modules should systematically collect caregiver characteristics to inform tailored recommendations, which could significantly improve both participation rates and intervention adherence. Establishing such foundational elements represents an important direction for future research development.

The caregiving role often generates substantial psychological burdens, stemming from challenging role transitions, difficulties adapting to medical environments, and the inherent stress of managing a loved one's illness. Compounding these issues, caregivers frequently suppress their emotions and neglect self-care, resulting in diminished QoL [37,38]. Research by Zhang et al [39] demonstrates that targeted interventions incorporating relaxation training can yield multiple benefits, such as regulating neuropsychological functions, alleviating chronic physical tension, improving physiological responses, and reducing illness-related stress, ultimately decreasing caregiver burden while promoting positive behavioral changes. Therefore, intervention content should integrate 3 key components: (1) cognitive restructuring to modify maladaptive thought patterns; (2) emotional support systems; and (3) practical strategies including context-appropriate health education and relaxation techniques. This comprehensive approach will better equip caregivers to manage their multifaceted challenges.

This study found that few interventions target caregivers' self-symptom management. While 2 studies assessed interventions for caregiver symptoms (eg, pain, fatigue, and sleep) [20,25], no significant improvements were observed, possibly due to differing intervention focuses. However, Shaffer et al [40] reported substantial reductions in insomnia among cancer caregivers through an internet-based program, even though it was not caregiver-specific. Similarly, Ye et al [41] demonstrated that ICBT improved sleep, anxiety, and depression in patients with insomnia . Future research should develop symptom-specific interventions tailored to caregivers to validate their efficacy.

Our study found that online caregiver interventions typically involve real-time psychologist support or email feedback. To optimize website- or WeChat-based interventions, integrating interactive features (eg, chat functions and message boards) is recommended. Caregivers should receive prompt module-specific feedback via email, with professionals addressing inquiries swiftly to build trust. Evidence confirms that therapist-guided ICBT outperforms unguided interventions [42], as therapist engagement sustains participant adherence. For caregivers with limited digital literacy, technical assistance and age-friendly design features (eg, larger subtitles and enhanced visibility) are essential [43]. While video or telephone interventions enable direct communication and emotional support, they risk inefficiency in handling repetitive queries and may compromise continuity. WeChat public platforms or online websites could serve as supplementary tools. Telemedicine and mobile health solutions should be leveraged to comprehensively address caregivers' diverse needs.

Limitations

This review acknowledges several limitations. First, the included studies exhibited significant variability in intervention measures, participant characteristics, sample sizes, and outcome indicators, potentially limiting the generalizability of ICBT efficacy findings. Second, the lack of long-term follow-up assessments in most studies necessitates future research with extended evaluation periods to better understand sustained intervention effects. Third, while we only assessed risk of bias in RCTs, more comprehensive bias evaluations using standardized tools would strengthen future findings. Finally, the included studies originated from various countries, highlighting certain cultural differences. Nonetheless, this diversity underscores the

feasibility and acceptability of ICBT among caregivers of patients with cancer.

Implications for Future Research

To optimize resource use and enhance support for caregivers seeking online assistance, it is imperative to establish a dedicated support team composed of oncologists, nurses, and psychotherapists. Oncologists can build trust with patients and caregivers, promoting engagement and addressing maladaptive cognitions, which has been shown to improve outcomes and reduce attrition [44]. Nurses can serve as interdisciplinary coordinators and health educators, identifying caregiver needs, delivering symptom management guidance, and providing holistic biopsychosocial support. By fostering collaboration among multidisciplinary professionals, caregivers can receive more comprehensive support, capitalizing on the diverse strengths of various specialists and enhancing overall caregiver satisfaction.

Conclusions

This study presents a scoping review of the application of ICBT among caregivers of patients with cancer. ICBT combines the advantages of CBT with the accessibility of smart devices. Preliminary evidence suggests that ICBT is effective in alleviating caregivers' negative emotions, reducing stress, and enhancing positive experiences. Nevertheless, significant shortcomings remain in the research concerning intervention content, sample size, adherence, evaluation criteria, software development, and intervention teams. Currently, internet-based cognitive-behavioral interventions for caregivers of patients with cancer in our country are still in the nascent stages. It is crucial to draw upon relevant international studies while considering the unique characteristics of domestic caregivers to develop tailored ICBT intervention programs. By leveraging artificial intelligence to create safe and effective online platforms and fostering multidisciplinary collaboration, we can provide comprehensive physical and mental health interventions for caregivers, thereby genuinely alleviating their caregiving burden, reducing negative emotions, and enhancing their QoL.

Acknowledgments

The authors thank AiMi Academic Services for English language editing and review services. No funding was received to assist with the preparation of this manuscript.

Authors' Contributions

JS and FXL contributed to the conceptualization of the study. CTS conducted investigation, data curation, and writing – original draft. XML handled investigation and data curation. All authors contributed to writing – review and editing and validation and read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[PDF File, 143 KB - cancer_v11i1e67131_app1.pdf]

Checklist 1
Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) checklist.
[DOC File, 355 KB - cancer_v11i1e67131_app2.doc]

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249. [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Ling HU, Ran HUO, Ke X, et al. Investigation of negative emotions in family members of cancer patients and analysis of comprehensive interventions. Special Health 2019(31):274-275. [doi: 10.3969/j.issn.2095-6851.2019.31.343]
3. Barello S, Castiglioni C, Bonanomi A, Graffigna G. The Caregiving Health Engagement Scale (CHE-s): development and initial validation of a new questionnaire for measuring family caregiver engagement in healthcare. BMC Public Health 2019 Nov 27;19(1):1562. [doi: 10.1186/s12889-019-7743-8] [Medline: 31771546]
4. Li J, Sun D, Zhang X, et al. The relationship between anticipatory grief and illness uncertainty among Chinese family caregivers of patients with advanced lung cancer: a cross-sectional study. BMC Palliat Care 2022 Mar 8;21(1):30. [doi: 10.1186/s12904-022-00925-4] [Medline: 35255876]
5. Qiuping LI, Yuen Loke A. Current research status of family caregivers for cancer patients. Chin J Nurs 2012;47(12):1132-1135. [doi: 10.3761/j.issn.0254-1769.2012.12.033]

6. Zhou BF, Cai YM, Wu XL, et al. Investigation on the sleep quality of primary caregivers for HCC patients receiving interventional therapy: the present situation and its influencing factors. J Interv Radiol 2022;31(2):189-192. [doi: 10.3969/j.issn.1008-794X.2022.02.016]
7. Yang Z, Jiang Z, Wang WH. Study of caregiver fatigue and coping styles of family caregivers of patients with advanced cancer. J Nurs Adm 2015;15(3):171-173.
8. Geng HM, Chuang DM, Yang F, et al. Prevalence and determinants of depression in caregivers of cancer patients: a systematic review and meta-analysis. Medicine (Baltimore) 2018 Sep;97(39):e11863. [doi: 10.1097/MD.0000000000011863] [Medline: 30278483]
9. BECK AT. Thinking and depression. I. Idiosyncratic content and cognitive distortions. Arch Gen Psychiatry 1963 Oct;9:324-333. [doi: 10.1001/archpsyc.1963.01720160014002] [Medline: 14045261]
10. Pan MY, Zhao JX, Huang HQ, et al. Research progress of internet-based cognitive behavioral therapy in breast cancer patients. Chin J Nurs 2022;57(3):363-367. [doi: 10.3761/j.issn.0254-1769.2022.03.018]
11. Webb CA, Rosso IM, Rauch SL. Internet-based cognitive-behavioral therapy for depression: current progress and future directions. Harv Rev Psychiatry 2017;25(3):114-122. [doi: 10.1097/HRP.0000000000000139] [Medline: 28475503]
12. Wang LF, Kong Q, Mi Y, et al. The psychological experience of caregivers of patients with advanced cancer: a meta-synthesis of qualitative studies. Chin J Nurs 2020;55(6):856-861. [doi: 10.3761/j.issn.0254-1769.2020.06.012]
13. Biliunaite I, Kazlauskas E, Sanderman R, Andersson G. Process evaluation of internet-based cognitive behavioral therapy intervention for informal caregivers. Front Med (Lausanne) 2021;8:725510. [doi: 10.3389/fmed.2021.725510] [Medline: 34869422]
14. Wang P. The impact of online cognitive behavioral therapy on the mental health of primary caregivers of cancer patients [Master's thesis]. : Shanxi Medical University; 2023.
15. Lockwood C, Dos Santos KB, Pap R. Practical guidance for knowledge synthesis: scoping review methods. Asian Nurs Res (Korean Soc Nurs Sci) 2019 Dec;13(5):287-294. [doi: 10.1016/j.anr.2019.11.002] [Medline: 31756513]
16. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0: The Cochrane Collaboration; 2011. URL: http://handbook-5-1.cochrane.org/ [accessed 2025-05-25]
17. Scott K, Beatty L. Feasibility study of a self-guided cognitive behaviour therapy internet intervention for cancer carers. Aust J Prim Health 2013;19(4):270-274. [doi: 10.1071/PY13025] [Medline: 23796231]
18. Chambers SK, Girgis A, Occhipinti S, et al. A randomized trial comparing two low-intensity psychological interventions for distressed patients with cancer and their caregivers. Oncol Nurs Forum 2014 Jul 1;41(4):E256-E266. [doi: 10.1188/14.ONF.E256-E266] [Medline: 24969260]
19. Mosher CE, Winger JG, Hanna N, et al. Randomized pilot trial of a telephone symptom management intervention for symptomatic lung cancer patients and their family caregivers. J Pain Symptom Manage 2016 Oct;52(4):469-482. [doi: 10.1016/j.jpainsymman.2016.04.006] [Medline: 27401514]
20. Kubo A, Kurtovich E, McGinnis M, et al. A randomized controlled trial of mhealth mindfulness intervention for cancer patients and informal cancer caregivers: a feasibility study within an integrated health care delivery system. Integr Cancer Ther 2019;18:1534735419850634. [doi: 10.1177/1534735419850634] [Medline: 31092044]
21. Biliunaite I, Kazlauskas E, Sanderman R, Truskauskaite-Kuneviciene I, Dumarkaite A, Andersson G. Internet-based cognitive behavioral therapy for informal caregivers: randomized controlled pilot trial. J Med Internet Res 2021 Apr 7;23(4):e21466. [doi: 10.2196/21466] [Medline: 33825687]
22. Luo Y, Xia W, Cheung AT, et al. Effectiveness of a mobile device-based resilience training program in reducing depressive symptoms and enhancing resilience and quality of life in parents of children with cancer: randomized controlled trial. J Med Internet Res 2021 Nov 29;23(11):e27639. [doi: 10.2196/27639] [Medline: 34847060]
23. Wakefield CE, Sansom-Daly UM, McGill BC, et al. Providing psychological support to parents of childhood cancer survivors: "Cascade" intervention trial results and lessons for the future. Cancers (Basel) 2021 Nov 9;13(22):5597. [doi: 10.3390/cancers13225597] [Medline: 34830752]
24. Trevino KM, Stern A, Hershkowitz R, et al. Managing Anxiety from Cancer (MAC): a pilot randomized controlled trial of an anxiety intervention for older adults with cancer and their caregivers. Palliat Support Care 2021 Apr;19(2):135-145. [doi: 10.1017/S1478951521000286] [Medline: 33818370]
25. Kaiser J, Nagl M, Hoffmann R, Linde K, Kersting A. Therapist-assisted web-based intervention for prolonged grief disorder after cancer bereavement: randomized controlled trial. JMIR Ment Health 2022 Feb 8;9(2):e27642. [doi: 10.2196/27642] [Medline: 35133286]
26. Yang B, Feng XL, Li XL, et al. Study on couple-focused internet cognitive behavioral intervention for young cervical cancer patients undergoing chemotherapy. J Nurs Sci 2022;37(1):64-67. [doi: 10.3870/j.issn.1001-4152.2022.01.064]
27. Carr AL, Bilenduke E, Adolf E, et al. A pilot randomized study of a telephone-based cognitive-behavioral stress-management intervention to reduce distress in phase 1 oncology trial caregivers. Palliat Support Care 2023 Oct;21(5):820-828. [doi: 10.1017/S1478951523000196] [Medline: 36994841]
28. Leonardi M, Giovannetti AM, Pagani M, Raggi A, Sattin D, National Consortium Functioning And Disability In Vegetative And In Minimal Conscious State Patients. Burden and needs of 487 caregivers of patients in vegetative state and in minimally

conscious state: results from a national study. Brain Inj 2012;26(10):1201-1210. [doi: 10.3109/02699052.2012.667589] [Medline: 22571687]
29. Winger JG, Rand KL, Hanna N, et al. Coping skills practice and symptom change: a secondary analysis of a pilot telephone symptom management intervention for lung cancer patients and their family caregivers. J Pain Symptom Manage 2018 May;55(5):1341-1349. [doi: 10.1016/j.jpainsymman.2018.01.005] [Medline: 29366911]
30. Meichsner F, Theurer C, Wilz G. Acceptance and treatment effects of an internet-delivered cognitive-behavioral intervention for family caregivers of people with dementia: a randomized-controlled trial. J Clin Psychol 2019 Apr;75(4):594-613. [doi: 10.1002/jclp.22739] [Medline: 30597537]
31. Tur C, Campos D, Suso-Ribera C, et al. An internet-delivered cognitive-behavioral therapy (iCBT) for prolonged grief disorder (PGD) in adults: a multiple-baseline single-case experimental design study. Internet Interv 2022 Sep;29:100558. [doi: 10.1016/j.invent.2022.100558] [Medline: 35865996]
32. Titov N, Andrews G, Johnston L, Robinson E, Spence J. Transdiagnostic Internet treatment for anxiety disorders: a randomized controlled trial. Behav Res Ther 2010 Sep;48(9):890-899. [doi: 10.1016/j.brat.2010.05.014] [Medline: 20561606]
33. Li WHC, Ho KY, Lam KKW, et al. Adventure-based training to promote physical activity and reduce fatigue among childhood cancer survivors: a randomized controlled trial. Int J Nurs Stud 2018 Jul;83:65-74. [doi: 10.1016/j.ijnurstu.2018.04.007] [Medline: 29689482]
34. He XF, Luo L, Jiang ZL. Factors influencing the quality of life of caregivers of cancer patients and intervention measures. Chongqing Medical Journal 2017;46(29):4166-4168. [doi: 10.3969/j.issn.1671-8348.2017.29.044]
35. Li Q, Loke AY. A literature review on the mutual impact of the spousal caregiver-cancer patients dyads: “communication”, “reciprocal influence”, and “caregiver-patient congruence”. Eur J Oncol Nurs 2014 Feb;18(1):58-65. [doi: 10.1016/j.ejon.2013.09.003] [Medline: 24100089]
36. Hu T, Li YN, Wang JT. The use of information and communication technology in supporting the caregivers of children with cancer:a scoping review. Nurs J Chin 2019(12):16-20. [doi: 10.3969/j.issn.1008-9993.2019.12.005]
37. Yang RM. Case study on emotional counseling in temporary assistance for homeless psychiatric patients during rehabilitation period. Tongfang Knowledge Network, Beijing 2023. [doi: 10.27047/d.cnki.ggudu.2023.001695]
38. Jiao J, Nian WY, Ren HL, et al. A qualitative study on the expected sadness experience of spouses of young late stage cancer patients. Journal of Nurses Training 2021;36(2):179-184. [doi: 10.16821/j.cnki.hsjx.2021.02.020]
39. Zhang C, Jin YX, Zhang J, et al. The effect of self relaxation training on the caregiving burden of caregivers of glioma chemotherapy patients. Shanghai Nursing Journal 2018:38-41. [doi: 10.3969/j.issn.1009-8399.2018.04.010]
40. Shaffer KM, Glazer JV, Chow PI, Ingersoll KS, Ritterband LM. Family cancer caregiver use of and benefit from an internet-delivered insomnia intervention: results from a single-group feasibility trial. J Psychosoc Oncol Res Pract 2024;6(2):129. [doi: 10.1097/or9.0000000000000129] [Medline: 38962167]
41. Ye YY, Yang XJ, Jiang XJ. Internet-based cognitive behavioral therapy for insomnia disorder. Chinese Journal of Clinical Neurosciences 2015;23(5):580-584.
42. Nordh M, Wahlund T, Jolstedt M, et al. Therapist-guided internet-delivered cognitive behavioral therapy vs internet-delivered supportive therapy for children and adolescents with social anxiety disorder: a randomized clinical trial. JAMA Psychiatry 2021 Jul 1;78(7):705-713. [doi: 10.1001/jamapsychiatry.2021.0469] [Medline: 33978699]
43. CUI C, LI YL, YANG Y. Research progress on intelligent cognitive behavioral therapy in symptom management of cancer patients. Chinese Journal of Modern Nursing 2022;28(17):2365-2370. [doi: 10.3760/cma.j.cn115682-20210815-03614]
44. Gellatly J, Bower P, Hennessy S, Richards D, Gilbody S, Lovell K. What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. Psychol Med 2007 Sep;37(9):1217-1228. [doi: 10.1017/S0033291707000062] [Medline: 17306044]

Abbreviations

CBT: cognitive-behavioral therapy
CCO: cancer coping online
ICBT: internet-based cognitive behavioral therapy
PTGI: posttraumatic growth inventory
QoL: quality of life
RCT: randomized controlled trial
TSM: telephone-based symptom management

Edited by N Cahill; submitted 03.10.24; peer-reviewed by KD Boer, L Ming; revised version received 15.04.25; accepted 20.04.25; published 04.06.25.

Please cite as:
Shen CT, Shi J, Liu FX, Lu XM
Internet-Based Cognitive Behavioral Therapy Interventions for Caregivers of Patients With Cancer: Scoping Review
JMIR Cancer 2025;11:e67131
URL: https://cancer.jmir.org/2025/1/e67131
doi:10.2196/67131

Family Experiences, Needs, and Perceptions in Home-Based Hospice Care for Patients With Terminal Cancer: Meta-Synthesis and Systematic Review

Xin Ming Deng[1], BSc; Kanokwan Hounsri[1], MNS; Violeta Lopez[2,3], PhD; Wilson Wai-San Tam, PhD

[1]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, 10 Medical Dr, Singapore, Singapore

[2]School of Nursing, Midwifery and Social Sciences, Central Queensland University, Rockhampton, Australia

[3]Department of Nursing, School of Nursing and Allied Medical Sciences, Holy Angel University, Angeles City, Philippines

Corresponding Author:
Xin Ming Deng, BSc
Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, 10 Medical Dr, Singapore, Singapore

Abstract

Background: Home-based hospice care offers patients with terminal cancer the comfort of receiving care in a familiar environment while enabling family members to provide personalised support. Despite the critical role families play, the literature remains underexplored in terms of their experiences, needs, and perceptions. A robust qualitative synthesis is needed to inform improvements in palliative care services.

Objective: This meta-synthesis aims to systematically review and synthesize qualitative evidence regarding the experiences, needs, and perceptions of family caregivers in home-based hospice care for patients with terminal cancer. The goal is identifying key themes that can improve caregiver support and service delivery.

Methods: A systematic search was conducted across MEDLINE, Embase, Scopus, PsycINFO, CINAHL, Google Scholar, and relevant gray literature sources up to March 14, 2025. Studies were included if they focused on family caregivers' experiences in home-based hospice care settings, excluding those that addressed only patients or health care providers. Two independent reviewers performed study selection, data extraction, and quality assessment using the Critical Appraisal Skills Programme checklist. Data were synthesized using a 3-step thematic synthesis approach, and the confidence in the findings was assessed via the GRADE-CERQual (Grading of Recommendations Assessment, Development, and Evaluation–Confidence in the Evidence from Reviews of Qualitative Research) framework.

Results: Five studies published between 1989 and 2022 from diverse geographical regions (including Asia and Western settings) met the inclusion criteria. Two major themes emerged: (1) being physically and emotionally present, where caregivers expressed a strong commitment to remain with their loved ones, providing emotional support and maintaining a sense of control; and (2) sharing responsibilities, which underscored the importance of both formal support from palliative care teams and informal support from family and friends in mitigating caregiver burden. These findings directly address the study's aims by illustrating how caregivers balance emotional commitment with the practical challenges of providing home-based care.

Conclusions: Although family caregivers are dedicated to delivering high-quality, personalized care, they encounter significant emotional and logistical challenges. Variability in study settings, potential recall bias from retrospective interviews, and limited gray literature access may affect the generalizability of the findings. This meta-synthesis underscores the essential role of family involvement in home-based hospice care for patients with terminal cancer. The combined reliance on emotional commitment and shared responsibilities—with support from professional care teams—is vital for optimal care delivery. Future interventions should enhance formal and informal support systems to meet family caregivers' diverse needs better.

Trial Registration: PROSPERO CRD42023486012; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023486012

(JMIR Cancer 2025;11:e71596) doi:10.2196/71596

KEYWORDS

palliative; hospice; home care; cancer; meta-synthesis.

Introduction

Background

Palliative care is an active, holistic approach aimed at relieving severe or chronic suffering and enhancing the quality of life for individuals with life-threatening illnesses at any stage of the illness trajectory, from diagnosis onward, whether curative treatments continue or not [1-3]. Hospice care, however, refers explicitly to palliative care provided when patients no longer pursue curative treatments, typically with a prognosis of 6 months or less, focusing on comfort, dignity, and quality of life during the end-of-life stages [2].

Cancer remains the second leading cause of death globally, with continually rising incidence rates each year [4]. Advances in cancer treatment and early detection have contributed to prolonged survival, even for patients in advanced stages. Nevertheless, extended survival often results in protracted suffering, posing significant physical, emotional, social, and spiritual challenges for both patients and their families, especially during the end-of-life phase [5]. Terminal cancer typically refers to patients with cancer with a prognosis of 6 months or less to live, at which point curative treatments are usually ceased and care transitions toward symptom management and comfort measures [2].

Home-based hospice care allows patients with terminal cancer to spend their final days at home, as many of them prefer [6,7]. This care typically comprises scheduled visits by health care professionals and 24-hour on-call support rather than continuous, around-the-clock in-home care [8]. Such services often include symptom management, holistic nursing care, and psychosocial and spiritual support tailored to individual family needs [9,10]. The popularity of home-based hospice care is increasing due to multiple factors, such as overcrowded hospital environments [11], rising complexity of symptoms and treatments [12], improvements in living standards and education [13], and a growing emphasis on maintaining quality of life at the end of life [14]. The aging population and a shift toward value-based care models have also contributed to the rising demand for home-based hospice services [15].

Family caregivers are pivotal in home-based hospice care, often providing daily care and managing multifaceted emotional and logistical responsibilities. In many cultural contexts, caregiving is perceived as a moral or filial obligation, significantly influenced by cultural norms and values that shape caregiver expectations and decision-making processes [16]. For example, in many Asian cultures, caregiving at home is deeply rooted in filial piety, emphasizing familial responsibility and moral duty toward elders [16,17]. Understanding caregivers' culturally influenced experiences, attitudes, perceptions, and unique support needs is essential for effective and culturally competent interventions.

Given these considerations, home-based hospice care is experiencing increasing demand and attention, primarily due to its valuable support for family caregivers who assume multifaceted responsibilities involving intensive physical care, emotional support, and complex decision-making. These caregivers' experiences, perceptions, and needs vary significantly, influenced by personal, cultural, and contextual factors. Therefore, understanding caregivers' perspectives is essential for assessing the effectiveness of current hospice services and identifying opportunities for enhancing family support at home, ultimately leading to improved patient and caregiver experiences during this critical period.

Objectives

The overall aim of this study was to update and synthesize qualitative research on home-based hospice care based on the experiences of family caregivers of patients with cancer.

The three specific objectives for this review were as follows:

1. To explore the experiences of families of patients with terminal cancer receiving home-based hospice care.
2. To examine attitudes and perceptions of families toward home-based hospice care.
3. To identify key needs within the context of home-based hospice care services.

This meta-synthesis seeks to address a significant gap in the current literature by conducting a comprehensive review of the experiences of family caregivers supporting patients with terminal cancer in home-based hospice care settings. By examining the caregiver experience, this study aims to assess whether existing palliative care provisions sufficiently meet their needs and provide insights for future improvements. Ultimately, this research will ensure that caregivers receive holistic and compassionate support during this critical phase of the illness trajectory.

Methods

Overview

The qualitative evidence from primary qualitative studies and mixed-methods studies were synthesized and integrated using the thematic synthesis method. The meta-synthesis protocol was reported following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist [18] (Checklist 1). The meta-synthesis was guided by the 6 steps of qualitative research synthesis developed by Major and Savin-Baden [19], including formulating the questions, identifying studies, selecting studies and extracting data, appraising studies, synthesizing and finalizing data, and reflecting upon the process, which was based on the step-by-step qualitative research synthesis approaches by Noblit and Hare [20] and Sandelowski and Barroso [21]. The study protocol has been registered in PROSPERO (Prospective Register of Systematic Reviews) under registration number CRD42023486012.

Eligibility Criteria

Identifying appropriate studies is crucial in alignment with step 2 of the Major and Savin-Baden [19] approach. Essential components in this identification process include (1) the scope of included studies, (2) inclusion and exclusion criteria, (3) quality assessment, (4) data synthesis method, and (5) criteria for reporting findings [22,23].

The criteria for considering studies for this review were based on the PICo (Population, Interest, Context) framework [24]. The inclusion criteria were as follows: (1) population: the studies involving adults who are family members of patients with terminal cancer; (2) interest: the experiences, attitudes, and needs regarding home-based hospice care services; (3) context: under the care of home-based hospice care service, particularly those with a physical home visit; and (4) the research design was qualitative or mixed methods. The exclusion criteria were as follows: (1) studies focusing on patients themselves, health care providers, or non–home-based hospice services (such as inpatient hospice or nursing home care), and studies involving only telemedicine visits, and (2) language is not English. The eligibility criteria are outlined in Multimedia Appendix 1.

Information Sources

We searched 5 electronic databases—Scopus, Embase, MEDLINE, CINAHL, and PsycINFO—from inception until March 14, 2025. These databases were chosen for their relevance to qualitative research in various health care settings [25]. To maximize the range of articles retrieved, we searched Google Scholar and gray literature sources, including ProQuest, for unpublished dissertations. The complete search strategy can be found in Multimedia Appendix 2.

Search Strategy

The search began by defining the scope of the study and addressing the research questions focusing on terminal cancer, home-based care, and palliative care. Broad search terms and synonyms were used to create a comprehensive search string encompassing all relevant keywords. The PICo framework [24] was used to guide search term generation. The full search strategies included an initial search of MEDLINE, Embase, Scopus, PsycINFO, and ProQuest up to September 13, 2023. Following our initial database search, we conducted an additional search in CINAHL on March 13, 2025, and Google Scholar on March 14, 2025, to ensure comprehensive coverage of relevant literature.

Controlled vocabulary terms were used for each database: Medical Subject Headings (MeSH) for PubMed and Emtree terms for Embase. Boolean operators and truncation symbols combined the terms according to each database's specifications. This initial step helped us develop effective search strategies, become familiar with the terminology, and conduct preliminary searches. We then consulted a subject librarian to further refine these search terms and strategy. Subsequently, a formal literature search was conducted to identify and compile eligible studies, with language limited to English only.

Selection Process

Adhering to the third step of the Major and Savin-Baden [19] meta-synthesis method, we used EndNote 21 (version 21.2.0.19537; Clarivate Plc) [26] to import articles and find duplicates. Subsequently, the titles and abstracts of the imported articles were screened for relevance using RAYYAN [27], using the blinding function to mitigate bias. This initial screening, followed by full-text screening, was independently conducted by 2 reviewers (XMD and KH), with discrepancies resolved through discussion to reach a consensus.

Data Extraction

As the third step of Major and Savin-Baden's [19] meta-synthesis approach, we meticulously extracted the qualitative data from the included articles. This process was conducted in 2 stages. First, a pilot test of the data-extraction form was performed by 2 reviewers (XMD and KH) and validated by the third reviewer (WWST) before extracting relevant information. Subsequently, the data-extraction form was applied, which included author, year of publication, study setting, aim, sample characteristics, methodology (population characteristics, sampling method, data collection, and data analysis), and key findings.

Quality Assessment

For the quality appraisal, 2 investigators (XMD and KH) independently assessed each included study using the Critical Appraisal Skills Programme (CASP) checklist (2019) [28]. All discrepancies between the 2 investigators were resolved through discussion. The third reviewer (WWST) counterchecked the results to ensure accuracy and consistency. This assessment included statements of research aims, appropriate qualitative methodology, research designs, recruitment strategies, data collection, adequate relationship between researcher and participants, ethical issues consideration, the rigor of data analysis, statement of findings, and value of the study.

Synthesis Methods

The data analysis used the 3-step thematic synthesis method [25] as the fifth step in the Major and Savin-Baden [19] approach. This method, based on Braun and Clarke's thematic analysis techniques [29], was extensively used in nursing and medical research to identify intervention needs, appropriateness, acceptability, and factors influencing implementation. Thematic synthesis integrates findings from primary studies to identify prominent or recurrent themes within the relevant literature.

In the initial step of thematic synthesis, findings were extracted and coded line-by-line using Excel (Microsoft Corp). Reviewer XMD conducted the line-by-line coding, maintaining fidelity to the data and preserving the original concepts. Reviewer KH subsequently verified the codes to ensure alignment and completeness, facilitating identifying and categorizing key elements within the data.

The primary codes were then grouped based on conceptual similarities upon mutual agreement by reviewers (XMD and KH), resulting in a structured interpretation of the findings through the development of descriptive themes. These descriptive themes were subsequently synthesized into higher-level analytical themes. These analytical themes represented the key outcomes relevant to our meta-synthesis topic, achieved by merging and summarizing similar descriptive themes to highlight core insights and conclusions drawn from the data. The final process and results were screened thoroughly and confirmed by all reviewers (XMD, KH, WWST, and VL), with any disagreements resolved through consensus.

Third, the use of the approach by Major and Savin-Baden [19] involved 5 steps; the last was adopting the 3-step thematic synthesis method [25]. This nursing and medical method was

used massively using Braun and Clarke's thematic analysis methods [29] to determine intervention needs, appropriateness, acceptability, and factors regarding implementation.

Confidence Measurement

After generating the analytical themes, a further quality appraisal stage using the GRADE-CERQual (Grading of Recommendations Assessment, Development, and Evaluation–Confidence in the Evidence from Reviews of Qualitative Research) approach [30] was used to evaluate the confidence level of our findings (Table 1). This aligned with the final step of Major and Savin-Baden's [19] approach of "reflecting upon the process."

Table . Evidence profile table.

Major theme and summarized review finding	Methodological limitations	Coherence	Adequacy	Relevance	GRADE-CERQual[a] assessment of confidence	References
Being physically and emotionally present						
• Belief that home palliative care provides better care than hospitals • Commitment to care at home • Cultural and moral obligations • Personal reflections and challenges	No or very minor concerns	No or very minor concerns	Minor concern	Minor concerns	High confidence	[16,31-34]
Sharing responsibilities						
• Challenges in caregiving • Formal support needs • Informal support needs	No or very minor concerns	No or very minor concerns	No or very minor concerns	Minor concerns	High confidence	[16,31-34]

[a]GRADE-CERQual: Grading of Recommendations Assessment, Development, and Evaluation–Confidence in the Evidence from Reviews of Qualitative Research.

Two reviewers (XD and KH) conducted independent reviews and discussed discrepancies to reach a consensus. The GRADE-CERQual approach evaluates confidence based on 4 components: methodological limitations, coherence, adequacy of data, and relevance of included studies. Each element was categorized as having "no or very minor concerns," "minor concerns," "moderate concerns," or "serious concerns," leading to varying grades of confidence [35].

The assessment of methodological limitations aligns with the CASP appraisal, evaluating the trustworthiness of study findings by examining the appropriateness of the research methodology, which is closely related to the quality of the results [36]. Our assessment indicated that the findings were supported by articles with no to very minor concerns regarding methodological limitations.

The coherence assessment measures the relevance of the data from the included studies to the review findings [37]. Based on key findings extracted from the included studies, we confirmed that the data were relevant to the review findings (Table 2).

To assess the adequacy of the data, we evaluated both the quantity and richness of the data in relation to the review findings, in line with the GRADE-CERQual approach [38]. Studies with limited data, particularly those with findings supported by only 1 or 2 participant voices, were noted as having insufficient depth and quantity to robustly support specific review findings [31,39].

Lastly, we assessed relevance by evaluating the extent to which the data from the primary studies apply to the context outlined in the synthesis results [32].

Table . Characteristics of included studies.

Study	Albert et al [33]	Milberg and Strang [34]	Hull [32]	Lee et al [16]	Barlund et al [31]
Setting	Malaysia, palliative care center, Kota Kinabalu, Sabah	Sweden, hospital-based home care	United States, combined hospice program, home care	Taiwan, hospice home care, northern Taiwan	Norway, municipalities, Førde Central Hospital, Sogn og Fjordane
Aim	To explore the suffering experienced by Malaysian family members caring for patients with advanced cancer nearing end-of-life.	To describe and interpret comprehensibility and manageability experiences of informal caregivers of advanced patients with cancer in palliative home care.	To explore hospice home care experiences and perceptions of family caregivers of dying relatives.	To examine family experiences and needs when providing hospice home care to older adults with terminal cancer.	To explore factors influencing caregivers' sense of security and facilitators for home deaths among dying patients with cancer.
Sample	First-degree relatives living with patients; primary caregivers for ≥8 hours/day	Primary caregivers at home of patients with cancer receiving hospital-based home-based palliative care	Primary home caregivers; adults living at home (with/without patient responsibilities); adults living outside home with regular care duties	Caregivers of patients with advanced cancer receiving home hospice care	Parents, children, or spouses of deceased patients with cancer
Method	Purposive sampling; semistructured in-depth interviews; thematic analysis	Maximum variation sampling; semistructured interviews; qualitative hermeneutic approach	Convenience sampling; semistructured interviews and field observation; thematic analysis	Consecutive sampling; in-depth semistructured interviews; qualitative inductive content analysis	Purposive sampling; semistructured in-depth interviews; thematic analysis
Key findings	• Empathic suffering: Witnessing functional decline; fear of discomfort; receiving bad news; duties. • Powerless and hopeless suffering • Predictive suffering • Compliance suffering: Burden of caregiving and social • Barriers' wrath: Patient-related barriers; family-related barriers; health care–related hurdles.	• Comprehensibility: Congruent inner reality through open information, symbols, basic life assumptions, previous knowledge • Manageability: Togetherness/isolation involving power, competence, accessibility, and support.	• Confronting reality: Acute health changes, physician prognosis, treatment refusal. • Surveying options: Gathering and reviewing alternatives, clarifying values, learning about hospice. • Immersion: Disengaging from other tasks, shifting priorities, acquiring patient care skills. • Refocusing	• Hoping for cure: Concealing diagnoses from patients; expectations for prolonged life. • Fluctuating emotions: Positive, negative, and difficult emotions. • Accepting death: Fulfilled duties, acceptance. • Perceptions of a good death: Smooth, painless, peaceful. • Needs: Emotional support, information.	• Personal factors • Health care professionals • Organizational factors

Ethical Considerations

This meta-synthesis did not involve primary data collection, and thus ethical approval was not required. However, ethical rigor was maintained by ensuring that all included studies adhered to standard ethical guidelines, such as obtaining informed consent from participants and safeguarding participant confidentiality. Additionally, no unpublished or personally identifiable information was included in the synthesis. All sources were appropriately cited, and transparency regarding any potential conflicts of interest or funding has been maintained throughout the study.

Positionality Statement

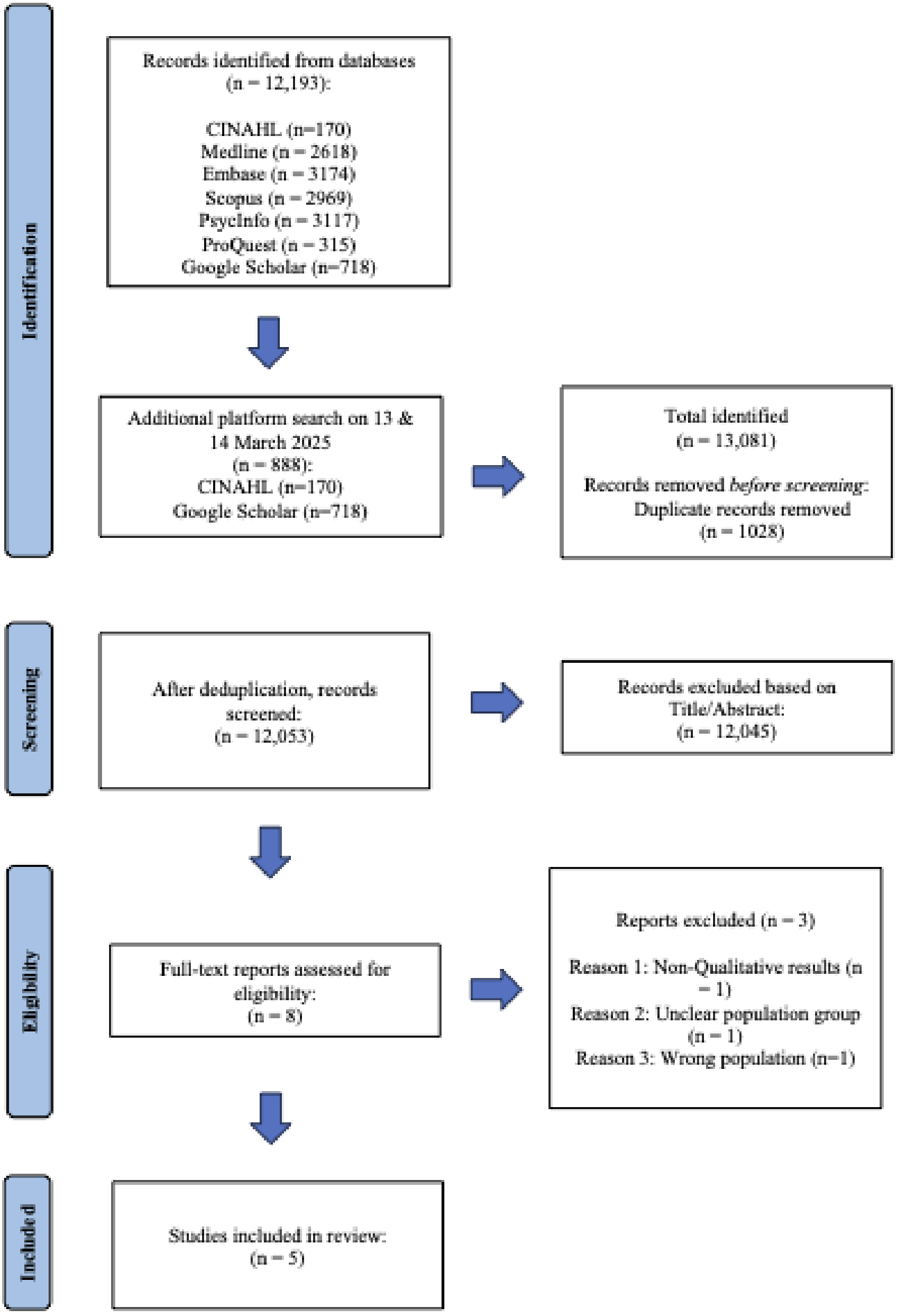
The reviewers overseeing the analysis and synthesis process (VL and WWST) are experienced qualitative researchers. The initial drafting, coding, and synthesis were primarily conducted by 2 researchers (XMD and KH) who have clinical backgrounds relevant to the review topic. All reviewers acknowledged that their personal characteristics, values, and beliefs, shaped by their clinical and academic experiences, could influence the synthesis process. Thus, regular discussions were held among the team to examine assumptions, challenge interpretations, and ensure a balanced and rigorous analysis of the findings to enhance reflexivity and reduce potential bias.

Results

Study Selection

A total of 13,081 articles were identified through our comprehensive search. This included 12,193 articles from the initial search and an additional 888 articles from the supplementary search. A total of 1028 duplicates were removed. Then, 12,053 titles and abstracts were reviewed, and 12,045 were excluded according to the eligibility criteria. Subsequently, 8 articles were read in full text, and 3 studies were excluded. Two articles were excluded due to the inability to confirm the cancer stage or to separate qualitative findings within mixed method studies, and attempts to contact the authors for verification were unsuccessful. An additional article was excluded after discussion because it predominantly contained patients' findings, which could not be distinctly separated to focus solely on caregivers' perspectives. Finally, 5 full-text articles met the eligibility criteria. The result of the selection process is presented in Figure 1.

Figure 1. PRISMA (2020) flowchart of search results and study selection. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Study Characteristics

The 5 studies included in this meta-synthesis were published between 1989 and 2022, with 2 studies [31,33] published in the last 5 years. The research was conducted across diverse geographical settings: Malaysia (n=1), Sweden (n=1), the United States (n=1), Taiwan (n=1), and Norway (n=1). Study settings varied, including a palliative care center, hospital-based home care, hospice programs, and municipal or hospice-affiliated home care services.

Sample sizes across the studies ranged from small, focused groups of 12 to larger caregiver cohorts of up to 44 participants. Most studies involved between 12 and 19 participants, while one study included a substantially larger sample. One study did not report the exact number of participants. These variations reflect the diversity in study aims and sampling strategies, as well as the depth of qualitative inquiry. All participants were family caregivers of individuals with advanced cancer receiving home-based palliative care.

Sampling strategies included purposive sampling (n=3), maximum variation sampling (n=1), and consecutive or convenience sampling (n=2). Data were primarily collected through semistructured, in-depth interviews (n=5), with one study also incorporating field observation. Thematic analysis was the predominant analytical approach, although one study used a qualitative hermeneutic method and another used inductive content analysis.

Although the articles had slightly different objectives, all contributed to our understanding of the experiences, needs, and perceptions of families of patients with terminal cancer in home-based hospice care. Sampling and data collection methods varied across the studies due to population and setting differences. Four of the 5 studies were qualitative, with Hull [32] being the exception, including a quantitative portion as part of the dissertation. However, this quantitative aspect did not influence our findings, as the qualitative portion was clearly differentiated and extracted.

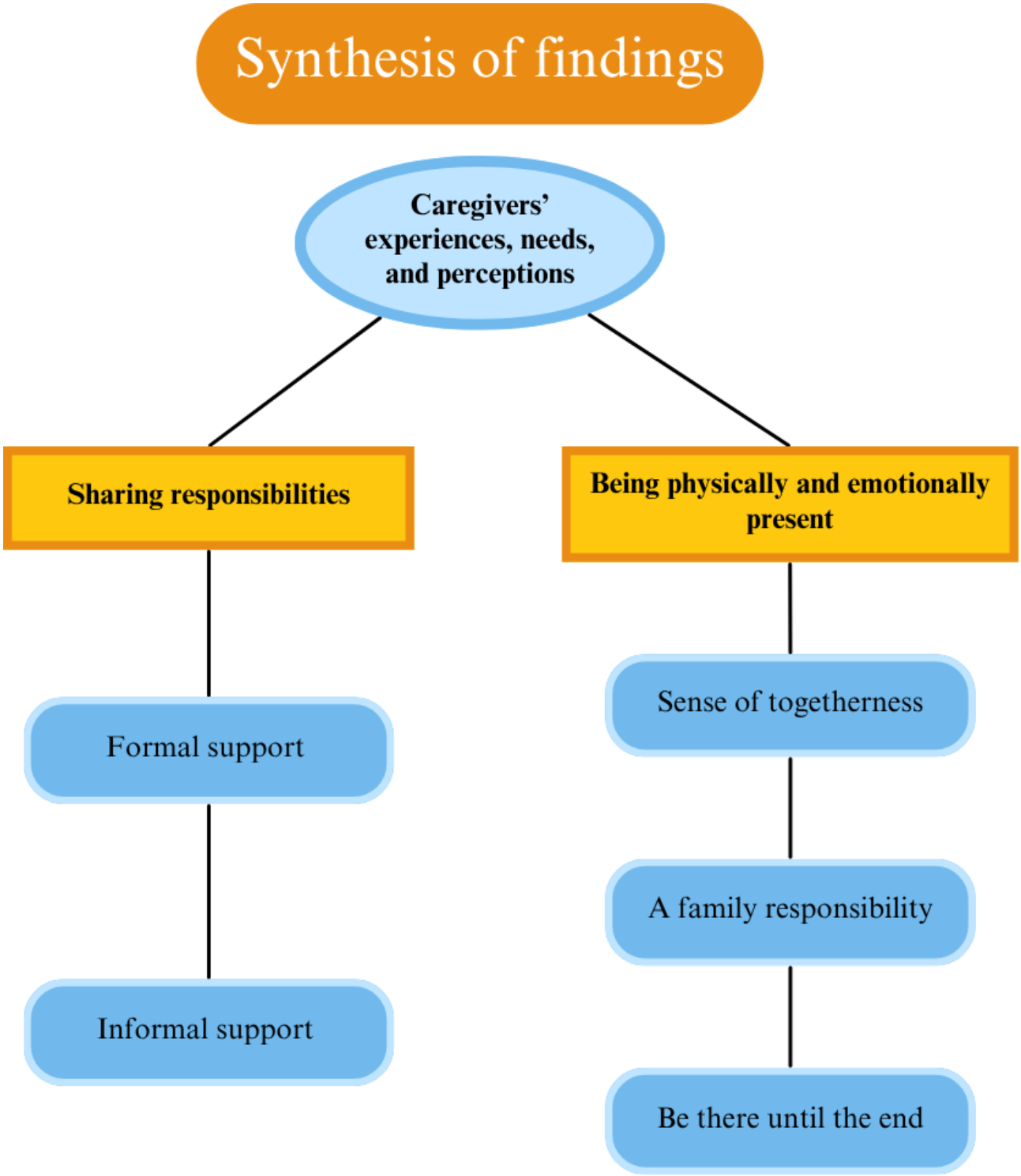
We noticed that one significant contributing study was notably outdated. After thorough discussion between 2 reviewers (XMD and KH), it was decided to include this study because it provided findings that remain relevant and fit well within the overall research context. Despite being published over 30 years ago in 1989 [32], the foundational concepts and findings still offer valuable insights into home-based hospice care and address aspects that are not sufficiently covered by more recent studies. The inclusion of this study ensured a comprehensive understanding of the evolution and continuity of caregiving practices. The characteristics of the included studies are presented in Table 2.

Synthesis Findings

Overview

A total of 17 groups of primary codes and corresponding descriptive themes were identified. These were further synthesized into 5 overarching analytical themes, culminating in 2 central themes that represent the final synthesis. These include (1) being physically and emotionally present and (2) sharing responsibilities. The 3 subthemes that support the main theme of “being physically and emotionally present” were a sense of togetherness, family responsibility, and being there until the end. Similarly, for the main theme of “sharing responsibilities”, we synthesized the findings based on 2 subthemes: formal support and informal support (Figure 2).

Figure 2. Synthesis of findings, including subthemes and main themes.



Family caregivers of patients with terminal cancer made a deliberate commitment to be physically and emotionally present to care for their loved ones at home. They were motivated by the belief that they could provide a higher standard of care than what was offered in the hospital, where they observed their dying relatives being treated more as objects than as individuals with dignity. One participant expressed this sentiment: "When my dad was in the

hospital, they bathed him, fed him, and gave him his meals, and he just lay there. After these, nobody bothered him, nobody did anything more. He was nothing" [32]. At home, caregivers felt they could better meet their loved ones' needs, taking comfort in the familiar environment and the reduced stress for both patient and caregiver. This approach offered a sense of control and personal involvement. For example, a 64-year-old daughter noted: "It didn't take long to learn how to care for Mum. Being a mother and a wife, it was easy enough to pick up those skills, so there were no problems. I could go into nursing now (laughs); I've had practical experience" [32]. Similarly, a 75-year-old wife reflected on her experience: "There was no issue for me in knowing how to care for my husband. I kept him clean, the

bed clean, and his pyjamas were fresh every day. It was really no problem for me" [32]. In another instance, a 35-year-old son demonstrated his capability by innovatively managing bedsores using a combination of a water mattress and an alternating mattress [32].

The 3 subthemes supported this theme: a sense of togetherness, a family responsibility, and being there until the end.

Sense of Togetherness

A sense of togetherness is a crucial emotional factor in home-based hospice care. Being at home made caregivers and patients feel safe and secure, knowing they were close and always within reach. One participant shared: "It was important for me to be with him... to be there all the time. And it wasn't difficult; it felt natural. It was safe for me to be there with him" [31]. Another participant described the comfort of staying connected through a monitor: "I always had the monitor with me, even when I was out front talking to a neighbour. I kept my little walkie-talkie with me so I could always keep tabs on him" [32]. Being at home symbolized the caregivers' commitment to staying close to the patient in the home they had built together and filled with memories, despite the good or bad times. One participant poignantly remarked:

> *This is our home, and this is where she should be as long as she lives. She's receiving better care here than she would anywhere else, and as long as I'm here, she'll have that care as long as it's needed. This is where she belongs. This is her life's work. We raised our family here for 38 years. I lived through the good years; now I want to live through the bad. What's one bad year out of 47? [32].*

A Family Responsibility

From a cultural perspective, caregiving is a natural extension of one's commitment to loved ones and a moral obligation. For many children caregivers, taking on the responsibility of caring for their parents with terminal illness at home is deeply rooted in cultural and ethical values, often viewed as a way to repay their parents. In this cultural context, there is a strong belief in managing care independently rather than relying on nursing homes to express filial duty. This sentiment is reflected in the words of a daughter from Taiwan: "Because my father's condition was so bad, it made my heart ache to watch him suffering from pain...I wanted to care for my father on my own and did not need anyone else to bear my responsibility" [16]. Another participant echoed this sense of responsibility: "The most important thing for me to learn is that you have to give yourself an opportunity to take care of your sick parents who have taken care of you since you were a child...I was really proud of becoming familiar with the skills of care" [16]. This highlights the emotional significance of caregiving, which carries a deep sense of responsibility. Many caregivers, particularly those who were the sons or daughters of patients with cancer, believed they had fulfilled their filial duties and wanted their loved ones to have a natural death and a proper funeral. As one son expressed: "We will not have regrets if our father dies tomorrow...We believe we demonstrated the value of filial piety as much as we could by taking good care of him at home without sending him to a nursing home...we knew he wanted to die at home" [16].

Be There Until the End

However, families often face significant emotional challenges when discussing topics related to death or illness while striving to be emotionally present for their loved ones. In Sweden, being present was crucial, even when it was difficult to discuss such matters. One participant shared: "She [mother] is talking a bit too much about the funeral and such things. I find it quite burdensome. But it is good for her, so we talk" [34]. Some family members chose to hide terminal diagnoses from their loved ones in an attempt to protect them from losing hope, believing that revealing the full extent of the illness might cause them to give up on life [16]. In Malaysia, caregivers also expressed a strong desire to remain with their loved ones and provide care until the end, keeping them happy and shielding them from their emotions. One caregiver shared: "I do not want him to see me crying. I want him to be happy. I know that I am the only one next to him. If I were there with a sad feeling...I think it might make him sad" [33]. The complexity of providing compassionate care while managing one's emotional burdens is evident in these experiences, underscoring families' challenges in navigating the end-of-life journey.

Sharing Responsibilities

Sharing responsibilities in caregiving can involve the patient, family, friends, home palliative care teams, and other community organizations. This collaborative approach helped family caregivers feel less isolated when managing challenging situations. Throughout the caregiving process, these groups play a vital role in supporting family caregivers of patients with terminal cancer. It is essential to recognize that family caregivers often require various types of support. This theme was further supported by 2 subthemes: support from a palliative home-based care team and support from others.

Formal Support

The involvement of a palliative home care team can help caregivers perceive caregiving as more manageable [34], boosting their confidence in providing care for their loved ones at home [16]. A Swedish participant shared a poignant experience:

> *My husband was going to be discharged from the hospital [with a percutaneous endoscopic gastrostomy], but I said I can't take care of such things. But they said it was very easy to learn how to use it. 'No,' I said. 'I can't take that responsibility'...Then the dietician came and said, 'It is so easy.' I felt I was going to be ill because I could not do this. (Sighs) And later on that afternoon, I had diarrhoea. I was not feeling well and was terribly worried...Then the palliative doctor came, and he was almost like an angel. He presented all the things the palliative home care team could offer. And then I felt that this was a support for us [34].*

The need for qualified health care professionals is often emphasized, as they are the support personnel to whom caregivers can express their concerns, thoughts, and worries.

This interaction gave caregivers confidence and security, reassuring them that they appropriately fulfilled their caregiving role. Caregivers welcome palliative care nurses or coordinators, as one participant reported:

> *"It felt safe and secure for us to know that they were visiting us (ie, home nursing care)" [31].*

Some caregivers acknowledged that having palliative home care made it easier to fulfill their wish to care for patients at home during the final stage of life, facilitating a dignified end-of-life experience. As one son reflected:

> *"I thought my mother's death was a good death because she passed away without pain or any distressing symptoms from cancer. It was really important for us, and we appreciated what the hospice home-based care team did for us" [16].*

Informal Support

Other family members, relatives, and friends play a crucial role in supporting family caregivers who provide care for patients with terminal cancer. In families with children, caregivers were glad to share their responsibility with their offspring, viewing it as an expression of filial piety. One participant said:

> *"I was glad to see my son helping me care for his dad...I thought that if I'm sick someday he would care for me like now and our relationship gave me the energy to care for my husband" [16].*

Managing the financial responsibilities of caregiving often necessitates sharing the burden with other family members or government organizations. The emotional and physical toll of caregiving can be overwhelming, as one participant poignantly expressed:

> *"I declined the offer of attendance allowance; I wanted another sister to do this. I didn't want to be alone in this [ie, follow the patient in the last phase]. I wanted more people to be involved because I had... (Sighs) I had been doing this alone for so long" [31].*

This highlights the deep need for collective involvement, especially in the final stages of care, which can be particularly draining when borne alone.

Quality Appraisal of Included Studies

The methodological quality of the included studies was assessed using the CASP Qualitative Checklist. All 5 included studies demonstrated generally high methodological quality, with scores ranging from 29 to 30 points out of a possible maximum score of 30 points (Table 3). All studies clearly stated their research aims, adopted appropriate qualitative methodologies, applied suitable research designs, and used adequate recruitment strategies and data collection methods. Four studies clearly considered ethical issues, while 2 lacked explicit discussion regarding the relationship between researchers and participants, introducing minor ambiguity. Nevertheless, all included studies clearly articulated their findings and demonstrated valuable contributions to the topic. A detailed summary of the quality appraisal results is presented in Table 3.

Table . Quality appraisal of the included studies using the Critical Appraisal Skills Programme (2019).

	Albert et al, 2022 [33]	Milberg and Strang, 2004 [34]	Hull, 1989 [32]	Lee et al, 2014 [16]	Barlund et al, 2021 [31]
Was there a clear statement of the aims of the research?	+	+	+	+	+
Is a qualitative methodology appropriate?	+	+	+	+	+
Was the research design appropriate to address the aims of the research?	+	+	+	+	+
Was the recruitment strategy appropriate to the aims of the research?	+	+	+	+	+
Was the data collected in a way that addressed the research issue?	+	+	+	+	+
Has the relationship between researcher and participants been adequately considered?	+	+/–	+	+	+/–
Have ethical issues been taken into consideration?	+	+	+/–	+/–	+
Was the data analysis sufficiently rigorous?	+	+	+	+	+
Is there a clear statement of findings?	+	+	+	+	+
Is the research valuable?	+	+	+	+	+
Total points	30	29	29	29	29

Confidence of Evidence

Two major themes emerged from the thematic synthesis: (1) being physically and emotionally present and (2) sharing responsibilities. Both themes were assessed with high confidence according to the GRADE-CERQual framework (Table 1). The theme “being physically and emotionally present” had no or very minor concerns regarding methodological limitations and coherence, minor concerns related to data adequacy, and minor concerns regarding relevance due to the partial inclusion of studies addressing home- and hospital-based settings. Similarly, the theme “sharing responsibilities” had no or minor problems related to methodological limitations, coherence, and adequacy, with minor concerns regarding relevance, given the partial relevance of 2 included studies that covered broader contexts beyond home-based care. These results demonstrate robust qualitative evidence reflecting the experiences, perceptions, and needs of family caregivers in home-based hospice care for patients with terminal cancer.

Discussion

Principal Findings

This review synthesized findings from 5 qualitative studies exploring family caregiver experiences, perceptions, and needs in home-based hospice care for patients with terminal cancer. Two prominent themes emerged: (1) being physically and emotionally present, highlighting caregivers’ dedication and the emotional complexities involved in caregiving, and (2) sharing responsibilities, demonstrating the importance of formal and informal support systems. These themes reflect the complex emotional and practical challenges caregivers encounter while striving to provide high-quality care aligned with their loved ones’ wishes.

Comparison to Prior Work

Our findings align with previous research highlighting family caregivers’ essential role in delivering compassionate and dignified end-of-life care [40]. The emotional dedication caregivers demonstrate, often deeply embedded in cultural expectations such as filial piety, confirms existing literature emphasizing the profoundly personal nature of caregiving.

However, this synthesis also underscores specific challenges faced by caregivers, including managing difficult conversations about death, balancing caregiving responsibilities, and navigating cultural norms, consistent with prior research identifying caregiver stress and potential burnout risks.

These findings align with established social support theories, confirming that emotional, informational, and instrumental support from formal (health care professionals) and informal networks (family and friends) substantially alleviates caregiving stress [41]. Nonetheless, our analysis revealed a significant gap in integrating these support systems, leaving caregivers vulnerable to isolation and overwhelm.

Strengths and Limitations

The strength of this review includes the rigorous methodological approach followed, adherence to PRISMA guidelines, quality appraisal using the CASP checklist, and confidence assessment using GRADE-CERQual. Including diverse cultural contexts from Asian and Western countries further enhances the generalizability and applicability of our findings.

However, this review has several limitations. First, limiting the search to studies published in English may have restricted the inclusion of research conducted in non–English-speaking regions, thereby reducing the diversity in cultural contexts and settings across the included studies and potentially affecting the generalizability of the findings [42]. Second, despite the growing global reliance on family caregivers in home-based hospice care [43], this review identified only 5 eligible studies, with just 2 published in the past 5 years. This limited number may reflect the specificity of our inclusion criteria, which focused solely on caregivers of patients with terminal cancer receiving home-based palliative care. Many existing end-of-life care studies include a broader population with varied diagnoses and may not isolate the unique caregiving experience related to cancer. This scarcity highlights a critical gap and the need for more focused, culturally diverse qualitative research in this area [44]. Third, the retrospective nature of some studies introduced a potential recall bias, affecting the accuracy of reported experiences. This could be addressed in future research through prospective study designs. Fourth, our review's reliance on specific databases potentially limited the comprehensiveness of identified literature, mainly gray literature. Future studies should broaden database searches and proactively include unpublished literature to enhance comprehensiveness. Lastly, including older literature (more than 5 years old) may limit the review's alignment with the most current evidence. However, older studies were included due to their foundational insights, which remain relevant to the current practice context. Future research should emphasize more recent publications to align closely with contemporary practices and emerging evidence.

Future Directions

The findings indicate several avenues for future research and practice. There is a clear need to develop and evaluate culturally sensitive interventions tailored to both formal and informal caregiver support needs. Initiatives should aim to bridge existing gaps in caregiver support through integrated services that alleviate isolation and promote emotional and practical caregiving capacities. Structured educational programs and support groups designed to improve communication around end-of-life topics could also substantially enhance caregiver experiences.

Conclusion

Home-based hospice care has a significant impact on patients, but the experience of family members who support them remains neglected in literature and daily practices. The results highlight both the emotional rewards and daunting challenges caregivers encounter and point to the need for systemic, culturally competent strategies to support this population. By bridging the gaps within formal and informal support systems and encouraging open communication, these health care providers can enable caregivers to maneuver their roles effectively. These findings present critical implications for the delivery of hospice care as the world learns to provide holistic, compassionate care, without fail, for patients and their families. Future research should investigate diverse caregiver experiences and further inform the refinement of focused interventions to improve home-based hospice care.

In conclusion, this review highlights the critical role of caregivers in home-based hospice care for patients with terminal cancer, emphasizing the need for culturally competent, comprehensive caregiver support strategies. Addressing identified gaps can significantly improve caregivers' experiences and ultimately enhance the quality of hospice care services delivered at home.

Acknowledgments

We sincerely thank VL from the School of Nursing, Midwifery & Social Sciences, Central Queensland University, Australia, for her invaluable guidance. We also wish to thank KH for her contributions to data curation and formal synthesis and WWST for contributing to conceptualization, methodology, supervision, and validation.

Authors' Contributions

XMD contributed to writing – original draft, writing – review and editing, visualization, methodology, formal synthesis, data curation, and conceptualization. KH contributed to writing – review and editing, data curation, and formal synthesis. VL contributed to writing – review and editing and formal synthesis. WWST contributed to writing – review and editing, visualization, validation, supervision, methodology, formal synthesis, data curation, and conceptualization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Inclusion and exclusion criteria.
[DOCX File, 16 KB - cancer_v11i1e71596_app1.docx]

Multimedia Appendix 2
Full search strategy.
[DOCX File, 22 KB - cancer_v11i1e71596_app2.docx]

Checklist 1
PRISMA checklist. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
[DOCX File, 25 KB - cancer_v11i1e71596_app3.docx]

References

1. Radbruch L, De Lima L, Knaul F, et al. Redefining palliative care-a new consensus-based definition. J Pain Symptom Manage 2020 Oct;60(4):754-764. [doi: 10.1016/j.jpainsymman.2020.04.027] [Medline: 32387576]
2. 2023 edition: hospice facts and figures. : National Hospice and Palliative Care Organization; 2023 URL: https://allianceforcareathome.org/wp-content/uploads/NHPCO-Facts-Figures-2023.pdf [accessed 2025-06-05]
3. Palliative care. World Health Organization. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/palliative-care [accessed 2020-08-05]
4. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424. [doi: 10.3322/caac.21492] [Medline: 30207593]
5. Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012: estimated cancer incidence, mortality and prevalence worldwide in 2012. Int J Cancer 2015;136(5):E359-E386. [doi: 10.1002/ijc.29210]
6. Higginson IJ, Sen-Gupta GJ. Place of care in advanced cancer: a qualitative systematic literature review of patient preferences. J Palliat Med 2000;3(3):287-300. [doi: 10.1089/jpm.2000.3.287] [Medline: 15859670]
7. Earle CC, Park ER, Lai B, Weeks JC, Ayanian JZ, Block S. Identifying potential indicators of the quality of end-of-life cancer care from administrative data. J Clin Oncol 2003 Mar 15;21(6):1133-1138. [doi: 10.1200/JCO.2003.03.059] [Medline: 12637481]
8. What is hospice care? American Cancer Society. URL: https://www.cancer.org/cancer/end-of-life-care/hospice-care/what-is-hospice-care.html [accessed 2024-05-07]
9. Chitnis XA, Georghiou T, Steventon A, Bardsley MJ. Effect of a home-based end-of-life nursing service on hospital use at the end of life and place of death: a study using administrative data and matched controls. BMJ Support Palliat Care 2013 Dec;3(4):422-430. [doi: 10.1136/bmjspcare-2012-000424] [Medline: 24950522]
10. Luckett T, Davidson PM, Lam L, Phillips J, Currow DC, Agar M. Do community specialist palliative care services that provide home nursing increase rates of home death for people with life-limiting illnesses? A systematic review and meta-analysis of comparative studies. J Pain Symptom Manage 2013 Feb;45(2):279-297. [doi: 10.1016/j.jpainsymman.2012.02.017] [Medline: 22917710]
11. Leff B, Montalto M. Hospitals at home: patients receive care at home to ease overcrowding. CBS News. 2023. URL: https://www.cbsnews.com/news/hospitals-at-home-patients-reduce-overcrowding [accessed 2024-05-07]
12. Lustbader D, Mudra M, Romano C, et al. The impact of a home-based palliative care program in an accountable care organization. J Palliat Med 2017 Jan;20(1):23-28. [doi: 10.1089/jpm.2016.0265] [Medline: 27574868]
13. Lai YJ, Chen YY, Ko MC, et al. Low socioeconomic status associated with lower utilization of hospice care services during end-of-life treatment in patients with cancer: a population-based cohort study. J Pain Symptom Manage 2020 Aug;60(2):309-315. [doi: 10.1016/j.jpainsymman.2020.03.015] [Medline: 32240750]
14. Feliciano DR, Reis-Pina P. Enhancing end-of-life care with home-based palliative interventions: a systematic review. J Pain Symptom Manage 2024 Nov;68(5):e356-e372. [doi: 10.1016/j.jpainsymman.2024.07.005] [Medline: 39002710]
15. 2024 outlook: home health, hospice and personal care. HealthCare Appraisers. URL: https://healthcareappraisers.com/2024-outlook-home-health-hospice-and-personal-care [accessed 2024-05-07]
16. Lee HTS, Melia KM, Yao CA, Lin CJ, Chiu TY, Hu WY. Providing hospice home care to the terminally ill elderly people with cancer in Taiwan: family experiences and needs. Am J Hosp Palliat Care 2014 Sep;31(6):628-635. [doi: 10.1177/1049909113499603] [Medline: 23921288]
17. Li WW, Singh S, Keerthigha C. A cross-cultural study of filial piety and palliative care knowledge: moderating effect of culture and universality of filial piety. Front Psychol 2021;12:787724. [doi: 10.3389/fpsyg.2021.787724] [Medline: 34925189]

18. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
19. Major C, Savin-Baden M. An Introduction to Qualitative Research Synthesis: Managing the Information Explosion in Social Science Research: Routledge; 2010. [doi: 10.4324/9780203497555]
20. Noblit GW, Hare RD. Meta-Ethnography: Synthesizing Qualitative Studies: Sage Publications; 1988.
21. Sandelowski M, Barroso J. Handbook for Synthesizing Qualitative Research: Springer Publishing Company; 2007.
22. Campbell R, Pound P, Pope C, et al. Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. Soc Sci Med 2003 Feb;56(4):671-684. [doi: 10.1016/s0277-9536(02)00064-3] [Medline: 12560003]
23. Atkins S, Lewin S, Smith H, Engel M, Fretheim A, Volmink J. Conducting a meta-ethnography of qualitative literature: lessons learnt. BMC Med Res Methodol 2008 Apr 16;8:21. [doi: 10.1186/1471-2288-8-21] [Medline: 18416812]
24. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012 Oct;22(10):1435-1443. [doi: 10.1177/1049732312452938] [Medline: 22829486]
25. Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions, Version 6.4: Cochrane; 2023. URL: https://training.cochrane.org/handbook [accessed 2025-06-05]
26. EndNote. URL: https://endnote.com [accessed 2024-05-07]
27. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
28. CASP qualitative checklist. Critical Appraisal Skills Programme (CASP). 2019. URL: https://casp-uk.net/casp-tools-checklists/qualitative-studies-checklist [accessed 2024-05-07]
29. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
30. Lewin S, Booth A, Glenton C, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implement Sci 2018 Jan 25;13(Suppl 1):2. [doi: 10.1186/s13012-017-0688-3] [Medline: 29384079]
31. Barlund AS, André B, Sand K, Brenne AT. A qualitative study of bereaved family caregivers: feeling of security, facilitators and barriers for rural home care and death for persons with advanced cancer. BMC Palliat Care 2021 Jan 8;20(1):7. [doi: 10.1186/s12904-020-00705-y] [Medline: 33419428]
32. Hull MM. A family experience: hospice supported home care of a dying relative [Dissertation]. Rochester, NY: University of Rochester; 1989.
33. Albert WKG, Fahrudin A, Lindong S, Yusuf H. The suffering experiences: family caregiver of home based palliative care in Malaysia. Open Access Maced J Med Sci 2022;10(E):1623-1629. [doi: 10.3889/oamjms.2022.8573]
34. Milberg A, Strang P. Exploring comprehensibility and manageability in palliative home care: an interview study of dying cancer patients' informal carers. Psychooncology 2004 Sep;13(9):605-618. [doi: 10.1002/pon.774] [Medline: 15334529]
35. Fernandes LG, Devan H, Kamper SJ, et al. Enablers and barriers of people with chronic musculoskeletal pain for engaging in telehealth interventions: protocol for a qualitative systematic review and meta-synthesis. Syst Rev 2020;9:122. [doi: 10.1186/s13643-020-01388-2]
36. Munthe-Kaas H, Bohren MA, Glenton C, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 3: how to assess methodological limitations. Implement Sci 2018 Jan 25;13(Suppl 1):9. [doi: 10.1186/s13012-017-0690-9] [Medline: 29384078]
37. Colvin CJ, Garside R, Wainwright M, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 4: how to assess coherence. Implement Sci 2018 Jan 25;13(Suppl 1):13. [doi: 10.1186/s13012-017-0691-8] [Medline: 29384081]
38. Glenton C, Bohren MA, Downe S, Paulsen EJ, Lewin S. Cochrane Qualitative Evidence Synthesis: Protocol and Review Template Version 14b: Cochrane; 2023. [doi: 10.5281/zenodo.5101812]
39. Noyes J, Booth A, Cargo M, et al. Qualitative evidence. In: Higgins JPT, Thomas J, Chandler J, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions: Wiley; 2019:525-545. [doi: 10.1002/9781119536604]
40. Ullrich A, Ascherfeld L, Marx G, Bokemeyer C, Bergelt C, Oechsle K. Quality of life, psychological burden, needs, and satisfaction during specialized inpatient palliative care in family caregivers of advanced cancer patients. BMC Palliat Care 2017 May 10;16(1):31. [doi: 10.1186/s12904-017-0206-z] [Medline: 28486962]
41. Thoits PA. Mechanisms linking social ties and support to physical and mental health. J Health Soc Behav 2011 Jun;52(2):145-161. [doi: 10.1177/0022146510395592] [Medline: 21673143]
42. Koffman J, Morgan M, Edmonds P, Speck P, Higginson IJ. Cultural meanings of pain and palliative care: a systematic review and implications for practice. J Palliat Med 2016;19(5):495-502. [doi: 10.1089/jpm.2015.0322]
43. Liu Y, Fang H, Lim CJ, Chong MF. Family caregivers' support needs in home-based palliative care: a qualitative study. BMC Palliat Care 2024;23(1):48. [doi: 10.1186/s12904-024-01308-1]
44. Brandt Å, Søndergaard J, Ellegaard H. Palliative home care: a systematic review of caregivers' experiences. BMJ Support Palliat Care 2017;7(4). [doi: 10.1136/bmjspcare-2017-001407.2]

Abbreviations

CASP: Critical Appraisal Skills Programme
GRADE-CERQual: Grading of Recommendations Assessment, Development, and Evaluation–Confidence in the Evidence from Reviews of Qualitative Research
MeSH: Medical Subject Headings
PICo: Population, Interest, Context
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: Prospective Register of Systematic Reviews

Edited by N Cahill; submitted 22.01.25; peer-reviewed by D Hansen, LT Schmidt; revised version received 10.04.25; accepted 15.04.25; published 19.06.25.

Please cite as:
Deng XM, Hounsri K, Lopez V, Tam WWS
Family Experiences, Needs, and Perceptions in Home-Based Hospice Care for Patients With Terminal Cancer: Meta-Synthesis and Systematic Review
JMIR Cancer 2025;11:e71596
URL: https://cancer.jmir.org/2025/1/e71596
doi:10.2196/71596

Clinical Prediction Models Incorporating Blood Test Trend for Cancer Detection: Systematic Review, Meta-Analysis, and Critical Appraisal

Pradeep S Virdee[1]; Kiana K Collins[1]; Claire Friedemann Smith[1]; Xin Yang[2]; Sufen Zhu[1]; Nia Roberts[3]; Jason L Oke[1]; Clare Bankhead[1]; Rafael Perera[1]; FD Richard Hobbs[1]; Brian D Nicholson[1]

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Primary Care Building, Woodstock Road, Oxford, United Kingdom

[2]St Edmund Hall, University of Oxford, Oxford, United Kingdom

[3]Bodleian Health Care Libraries, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Pradeep S Virdee
Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Primary Care Building, Woodstock Road, Oxford, United Kingdom

Abstract

Background: Blood tests used to identify patients at increased risk of undiagnosed cancer are commonly used in isolation, primarily by monitoring whether results fall outside the normal range. Some prediction models incorporate changes over repeated blood tests (or trends) to improve individualized cancer risk identification, as relevant trends may be confined within the normal range.

Objective: Our aim was to critically appraise existing diagnostic prediction models incorporating blood test trends for the risk of cancer.

Methods: MEDLINE and EMBASE were searched until April 3, 2025 for diagnostic prediction model studies using blood test trends for cancer risk. Screening was performed by 4 reviewers. Data extraction for each article was performed by 2 reviewers independently. To critically appraise models, we narratively synthesized studies, including model building and validation strategies, model reporting, and the added value of blood test trends. We also reviewed the performance measures of each model, including discrimination and calibration. We performed a random-effects meta-analysis of the c-statistic for a trends-based prediction model if there were at least 3 studies validating the model. The risk of bias was assessed using the PROBAST (prediction model risk of bias assessment tool).

Results: We included 16 articles, with a total of 7 models developed and 14 external validation studies. In the 7 models derived, full blood count (FBC) trends were most commonly used (86%, n=7 models). Cancers modeled were colorectal (43%, n=3), gastro-intestinal (29%, n=2), nonsmall cell lung (14%, n=1), and pancreatic (14%, n=1). In total, 2 models used statistical logistic regression, 2 used joint modeling, and 1 each used XGBoost, decision trees, and random forests. The number of blood test trends included in the models ranged from 1 to 26. A total of 2 of 4 models were reported with the full set of coefficients needed to predict risk, with the remaining excluding at least one coefficient from their article or were not publicly accessible. The c-statistic ranged 0.69 - 0.87 among validation studies. The ColonFlag model using trends in the FBC was commonly externally validated, with a pooled c-statistic=0.81 (95% CI 0.77-0.85; n=4 studies) for 6-month colorectal cancer risk. Models were often inadequately tested, with only one external validation study assessing model calibration. All 16 studies scored a low risk of bias regarding predictor and outcome details. All but one study scored a high risk of bias in the analysis domain, with most studies often removing patients with missing data from analysis or not adjusting the derived model for overfitting.

Conclusions: Our review highlights that blood test trends may inform further investigation for cancer. However, models were not available for most cancer sites, were rarely externally validated, and rarely assessed calibration when they were externally validated.

Trial Registration: PROSPERO CRD42022348907; https://www.crd.york.ac.uk/PROSPERO/view/CRD42022348907

(JMIR Cancer 2025;11:e70275) doi:10.2196/70275

KEYWORDS

blood test; hematologic tests; trend; prediction model; primary health care; cancer; neoplasms; systematic review

Introduction

Cancer incidence trends are projected to increase globally: 18 million new cases diagnosed in 2020 versus 28 million projected in 2040 [1]. The likelihood of survival improves by cancer detection at earlier stages [2-7]. Earlier detection is crucial to improve patient outcomes and reduce cancer-related mortality [8]. Screening programs may contribute to early detection but have been implemented for a minority of countries and cancers [9]. Risk prediction models for cancer could improve early detection rates. These models combine patient data, such as patient demographics, medical history, or cancer symptoms, to identify patients with an increased risk of undiagnosed cancer.

Blood tests commonly performed in clinical practice, including full blood count (FBC) and liver function tests, are often included in cancer risk prediction models, as they have an important role in risk-stratifying symptomatic patients for cancer investigation [10,11]. Blood tests are commonly requested by clinicians, with rates of testing increasing yearly. Despite panels of blood tests being taken together, blood tests are almost entirely interpreted in isolation in current clinical guidance [11,12]. In the United Kingdom, the National Institute for Health and Care Excellence (NICE) suspected cancer guidelines recommend referral for urgent investigation if low albumin, low hemoglobin, raised platelets, raised bilirubin, raised calcium, or raised inflammatory markers are observed, as these increase risk of cancer [11]. Monitoring temporal trends (ie, changes over time) in repeated blood tests may improve risk stratification, by incorporating an individual's trajectory from which to identify change. For example, declining hemoglobin confined within the normal range would be a relevant cancer-related trend, but missed in practice as the results appear normal. Our recent systematic review on the association between blood test trends and cancer diagnosis identified many trends that have the potential to improve cancer risk stratification [13]. However, the potential benefits and challenges and methodological considerations of incorporating combinations of trends into cancer risk prediction models remain unrealized.

Recent methodological advancements in both traditional statistical and machine-learning methods allow for the development of dynamic prediction models, which incorporate repeated measures data for clinical risk prediction and may hold greater potential to rule-in and rule-out referral for cancer investigation. We aimed to conduct a systematic review to critically appraise diagnostic clinical prediction models using trends in blood tests commonly used in primary care for the risk of undiagnosed cancer.

Methods

Overview

We followed the PRISMA (Preferred Reporting Items for Systematic review and Meta-Analysis) guidelines (Checklist 1) for reporting the findings of this review [14]. Ethical approval was not required, as there were no direct patient investigations in this study and only published articles were systematically reviewed. The review protocol was registered with the International PROSPERO (Prospective Register of Systematic Reviews) database on July 25, 2022 (CRD42022348907). There were no deviations to the protocol.

Participants

We included studies of participants aged 18 years or older reporting prediction models incorporating trends in blood tests commonly available in primary care and cancer diagnosis in any clinical setting. We excluded blood tests taken after cancer diagnosis, such as to predict prognosis or monitor treatment.

Outcome

The main outcome was a first diagnosis of cancer across all cancer sites, including composite cancer sub-groupings and all cancers combined. Cancer diagnosis was defined as per the individual studies, such as confirmed cancer via laboratory tests/radiology in clinical/prospective studies or the use of ICD10 (*International Statistical Classification of Diseases and Related Health Problems 10th Revision)* codes [15] in studies of eHealth records.

Search Strategy

We worked with our review specialist (NR) to derive a comprehensive search strategy. The MEDLINE (OVID) (1946-present) and EMBASE (OVID) (1974-present) databases were searched from inception to April 3, 2025 to identify articles that report on the association between trends in blood tests commonly available in clinical practice and a cancer diagnosis. The initial search was conducted in June 2022, with a full update in February and May 2023 and April 2025. Search terms included MeSH headings and title, abstract, and author keywords for blood tests, cancer diagnosis, and prediction or risk. Cancer-related terms included "tumor" and "cancer". However, some cancers are not usually paired with these terms, such as "leukaemia" or "lymphoma", so it was important to include such cancer types explicitly to ensure they were captured. No language or other limits were applied to the search. The full search strategy for each database is provided in Table S1 (MEDLINE) and Table S2 (EMBASE) in Multimedia Appendix 1. In the eligible studies, we actively searched through each article's reference list to find eligible studies that were not identified by the search strategy.

Study Selection

All references initially underwent de-duplication in Endnote 20 [16] (by NR). Abstract and title screening was performed in Endnote 20 and Rayyan [17] (by PSV, KKC, CFS, and XY). The retrieved articles were initially split among the reviewers for screening, with a sample of 1000 from each of the three reviewers (KKC, CFS, and XY) independently screened by a second reviewer (PSV) to assess agreement, with discrepancies discussed until an agreement was reached. The full-text screening was subsequently performed independently by two reviewers (by PSV and SZ) to identify eligible articles for data extraction and analysis, with discrepancies discussed until agreement was reached. We included any in-human primary research article reporting the development or validation of a diagnostic clinical risk prediction model using a prediagnostic trend over repeat measurements of at least one blood test parameter (Table 1) for subsequent diagnosis of cancer. A prediction model was defined as any multivariable model

designed to predict the presence of undiagnosed cancer (outcome), where at least one predictor in the model was a blood test trend. A model was considered to include "trend" if it included temporal changes in the quantitative blood test result over repeatedly measured tests per patient as a predictor. The blood tests in Table 1 are nonspecific (ie, not cancer-specific) blood tests that are commonly available in primary care settings. Recent evidence highlighted trends in many of these common tests as risk factors for cancer diagnosis [13]. Using these blood tests provides an opportunity to use commonly available data to support cancer detection.

Table . Blood tests included in this review.

Blood test	Blood level
Full blood count	Red blood cell count, hemoglobin, hematocrit, mean cell volume, mean cell hemoglobin, mean cell hemoglobin concentration, red blood cell distribution width, platelet count, mean platelet volume, white blood cell count, basophil count, eosinophil count, lymphocyte count, monocyte count, neutrophil count, basophil %, eosinophil %, lymphocyte %, monocyte %, neutrophil %
Liver function tests	Alanine aminotransaminase, albumin, alkaline phosphatase, aspartate transaminase, bilirubin
Renal function	Sodium, potassium, creatinine, urea
Inflammatory markers	C-reactive protein, erythrocyte sedimentation rate, plasma viscosity
Other tests	Amylase, HbA_{1c}[a], calcium, calcium adjusted, total protein, blood glucose, fasting glucose, thyroid stimulating hormone

[a]HbA_{1c}: hemoglobin A1c.

We excluded abstracts and conference proceedings, as they produce incomplete data for a thorough review. Studies using a cross-sectional design were excluded, as the data reflects a "snapshot" at a certain time so cannot assess risk over time. Clinical trials of treatment intervention were excluded to reduce the influence of treatments on blood test data. Existing systematic reviews, correspondence, and case studies pertaining to<5 individuals were excluded. Non-English full-texts without English versions available or nontranslatable were excluded.

Data Extraction

Data was extracted using an extraction form designed in Microsoft Excel and piloted on 3 randomly selected eligible articles. Data items included study design and population, blood test trends studied, analytic methods, cancer site, and predictive performance measures. Data extraction from each eligible article was performed by 2 reviewers independently (PSV, KKC, CFS, XY, and SZ), with disagreements discussed until agreement was reached.

Data Analysis and Synthesis

Quantitative data were summarized using means with SD for continuous data and counts with proportions for categorical data. We narratively described and critically appraised prediction models incorporating prediagnostic blood test trend. We performed a random-effects meta-analysis of the c-statistic (or area under the curve) for prediction models externally validated by at least 3 studies. The τ^2 statistic was used to describe heterogeneity and I^2 statistic to assess the proportion of heterogeneity explained by between-study differences. We also conducted a post hoc analysis, repeating the meta-analysis by including only studies using primary care data and again using only other studies, to assess if findings differed between underlying populations of care. Analyses were performed in Stata/SE 17.0.

Risk of Bias Assessment

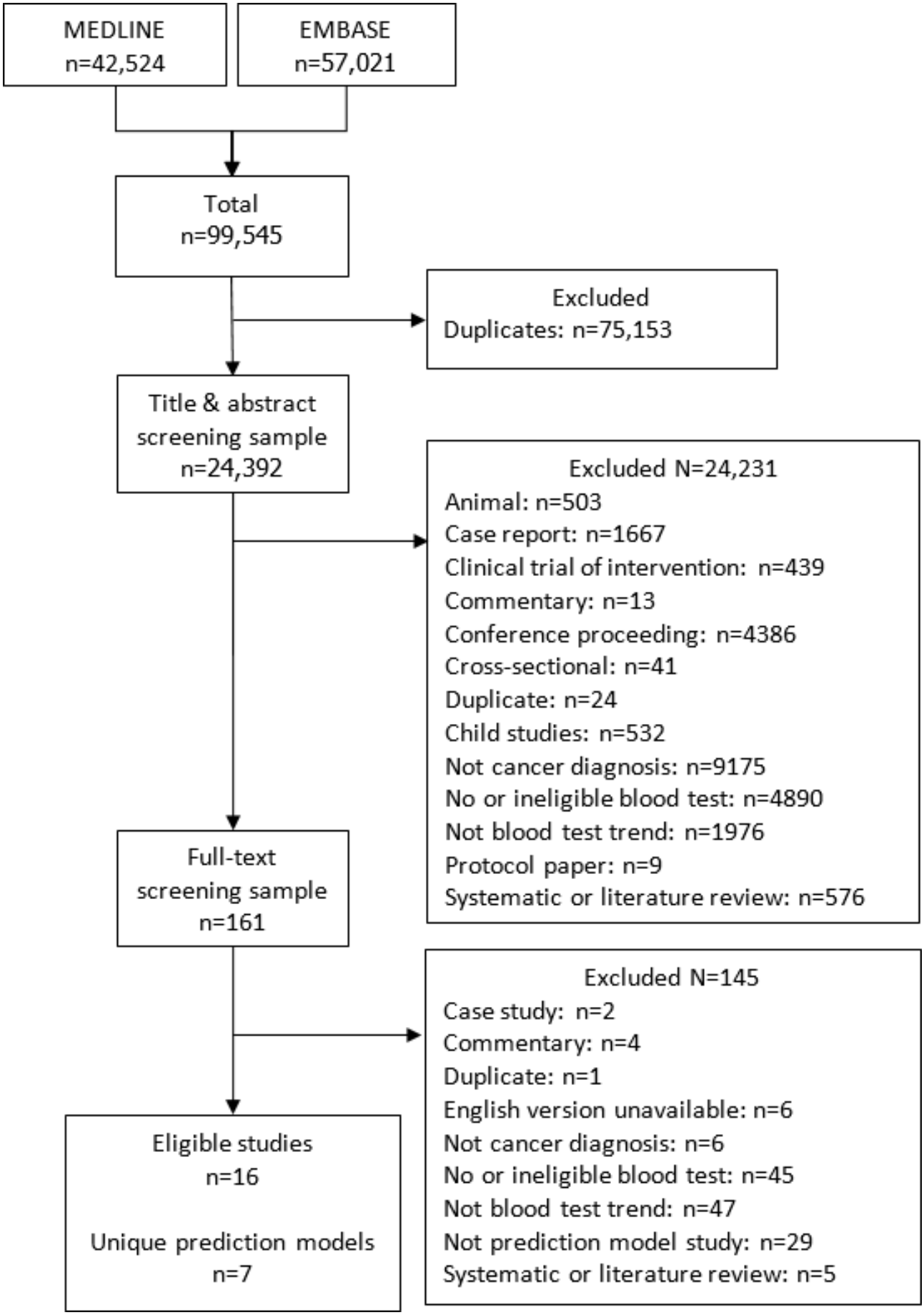
Risk of bias in each study was assessed using the Cochrane Prediction model Risk Of Bias Assessment Tool (PROBAST) [18]. Each study was assessed by two reviewers independently (PSV, KKC, CFS, XY, and SZ), with disagreements discussed until agreement was reached. Articles coauthored by a reviewer were assessed by other reviewers.

Results

Overall Summary

In total, 99,545 references were identified, of which 24,392 were unique after deduplication (Figure 1). A total of 16 studies met the eligibility criteria and were included in the review [19-34]. A total of 7 blood test trend-based prediction models were developed in total among 5 studies [23,27,28,30,31] and the remaining 11 studies [19-22,24-26,29,32-34] externally validated existing prediction models. In total, there were 14 external validations of 2 models (ColonFlag by Kinar et al [27] and ENDPAC (Enriching New-Onset Diabetes for Pancreatic Cancer) by Sharma et al [30]).

Figure 1. PRISMA (preferred reporting items for systematic review and meta-analysis) diagram.



Description of Studies

Study Design

A description of each study is provided in Table S3 in Multimedia Appendix 1. Of the 16 studies, a case-control design was used by 19% (n=3) [23,25,29] and cohort design by 81% (n=13) [19-22,24,26-28,30-34]. In addition, 25% (n=4) [19,20,22,24] used prospectively-collected data and 75% (n=12) [21,23,25-34] used retrospective data. Furthermore, 19% (n=3) [19,20,28] collected data at clinical centers, 75% (n=12) [21-23,25-27,29-34] used eHealth record databases, and 6% (n=1) [24] used both. All studies used opportunistic tests (ie, performed for any reason excluding screening for cancer, such as to monitor symptoms or comorbidity).

Participants

The mean number of participants recruited was 23,896 among prospective studies and 502,730 among retrospective studies, ranging from 617 to 2,914,589 participants over all the studies. The 16 articles spanned 4 different countries: the United States of America (44%, n=7) [23,25,28-30,33,34], the United Kingdom (25%, n=4) [19-21,31], Israel (25%, n=4) [22,26,27,32], and Canada (6%, n=1) [24]. The period of recruitment ranged from 1996 to 2020 in all studies. There were 38% (n=6) [21,26-28,31,32] studies conducted in primary care, 12% (n=2) [19,20] in secondary care, and 31% (n=5) in other settings: community-based insured adults (n=1) [25], endoscopy unit (n=1) [24], and insured individuals (n=3) [23,29,33]. It was unclear in 18% (n=3) [22,30,34]. One study [24] (6%) was limited to asymptomatic patients, including only patients without symptoms, and the remaining 94% (n=15) [19-23,25-34] included participants regardless of whether they experienced symptoms or not. A total of 6 studies [20,21,24,26,28,31] reported age, with a mean age 58.1 years (SD 5.2) among them. A total of 7 studies [21,25,27-29,31,32] reported sex, with mean 54.9% (SD 3.9) of females among them.

Model Building Strategy

Characteristics of the 7 models are in Table 2. A total of 4 models (57%) were developed in the USA population [23,28,30], 2 (29%) in United Kingdom [31], and 1 (14%) in Israel [27]. A total of 3 models (43%) were developed for risk of colorectal cancer [27,31], 2 (29%) for gastro-intestinal cancer (defined by Read as cancer of the esophagus, stomach, small intestine, colon, rectum, or anus) [28], 1 (14%) for nonsmall cell lung cancer [23], and 1 (14%) for pancreatic cancer [30]. A total of 6 models assessed cancer risk from the time of the latest blood test included and it was unclear in one study [23].

Table . Characteristics of 7 trend-based prediction models for cancer diagnosis.

Article	Country	Model (name, if assigned)	Outcome	Outcome risk window	Patient setting	Blood level(s) trend	Number of cases/total	Predictors in the final model
Gould et al [23]	United States of America	MES	Nonsmall cell lung cancer	Diagnosis	Other – insured individuals	ALT[a], creatinine, blood glucose, MCHC[b], platelets, RDW[c], WBC[d]	3942/117669	Age, sex, education, race, marital status, smoking status, smoking pack year, smoking years, smoking intensity, days since quitting, Hospitalization due to COPD and allied conditions, Diagnosis of COPD and allied conditions, Hospitalization due to Cancer, Diagnosis of Cancer, ALT, Creatinine, Glucose, MCHC, Platelets, RDW, WBC
Kinar et al [27]	Israel	ColonFlag	Colorectal cancer	3 - 6 months	Primary care	RBC[e], hemoglobin, hematocrit, MCV[f], MCH[g], MCHC, RDW, platelets, MPV[h], WBC, basophil#, basophil%, eosinophil#, eosinophil%, lymphocyte#, lymphocyte %, monocyte#, monocyte %, neutrophil#, neutrophil %	2437/466107	RBC, hemoglobin, hematocrit, MCV, MCH, MCHC, RDW, platelets, MPV, WBC, basophil#, basophil%, eosinophil#, eosinophil%, lymphocyte#, lymphocyte %, monocyte#, monocyte %, neutrophil#, neutrophil %, age, sex

Article	Country	Model (name, if assigned)	Outcome	Outcome risk window	Patient setting	Blood level(s) trend	Number of cases/total	Predictors in the final model
Read et al [28]	United States of America	Logistic model	Gastrointestinal cancer (esophagus, stomach, small intestine, colon, rectum, or anus)	6 months	Primary care	RBC, hemoglobin, hematocrit, MCV, MCH, MCHC, RDW, platelets, MPV, WBC, basophil#, basophil%, eosinophil#, eosinophil%, lymphocyte#, lymphocyte %, monocyte#, monocyte %, neutrophil#, neutrophil %	1025/148158	Age, sex, race, BMI, RBC, hemoglobin, hematocrit, MCV, MCH, MCHC, RDW, platelets, MPV, WBC, basophil#, basophil%, eosinophil#, eosinophil%, lymphocyte#, lymphocyte %, monocyte#, monocyte %, neutrophil#, neutrophil %, most recent BMP (8 components)
Read et al [28]	United States of America	Machine learning model	Gastrointestinal cancer (esophagus, stomach, small intestine, colon, rectum, or anus)	6 months	Primary care	RBC, hemoglobin, hematocrit, MCV, MCH, MCHC, RDW, platelets, MPV, WBC, basophil#, basophil%, eosinophil#, eosinophil%, lymphocyte#, lymphocyte %, monocyte#, monocyte %, neutrophil#, neutrophil %	1025/148158	Age, sex, race, BMI, RBC, hemoglobin, hematocrit, MCV, MCH, MCHC, RDW, platelets, MPV, WBC, basophil#, basophil%, eosinophil#, eosinophil%, lymphocyte#, lymphocyte %, monocyte#, monocyte %, neutrophil#, neutrophil %, most recent BMP (8 components)
Sharma et al [30]	United States of America	ENDPAC[i]	Pancreatic cancer	3 years	Unclear	Blood glucose	16/256	Change in weight, change in blood glucose category, age, change in blood glucose
Virdee et al [31]	United Kingdom	BLOOD-TRACC[j] Colorectal (females)	Colorectal cancer	2 years	Primary care	Hemoglobin, MCV, platelets	677/246695	Age, hemoglobin trend, MCV trend, platelets trend

Article	Country	Model (name, if assigned)	Outcome	Outcome risk window	Patient setting	Blood level(s) trend	Number of cases/total	Predictors in the final model
Virdee [31]	United Kingdom	BLOOD-TRACC Colorectal (males)	Colorectal cancer	2 years	Primary care	Hemoglobin, MCV, platelets	865/250716	Age, hemoglobin trend, MCV trend, platelets trend

[a]ALT: alanine aminotransaminase.

[b]MCHC: mean cell hemoglobin concentration.

[c]RDW: red blood cell distribution width.

[d]WBC: white blood cell count.

[e]RBC: red blood cell count.

[f]MCV: mean cell volume.

[g]MCH: mean cell hemoglobin.

[h]MPV: mean platelet volume.

[i]ENDPAC: enriching new-onset diabetes for pancreatic cancer.

[j]BLOODTRACC: full blood count trends for colorectal cancer detection.

In total, 2 models were developed using multivariate joint modeling [31], 2 using logistic regression [28,30], and 1 using each of XGBoost [23], decision trees [27], and random forests [28]. A total of 3 models (43%) were built by including all candidate predictors [27,28], 2 (29%) included clinically relevant predictors that were commonly available in practice [31], 1 (14%) included statistically significant variables in univariable analysis [30], and the model building process was unclear for 1 (14%) model [23]. To address missing blood test data, 2 (29%) models derived missing blood levels from other available blood levels using known mathematical relationships (eg mean cell hemoglobin=hemoglobin/red blood cell count) [31], 2 (29%) used imputation methods [28], 1 (14%) analyzed the blood test data as-is (without altering missing data) [23], and 1 (14%) used other methods (linear models to replace missing values using historical blood tests or mean value across all blood tests if no historic blood tests were present) [27]. Methods for handling missing blood test data were not discussed in 1 (14%) study [30].

Modeling Blood Test Trends

A total of 3 models (43%) assessed trends over repeated quantitative blood test results; Kinar et al [27] used ensembles of decision trees for the ColonFlag model, modeling changes over tests measured at 3 - 6 months before diagnosis and 18 and 36 months before that for each patient in the ensemble model, and Virdee et al [31] used multivariate joint modeling, which uses mixed-effects modeling to account for differing numbers of tests and the time between them in sporadically available repeated measures data between patients, for both BLOODTRACC models. One model (14%), by Sharma et al[30], calculated the difference between tests and included this as a single continuous variable in a logistic regression model to determine risk. It was unclear how trends were included in 3 (43%) models to predict risk [23,28].

The number of repeat blood tests used to define trend varies between models. Read et al [28] calculated the change in slope (reflecting the trend/trajectory) over at least 2 repeated tests sporadically measured over 3 years, Sharma et al [30] calculated the difference between blood tests measured at 18-3 months before new-onset diabetes and included this in their model, and Virdee et al[35] included the change in slope across all available blood tests (median=3 per patient) sporadically measured over 5 years to predict risk. The number of repeated blood tests used to derive trends was not reported for 3 models (43%) but the period of repeated testing among them ranged between 18 months and 5 years [23,27,30]. See Table S4 in Multimedia Appendix 1 for further details.

A total of 6 models (86%) used combinations of blood test trends and 1 model (14%) used trend in a single blood test (plus with other patient data) to predict cancer risk. The logistic model and random forests model by Read et al [28] combined trends in 28 blood tests Kinar et al [27]. combined trends in 20 blood tests (that make up the FBC) using decision trees, and Gould et al [23] combined trends in 7 blood tests using XGBoost. Virdee et al [35] combined 3 blood test trends (hemoglobin, mean corpuscular volume, and platelets) using multivariate joint modeing.

Model Reporting

Total 3 (43%) models were reported using appropriate reporting guidelines to report model findings (TRIPOD [Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis] guidelines [28,31,36]). For 3 (43%) models, justification for their choice of outcome risk window was provided [23,31]. In addition, 2 (29%) models were reported to be sufficiently powered, having provided a sample size calculation to show the number of patients and events needed to ensure reliable predictions and minimize optimistic performance [31].

Read et al [28] did not report the coefficients from their logistic model and Sharma et al [30] did not report the intercept from their logistic model. The full risk equation needed to derive an individual's risk of diagnosis was only reported for 2 models [31]. The models developed using XGBoost, decision trees, and random forests were not reported, due to the nature of machine

learning, and a reference to publicly available models was not provided [23,27,28].

Internal Validation

A total of 6 (86%) models underwent internal validation and one (14%) (by Sharma [30]) did not (Table 3). The internal validation sample was obtained using random data splitting for 4 (57%) models [23,27,31] and cross-validation for 2 (29%) models [23,,2828]. On average, there were 214,883 participants in the validation samples, ranging from 78,433 to 462,900. A total of 4 (57%) models were adjusted for overestimated performance [27,28,31] and it was unclear for 2 (29%) models [23,28].

Table . Performance statistics from internal and external validations of the final models, which include trends and other patient data.

Article	Model name/description	Outcome risk window	Overall performance		Discrimination		Calibration	
			Method	Result	Method	Result (95% CI)	Method	Result
Internal validation								
Gould et al [23]	MES	3 - 6 months	No		AUC/C-statistic	0.870 (0.856 - 0.886)	Isotonic regression	
Gould et al [23]	MES	6 - 9 months	No		AUC/C-statistic	0.862 (0.845 - 0.878)	No	
Gould et al [23]	MES	9 - 12 months	No		AUC/C-statistic	0.856 (0.840 - 0.872)	No	
Kinar et al [27]	ColonFlag	1 month	No		AUC/C-statistic	0.84	No	
Kinar et al [27]	ColonFlag	3 - 6 months	No		AUC/C-statistic	0.82	Hosmer-Lemeshow test	*P*=.47
Read et al [28]	Logistic regression	6 months	Brier score	0.008	AUC/C-statistic	0.711 (0.691-0.731)	No	
Read et al [28]	Machine-learning (random forest)	6 months	Brier score	0.092	AUC/C-statistic	0.713 (0.689-0.737)	No	
Virdee et al [35]	BLOOD-TRACC[a] Colorectal (females)	2 years	Brier score	0.0028	AUC/C-statistic	0.763 (0.753 - 0.775)	Calibration slope	1.05
Virdee et al [35]	BLOOD-TRACC Colorectal (males)	2 years	Brier score	0.0033	AUC/C-statistic	0.751 (0.739 - 0.764)	Calibration slope	1.06
External validation								
Ayling et al [19]	ColonFlag	Diagnosis	No		No		No	
Ayling et al [20]	ColonFlag	6 months	No		No		No	
Birks et al [21]	ColonFlag	3 - 6 months	No		AUC/C-statistic	0.844 (0.839 - 0.849)	No	
Birks et al [21]	ColonFlag	6 - 12 months	No		AUC/C-statistic	0.813 (0.809 - 0.818)	No	
Birks et al [21]	ColonFlag	12 - 24 months	No		AUC/C-statistic	0.791 (0.786 - 0.796)	No	
Birks et al [21]	ColonFlag	18 - 24 months	No		AUC/C-statistic	0.776 (0.771 - 0.781)	No	
Birks et al [21]	ColonFlag	24 - 36 months	No		AUC/C-statistic	0.751 (0.746 - 0.756)	No	
Goshen et al [22]	ColonFlag	Diagnosis	No		No		No	
Hilsden et al [24]	ColonFlag	1 year	No		No		No	

Article	Model name/description	Outcome risk window	Overall performance		Discrimination		Calibration	
			Method	Result	Method	Result (95% CI)	Method	Result
Hornbrook et al [25]	ColonFlag	6 months	No		AUC/C-statistic	0.80 (0.79 - 0.82)	No	
Kinar et al [27]	ColonFlag	1 month	No		AUC/C-statistic	0.84 (0.82 - 0.86)	No	
Kinar et al [27]	ColonFlag	3 - 6 months	No		AUC/C-statistic	0.81 (0.80 - 0.83)	Hosmer-Lemeshow test	*P*<.001
Kinar et al [26]	ColonFlag	12 - 18 months	No		No		No	
Schneider et al [29]	ColonFlag	6 months	No		AUC/C-statistic	0.78 (0.77 - 0.78)	No	
Virdee et al [31](Females)	ColonFlag	2 years	No		AUC/C-statistic	0.761 (0.744 - 0.768)	No	
Virdee et al [31] (Males)	ColonFlag	2 years	No		AUC/C-statistic	0.762 (0.749 - 0.774)	No	
Boursi et al [32]	ENDPAC[b]	3 years	No		AUC/C-statistic	0.69	No	
Chen et al [33]	ENDPAC	3 years	No		AUC/C-statistic	0.75	No	
Khan et al [34]	ENDPAC	4 years	No		AUC/C-statistic	0.72	No	
[30] Sharma et al [30]	ENDPAC	Diagnosis	No		No		No	

[a]BLOODTRACC: Full blood count trends for colorectal cancer detection.

[b]ENDPAC: enriching new-onset diabetes for pancreatic cancer.

Only 4 (57%) models assessed overall performance. Virdee et al [31], derived Brier scores of 0.0028 (men) and 0.0033 (women) for 2-year risk of colorectal cancer and Read et al [28] derived Brier scores of 0.008 (logistic regression) and 0.092 (random forests) for 6-month risk of GI cancer28.

A total of 6 (86%) models (100% of those internally validated) assessed discrimination, each using the c-statistic. Gould 2021 [23] and Kinar 2016 [27] reported c-statistic=0.87 and 0.82 for 3 - 6-month risk of nonsmall cell lung cancer in the United States of America and Israel based on various blood test trends measured over 5 years combined with other patient data and colorectal cancer based on all FBC parameters over 3 years combined with other patient data, respectively. Read 2023 [28] reported c-statistic=0.711 (logistic regression) and 0.713 (random forests) for 6-month risk of GI cancer based on FBC trends combined with other patient data. Virdee et al [31] reported c-statistic=0.75 (men) and 0.76 (women) for 2-year risk of colorectal cancer following trends in hemoglobin, mean cell volume, and platelets, together with age, measured over 5 years in UK primary care patients.

A total of 4 (57%) models were assessed for calibration. Gould 2021 [23] used isotonic regression to assess calibration, but did not report the corresponding results. Kinar 2016 [27] used the Hosmer-Lemeshow test and reported *P*=.47 for 3 - 6 month risk of colorectal cancer. Virdee et al [31] derived calibration slopes of 1.06 (men) and 1.05 (women) for 2-year risk of colorectal cancer and presented calibration plots.

External Validation

Fourteen external validation studies were performed in total for 2 models (Table 3): the ColonFlag by [27] was externally validated by 10 studies and the ENDPAC model by [30] by 4 studies. There were on average 244,580 participants included in the external validation studies, ranging from 532 to 2,225,249. Overall performance, discrimination, and calibration are all essential assessments to assess external validity of prediction models [37]. Overall performance of the ColonFlag or ENDPAC model was not assessed during external validation.

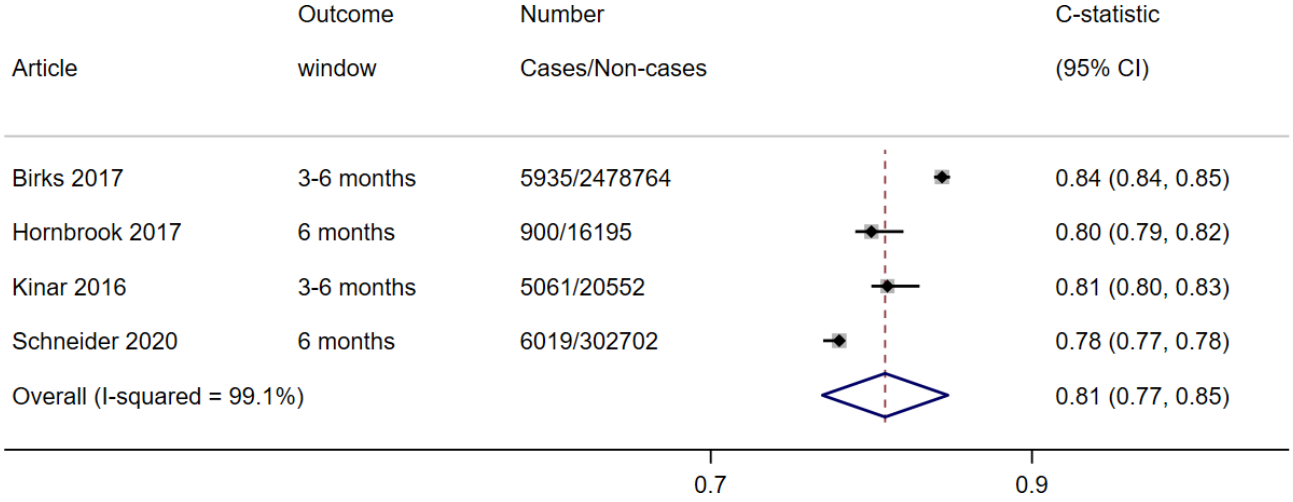
A total of 6 (29%) of the 14 external validations assessed discrimination, with all using the c-statistic. Birks et al [21] externally validated ColonFlag at multiple time intervals between the most recent blood test and diagnosis in a UK sample, reporting c-statistic=0.844 at 3 - 6 months, which reduced to 0.751 at 23 - 36 months [21]. Kinar et al [27] also externally validated the ColonFlag using UK data and reported a similar c-statistic (0.81) at 3 - 6 months before colorectal cancer diagnosis [27]. However, Kinar et al [27] removed the red blood cell distribution width blood level from the model and assessed predictive performance of the resulting model.

This was because the UK dataset did not include red blood cell distribution width, but the removal of a predictor from the model consequently means the external validation is incomplete.

A total of 4 studies with available data assessed <6-month risk of colorectal from ColonFlag and were included in a random-effects meta-analysis [21,25,27,29]. The pooled estimate indicated c-statistic=0.81 (95% CI 0.77 - 0.85) (τ^2=0.0016), with 99.1% (I^2) of the heterogeneity attributable to between-study differences (Figure 2). Our post hoc meta-analyses including only primary care populations and nonprimary care populations separately reduced heterogeneity, but this remained high (Figure S1 in Multimedia Appendix 1).

Calibration was assessed by Kinar et al [27]2016 only, using the Hosmer-Lemeshow test for the ColonFlag. They reported weak calibration at 3 - 6 months in the UK dataset (P<.001).

Figure 2. Forest plot of c-statistic for risk of colorectal cancer from ColonFlag external validations [21,25,27,29].



Added Value of Trend

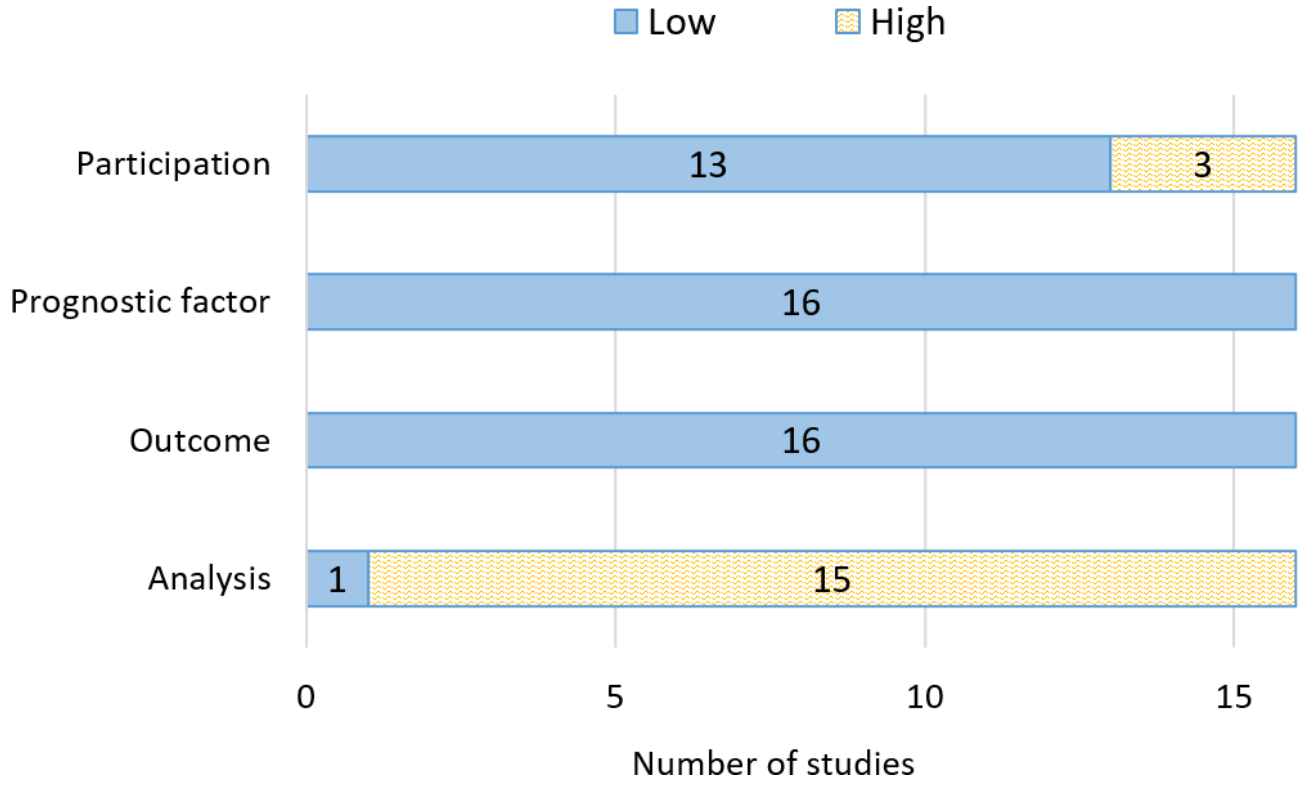
Kinar et al [27] assessed which blood test trends contributed most to the c-statistic of their prediction model for 3 - 6 month risk of colorectal cancer. Their model included trend in 20 FBC parameters, age, and sex. Red blood cell-related parameters contributed the most to the c-statistic, with trend in hemoglobin contributing the most (around 0.11) when added to age and sex. White blood cell-related parameters added the least to the c-statistic when combined with age and sex, such as adding around 0.03 AUC with the inclusion of monocyte count trend.

Read et al [28] used logistic regression to develop prediction models for the 6-month risk of gastro-intestinal cancer, including age, sex, BMI, blood test trends, and further covariates. They compared the c-statistic of their final model to one including blood tests measured at a single time point (the last test prior to the prediction interval). They report a higher c-statistic for their model including blood test trends (0.711, 95% CI 0.691 - 0.731) compared with the model including blood tests from a single time point (0.697, 95% CI 0.679 - 0.715). As secondary analyses, they assessed the c-statistic for one-, three-, and five-year risk, reporting higher c-statistics for models including blood test trends compared to models including single blood tests for one- (0.705, 95% CI 0.689 - 0.722 trend and 0.693, 95% CI 0.675 - 0.710 single) and three-year (0.735, 95% CI 0.713 - 0.757 trend and 0.683, 95% CI 0.665 - 0.701 single) risk but a lower c-statistic for their model including trends for five-year risk (0.672, 95% CI 0.653 - 0.691 trend and 0.703, 95% CI 0.686 - 0.720 single). No other study reported the added benefit of blood test trend to the prediction models.

Risk of Bias

Risk of bias for each domain is summarised in Figure 3 and per study in Table S5 in Multimedia Appendix 1. All 16 studies scored a low risk of bias in the predictors and outcome domains. All but 3 studies in the participant domain scored low risk of bias, with (Gould et al, Hornbrook et al, and, Schneider et al [23,25,29]) scoring high risk of bias for not including all eligible patients in their analyses. All but one study scored a high risk of bias in the analysis domain, commonly due to studies removing patients with missing data from all their analyses, not adjusting the developed model for under or overfitting, or not accounting for complexities in the data, such as censoring.

Figure 3. Summary of risk of bias scores, assessed using the prediction model risk of bias assessment tool.



Discussion

Principal Findings

This systematic review builds on our recent review on the association between blood test trend and cancer diagnosis [13] by highlighting the potential for risk stratification and methodological considerations of incorporating combinations of trends into cancer risk prediction models for use in practice. Our review identified logistic regression (incorporating the difference between 2 blood tests as a single variable) and multivariate joint modeling as the most commonly used modeling techniques. Models were often developed using poor methods. For example, although all but one model underwent internal validation during model development, model performance was not adequately assessed, with calibration often ignored and recalibration rarely performed for overfitting [37-41]. Where calibration was assessed, the Hosmer-Lemeshow test was sometimes used, which is known to have limited power and poor interpretability [37]. Many models were inadequately reported, with only one study providing the full risk-equation needed to derive an individual's risk of diagnosis. Without the full risk equation being available, models are unlikely to be independently externally validated or easily embedded into practice. Although our primary focus was to critically appraise trend-based prediction models, it is important to also highlight caution in the interpretation of performance measures from the models, as these may be subject to publication bias. For example, a prediction model with a poorer c-statistic is less likely to be published.

The ColonFlag model was most commonly externally validated, although this model is commercially developed so not publicly available. This model uses trends in FBC parameters to predict a monotonic score confined between 0 - 100, where higher scores reflect a higher likelihood of colorectal cancer diagnosis [27]. A pooled c-statistic of 0.81 from 4 studies indicates that trends in the FBC could be generalizable to other clinical settings and geographical locations, with good predictive ability to distinguish between patients with and without colorectal cancer. Heterogeneity was however high. This was anticipated due to the variation between studies included in the meta-analysis, such as differing geographical settings, health care systems, and eHealth records used. Therefore, caution should be given in the interpretation of these results when making generalisations between different clinical settings. There were few studies demonstrating the external validity of other models including blood test trend. Predictive ability of models was not assessed by cancer characteristics, such as by cancer stage, in any study.

Comparison of Models

A total of 3 models were identified for colorectal cancer: the ColonFlag and sex-specific BLOODTRACC models. Both models include age and sex, with the ColonFlag also including trend in all 20 FBC parameters and the BLOODTRACC models including trend in only three FBC parameters (hemoglobin, mean cell volume, and platelets). The ColonFlag uses changes over tests measured at 36 and 18 months up to the current test, with all patients requiring a test at each time point, whereas the BLOODTRACC models use all available tests over a five-year period before the current test and takes into consideration the timing of tests, as blood tests are not performed routinely in the United Kingdom. Although the ColonFlag was developed for 3 - 6 month risk in Israeli primary care, external validation studies of this model for two-year risk found it performed similarly to the BLOODTRACC models for 2-year risk in UK

primary care. This suggests that the 17 additional blood test trends in the ColonFlag may not add further diagnostic benefit to the combination of hemoglobin, mean corpuscular volume, and platelet trends for colorectal cancer. This may suggest that the underlying methodology used to develop the models (decision trees for the ColonFlag and joint modeling for the BLOODTRACC models) does not affect discriminative performance, but this would need assessing on the same patient dataset and multiple study designs employed to reduce heterogeneity. This assessment was performed in the BLOODTRACC model derivation study, where both models derived comparable c-statistics in the same cohort, both overall and in subgroups of age, by number of blood tests used to derive trends, and by longitudinal period used to derive trends [31].

Read et al[28] developed two models for gastro-intestinal cancer, one using random forests and one using logistic regression. Both models were designed to be as similar as possible, such as using the same study sample, outcome window, longitudinal period to derive trends, and similar covariates, with the methodological approach used to derive the methods being the biggest difference. Both models achieved an AUC of 0.71, suggesting that the underlying methodological approach may not affect discriminative performance, although the logistic model had better overall performance (lower Brier score). Neither model was assessed for calibration so further testing is required.

The remaining 2 models were for lung and pancreatic cancer. These were not compared with other models, as no further models for lung or pancreatic cancer were identified.

Strengths and Limitations

To our knowledge, this is the first review of cancer prediction models that incorporate blood test trend. We performed a comprehensive search, developed with an information specialist, including full-length articles retrieved from MEDLINE and EMBASE. It is possible that additional relevant studies may be found exclusively in other databases and were missed by our review. However, it is likely that most relevant manuscripts were found, as MEDLINE and EMBASE had 97.5% coverage of articles in previous systematic reviews and we conducted citation searching of all included manuscripts [42]. Our review identified prediction models for only four cancer types, with two externally validated (colorectal and pancreatic). We were therefore unable to draw conclusions regarding external validity for many cancer types. One further limitation is that we were unable to draw conclusions regarding publication bias, assessing whether prediction models were more likely to be published if they had good predictive performance. Only five models had c-statistics with corresponding confidence intervals at internal validation, making it difficult to assess symmetry in a funnel plot and deduce any publication bias.

Comparison With Previous Work

To date, prediction models for cancer risk are most commonly developed using single blood test results (plus other predictors). These include the QCancer models for the 2-year risk of cancer [43,44] and unexpected weight loss models for the 6-month risk of cancer [45], which combine patient demographics, symptoms, and single blood test values for cancer risk in symptomatic patients in UK primary care practices. Collectively, these models have c-statistics ranging 0.79 - 0.92, comparable to 0.71 - 0.87 reported for the models included in this review, which often included only blood test trends, age, and sex and different outcome risk windows. Existing systematic reviews have identified prediction models for individual cancer sites, including lung, breast, colorectal, and prostate, but the focus of these reviews was not on the role of blood test trend [46-49]. Lung cancer prediction models in those reviews often included patient demographics, pneumonia, exposure to smoking, and single blood tests for one-year risk, with c-statistic ranging 0.66 - 0.91. In this review, Gould et al [23] reported 0.87 for six-month risk of lung cancer using similar predictors combined with trend in seven blood tests. Colorectal cancer prediction models in those reviews often included patient demographics and single blood tests, with c-statistic ranging from 0.82 - 0.84 for 6-month risk and 0.72 - 0.92 for 2-year risk. In this review, Kinar et al [27] and Birks et al [21] reported 0.82 - 0.84 for 6-month risk and Virdee et al [31] reported 0.75 - 0.76 for 2-year risk of colorectal cancer using trend in 20 and three blood tests, respectively, age, and sex. Although those reviews identified prediction models using single blood test results for breast and prostate cancer [46,49], we found no prediction models incorporating trends for these cancers in this systematic review.

Clinical and Research Implications

Thorough testing of prediction models is required before clinical guidelines for cancer investigation can incorporate blood test trends. This includes assessment for the predictive ability of blood test trend compared to single blood tests and symptoms and the potential for early detection of cancer. For example, in the cancer field, the NICE guidelines recommend primary care to refer for cancer investigation if a patient's risk is above 3%, which is often used to support referral of symptomatic patients, whose risk is likely higher than nonsymptomatic patients. For models derived for more general populations, such as the trend-based models included in this review, there is no clear cut-off. To assess the potential added benefit of trend, studies would need to compare the diagnostic accuracy of trend-based and static/single-test models. No study in our review performed such comparisons, so this potential remains unknown. Patient- and clinician-acceptability of blood test trend approaches for cancer detection also requires investigation to optimize uptake of such models in practice. As some clinicians order blood tests more than others, methods to standardize blood testing across practices may be warranted and could reduce practice-level variability through clinical guidelines on repeat blood testing. This additional testing may add burden to health care, but the balance of patient benefit and outcomes to health care burden would need investigation. In terms of reporting, prediction models were often not reported in full, which is required for implementation into clinical systems and use in practice. Future models should follow appropriate reporting guidelines to ensure they are appropriately reported, such as the TRIPOD [36] or TRIPOD-AI [50] guidelines.

Sub-optimal methods to analyse trends were often identified, such as logistic regression incorporating change between tests.

Recent technological advancements have allowed for dynamic models, which are designed for repeated measures data by appropriately accounting for nonindependent data sporadically recorded in routine clinical practice [51], to be incorporated into analysis software packages. These include models such as landmarking and joint modeling of longitudinal and time-to-event data [52-54]. Research is required to assess the implementation considerations of different methodological techniques. For example, the feasibility of incorporating computationally intensive approaches, such as joint modeling, or approaches that require larger datasets or are nontransparent, such as machine learning. Our ongoing research aims to develop and validate trend-based prediction models for cancer, with eventual integration of trend into risk stratification in clinical practice [55]. Future prediction model studies should employ appropriate validation metrics, as we found that most studies did not assess overall performance or calibration. Further sub-optimal analysis methods commonly used included removing patients with missing data from all their analyses, not adjusting the developed model for under or overfitting, or not accounting for complexities in the data, such as censoring. Future models should consider such points to reduce bias.

Conclusion

We highlight the cancers for which there is a reported prediction model incorporating changes in repeated blood tests over time and the cancers and blood tests with no published literature. We provide an overview of the predictive performance of prediction models incorporating blood test trends and highlight that further testing is needed for all models identified. This review lays the foundation for further research.

Acknowledgments

PSV and BDN are funded for this work by a Cancer Research UK Clinical Careers Committee Postdoctoral Fellowship (RCCPDF\100005). The authors would also like to thank patient and public involvement representatives Alton Sutton, Bernard Gudgin, Clara Martins de Barros, Emily Lam, Ian Blelloch, Julian Ashton, Margaret Ogden, Shannon Draisey, and Susan Lynne for applying a patient perspective on the relevance of blood test trends for cancer detection.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

PSV, JLO, CB, RP, RH, BDN – Conceptualization
PSV, KKC, CFS, XY, NR – Data curation
PSV – Formal analysis
PSV, BDN – Funding acquisition
PSV – Methodology
PSV – Project administration
PSV – Resources
PSV – Software
KKC, CFS, XY – Validation
PSV – Visualization
PSV – Writing – original draft
All authors – Writing – review & editing

Conflicts of Interest

None declared.

Multimedia Appendix 1
Final search strategy.
[DOCX File, 60 KB - cancer_v11i1e70275_app1.docx]

Checklist 1
PRISMA checklist.
[PDF File, 75 KB - cancer_v11i1e70275_app2.pdf]

References

1. Worldwide cancer incidence statistics. Cancer Research UK. 2023. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer/incidence#heading-One [accessed 2025-05-24]

2. Cancer statistics for the UK - cancer screening and diagnosis. Cancer Research UK. 2023. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk#heading-Four [accessed 2025-05-24]
3. Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. CA Cancer J Clin 2022 Jan;72(1):7-33. [doi: 10.3322/caac.21708] [Medline: 35020204]
4. Survival for lung cancer. Cancer Research UK. 2023. URL: https://www.cancerresearchuk.org/about-cancer/lung-cancer/survival [accessed 2025-05-24]
5. Survival for bowel cancer. Cancer Research UK. 2023. URL: https://www.cancerresearchuk.org/about-cancer/bowel-cancer/survival [accessed 2025-05-24]
6. Survival for breast cancer. Cancer Research UK. 2023. URL: https://www.cancerresearchuk.org/about-cancer/breast-cancer/survival [accessed 2025-05-24]
7. Survival of prostate cancer. Cancer Research UK. 2023. URL: https://www.cancerresearchuk.org/about-cancer/prostate-cancer/survival [accessed 2025-05-24]
8. Crosby D, Bhatia S, Brindle KM, et al. Early detection of cancer. Science 2022 Mar 18;375(6586):eaay9040. [doi: 10.1126/science.aay9040] [Medline: 35298272]
9. What is cancer screening. Cancer Research UK. 2022. URL: https://www.cancerresearchuk.org/about-cancer/cancer-symptoms/spot-cancer-early/screening/what-is-cancer-screening#screening20 [accessed 2025-05-24]
10. Rubin GP, Saunders CL, Abel GA, et al. Impact of investigations in general practice on timeliness of referral for patients subsequently diagnosed with cancer: analysis of national primary care audit data. Br J Cancer 2015 Feb 17;112(4):676-687. [doi: 10.1038/bjc.2014.634] [Medline: 25602963]
11. Watson J, Mounce L, Bailey SE, et al. Blood markers for cancer. BMJ 2019 Oct 14;367:l5774. [doi: 10.1136/bmj.l5774] [Medline: 31611235]
12. Suspected cancer: recognition and referral (NG12). NICE. 2015. URL: https://www.nice.org.uk/guidance/ng12 [accessed 2025-05-24]
13. Virdee PS, Collins KK, Friedemann Smith C, et al. The association between blood test trends and undiagnosed cancer: a systematic review and critical appraisal. Cancers (Basel) 2024 Apr 26;16(9):1692. [doi: 10.3390/cancers16091692] [Medline: 38730644]
14. Moher D, Liberati A, Tetzlaff J, PRISMA Group, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
15. International statistical classification of diseases and related health problems 10th revision (ICD-10). World Health Organisation. 2019. URL: https://icd.who.int/browse10/2019/en [accessed 2025-05-24]
16. EndNote 20. EndNote. 2023. URL: https://endnote.com [accessed 2025-05-24]
17. Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
18. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019 Jan 1;170(1):51-58. [doi: 10.7326/M18-1376] [Medline: 30596875]
19. Ayling RM, Lewis SJ, Cotter F. Potential roles of artificial intelligence learning and faecal immunochemical testing for prioritisation of colonoscopy in anaemia. Br J Haematol 2019 Apr;185(2):311-316. [doi: 10.1111/bjh.15776] [Medline: 30714125]
20. Ayling RM, Wong A, Cotter F. Use of ColonFlag score for prioritisation of endoscopy in colorectal cancer. BMJ Open Gastroenterol 2021 Jun;8(1):e000639. [doi: 10.1136/bmjgast-2021-000639] [Medline: 34083226]
21. Birks J, Bankhead C, Holt TA, et al. Evaluation of a prediction model for colorectal cancer: retrospective analysis of 2.5 million patient records. Cancer Med 2017 Oct;6(10):2453-2460. [doi: 10.1002/cam4.1183] [Medline: 28941187]
22. Goshen R, Choman E, Ran A, et al. Computer-assisted flagging of individuals at high risk of colorectal cancer in a large health maintenance organization using the ColonFlag test. JCO Clin Cancer Inform 2018 Dec;2:1-8. [doi: 10.1200/CCI.17.00130] [Medline: 30652563]
23. Gould MK, Huang BZ, Tammemagi MC, et al. Machine learning for early lung cancer identification using routine clinical and laboratory data. Am J Respir Crit Care Med 2021 Aug 15;204(4):445-453. [doi: 10.1164/rccm.202007-2791OC]
24. Hilsden RJ, Heitman SJ, Mizrahi B, et al. Prediction of findings at screening colonoscopy using a machine learning algorithm based on complete blood counts (ColonFlag). PLoS ONE 2018;13(11):e0207848. [doi: 10.1371/journal.pone.0207848]
25. Hornbrook MC, Goshen R, Choman E, et al. Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. Dig Dis Sci 2017 Oct;62(10):2719-2727. [doi: 10.1007/s10620-017-4722-8] [Medline: 28836087]
26. Kinar Y, Akiva P, Choman E, et al. Performance analysis of a machine learning flagging system used to identify a group of individuals at a high risk for colorectal cancer. PLoS One 2017;12(2):e0171759. [doi: 10.1371/journal.pone.0171759] [Medline: 28182647]
27. Kinar Y, Kalkstein N, Akiva P, et al. Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. J Am Med Inform Assoc 2016 Sep;23(5):879-890. [doi: 10.1093/jamia/ocv195] [Medline: 26911814]

28. Read AJ, Zhou W, Saini SD, et al. Prediction of gastrointestinal tract cancers using longitudinal electronic health record data. Cancers (Basel) 2023 Feb 22;15(5):1399. [doi: 10.3390/cancers15051399] [Medline: 36900192]
29. Schneider JL, Layefsky E, Udaltsova N, et al. Validation of an algorithm to identify patients at risk for colorectal cancer based on laboratory test and demographic data in diverse, community-based population. Clin Gastroenterol Hepatol 2020 Nov;18(12):2734-2741. [doi: 10.1016/j.cgh.2020.04.054]
30. Sharma A, Kandlakunta H, Nagpal SJS, et al. Model to determine risk of pancreatic cancer in patients with new-onset diabetes. Gastroenterology 2018 Sep;155(3):730-739. [doi: 10.1053/j.gastro.2018.05.023] [Medline: 29775599]
31. Virdee PS, Patnick J, Watkinson P, et al. Full blood count trends for colorectal cancer detection in primary care: development and validation of a dynamic prediction model. Cancers (Basel) 2022 Sep 29;14(19):4779. [doi: 10.3390/cancers14194779] [Medline: 36230702]
32. Boursi B, Patalon T, Webb M, et al. Validation of the enriching new-onset diabetes for pancreatic cancer model: a retrospective cohort study using real-world data. Pancreas 2022 Feb 1;51(2):196-199. [doi: 10.1097/MPA.0000000000002000] [Medline: 35404897]
33. Chen W, Zhou B, Luong TQ, et al. Prediction of pancreatic cancer in patients with new onset hyperglycemia: a modified ENDPAC model. Pancreatology 2024 Nov;24(7):1115-1122. [doi: 10.1016/j.pan.2024.09.015] [Medline: 39353843]
34. Khan S, Safarudin RF, Kupec JT. Validation of the ENDPAC model: Identifying new-onset diabetics at risk of pancreatic cancer. Pancreatology 2021 Apr;21(3):550-555. [doi: 10.1016/j.pan.2021.02.001] [Medline: 33583686]
35. Virdee PS, Patnick J, Watkinson P, et al. Trends in the full blood count blood test and colorectal cancer detection: a longitudinal, case-control study of UK primary care patient data. NIHR Open Res 2022;2(32):32. [doi: 10.3310/nihropenres.13266.2] [Medline: 37056715]
36. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 7;350:g7594. [doi: 10.1136/bmj.g7594] [Medline: 25569120]
37. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010 Jan;21(1):128-138. [doi: 10.1097/EDE.0b013e3181c30fb2] [Medline: 20010215]
38. Archer L, Snell KIE, Ensor J, et al. Minimum sample size for external validation of a clinical prediction model with a continuous outcome. Stat Med 2021 Jan 15;40(1):133-146. [doi: 10.1002/sim.8766] [Medline: 33150684]
39. Riley RD, Collins GS, Ensor J, et al. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. Stat Med 2022 Mar 30;41(7):1280-1295. [doi: 10.1002/sim.9275] [Medline: 34915593]
40. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med 2021 Aug 30;40(19):4230-4251. [doi: 10.1002/sim.9025] [Medline: 34031906]
41. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020 Mar 18;368:m441. [doi: 10.1136/bmj.m441] [Medline: 32188600]
42. Bramer WM, Giustini D, Kramer BMR. Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: a prospective study. Syst Rev 2016 Mar 1;5:39. [doi: 10.1186/s13643-016-0215-7] [Medline: 26932789]
43. Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify women with suspected cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract 2013 Jan;63(606):e11-e21. [doi: 10.3399/bjgp13X660733] [Medline: 23336450]
44. Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify men with suspected cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract 2013 Jan;63(606):e1-10. [doi: 10.3399/bjgp13X660724] [Medline: 23336443]
45. Nicholson BD, Aveyard P, Koshiaris C, et al. Combining simple blood tests to identify primary care patients with unexpected weight loss for cancer investigation: Clinical risk score development, internal validation, and net benefit analysis. PLoS Med 2021 Aug;18(8):e1003728. [doi: 10.1371/journal.pmed.1003728] [Medline: 34464384]
46. Aladwani M, Lophatananon A, Ollier W, et al. Prediction models for prostate cancer to be used in the primary care setting: a systematic review. BMJ Open 2020 Jul 19;10(7):e034661. [doi: 10.1136/bmjopen-2019-034661] [Medline: 32690501]
47. Toumazis I, Bastani M, Han SS, et al. Risk-Based lung cancer screening: a systematic review. Lung Cancer (Auckl) 2020 Sep;147:154-186. [doi: 10.1016/j.lungcan.2020.07.007] [Medline: 32721652]
48. Virdee PS, Marian IR, Mansouri A, et al. The full blood count blood test for colorectal cancer detection: a systematic review, meta-analysis, and critical appraisal. Cancers (Basel) 2020 Aug 19;12(9):2348. [doi: 10.3390/cancers12092348] [Medline: 32825191]
49. Zheng Y, Li J, Wu Z, et al. Risk prediction models for breast cancer: a systematic review. BMJ Open 2022 Jul;12(7):e055398. [doi: 10.1136/bmjopen-2021-055398]
50. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 16;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
51. Bull LM, Lunt M, Martin GP, et al. Harnessing repeated measurements of predictor variables for clinical risk prediction: a review of existing methods. Diagn Progn Res 2020;4:9. [doi: 10.1186/s41512-020-00078-z] [Medline: 32671229]

52. Lee C, Yoon J, Schaar MVD. Dynamic-DeepHit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. IEEE Trans Biomed Eng 2020 Jan;67(1):122-133. [doi: 10.1109/TBME.2019.2909027] [Medline: 30951460]
53. Paige E, Barrett J, Stevens D, et al. Landmark models for optimizing the use of repeated measurements of risk factors in electronic health records to predict future disease risk. Am J Epidemiol 2018 Jul 1;187(7):1530-1538. [doi: 10.1093/aje/kwy018] [Medline: 29584812]
54. Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. Biom J 2011 Sep;53(5):750-763. [doi: 10.1002/bimj.201100052] [Medline: 21834127]
55. Virdee PS, Bankhead C, Koshiaris C, et al. Blood test trend for cancer detection (BLOTTED): protocol for an observational and prediction model development study using English primary care electronic health record data. Diagn Progn Res 2023 Jan 10;7(1):1. [doi: 10.1186/s41512-022-00138-6] [Medline: 36624489]

Abbreviations

ENDPAC: Enriching New-Onset Diabetes for Pancreatic Cancer
FBC: full blood count
ICD10: International Statistical Classification of Diseases and Related Health Problems 10th Revision
NICE: National Institute for Health and Care Excellence
PRISMA: Preferred Reporting Items for Systematic review and Meta-Analysis
PROBAST: prediction model risk of bias assessment tool
PROSPERO: Prospective Register of Systematic Reviews
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

Edited by N Cahill; submitted 18.12.24; peer-reviewed by L Smith, V Moglia, Z He; revised version received 02.05.25; accepted 05.05.25; published 27.06.25.

Please cite as:
Virdee PS, Collins KK, Smith CF, Yang X, Zhu S, Roberts N, Oke JL, Bankhead C, Perera R, Hobbs FDR, Nicholson BD
Clinical Prediction Models Incorporating Blood Test Trend for Cancer Detection: Systematic Review, Meta-Analysis, and Critical Appraisal
JMIR Cancer 2025;11:e70275
URL: https://cancer.jmir.org/2025/1/e70275
doi:10.2196/70275

Benefits of Remote-Based Mindfulness on Physical Symptom Outcomes in Cancer Survivors: Systematic Review and Meta-Analysis

Maria Komariah[1*], BSN, RN, MHSc, PhD; Sidik Maulana[1*], BSN, RN; Shakira Amirah[2], BSc; Hesti Platini[1], BSN, RN, MNS, CNS; Laili Rahayuwati[1], MHSc, MSc, Prof Dr; Ah Yusuf[3], BSN, RN, MHSc, Prof Dr; Mohd Khairul Zul Hasymi Firdaus[4], BSN, RN, MNS, PhD

[1]Faculty of Nursing, Padjadjaran University, Bandung, Indonesia

[2]Faculty of Medicine, University of Indonesia, Depok, Indonesia

[3]Faculty of Nursing, Airlangga University, Surabaya, Indonesia

[4]Kulliyah of Nursing, International Islamic University Malaysia, Selangor, Malaysia

[*]these authors contributed equally

Corresponding Author:
Maria Komariah, BSN, RN, MHSc, PhD
Faculty of Nursing, Padjadjaran University, Bandung, Indonesia

Related Article:

This is a corrected version. See correction statement: https://cancer.jmir.org/2025/1/e71958

Abstract

Background: Many cancer survivors experience a wide range of symptoms closely linked to psychological problems, highlighting the need for psychological treatment, one of the most popular being mindfulness. The use of the internet has greatly increased in the last decade, and has encouraged the use of remote-based interventions to help people living with cancer access treatment remotely via devices.

Objective: The primary aim of this study was to explore the efficacy of internet-based mindfulness interventions on the physical symptoms of people living with cancer, where physical symptoms are defined as distressing somatic experiences (eg fatigue, insomnia, and pain) regardless of the underlying cause. The secondary aim was to investigate interventions for the quality of life (QoL).

Methods: This study followed the Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) guidelines. Relevant articles were systematically searched using electronic databases, namely Scopus, Medline through PubMed, Cumulated Index in Nursing and Allied Health Literature (CINAHL) through EBSCOhost, and Cochrane Central Database. Randomized controlled and pilot trials involving adults and/or older adults with cancer and using remote-based mindfulness interventions compared to usual care were included. The quality of the trials included in this study was assessed using the revised Cochrane risk of bias, version 2.0. This study estimated the standardized mean difference (SMD) and mean difference (MD) with 95% CI. The I^2 test was used to identify potential causes of heterogeneity. Publication bias was assessed using contour-enhanced funnel plots and the Egger linear regression test to reveal a small study effect.

Results: The initial search yielded 1985 records, of which 13 studies were ultimately included. After treatment, remote-based mindfulness significantly reduced fatigue (SMD −0.94; 95% CI: −1.56 to −0.33; *P*=.002), sleep disturbance (SMD −0.36; 95% CI: −0.60 to −0.12; *P*=.004), and improved physical function (SMD .25; 95% CI: 0.09 to 0.41; *P*=.002) compared to that observed before treatment. However, compared with usual care, remote-based mindfulness showed a statistically significant reduction only in sleep disturbance (SMD: −0.37; 95% CI: −0.58 to −0.16; *P*=.0006) after treatment. Moreover, remote-based mindfulness was not statistically significant in reducing pain both within and between groups.

Conclusions: Remote-based mindfulness shows promise in reducing sleep disturbances; however, its impact on fatigue, pain, and physical function may be limited.

(JMIR Cancer 2025;11:e54154) doi:10.2196/54154

KEYWORDS

cancer; physical symptoms; mindfulness; remote-based intervention; quality of life

Introduction

Advancements in cancer medication have extended the life expectancy of cancer patients in recent years [1]. However, more cancer survivors undergo cancer treatment for a longer period. Chronic treatment has been shown to increase symptom burden and reduce the quality of life (QoL) of cancer survivors [2-7]. More than two-thirds of cancer survivors with advanced disease are symptomatic [8]. Cancer survivors receive supportive care focused on relieving symptoms at all stages of their illness [9-11].

Most cancer survivors frequently experience physical symptoms such as pain and fatigue. Physical symptoms are defined as the subjective experiences of distressing somatic symptoms (eg fatigue, insomnia, pain, and nausea), regardless of the cause [12]. In most cancer survivors, pain may be managed with relatively standard treatment [13]. Recent suggestions include a multimodal approach with tailored therapy, including perceptual, homeostatic, and behavioral reactions to chronic illness. This approach allows healthcare professionals to dynamically manage pain by integrating pharmacological and nonpharmacological strategies (eg, acupuncture and psychotherapy) based on pain pathophysiology and characteristics. Following pain symptoms, 50 - 90% of patients experience fatigue, which negatively affects their daily activities and QoL [14]. Insomnia is also a common symptom in cancer survivors and can have a systematic effect on psychological burdens, such as stress, fatigue, and depression [15,16].

The symptoms experienced by cancer survivors and their relationship with psychological problems often benefit from psychotherapy. The benefits of psychotherapy can be explained by the body-mind-spirit model [17], which highlights the interconnectedness of physical, mental, and spiritual health [18]. Commonly used psychotherapies include mindfulness-based stress reduction-based interventions and cognitive behavioral therapy (CBT). These therapies are effective in reducing symptoms in cancer survivors, particularly chronic pain and stress [19-21]. CBT is considered beneficial for alleviating pain and other symptoms by reducing catastrophic thinking and enhancing self-efficacy in coping with symptoms such as pain [22]. Similarly, mindfulness-based interventions are considered beneficial for chronic pain by promoting mindfulness and promoting greater acceptance of pain or other symptoms [22]. Unlike traditional psychotherapies, such as CBT, which primarily focus on cognitive restructuring, mindfulness interventions offer the unique benefit of directly enhancing patients' capacity for present-moment awareness and acceptance of their experiences.

Advancements in healthcare information technology along with the broader accessibility of healthcare services have driven the rapid growth of remote-based interventions. The intervention spans a wide array of practices and specialties, facilitating interactions through various modalities such as telephone, email, video conferencing, online platforms, and remote monitoring devices. The rapid growth of remote-based methods has led to the delivery of mindfulness through the internet. Remote-based interventions have been integrated into cancer care and treatment, which suggests a benefit in treatment outcomes [23]. Remote-based mindfulness is defined as a psychotherapy program that uses a technological device that ensures interactive and immediate communication and does not require the patient to be present with the therapist [24].

Recent evidence suggests the benefits of remote-based interventions using a website on psychological well-being, such as reducing distress, depression, and anxiety [25-27]. Remote-based interventions may be more suitable for patients who experience weakness and fatigue due to physical limitations, such as cancer survivors. A study conducted by Schellekens et al suggested the benefit of web-based mindfulness-based cognitive therapy programs for improving care outcomes in patients with chronic cancer-related fatigue [28]. While a previous meta-analysis has evaluated the benefit of remote-based mindfulness for cancer survivors [29,30], its focus on physical symptom outcomes remains limited. Therefore, this study aimed to explore the benefit of remote-based mindfulness interventions on physical symptom outcomes as a primary and/or secondary outcome of trial studies in cancer survivors.

Methods

Study Design

This study was a systematic review and meta-analysis. This study was presented in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA; Checklist 1) [31]. The protocol was not prospectively registered in any database such as PROSPERO (Prospective Register of Systematic Reviews).

Eligibility Criteria

The inclusion criteria were defined according to the Population, Intervention, Comparison, Outcome (PICO) framework. The population of the included studies was diagnosed with cancer through imaging, laboratory tests (including tumor marker tests), tumor biopsies, endoscopic examinations, surgeries, and genetic testing. Interventions were remote-based mindfulness interventions defined as mindfulness interventions that integrated information and communication technology, such as mobile phones, websites, mobile apps, and asynchronous instruction with text-based reminder messages. Comparisons were defined as standard or usual care with face-to-face mindfulness interventions, or standard cancer care. The outcomes of this study included the physical symptoms related to cancer outcomes. Physical symptoms were defined as the subjective experiences of distressing somatic symptoms (eg, fatigue, insomnia, and pain). The outcomes were measured using self-reports or standard questionnaires. The exclusion criteria were the types of articles, such as case reports, editorials, invited commentary, reviews, non-research letters, and

abstract-only articles. To prevent bias, articles published before 2012 and those written in a language other than English as an international language were excluded from this study. This review focused on studies published after 2012 to ensure that the findings represented the most recent advancements in technology, healthcare practices, and guidelines that have progressed markedly over the past decade.

Study Search Strategy and Selection Process

The selection process for this study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocol. This review systematically searched electronic databases, namely Scopus, Medline, PubMed, Cumulated Index in Nursing and Allied Health Literature (CINAHL), EBSCOhost, and the Cochrane Central Database. The search was conducted until December 2022. The following keywords were used. (All Fields] OR Internet-based intervention "web-based"[All Fields] OR "internet-based intervention"[All Fields] OR "online based"[All Fields]) AND ("mind s"[All Fields] OR "minded"[All Fields] OR "mindful"[All Fields] OR "mindfulness"[MeSH Terms] OR "mindfulness"[All Fields] OR "mindfulness intervention"[All Fields] OR "mindfulness-based stress reduction"[All Fields] OR "mindfulness- based cognitive therapy"[All Fields]) AND ("cancer s"[All Fields] OR "cancer"[All Fields] OR "cancers"[All Fields] OR "oncology patients"[All Fields] OR "Patients with cancer" [All Fields]). The detailed search strategy can be found in Multimedia Appendix 1. In addition, we used a hand-searched reference list of the included studies to expand the number of additional studies.

The reference manager automatically removed duplicate articles using Mendeley (Mendeley Ltd.). Two independent authors (SM and SA) initially screened the text (eg, title and abstract). The full text of the articles that met the eligibility criteria were independently assessed by two independent authors. At this stage, the articles were meticulously evaluated based on predetermined inclusion and exclusion criteria, and irrelevant studies were excluded. Discrepancies were resolved by a third reviewer (MK).

Data Extraction

Two authors (MK and SM) independently extracted data using standard tabulation tables (spreadsheets). The following data were included: study characteristics (ie, author, year, study design, country, model intervention, and follow-up duration); participant characteristics (ie, average age, education level, number of participants, and cancer site); and physical symptoms (eg pain, fatigue, and insomnia). Data extraction was performed independently and disagreements were resolved through discussion and consensus among the authors.

This study assessed the quality of this randomized-controlled trial (RCT) using the Cochrane risk of bias, version 2.0. Three authors (MK, SM, and HP) evaluated the enrolled studies separately. The following factors were considered in the assessment: bias arising from random processes, bias due to deviation from the intended intervention, bias due to missing outcome data, bias in outcome measures, and bias in selection of reported outcomes. This discourse resolved the differences in perceptions regarding the quality of the research.

Statistical Analysis

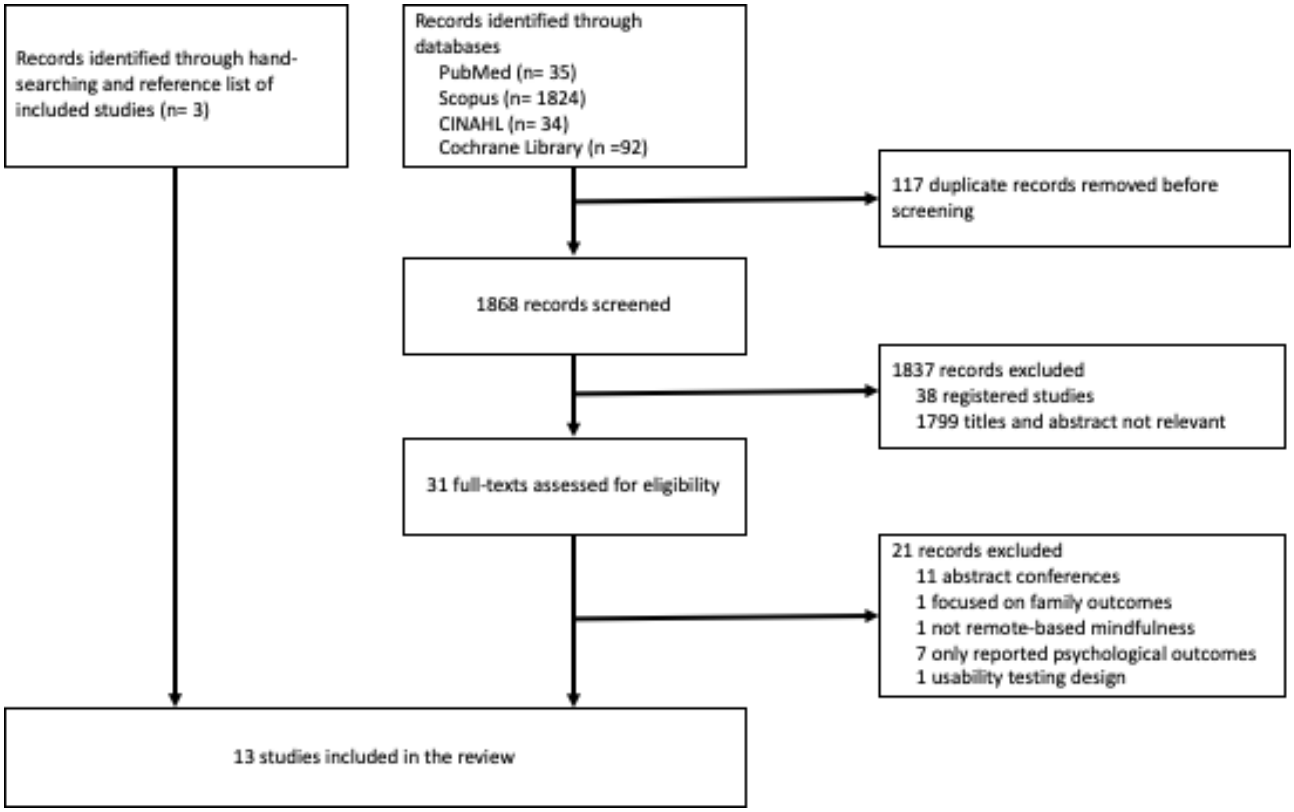
All statistical analyses were performed using Review Manager version 5.4.1 (RevMan) [32]. This study estimated the effect size in the form of the standardized mean difference (SMD) for the outcome and the mean difference (MD), with the 95% CI. The SMD was used when the outcomes were measured in different units across studies. The MD was used when the outcomes were measured in the same unit across studies. The SMD criteria were divided into three categories: low, medium, and large effects, with values of <0.5, ≥05, and ≥0.8, respectively [33]. This review conducted posttreatment analysis that reported pre- and post–remote-based intervention. We also conducted a comparison between remote-based intervention and usual care after treatment. The inconsistency index (I^2) and subgroup analysis using the I^2 test were used to identify potential causes of heterogeneity. An I^2 value of >50% and a P-value of <.05 were considered statistically significant for heterogeneity [34]. A random-effects model was applied despite the study heterogeneity to account for interstudy variability [35]. In this study, a two-tailed P value of .05 was considered statistically significant. Publication bias was analyzed qualitatively using a contour-enhanced funnel plot and quantitatively using the Egger linear regression test.

Results

Study Selection

The process of selecting the studies for inclusion in the review is presented in Figure 1. An initial search across PubMed, CINAHL, Scopus, and Cochrane Library databases yielded 1985 articles. A total of 177 duplicate articles were removed before screening, resulting in 1868 articles. After screening, 1837 studies were excluded because of 38 preregistered studies (eg ClinicalTrials.gov), and 1799 titles and abstracts were not relevant. After assessing 31 full-text articles for eligibility, 21 studies were conference abstracts, focused on family outcomes, not remote-based mindfulness or usability testing, and did not report the physical outcomes. Ten studies met the criteria identified through the database, and 3 studies were identified through manual searches and reference lists of the included studies. Hence, 13 studies were included in the systematic review and meta-analysis [26,36-47].

Figure 1. Study selection process.



Characteristics of the Included Studies

The mean age of the participants was <60 years in 10 of the included studies and ≥60 years in 2 studies. Most participants were female, with 74.38% (572/894) in the remote-based intervention group and 70.61% (322/894) in the usual care group. The studies were conducted across several countries, with most studies conducted in the United States (n=4) and the Netherlands (n=4), followed by China (n=2), and 1 each in Ireland, Denmark, and Iran. Regarding the study design, 11 studies were RCTs and 2 were pilot RCTs. The mindfulness type included web-based interventions, mobile apps, and virtual meetings, whereas the control groups included treatment as usual, wait-list controls, face-to-face mindfulness, and interventions without a control group. The outcome measurements included assessments of fatigue, sleep disturbances, and physical function. Fatigue was measured in 5 studies by using different instruments, such as the checklist of individual strength (CIS)-fatigue, BFI-9, fatigue symptom inventory, and cancer quality of life questionnaire (QLQ)-C30. Sleep disturbance was evaluated in 6 studies using tools such as the patient-reported outcomes measurement information system (PROMIS), Pittsburgh sleep quality index (PSQI), and insomnia severity index (ISI). Physical function was measured in 7 studies, most frequently using the short form (SF)-12, functional assessment of chronic illness therapy (FACIT), and QLQ-C30. The details of these characteristics are presented in Table 1 and Multimedia Appendix 2.

Table . Characteristic of the included studies (n=13).

Characteristics		
Mean age (years)	Number of studies (n=13)	Reference
<60	10	[36-41,43-45,47]
≥60	2	[26,42]
Data not available	1	[46]
Sex (Female)	n (%)	Reference
Remote-based group	572 (74.38)	NA
Usual care	322 (70.61)	NA
Country	Number of studies (n=13)	Reference
United States of America	4	[41,42,44,45]
Ireland	1	[40]
Netherlands	4	[36-39]
Denmark	1	[26]
China	2	[43,46]
Iran	1	[47]
Study design		Reference
RCT[a]	11	[26,37-43,45-47]
Pilot-RCT	2	[36,44]
Type of mindfulness delivered	Number of studies (n=13)	Reference
Web-based	5	[36-39,43]
Mobile apps	4	[26,41,42,44]
Virtual meeting	2	[46,47]
Unspecified	2	[40,45]
Type of control group	Number of studies (n=13)	Reference
Treatment as usual	6	[38-40,45-47]
Wait-list control	4	[26,41-43]
Face-to-face mindfulness	1	[37]
Without control	2	[36,44]
Fatigue measurement	Number of studies (n=5)	Reference
CIS-Fatigue[b]	1	[36]
BFI-9	1	[41]
FSI[c]	2	[44,45]
QLQ-30[d]	1	[47]
Sleep disturbance measurement	Number of studies (n=6)	Reference
PROMIS[e]	1	[41]
PSQI[f]	3	[43-45]
ISI[g]	2	[26,47]
Physical function measurement	Number of studies (n=7)	Reference
SF-12[h]	4	[37-39,44]
FACIT[i]	2	[41,42]
QLQ-30	1	[46]

[a]RCT: randomized-controlled trial.

[b]CIS-fatigue: checklist individual strength for fatigue.

[c]FSI: fatigue symptom inventory.

[d]QLQ-C30: Cancer Quality of Life Questionnaire- C30.

[e]PROMIS: patient-reported outcome measurement information system.

[f]PSQI: Pittsburgh Sleep Quality Index.

[g]ISI: insomnia severity index.

[h]SF-12: short form-12 items.

[i]FACIT: Functional Assessment of Chronic Illness Therapy.

Study Outcomes

A meta-analysis of remote-based mindfulness revealed 4 physical outcomes. The outcomes included fatigue (n=5), sleep disturbance (n=6), pain (n=3), and physical function (n=6). The outcome measurements varied, as shown in Table 1. The effect sizes for each outcome are listed in Table 2.

Table . Effect size of mobile-based mindfulness on physical symptoms in cancer survivors.

Outcome	Number of studies	Effect size	95% CI	*P* value	Heterogeneity	Reference
Pre- and postintervention						
Fatigue	5	SMD[a] −0.94	−1.56 to −0.33	.002*[b]	85%	[36,41,44,45,47]
Sleep disturbance	6	SMD −0.36	−0.60 to −0.12	.004*	31%	[26,41,43-45,47]
Pain	3	MD[c] −5.33	−10.90 to 0.25	.06	85%	[40,41,44]
Physical function	6	SMD 0.25	0.09 to 0.41	.002*	0%	[37-39,41,44,46]
Controlled intervention						
Fatigue	3	SMD −1.09	−2.87 to 0.68	.23	95%	[41,45,47]
Sleep disturbance	5	SMD −0.37	−0.58 to −0.16	.006*	46%	[26,41,43,45,47]
Pain	2	MD −0.90	−2.31 to 0.52	.21	0%	[40,41]
Physical function	5	SMD 0.59	−0.06 to 1.24	.08	92%	[38,39,41,42,46]

[a]SMD: Standard mean difference.

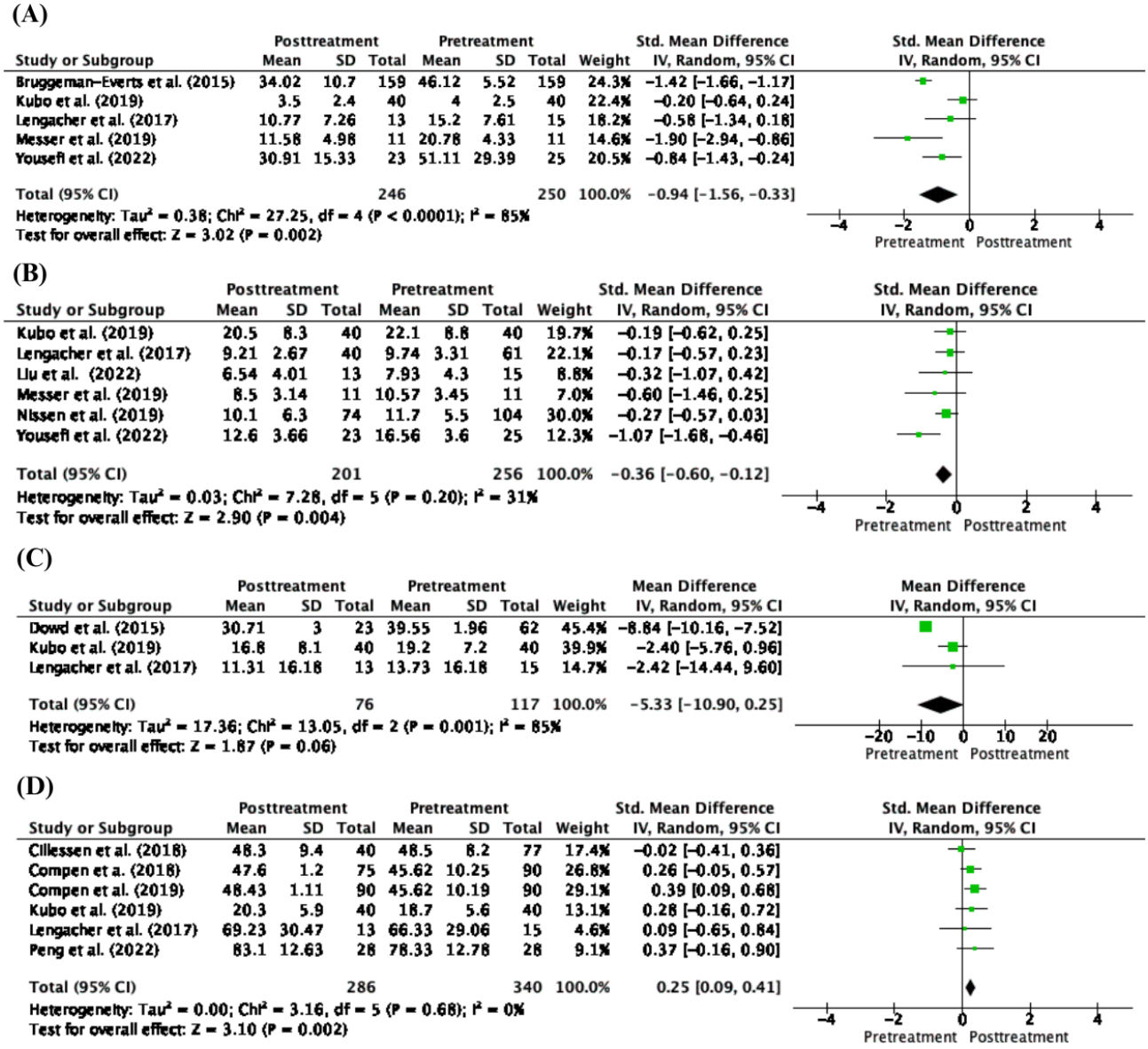
[b]The asterisk indicates statistical significance (P<.05)

[c]MD: Mean difference.

Pre- and Postanalysis of Remote-Based Mindfulness to Physical Outcomes After Treatment

After remote-based mindfulness treatment, cancer survivors showed a significant reduction in fatigue (SMD −0.94; 95% CI: −1.56 to −0.33; P=.002), sleep disturbance (SMD −0.36; 95% CI: −0.60 to −0.12; P=.004), and improvement in physical function (SMD 0.25; 95% CI: 0.009 to 0.41; P=.002) compared with baseline or pretreatment values. Although posttreatment outcomes were more favorable compared to baseline values, there was no statistically significant difference in pain reduction (MD −5.33; 95% CI: −10.90 to 0.25; P=.06; Table 2). A forest plot of the pre- and posttreatment meta-analyses conducted on the remote-based mindfulness group is shown in Figure 2. Among these 4 outcomes, fatigue and pain showed significant heterogeneity (I^2=85%).

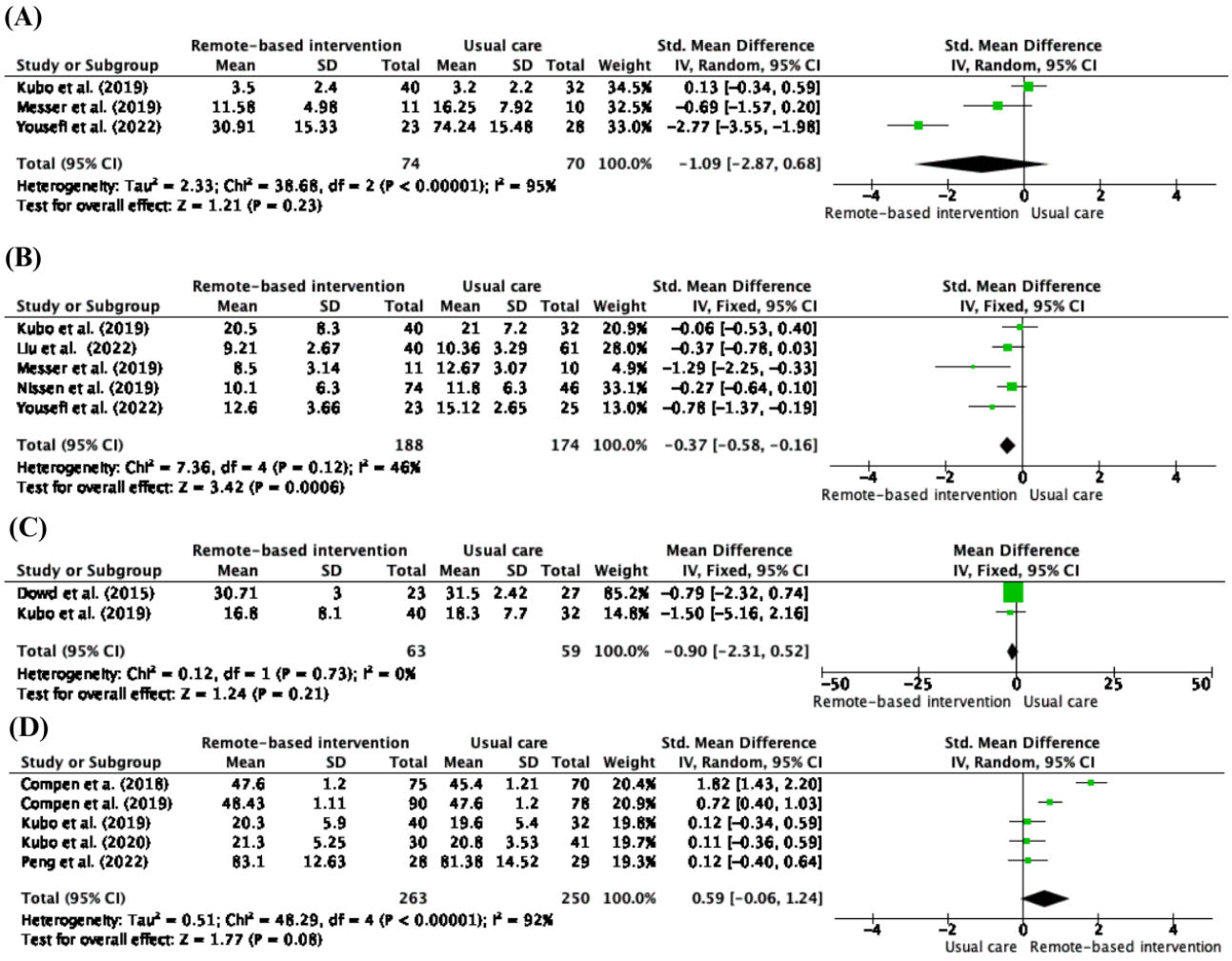
Figure 2. Meta-analysis of the benefits of remoted-based mindfulness intervention on physical symptoms after treatment. (**A**) Fatigue outcome. (**B**) Sleep disturbance outcome. (**C**) Pain outcome. (**D**) Physical function.



Benefits of Remote-Based Mindfulness on Physical Symptoms Compared to Usual Care After Treatment

Despite the small effect, the meta-analysis showed that remote-based mindfulness significantly reduced sleep disturbance (SMD −0.37; 95% CI: −0.58 to −0.16, *P*=.0006) compared with usual care after treatment. There were no statistically significant differences in the reduction of fatigue, pain, or improvement of physical function between the remote-based mindfulness and usual care groups (Table 2). Although not statistically significant, the remote-based mindfulness group had reduced fatigue, sleep disturbance, and pain compared with the usual care group after treatment. The forest plot of the meta-analysis of the benefits of remote-based mindfulness compared to usual care after treatment is shown in Figure 3.

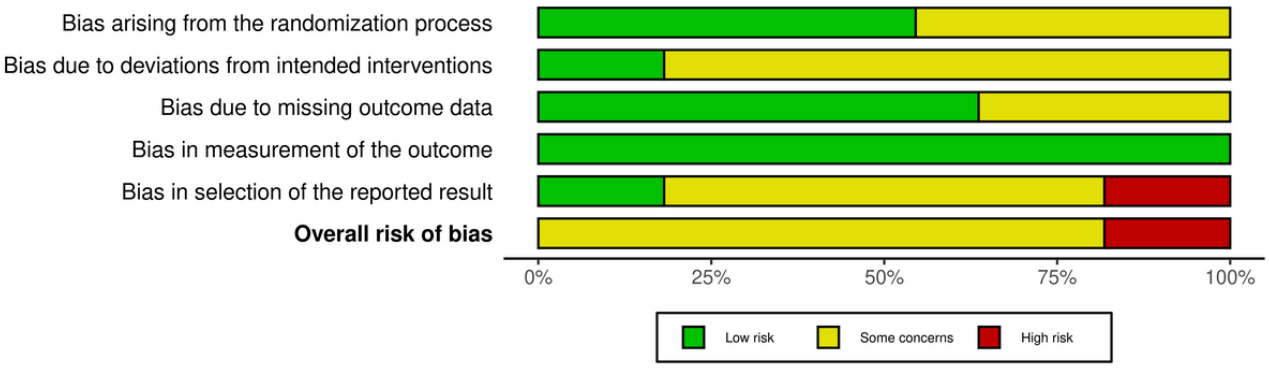
Figure 3. Meta-analysis of the benefits of remote-based mindfulness intervention on physical symptoms compared to usual care. (**A**) Fatigue outcome. (**B**) Sleep disturbance outcome. (**C**) Pain outcome. (**D**) Physical function.



Quality Assessment

Over 75% of the studies showed some concerns in at least 1 domain, but no study was rated as high risk considering the measurement of the outcomes (Figure 4). Most studies showed a low risk of bias across most domains, particularly for bias in the measurement of outcomes and missing outcome data. However, some concerns were found regarding the bias arising from the randomization process and deviations from intended interventions, with several studies lacking sufficient details on allocation concealment or participant adherence. Two studies, notably those by Cillessen et al (2018) and Nissen et al (2019), demonstrated a high risk of bias in the selection of the reported results. These studies may have selectively reported favorable outcomes, raising concerns about the validity of their findings. A detailed assessment of each included study can be found in the traffic-light plot provided in Multimedia Appendix 3.

Figure 4. Summary risk of bias.



Publication Bias

We evaluated the likelihood of publication bias by analyzing funnel plots and using the Egger test. We did not conduct statistical tests or create funnel plots for any outcome because each outcome had fewer than 10 studies, which is necessary to ensure sufficient power for detecting asymmetry [29,48].

Discussion

Study Findings and Comparison With Previous Works

To the best of our knowledge, this systematic review and meta-analysis is the first to assess the benefits of remote-based mindfulness interventions on physical outcomes in individuals living with cancer. This study has yielded several findings. First, the meta-analysis concluded that a significant effect was observed in reducing fatigue and sleep disturbance after treatment. Second, remote-based mindfulness was significantly more effective in reducing sleep disturbances compared to usual care. Third, remote-based mindfulness was not significantly effective at reducing pain. Finally, a significant improvement in physical function was observed after treatment.

The present meta-analysis suggests that remote-based mindfulness is beneficial for improving physical outcomes. The present study adds to the knowledge regarding the benefits of remote-based mindfulness in cancer survivors. A previous meta-analysis suggested that remote-based mindfulness reduces psychological symptoms in cancer survivors, such as depression, distress, and perceived stress [30,49]. Another meta-analysis observed a significant effect of remote-based mindfulness with a specific web-based platform in reducing anxiety, depression, and distress [29].

The biological mechanisms underlying the benefits of mindfulness treatments suggest additional pathways that may strengthen evidence-based understanding of their physical health effects. Preliminary supporting studies indicate that mindfulness interventions promote two pathways of stress resilience in the brain (the regulatory and reactivity pathways) and may enhance the regulation of the stress reactivity of the hypothalamic-pituitary-adrenal and sympathetic-adrenal-medullary axes, thereby elucidating the effects of mindfulness interventions on stress-related health and disease outcomes over time [50]. The effectiveness of remote-based mindfulness can be understood through the body-mind-spirit model, in which physical health is influenced by the interconnectedness of biological and psychological factors involving self-regulation [17,18]. This self-regulation encompasses the release of dopamine, endocannabinoids, endorphins, and stress hormones in addition to the signaling pathways of oxytocin and serotonin [51].

The present meta-analysis showed a significant effect in reducing sleep disturbance compared with usual care, which is consistent with the findings of a previous meta-analysis [29]. Mindfulness treatment has the potential to alleviate sleep disturbances because mindfulness practice enables individuals to observe their thoughts, emotions, and bodily sensations without emotional involvement or judgment [52]. It also seeks to increase an individual's awareness and acceptance of their thoughts, emotions, and physiological sensations. This treatment improves cognitive flexibility and cultivates a more comprehensive understanding of sleep, thereby alleviating anxiety or arousal, which may exacerbate sleep disturbances [30].

Despite the present meta-analysis showing that remote-based mindfulness significantly reduced fatigue after treatment, the results showed no significant difference when compared with usual care. Consistent with a previous meta-analysis of web-based mindfulness, there was no significant effect compared to usual care [30]. This may align with the different types of cancer and stages, types of technological intervention, treatment duration, and diverse measurement instruments within the studied population. Despite this, remote-based mindfulness showed high effectiveness after treatment, which aligns with a previous meta-analysis of face-to-face mindfulness [53]. A meta-analysis conducted by Johns et al showed a moderate effect after treatment and a small effect at the first-month follow-up [53]. Remote-based mindfulness is well-documented for its efficacy in reducing and managing stress, which may subsequently impact fatigue. Furthermore, fatigue may be alleviated by enhancing insomnia, as better sleep quality leads to increased freshness [47]. Peripheral inflammatory cytokines can communicate with the central nervous system to induce cancer-related fatigue [54]. Mindfulness, such as the body-mind-spirit technique, may reduce NF-kB signaling, a major regulator of inflammatory activity [55].

This meta-analysis showed no significant difference in pain reduction compared to usual care. This outcome may be attributed to the fact that both the remote-based and control groups were provided with standard care, which included adequate analgesic administration as part of their standard treatment protocol [56]. Mindfulness-based interventions may have been marked by the high efficacy of analgesics in alleviating chronic pain in cancer survivors. A previous meta-analysis of face-to-face mindfulness showed only a small effect in reducing chronic pain in various health conditions [57]. A psychotherapy form similar to online-based acceptance and commitment therapy showed moderately reduced chronic pain in various health conditions [58].

Evidence suggests that remote-based mindfulness improves QoL [29], with no exception to the present meta-analysis, which showed that remote-based mindfulness significantly improved the physical function of QoL after treatment. By reducing cancer-related symptoms, including physical symptoms, remote-based mindfulness can improve physical function. However, the present meta-analysis concluded that there was no significant improvement in physical function compared with usual care. This result may largely benefit psychological outcomes rather than physical health outcomes.

Future Direction

This evidence suggests a potential remote-based mindfulness intervention to alleviate physical symptoms (eg, sleep disturbance and fatigue) and improved physical function. The understanding of mindfulness interventions, including remote-based mindfulness, and their benefit on physical health remains insufficient considering the large RCT literature

associating mindfulness interventions with psychological outcomes [50,59]. Further research is needed to evaluate the efficacy of remote-based mindfulness in improving physical outcomes (eg, blood pressure, weight loss, and biomarkers of health). Integrating mindfulness practices into supportive care programs acknowledges the importance of addressing multidimensional aspects of a patient's experience. This personalized and holistic approach aligns with the principles of patient-centered care, recognizing the unique needs and challenges faced by individuals undergoing cancer treatment.

Despite the small number of included studies, the evidence of the pain outcomes suggests the limited benefit of remote-based mindfulness intervention due to the administration of standard analgesics in both groups [56]. Considering the analgesic effects induced within the central nervous system, the common adverse effects of opioids include nausea, vomiting, constipation, drowsiness, disorientation, hallucinations, and respiratory depression. Other adverse effects include endocrine alterations (eg, androgen insufficiency and bone demineralization) and the risk of depression due to long-term opioid prescriptions [51]. Owing to the growing "opioid crisis," the use of opioids as a psychotherapy option is now being recommended as a complementary treatment. Hence, further research and modification of mindfulness interventions with other psychotherapies is needed to enhance the benefits and evidence of remote-based mindfulness on pain.

Limitations

Despite this present study indicating the potential effects of remote-based mindfulness on physical health outcomes and physical status, our study has several limitations. This meta-analysis was not registered prospectively in any registered database such as PROSPERO. The transparency of this meta-analysis was limited because of the minimized risk of selective reporting. A few studies included in the meta-analysis had a high bias in the selection of the reported results that influenced the concern that positive results are more likely to be published. Meta-regression was not performed in the present meta-analysis to assess potential moderating factors such as participant characteristics, intervention components, or variations in study design. Moreover, this systematic review and meta-analysis assessed mindfulness as psychotherapy, and the included studies were unlikely to evaluate physical health outcomes as primary outcomes.

Conclusion

This meta-analysis provided evidence regarding remote-based mindfulness interventions to alleviate physical symptoms in cancer survivors. The findings of this study suggest that remote-based mindfulness interventions may be effective in reducing sleep disturbances in clinical practice. Despite limited evidence regarding its benefits compared with usual care, the effect of remote-based mindfulness on fatigue and physical function was observed after treatment. Due to the limited number of included studies and the heterogeneity of the included studies, the conclusions must be considered along with these limitations. Therefore, well-designed trials are required to obtain robust evidence.

Acknowledgments

We would like to express our heartfelt gratitude to Universitas Padjadjaran for providing a grant to conduct this systematic review and meta-analysis. We also thank you for the financial support provided by the article processing charges (APC).

Data Availability

The published article and its supplementary information files contain all study data.

Authors' Contributions

MK, SM, HP, and LR were involved in the conceptualization. SM and SA performed the formal investigation and analysis. All authors contributed to writing the review and editing. All authors contributed to and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 36 KB - cancer_v11i1e54154_app1.docx]

Multimedia Appendix 2
Characteristic of the included studies.
[DOCX File, 19 KB - cancer_v11i1e54154_app2.docx]

Multimedia Appendix 3
Traffic light plot.

[DOCX File, 418 KB - cancer_v11i1e54154_app3.docx]

Checklist 1
PRISMA Checklist.
[PDF File, 110 KB - cancer_v11i1e54154_app4.pdf]

References

1. De Angelis R, Sant M, Coleman MP, et al. Cancer survival in Europe 1999–2007 by country and age: results of EUROCARE-5—a population-based study. Lancet Oncol 2014 Jan;15(1):23-34. [doi: 10.1016/S1470-2045(13)70546-1] [Medline: 24314615]
2. Reilly CM, Bruner DW, Mitchell SA, et al. A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment. Supp Care Cancer 2013 Jun;21(6):1525-1550. [doi: 10.1007/s00520-012-1688-0] [Medline: 23314601]
3. Färkkilä N, Torvinen S, Roine RP, et al. Health-related quality of life among breast, prostate, and colorectal cancer patients with end-stage disease. Qual Life Res 2014 May;23(4):1387-1394. [doi: 10.1007/s11136-013-0562-y] [Medline: 24178630]
4. Färkkilä N, Sintonen H, Saarto T, et al. Health‐related quality of life in colorectal cancer. Colorectal Dis 2013 May;15(5):e215-e222. [doi: 10.1111/codi.12143] [Medline: 23351057]
5. Kokkonen K, Saarto T, Mäkinen T, et al. The functional capacity and quality of life of women with advanced breast cancer. Breast Cancer (Auckl) 2017 Jan;24(1):128-136. [doi: 10.1007/s12282-016-0687-2] [Medline: 27002988]
6. Adamowicz K. Assessment of quality of life in advanced, metastatic prostate cancer: an overview of randomized phase III trials. Qual Life Res 2017 Apr;26(4):813-822. [doi: 10.1007/s11136-016-1429-9] [Medline: 27738867]
7. Rautalin M, Färkkilä N, Sintonen H, et al. Health-related quality of life in different states of breast cancer – comparing different instruments. Acta Oncol 2018 May;57(5):622-628. [doi: 10.1080/0284186X.2017.1400683] [Medline: 29140139]
8. Singer AE, Goebel JR, Kim YS, et al. Populations and interventions for palliative and end-of-life care: a systematic review. J Palliat Med 2016 Sep;19(9):995-1008. [doi: 10.1089/jpm.2015.0367] [Medline: 27533892]
9. Gaertner J, Siemens W, Meerpohl JJ, et al. Effect of specialist palliative care services on quality of life in adults with advanced incurable illness in hospital, hospice, or community settings: systematic review and meta-analysis. BMJ 2017 Jul 4;357:j2925. [doi: 10.1136/bmj.j2925] [Medline: 28676557]
10. Kavalieratos D, Corbelli J, Zhang D, et al. Association between palliative care and patient and caregiver outcomes: a systematic review and meta-analysis. JAMA 2016 Nov 22;316(20):2104-2114. [doi: 10.1001/jama.2016.16840] [Medline: 27893131]
11. Jordan K, Aapro M, Kaasa S, et al. European Society for Medical Oncology (ESMO) position paper on supportive and palliative care. Ann Oncol 2018 Jan 1;29(1):36-43. [doi: 10.1093/annonc/mdx757] [Medline: 29253069]
12. Löwe B, Toussaint A, Rosmalen JGM, et al. Persistent physical symptoms: definition, genesis, and management. Lancet 2024 Jun 15;403(10444):2649-2662. [doi: 10.1016/S0140-6736(24)00623-8] [Medline: 38879263]
13. WHO guidelines for the pharmacological and radiotherapeutic management of cancer pain in adults and adolescents. : World Health Organization; 2019 URL: https://www.who.int/publications/i/item/9789241550390 [accessed 2025-01-10]
14. Hassan BJ, Riechelmann R, Del Giglio A. Cancer-related fatigue: a review. Rev Assoc Med Bras 1992;57(2). [doi: 10.1590/S0104-42302011000200021]
15. Büttner-Teleagă A, Kim YT, Osel T, Richter K. Sleep disorders in cancer-a systematic review. Int J Environ Res Public Health 2021 Nov 7;18(21):11696. [doi: 10.3390/ijerph182111696] [Medline: 34770209]
16. Bortsov AV, Millikan RC, Belfer I, Boortz-Marx RL, Arora H, McLean SA. Mu-opioid receptor gene A118G polymorphism predicts survival in patients with breast cancer. Anesthesiology 2012 Apr;116(4):896-902. [doi: 10.1097/ALN.0b013e31824b96a1] [Medline: 22433205]
17. Komariah M, Agustina HR, Rahayuwati L, et al. Strengthening psychological well-being of Indonesian females with breast cancer through the religious-based caring program: A quasi-experimental study among Muslim population. Belitung Nurs J 2024;10(4):416-423. [doi: 10.33546/bnj.3452] [Medline: 39211456]
18. Mark G, Lyons A. Conceptualizing mind, body, spirit interconnections through, and beyond, spiritual healing practices. Expl NY 2014;10(5):294-299. [doi: 10.1016/j.explore.2014.06.003] [Medline: 25048457]
19. Dils AT, O'Keefe K, Dakka N, Azar M, Chen M, Zhang A. The efficacy of cognitive behavioral therapy for mental health and quality of life among individuals diagnosed with cancer: A systematic review and meta-analysis. Cancer Med 2024 Aug;13(16):e70063. [doi: 10.1002/cam4.70063] [Medline: 39165223]
20. Cillessen L, Johannsen M, Speckens AEM, Zachariae R. Mindfulness‐based interventions for psychological and physical health outcomes in cancer patients and survivors: A systematic review and meta‐analysis of randomized controlled trials. Psychooncol 2019 Dec;28(12):2257-2269. [doi: 10.1002/pon.5214] [Medline: 31464026]
21. Feng B, Hu X, Lu WW, Wang Y, Ip WY. Are mindfulness treatments effective for pain in cancer patients? A systematic review and meta‐analysis. Eur J Pain 2022 Jan;26(1):61-76. [doi: 10.1002/ejp.1849] [Medline: 34369040]

22. Turner JA, Anderson ML, Balderson BH, Cook AJ, Sherman KJ, Cherkin DC. Mindfulness-based stress reduction and cognitive behavioral therapy for chronic low back pain: similar effects on mindfulness, catastrophizing, self-efficacy, and acceptance in a randomized controlled trial. Pain 2016 Nov;157(11):2434-2444. [doi: 10.1097/j.pain.0000000000000635] [Medline: 27257859]
23. Komariah M, Maulana S, Platini H, Pahria T. A scoping review of telenursing's potential as a nursing care delivery model in lung cancer during the COVID-19 pandemic. J Multidiscip Healthc 2021;14:3083-3092. [doi: 10.2147/JMDH.S337732] [Medline: 34754196]
24. Comendador L, Jiménez-Villamizar MP, Losilla JM, et al. Effect of synchronous remote-based interventions on suicidal behaviours: protocol for a systematic review and meta-analysis. BMJ Open 2023 Dec 6;13(12):e075116. [doi: 10.1136/bmjopen-2023-075116] [Medline: 38056944]
25. van den Berg SW, Gielissen MFM, Custers JAE, van der Graaf WTA, Ottevanger PB, Prins JB. BREATH: web-based self-management for psychological adjustment after primary breast cancer—results of a multicenter randomized controlled trial. J Clin Oncol 2015 Sep 1;33(25):2763-2771. [doi: 10.1200/JCO.2013.54.9386] [Medline: 26169621]
26. Nissen ER, O'Connor M, Kaldo V, et al. Internet-delivered mindfulness-based cognitive therapy for anxiety and depression in cancer survivors: a randomized controlled trial. Psychooncol 2020 Jan;29(1):68-75. [doi: 10.1002/pon.5237] [Medline: 31600414]
27. McCann L, McMillan KA, Pugh G. Digital interventions to support adolescents and young adults with cancer: systematic review. JMIR Cancer 2019 Jul 31;5(2):e12071. [doi: 10.2196/12071] [Medline: 31368438]
28. Schellekens MPJ, Bruggeman-Everts FZ, Wolvers MDJ, Vollenbroek-Hutten MMR, van der Lee ML. What web-based intervention for chronic cancer-related fatigue works best for whom? Explorative moderation analyses of a randomized controlled trial. Supp Care Cancer 2022 Oct;30(10):7885-7892. [doi: 10.1007/s00520-022-07223-y] [Medline: 35726109]
29. Wang T, Tang C, Jiang X, Guo Y, Zhu S, Xu Q. Effectiveness of web-based mindfulness-based interventions for patients with cancer: systematic review and meta-analyses. J Med Internet Res 2024 Jun 25;26:e47704. [doi: 10.2196/47704] [Medline: 38917445]
30. Fan M, Wang Y, Zheng L, Cui M, Zhou X, Liu Z. Effectiveness of online mindfulness-based interventions for cancer patients: a systematic review and meta-analysis. Jpn J Clin Oncol 2023 Nov 5;53(11):1068-1076. [doi: 10.1093/jjco/hyad101]
31. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
32. Review Manager (RevMan). The Cochrane Collaboration. 2020. URL: https://login.cochrane.org/realms/cochrane/protocol/openid-connect/auth?client_id=revman-web&redirect_uri=https%3A%2F%2Frevman.cochrane.
[accessed 2024-02-14]
33. Andrade C. Mean difference, standardized mean difference (smd), and their use in meta-analysis: as simple as it gets. J Clin Psychiatry 2020 Sep 22;81(5):20f13681. [doi: 10.4088/JCP.20f13681] [Medline: 32965803]
34. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560. [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
35. Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ 2019 Jan 30;364:k4597. [doi: 10.1136/bmj.k4597] [Medline: 30700442]
36. Bruggeman Everts FZ, van der Lee ML, de Jager Meezenbroek E. Web-based individual mindfulness-based cognitive therapy for cancer-related fatigue — A pilot study. Internet Interv 2015 May;2(2):200-213. [doi: 10.1016/j.invent.2015.03.004]
37. Cillessen L, Schellekens MPJ, Van de Ven MOM, et al. Consolidation and prediction of long-term treatment effect of group and online mindfulness-based cognitive therapy for distressed cancer patients. Acta Oncol 2018 Oct;57(10):1293-1302. [doi: 10.1080/0284186X.2018.1479071] [Medline: 29932784]
38. Compen F, Bisseling E, Schellekens M, et al. Face-to-face and internet-based mindfulness-based cognitive therapy compared with treatment as usual in reducing psychological distress in patients with cancer: a multicenter randomized controlled trial. J C O 2018 Aug 10;36(23):2413-2421. [doi: 10.1200/JCO.2017.76.5669]
39. Compen F, Bisseling E, Schellekens M, Donders R, Carlson L. RCT naar mindfulness-based cognitieve therapie voor psychische klachten bij oncologische patiënten. Tijdschr Psychiatr 2019.
40. Dowd H, Hogan MJ, McGuire BE, et al. Comparison of an online mindfulness-based cognitive therapy intervention with online pain management psychoeducation. Clin J Pain 2015 Jun;31(6):517-527. [doi: 10.1097/AJP.0000000000000201] [Medline: 25565584]
41. Kubo A, Kurtovich E, McGinnis M, et al. A randomized controlled trial of mHealth mindfulness intervention for cancer patients and informal cancer caregivers: a feasibility study within an integrated health care delivery system. Integr Cancer Ther 2019;18:1534735419850634. [doi: 10.1177/1534735419850634] [Medline: 31092044]
42. Kubo A, Kurtovich E, McGinnis M, et al. Pilot pragmatic randomized trial of mHealth mindfulness-based intervention for advanced cancer patients and their informal caregivers. Psychooncol 2024 Feb;33(2):e5557. [doi: 10.1002/pon.5557] [Medline: 32979294]

43. Liu Z, Li M, Jia Y, et al. A randomized clinical trial of guided self-help intervention based on mindfulness for patients with hepatocellular carcinoma: effects and mechanisms. Jpn J Clin Oncol 2022 Mar 3;52(3):227-236. [doi: 10.1093/jjco/hyab198] [Medline: 35088079]
44. Lengacher CA, Reich RR, Ramesar S, et al. Feasibility of the mobile mindfulness‐based stress reduction for breast cancer (mmbsr(bc)) program for symptom improvement among breast cancer survivors. Psychooncol 2018 Feb;27(2):524-531. [doi: 10.1002/pon.4491] [Medline: 28665541]
45. Messer D, Horan JJ, Larkey LK, Shanholtz CE. Effects of internet training in mindfulness meditation on variables related to cancer recovery. Mindfulness (NY) 2019 Oct;10(10):2143-2151. [doi: 10.1007/s12671-019-01182-y]
46. Peng L, Yang Y, Chen M, et al. Effects of an online mindfulness-based intervention on Fear of Cancer Recurrence and quality of life among Chinese breast cancer survivors. Complement Ther Clin Pract 2022 Nov;49:101686. [doi: 10.1016/j.ctcp.2022.101686] [Medline: 36347151]
47. Yousefi A, Arani AM, Ghadiany M, et al. The effectiveness of online mindfulness-based cancer recovery program on psychological Variables of colorectal and stomach cancer patients: A randomized control trial. Med Sci 2022;26(120):1. [doi: 10.54905/disssi/v26i120/ms55e2021]
48. Oberoi S, Yang J, Woodgate RL, et al. Association of mindfulness-based interventions with anxiety severity in adults with cancer: a systematic review and meta-analysis. JAMA Netw Open 2020 Aug 3;3(8):e2012598. [doi: 10.1001/jamanetworkopen.2020.12598] [Medline: 32766801]
49. Fung JYT, Lim H, Vongsirimas N, Klainin-Yobas P. Effectiveness of eHealth mindfulness-based interventions on cancer-related symptoms among cancer patients and survivors: A systematic review and meta-analysis. J Telemed Telecare 2024 Apr;30(3):451-465. [doi: 10.1177/1357633X221078490] [Medline: 35212553]
50. Creswell JD, Lindsay EK, Villalba DK, Chin B. Mindfulness training and physical health: mechanisms and outcomes. Psychosom Med 2019 Apr;81(3):224-232. [doi: 10.1097/PSY.0000000000000675] [Medline: 30806634]
51. Esch T, Stefano GB. The BERN framework of mind-body medicine: integrating self-care, health promotion, resilience, and applied neuroscience. Front Integr Neurosci 2022;16:913573. [doi: 10.3389/fnint.2022.913573] [Medline: 35910341]
52. Ong JC, Ulmer CS, Manber R. Improving sleep with mindfulness and acceptance: a metacognitive model of insomnia. Behav Res Ther 2012 Nov;50(11):651-660. [doi: 10.1016/j.brat.2012.08.001] [Medline: 22975073]
53. Johns SA, Tarver WL, Secinti E, et al. Effects of mindfulness-based interventions on fatigue in cancer survivors: A systematic review and meta-analysis of randomized controlled trials. Crit Rev Oncol Hematol 2021 Apr;160:103290. [doi: 10.1016/j.critrevonc.2021.103290] [Medline: 33675902]
54. Bower JE. Cancer-related fatigue--mechanisms, risk factors, and treatments. Nat Rev Clin Oncol 2014 Oct;11(10):597-609. [doi: 10.1038/nrclinonc.2014.127] [Medline: 25113839]
55. Bower JE, Greendale G, Crosswell AD, et al. Yoga reduces inflammatory signaling in fatigued breast cancer survivors: a randomized controlled trial. Psychoneuroendocrinology 2014 May;43:20-29. [doi: 10.1016/j.psyneuen.2014.01.019] [Medline: 24703167]
56. Mestdagh F, Steyaert A, Lavand'homme P. Cancer pain management: a narrative review of current concepts, strategies, and techniques. Curr Oncol 2023 Jul 18;30(7):6838-6858. [doi: 10.3390/curroncol30070500] [Medline: 37504360]
57. Hilton L, Hempel S, Ewing BA, et al. Mindfulness meditation for chronic pain: systematic review and meta-analysis. Ann Behav Med 2017 Apr;51(2):199-213. [doi: 10.1007/s12160-016-9844-2] [Medline: 27658913]
58. Trindade IA, Guiomar R, Carvalho SA, et al. Efficacy of online-based acceptance and commitment therapy for chronic pain: a systematic review and meta-analysis. J Pain 2021 Nov;22(11):1328-1342. [doi: 10.1016/j.jpain.2021.04.003] [Medline: 33892153]
59. Creswell JD. Mindfulness Interventions. Annu Rev Psychol 2017 Jan 3;68(1):491-516. [doi: 10.1146/annurev-psych-042716-051139] [Medline: 27687118]

Abbreviations

CBT: cognitive behavioral therapy
CINAHL: Cumulated Index in Nursing and Allied Health Literature
FACIT: functional assessment of chronic illness therapy
ISI: insomnia severity index
MD: mean difference
PICO: Population, Intervention, Comparison, Outcome
PROMIS: patient-reported outcomes measurement information system
PROSPERO: Prospective Register of Systematic Reviews
PSQI: Pittsburgh sleep quality index
QLQ: quality of life questionnaire
QoL: quality of life
RCT: randomized-controlled trial
SF: short form

SMD: standardized mean difference

Edited by TDA Cardoso; submitted 31.10.23; peer-reviewed by FM Carr, N Lau, T Pace; revised version received 12.12.24; accepted 13.12.24; published 16.01.25.

Please cite as:
Komariah M, Maulana S, Amirah S, Platini H, Rahayuwati L, Yusuf A, Firdaus MKZH
Benefits of Remote-Based Mindfulness on Physical Symptom Outcomes in Cancer Survivors: Systematic Review and Meta-Analysis
JMIR Cancer 2025;11:e54154
URL: https://cancer.jmir.org/2025/1/e54154
doi:10.2196/54154

Review

The Effect of Nutritional Mobile Apps on Populations With Cancer: Systematic Review

Krystal Lu Shin Ng[1], MSc; Murallitharan Munisamy[2], PhD; Joanne Bee Yin Lim[3], PhD; Mustafa Alshagga[1], PhD

[1]Division of Biomedical Sciences, Faculty of Science and Engineering, University of Nottingham Malaysia Campus, Semenyih, Malaysia

[2]National Cancer Society Malaysia, Kuala Lumpur, Malaysia

[3]School of Humanities, Faculty of Art and Social Sciences, University of Nottingham Malaysia Campus, Semenyih, Malaysia

Corresponding Author:
Krystal Lu Shin Ng, MSc
Division of Biomedical Sciences
Faculty of Science and Engineering
University of Nottingham Malaysia Campus
Jalan Broga
Semenyih, 43500
Malaysia
Phone: 60 172306490
Email: krystal_1224@hotmail.com

Abstract

Background: Limited access to nutrition support among populations with cancer is a major barrier to sustainable and quality cancer care. Increasing use of mobile health in health care has raised concerns about its validity and health impacts.

Objective: This systematic review aimed to determine the effectiveness of commercial or cancer-specific nutritional mobile apps among people living with cancer.

Methods: A systematic search of the CENTRAL, Embase, PubMed (MEDLINE), and Scopus databases was carried out in May 2024. All types of intervention studies were included, except observational studies, gray literature, and reference lists of key systematic reviews. Studies were eligible for inclusion if they involved (1) patients with or survivors of cancer and (2) nutrition-related mobile apps. Studies were excluded if the nutrition intervention was not delivered via mobile app or the app intervention was accompanied by dietary counseling. The review process was conducted based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The Risk of Bias 2 and Risk of Bias in Nonrandomized Studies tools were used to assess the study quality. The Cochrane Review Manager (version 5.4) software was used to synthesize the results of the bias assessment.

Results: A total of 13 interventions were included, comprising 783 adults or teenagers with cancer. Most studies focused on breast cancer (6/13, 46%), overweight (6/13, 46%), and survivors (9/13, 69%). Data on anthropometry and body composition (7/13, 54%; 387 participants), nutritional status (3/13, 23%; 249 participants), dietary intake (7/13, 54%; 352 participants), and quality of life (6/13, 46%; 384 participants) were gathered. Experimental groups were more likely to report significant improvements in body weight or composition, dietary compliance, nutritional status, and quality of life than control groups.

Conclusions: Although mobile app platforms are used to deliver nutrition interventions, the evidence for long-term efficacy, particularly in populations with cancer, remains elusive. More robust randomized controlled trials with larger sample sizes, as well as more homogeneous population characteristics and outcome measures, are warranted.

Trial Registration: PROSPERO CRD42023330575; https://tinyurl.com/55v56yaj

(JMIR Cancer 2025;11:e50662) doi:10.2196/50662

KEYWORDS

cancer; mobile app; nutrition; body composition; quality of life; mobile health; mHealth; diet; intervention; mobile phone; PRISMA

Introduction

Background

More than 50% of patients with cancer are likely to develop undernutrition upon diagnosis [1]. At least 5% of patients with cancer who are malnourished report drastic weight loss [2]. In total, 3 out of 5 patients report a significant weight reduction ranging from 1 to 10 kg 6 months after a cancer diagnosis [3]. Surprisingly, approximately 20% of patients with cancer die of undernutrition and its complications [4]. It is necessary to implement an early screening and detection of undernutrition based on the parameters of dietary intake, biochemical indexes, and body weight and composition. The overall nutritional status can be evaluated using cancer-specific assessment tools such as the Subjective Global Assessment, Patient-Generated Subjective Global Assessment (PG-SGA), and Mini Nutritional Assessment [5].

European Society of Parenteral and Enteral Nutrition guidelines have highlighted the importance of a multidisciplinary approach in managing undernutrition among patients with cancer [4]. However, this nutritional issue is not considered as equally important as the cancer disease itself [6,7]. If undernutrition is left untreated, this can result in poor immune response, increased treatment toxicities, impaired quality of life (QoL), increased risk of infection, increased admission rates and hospital stays, and increases in health care costs [4,7,8].

Overnutrition or excessive body fatness is another nutritional disorder that should be gaining greater attention in survivorship care [9,10]. Approximately 1 in 3 survivors of cancer report having obesity and not meeting the American Cancer Society's BMI guidelines of <30 kg/m^2 [10]. It is highly recommended that those living with or free of cancer eat a balanced diet to reduce the risk of recurrence and promote healthy survivorship [11].

To sustain a normal body weight, patients with cancer are advised to consume enough food to meet their daily requirement of energy and protein. In view of the differences in energy expenditure, the European Society of Parenteral and Enteral Nutrition recommends that the total energy requirement of patients with cancer be similar to that of survivors or healthy populations [4]. This elucidates that focusing on the basic principle of a balanced diet could be a nutrition guideline for patients with cancer, particularly those who are undernourished.

Studies have shown that approximately 90% of patients with cancer perceive nutrition support as an essential component in oncology care. However, less than half of patients with cancer are seen by dietitians [12]. According to the PG-SGA score, in a study by Pinho et al [1], 45% of patients with cancer required dietary intervention. In spite of that, dietetic support is not readily accessible to patients throughout their cancer journey. The high prevalence of undernutrition is commonly observed in people with upper digestive cancer, head and neck cancer, and lung cancer [1,2]. Still, in a study by Deftereos et al [7], approximately 40% of patients with upper digestive cancer did not receive any dietetic intervention before surgery.

Poor access to dietary services can be attributable to several factors, including lack of qualified dietitian staffing, lack of integration of nutrition services, lack of medical reimbursement for nutrition services, lack of awareness of cancer-related malnutrition, and inconsistent practice of nutritional risk screening in oncology [8]. Without a professional evidence-based dietary intervention, patients are likely to obtain information from the media or their peers. Conflicting information about nutrition makes them confused about what they should eat to optimize their well-being after a cancer diagnosis. Due to fear of cancer recurring, survivors can be desperate to modify their dietary habits [13,14]. This results in the adoption of unproven dietary strategies, including fad diets, juicing, and herbs and supplements, as well as avoiding certain food groups that are essential to their health [12].

The World Health Organization has called for a global initiative to leverage the use of digital health in areas of clinical medicine and public health [15]. During the COVID-19 pandemic, the application of digital technology targeting from planning and tracking, medical supplying, and screening for infection to clinical management was successful [16]. The pandemic has brought about an accelerated growth of digital health use to deliver continuous health care services while reducing virus transmission. For instance, telemedicine allowed for appointment scheduling and enhanced feasible health care delivery during the pandemic [17]. In addition, the use of digital health encourages engagement between practitioners and patients, as well as ensuring a sustainable health care system [18,19].

Objectives

To date, the implications of mobile app use in cancer screening, prevention, and management have been greatly highlighted [20,21]. However, there is a lack of empirical evidence that focuses on populations with cancer [22-24] and mobile app platforms [22], particularly for healthy eating and nutritional management. This systematic review aimed to determine the effectiveness of commercial or cancer-specific nutritional apps in improving nutrition-related health outcomes for people receiving treatment for or living with cancer.

Methods

Study Protocol and Guidance

The protocol for this review was registered with PROSPERO (registration number: CRD42023330575) [25]. This review was reported based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (Multimedia Appendix 1) [26].

Databases and Search Strategy

Systematic searches were conducted across 4 databases—CENTRAL, Embase, PubMed, and Scopus—in May 2024. The search strategy incorporated Medical Subject Headings (MeSH), keywords, and free-text searches that related to the 3 main concepts: mobile apps, cancer, and nutrition. The search string used in the literature search was as follows: "Mobile Applications"[Mesh] OR "mobile application*"[tw] OR "mobile apps"[tw] OR "mobile app"[tw] OR "mobile

technolog*"[tw] OR "mobile health"[tw] OR mHealth[tw] OR smartphone[tw] OR "smart phone"[tw] OR telemedicine[tw] AND "Neoplasms"[Mesh] OR cancer*[tw], neoplasm*[tw] OR oncology[tw] OR tumour*[tw] OR tumor*[tw] OR malignant[tw] OR malignanc*[tw] AND "Diet, Food, and Nutrition"[Mesh] OR nutrition[tw] OR diet[tw] OR eat[tw] OR food[tw] (Multimedia Appendix 2). It included original articles published between January 2013 and December 2023 and in the English language. This is a change from the registered protocol [25].

Study Selection

First, EndNote (version 20.3; Clarivate Analytics) was used to identify and remove duplicates from the list. The titles and abstracts of articles were screened independently by 2 reviewers (KLSN and MA) to identify potential eligible studies. The references retrieved from the search were categorized as excluded or included based on the population, intervention, comparator, outcome, and study design criteria [27]:

- Population—this included individuals who had a cancer diagnosis or a history of cancer.
- Intervention—the studies included commercial or cancer-specific mobile apps and nutrition-related key functions, including recording or monitoring food intake and providing feedback, recommendations, or coaching. Due to limited studies that included stand-alone use of mobile apps, studies on multicomponent interventions, such as targeting sleep, physical activity, or psychosocial care, were included.
- Outcome—the measures included changes in nutritional-related health outcomes. Due to a lack of feasibility studies, data on the evaluation of the quality of the mobile apps were not included. This is a change from the registered protocol [25].
- Study design—all types of intervention studies were considered, such as pretest-posttest studies, pilot studies, quasi-experimental studies, and randomized controlled trials (RCTs). Observational studies, gray literature, expert recommendations, or references in articles were not included.

A full-text screening was carried out by reviewing in detail the studies that were not excluded at the first screening based on the inclusion criteria. Each full text was retrieved and assessed independently by the same authors before inclusion in the review. Non–English-language articles were excluded. Any disagreements during the selection process were resolved through consensus.

Data Extraction and Synthesis

The data were extracted systematically from each article by KLSN and then checked by MA. The data included were authors, publication date (year), country, study design, sample size, participant characteristics, and details on the mobile app intervention. Next, data were extracted based on the type of population (survivors or patients receiving treatment), components of the app (eg, diet alone or diet plus physical activity), duration of the intervention and follow-up, and outcome measures (body weight, body composition, QoL, and dietary factors). A comparison of the descriptive findings was made across the studies. The outcome data between groups and before and after the intervention within groups were compared using mean differences and significance values (*P* value). The heterogeneity of the interventions and measures precluded a statistical combination of the quantitative findings; therefore, a meta-analysis was not conducted.

Risk-of-Bias Assessment

Analysis of the risk of bias was conducted using the Review Manager (version 5.4; The Cochrane Collaboration) software. The Risk of Bias 2 (RoB 2) and Risk of Bias in Nonrandomized Studies (ROBINS) tools were used for RCTs and non-RCTs, respectively. The risk of bias assessment was carried out by 2 reviewers independently (KLSN and MA). All discrepancies were resolved through consensus.

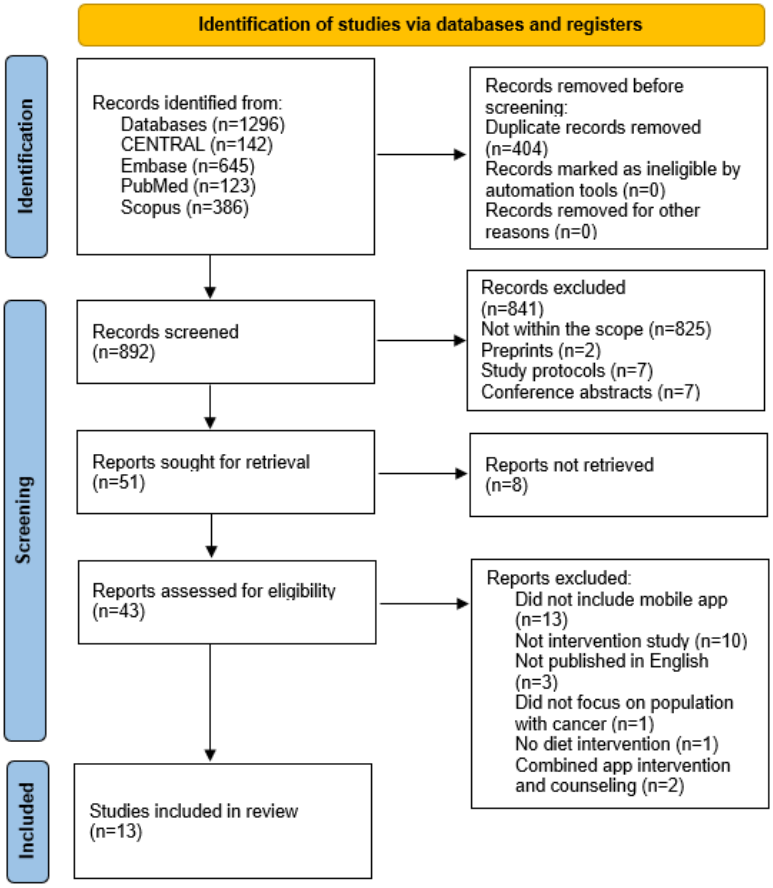
The seven areas included in the RoB 2 tool were (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personnel, (4) blinding of outcome assessment, (5) incomplete outcome data, (6) selective reporting, and (7) other bias. For the ROBINS tool, the seven areas included were (1) bias due to confounding, (2) bias in selection of participants for the study, (3) bias in classification of interventions, (4) bias due to deviations from the intended intervention, (5) bias due to missing data, (6) bias in measurement of outcomes, and (7) bias in selection of the reported results. According to the Cochrane Handbook for Systematic Reviews of Interventions, each area was assigned a classification of low, unclear, or high risk of bias [28].

Results

Overview

A total of 1296 articles were identified from all database searches. After 31.17% (404/1296) of duplicates were removed, the abstracts and titles of 68.83% (892/1296) of relevant articles were screened. The full texts of 5.7% (51/892) of these studies were retrieved and assessed for eligibility based on the population, intervention, comparator, outcome, and study design criteria. Finally, 13 articles were eligible to be included in this review. The procedure for article selection is shown in Figure 1 [26].

Figure 1. Study flowchart adapted from the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement.



Study Characteristics

Participants

Of the 13 included studies, there were 7 (54%) that were RCTs [29-35]; 4 (31%) that were single-arm, pretest-posttest studies [36-39]; and 2 (15%) that were quasi-experimental studies [40,41]. In total, 62% (8/13) of the included studies were conducted in the United States [29-31,35-38,40]; 15% (2/13) were conducted in South Korea [33,39]; and the remaining 23% (3/13) were conducted in Germany [41], Turkey [32], and Australia [34]. Most studies (9/13, 69%) were published within the past 5 years [29,30,32-35,38-40].

Among the 13 studies, a total of 783 participants with cancer aged 12 to 75 years was included. The sample sizes ranged from 22 to 127. In total, 15% (2/13) of the studies had no comparison groups [36,37]. A total of 8% (1/13) of the studies were conducted on teenagers [38], whereas the remaining 92% (12/13) of the studies were conducted on adults aged between 18 and 75 years. The most prevalent condition targeted in the studies was breast cancer (6/13, 46%) [29-32,37,40], followed by gastrointestinal cancer (2/13, 15%) [33,34], hematologic cancer (2/13, 15%) [35,38], mixed cancer (2/13, 15%) [36,41], and esophageal cancer (1/13, 8%) [39]. A total of 46% (6/13) of the studies were conducted among participants with overweight or obesity [29,35-38,40]. In total, 15% (2/13) of the studies were conducted among participants with body weight within the normal range [33,39], whereas 38% (5/13) of the studies did not state the weight status of the population [30-32,34,41]. Of the 13 studies, 9 (69%) recruited survivors [29-32,35-38,40], and the remaining 4 (31%) recruited patients with newly diagnosed cancer or receiving treatment [33,34,39,41]. Table 1 shows the summary of the study details, participant characteristics, and intervention types.

Table 1. Study characteristics.

Study and country	Study design	Sample size, N	Population characteristics	Cancer type	Intervention	Control	App features
Chow et al [35], 2020, United States	Pilot RCT[a]	Experimental group: 24; control group: 17	Adults diagnosed for ≥5 years; mean age 44 (range 20.9-54.0) years in the experimental group and 46.0 (range 20.2-54.8) years in the control group; mean BMI 28.6 (SD 6.5) kg/m^2 in the experimental group and 29.6 (SD 6.3) kg/m^2 in the control group	Hematologic	Had access to HealthWatch 360 (GB HealthWatch) and Fitbit Flex wristband (Google) with goal setting and peer support	Had access to Fitbit Flex wristband and HealthWatch 360 without goal setting and peer support	Commercial app; tracking of food intake
McCarroll et al [36], 2015, United States	Pretest-posttest study	50	Women with OW[b] or obesity diagnosed over the previous 3 years; mean age 58.4 (SD 10.3) years	Endometrium or breast cancer	Had access to LoseIt!	—[c]	Commercial app; logging of food, exercise, and BW[d] and provision of personalized feedback
Orlemann et al [41], 2018, Germany	Pilot; QED[e]	Experimental group: 12; control group: 12	Adults receiving treatment; mean age 58.4 (range 27-90) years	Mixed (GI[f] tumor; n=16)	Had access to OncoFood (Huawei Technologies Co Ltd)	Received nutrition counseling and therapy without app	Custom-developed app; recording of food intake and monitoring of nutritional goals and BW
Stubbins et al [37], 2018, United States	Prospective, single arm, and open label	33	Survivors with OW; mean age 57 (SD 9) years; mean BMI 32.7 (SD 5.7) kg/m^2	Breast cancer	Used MOCHA[g] for ≥5 days	—	Custom-developed app; access to sleep and mood data, provision of a list of cardiovascular and strength activities with amount of calories burned, logging of food, and monitoring of progress
Baik et al [30], 2020, United States	Pilot RCT	Experimental group: 39; control group: 39	Latina survivors; mean age 52.54 (SD 11.36) years	Breast cancer	Access to My Guide	Access to My Health	Custom-developed app; My Guide: focus on ways to cope with side effects of treatment, stress management, social support, and breast cancer–related knowledge; My Health: provides recommendations regarding nutrition, exercise, and prevention of chronic illnesses
Buscemi et al [31], 2019, United States	Pilot RCT	Experimental group: 40; control group: 40	Latina survivors; mean age 52.54 (SD 11.36) years	Breast cancer	Access to My Guide	Access to My Health	Custom-developed app; My Guide: focus on ways to cope with side effects of treatment, stress management, social support, and breast cancer–related knowledge; My Health: provides recommendations regarding nutrition, exercise, and prevention of chronic illnesses

Study and country	Study design	Sample size, N	Population characteristics	Cancer type	Intervention	Control	App features
Cairo et al [40], 2020, United States	Non-RCT	Experimental group: 66; control group: 61	Female survivors; mean age 51.4 (SD 8.1) years in the experimental group and 56.7 (SD 9.8) years in the control group; mean BMI 29.4 (SD 6.0) kg/m^2 in the experimental group and 30.2 (SD 7.3) kg/m^2 in the control group	Breast cancer	Access to Vida	Received self-guided nutrition "toolkit," exercise stretch band, pedometer, and self-guided walking DVD	Commercial app; tracking of medication, diet, exercise, sleep, and BW and pairing with a certified coach
Fuemmeler et al [38], 2020, United States	Single-group pretest-posttest design	15	Teenage survivors; mean age 14.8 (SD 1.97) years; mean BMI 22.6 (SD 4.1) kg/m^2 in the experimental group and 22.7 (SD 2.7) kg/m^2 in the control group (data from post hoc analysis)	Acute lymphoblastic leukemia or lymphoma	Used Mila Blooms for ≥4 weeks	Used Mila Blooms for <4 weeks	Custom-developed app; monitors progress, allows for autonomic feedback, and uses game mechanics to promote healthy eating and PA[h]
Allicock et al [29], 2021, United States	Pilot RCT	Experimental group: 13; control group: 9	African American women after treatment (except Herceptin therapy and endocrine pills); mean age 52.8 (SD 9.57) years in the experimental group and 51.44 (SD 9.18) years in the control group; mean BMI 33.26 (SD 5.42) kg/m^2 in the experimental group and 38.35 (SD 7.08) kg/m^2 in the control group	Breast cancer	Access to CHAT[i] and ActiGraph wGT3X-BT accelerometer plus tailored health messages	Access to CHAT and ActiGraph wGT3X-BT accelerometer	Custom-developed app; provision of suggestions about PA and healthy diet
Çınar et al [32], 2021, Turkey	Single-blinded, single-center, randomized pretest-posttest design	Experimental group: 31; control group: 33	Women receiving hormonal therapy; mean age 45.7 (SD 9.0) years	Breast cancer	Received routine care plus mobile app–based training	Received routine care	The nature of the app was not mentioned; provision of information about breast cancer, symptom diary, balanced diet, regular PA, and stress management
Keum et al [33], 2021, South Korea	Prospective, single-center, nonblinded RCT	Experimental group: 20; control group: 20	Patients scheduled for chemotherapy; median age 62 (range 45-70) years in the experimental group and 61 (range 34-78) years in the control group; mean BMI 21.91 (SD 1.57) kg/m^2 in the experimental group and 23.5 (SD 2.72) kg/m^2 in the control group	Pancreatic cancer	Access to Noom mobile app (Noom Inc)	Did not have access to the Noom app and received none of the nutrition intervention	Commercial app; logging of food, step count, and BW; provided coaching and allowed for messaging for tracking caloric intake and muscle gain

Study and country	Study design	Sample size, N	Population characteristics	Cancer type	Intervention	Control	App features
Yang et al [39], 2021, South Korea	Prospective, single-arm pilot study	Experimental group: 38; control group: 60	Men scheduled for neoadjuvant chemoradiotherapy; median age 59.2 (SD 6.5) years in the experimental group and 58.5 (SD 7.8) years in the control group; mean BMI 21.8 (SD 2.6) kg/m^2 in the experimental group and 22 (SD 6) kg/m^2 in the control group	Esophageal cancer	Access to Noom mobile app	Previous cohort: received usual care	Commercial app; recording, monitoring, and provision of recommendations about diet, exercise, and BW changes
Huggins et al [34], 2022, Australia	3-arm RCT	Mobile app group: 36; telephone group: 38; control group: 37	Adults newly diagnosed with cancer; mean age 63.2 (SD 9.9) years in the control group, 67.5 (SD 10.3) years in the telephone group, and 66.6 (SD 9.7) years in the mobile app group; mean BW 75.0 (SD 20.0) kg in the control group, 71.9 (SD 12.7) kg in the telephone group, and 76.4 (SD 14.7) kg in the mobile app group	Upper GI cancer	Mobile app group: received symptom-directed nutrition intervention via the internet-enabled mobile app “myPace”; telephone group: received symptom-directed nutrition intervention via telephone	Received usual care	Commercial app; self-monitoring of goal attainment and BW

[a]RCT: randomized controlled trial.

[b]OW: overweight.

[c]Not applicable.

[d]BW: body weight.

[e]QED: quasi-experimental design.

[f]GI: gastrointestinal.

[g]MOCHA: Methodist Hospital Cancer Health Application.

[h]PA: physical activity.

[i]CHAT: Creating Healthy Actions Through Technology.

Mobile Apps

Types

Most studies (10/13, 77%) included a multicomponent intervention that combined diet with physical activity, psychosocial support, sleep, or behavior modification. Specifically, 50% (5/10) of these studies involved a combination of diet and physical activity [29,35,36,38,39], with additional components in the other 50% (5/10) of the studies [30-32,37,40]. The remaining 23% (3/13) of the studies included a dietary intervention as a single component [33,34,41].

Duration

The duration of the interventions ranged from 4 weeks to 6 months. A total of 62% (8/13) of the studies lasted up to 8 weeks [29-31,36-39,41], with 75% (6/8) of these studies including anthropometry or body composition as outcome measures. A total of 23% (3/13) of the studies lasted between 12 and 16 weeks [32,33,35], with one of the studies mainly evaluating QoL. The remaining 15% (2/13) of the studies lasted up to 6 months [34,40] and included both anthropometry and QoL measures. A total of 38% (5/13) of the studies continued to evaluate the participants’ progress after the intervention by investigating changes in QoL or dietary intake [29-31,34,35].

Features

A total of 46% (6/13) of the studies included the common features of logging, tracking, or monitoring in the mobile apps [33-35,37,38,41]. In total, 31% (4/13) of the studies focused on the provision of dietary information [29-32], whereas the remaining 23% (3/13) of the studies allowed for logging and provision of guidance or coaching [36,39,40]. Table 1 provides a more detailed description.

Retention Rate

Of the 13 studies, 9 (69%) reported the percentage of participants who remained in the study over the intervention or follow-up periods. A total of 44% (4/9) of these studies reported a retention rate of >90% [29,31,35,40], 44% (4/9) reported retention rates of 70% to 90% [33,36,37,39], and 11% (1/9) reported a retention rate of <70% [34]. Table 2 provides a more detailed description.

Table 2. Key findings of the included studies.

Study	App intervention duration	Follow-up	Retention rate (%)	Outcome measures	Main findings
Diagnosis					
Huggins et al [34], 2022	18 weeks	30 weeks	49.5	• QALYs[a] (EQ-5D-5L tool) • QoL[b] (EORTC QLQ-C30[c] scale) • Nutritional status (PG-SGA[d]–Short Form) • Self-reported BW[e]	• Mean weight 75.6 (SD 20.3) kg at 3 months, 75.6 (SD 17.5) kg at 6 months, and 73.2 (SD 18.4) kg at 12 months in the control group; 71.7 (SD 11.8) kg at 3 months, 70.2 (SD 11.7) kg at 6 months, and 68.6 (SD 13.3) kg at 12 months in the telephone group; and 71.7 (SD 15.6) kg at 3 months, 68.7 (SD 14.1) kg at 6 months, and 68.5 (SD 14.1) kg at 12 months in the mobile app group; *P*=.08 for control group vs telephone group; *P*=.03 for mobile app group vs telephone group; *P*=.48 for mobile app group vs control group • Mean QoL score 54.3 (SD 25.1) at 3 months, 69.8 (SD 12.2) at 6 months, and 72.2 (SD 15.9) at 12 months in the control group; 66.4 (SD 19.7) at 3 months, 68.0 (SD 28.13) at 6 months, and 74.8 (SD 23.8) at 12 months in the telephone group; and 62.3 (SD 24.5) at 3 months, 59.25 (SD 21.1) at 6 months, and 73.5 (SD 20.5) at 12 months in the mobile app group; *P*=.22 for control group vs telephone group; *P*=.08 for mobile app group vs telephone group; *P*=.85 for mobile app group vs control group • Mean QALY score 0.55 (SD 0.28) at 12 months in the control group, 0.57 (SD 0.28) at 12 months in the telephone group, and 0.59 (SD 0.23) at 12 months in the mobile app group; *P*>.99 for control group vs telephone group; *P*=.71 for mobile app group vs telephone group; *P*=.14 for mobile app group vs control group • Mean PG-SGA score 7.5 (SD 5.0) at 3 months, 4.6 (SD 3.6) at 6 months, and 4.1 (SD 4.1) at 12 months in the control group; 7.8 (SD 5.7) at 3 months, 6.2 (SD 5.1) at 6 months, and 4.3 (SD 4.7) at 12 months in the telephone group; and 8.4 (SD 6.1) at 3 months, 7.2 (SD 4.0) at 6 months, and 4.9 (SD 3.6) at 12 months in the mobile app group; *P*=.35 for control group vs telephone group; *P*=.58 for mobile app group vs telephone group; *P*=.19 for mobile app group vs control group
Treatment					
Orlemann et al [41], 2018	4 weeks	—[f]	NR[g]	• BW, BMI, SMM[h], and FFM[i] (BIA[j]) • Nutritional goals (intake of protein, fibers, energy, carbohydrates, and fats)	• Mean change in BW 1.03 kg in the experimental group and –1.46 kg in the control group (*P*=.045) • Mean change in SMM 0.58 kg in the experimental group and –0.61 kg in the control group (*P*=.009); mean change in FFM after the intervention (*P*=.03) • *P*=.91 for difference in mean changes in the intake of protein and fats, *P*=.34 for difference in mean changes in the intake of fiber, *P*=.27 for difference in mean changes in the intake of carbohydrates, and *P*=.42 for difference in mean changes in the intake of energy in the control group after the intervention; mean values NR

Study	App intervention duration	Follow-up	Retention rate (%)	Outcome measures	Main findings
Çınar et al [32], 2021	12 weeks	—	NR	• QoL (FACT-ES[k]) • Symptom distress (NCCN[l] Distress Thermometer)	• QoL: t_{30}=–5.13 and *P*<.001 in the experimental group and t_{32}=3.25 and *P*=.003 in the control group; physical well-being: t_{30}=–4.60 and *P*<.001 in the experimental group and t_{32}=1.13 and *P*=.27 in the control group; emotional well-being: t_{30}=–2.58 and *P*=.02 in the experimental group and t_{32}=2.88 and *P*=.007 in the control group; functional well-being: t_{30}=–1.01 and *P*=.32 in the experimental group and t_{32}=2.67 and *P*=.01 in the control group; endocrine symptoms: t_{30}=–6.49 and *P*<.001 in the experimental group and t_{32}=3.08 and *P*=.004 in the control group; pretest distress score: 1003 (*P*=.32); posttest distress score: –2265 (*P*=.03)
Keum et al [33], 2021	12 weeks	—	82.5	• QoL (EORTC QLQ-C30) • Nutritional status (PG-SGA) • SMI[m] (CT[n]) • Total protein and energy intake	• Reduced SMI: –3.27 in the experimental group and –13.96 in the control group (*P*=.11) • Improved GHS[o] and QoL in experimental group compared to control group (*P*=.004) • Mean protein intake after the intervention: 1.3 g per kg per day in the experimental group and 1 g per kg per day in the control group (*P*=.02); mean energy intake after the intervention: 25.2 kcal per kg per day in the experimental group and 17.7 kcal per kg per day in the control group (*P*=.04) • Improved PG-SGA score in both groups (*P*<.001)
Yang et al [39], 2021	8 weeks	—	78.9	• SMI (CT) • NLR[p], PLR[q], and PNI[r]	• Mean change in SMI after the intervention –7.4% (SD 6.5%) in the experimental group and –8.1% (SD 5.3%) in the control group (*P*=.57) • PNI: mean change –9.8 (SD 6) in the experimental group and –6.7 (SD 7.5) in the control group (*P*=.04); NLR: mean change 0.4 (SD 3.9) in the experimental group and 0.6 (SD 5.1) in the control group (*P*=.82); PLR: mean change 84.1 (SD 157.6) in the experimental group and 62.4 (SD 173.4) in the control group (*P*=.55)
Survivorship					

Study	App intervention duration	Follow-up	Retention rate (%)	Outcome measures	Main findings
Chow et al [35], 2020	16 weeks	8 weeks	90.2	• PA[s] • Daily percentage of added sugar, saturated fat, and sodium (HEI[t]-2015) • Physical health and mental health (PROMIS[u] Global Health–10) • Health-related self-efficacy score	• Physical health: mean 2.7 (95% CI 0.7-4.6) in the experimental group and 1.8 (95% CI –0.3 to 3.8) in the control group (between-group *P*=.52); mental health: mean 4.2 (95% CI 1.5-6.9) in the experimental group and 1.8 (95% CI –1.1 to 4.8) in the control group (between-group *P*=.24) • HEI-2015 score: mean 1.6 (95% CI –1.5 to 4.6) in the experimental group and 0.6 (95% CI –2.8 to 4.0) in the control group (between-group *P*=.67); daily percentage of added sugar: mean –0.8 (95% CI –2.2 to 0.5) in the experimental group and 0.1 (95% CI –1.5 to 1.6) in the control group (between-group *P*=.39); daily percentage of saturated fat: mean –0.3 (95% CI –1.5 to 0.9) in the experimental group and –0.8 (95% CI –2.2 to 0.6) in the control group (between-group *P*=.60); sodium intake: mean –832 (95% CI –1421 to –243) mg per day in the experimental group and –279 (95% CI –937 to 379) mg per day in the control group (between-group *P*=.22)
McCarroll et al [36], 2015	4 weeks	—	70	• BW, BMI, and WC[v] • QoL and self-efficacy (FACT-G[w] and Weight Efficacy Lifestyle Questionnaire) • Minutes spent in PA • Weekly intake of carbohydrates, fats, protein, fiber, and calories	• Mean pretest BW 97.3 (SD 22.5) kg and mean posttest BW 95.0 (SD 22.1) kg (*P*<.001); mean pretest BMI 36.4 (SD 8.1) kg/m^2 and mean posttest BMI 35.6 (SD 8.0) kg/m^2 (*P*<.001); mean pretest WC 106.6 (SD 16.8) cm and mean posttest WC 103.4 (SD 17.4; *P*<.001) cm • Mean pretest FACT-G score 50.47 (SD 13.3) and mean posttest FACT-G score 44.35 (SD 19.9; *P*=.15) • Carbohydrates: mean pretest intake 120.6 (SD 69.3) g and mean posttest intake 124.0 (SD 120.3) g (*P*=.73); fat: mean pretest intake 44.1 (SD 23.4) g and mean posttest intake 58.2 (SD 60.0) g (*P*=.18); protein: mean pretest intake 55.2 (SD 26.6) g and mean posttest intake 65.4 (SD 62.3) g (*P*=.23); fiber: mean pretest intake 11.0 (SD 6.3) g and mean posttest intake 13.3 (SD 13.6) g (*P*=.28); calories: mean pretest intake 1022.6 (SD 494.4) kcal and mean posttest intake 1281.1 (SD 1130.6) kcal (*P*=.26)
Stubbins et al [37], 2018	4 weeks	—	75.8	• Adherence to the MOCHA[x] app • System Usability Scale score • Weight loss • Dietitian-participant interaction	• Mean reduced BW 2 (range +4 to –10.6) lbs after the intervention; *P* value NR
Baik et al [30], 2020	6 weeks	2 weeks	NR	• QoL (FACT-B[y]) • Symptom burden (25-item Breast Cancer Prevention Trial questionnaire) • Cancer-specific distress (15-item Impact of Event Scale) • Cancer-relevant self-efficacy (CASE-Cancer[z]) • Breast cancer knowledge (16-item Knowledge About Breast Cancer questionnaire)	• Experimental group—breast cancer well-being score for low app users: mean pretest score 23.47 (range 12-36) and mean posttest score 26.13 (range 14-35); control group—social well-being, score: mean pretest score 20.74 (range 5-28) and mean posttest score 22.52 (range 11-28); *P* value NR

Study	App intervention duration	Follow-up	Retention rate (%)	Outcome measures	Main findings
Buscemi et al [31], 2019	6 weeks	2 weeks	>90	• Daily intake of fat and FV[aa] • PA level	• Fat sources: EMM[ab] 2.38 (SE 0.21) in the experimental group and 2.86 (SE 0.21) in the control group at baseline, 2.42 (SE 0.22) in the experimental group and 2.38 (SE 0.21) in the control group at 6 weeks, and 2.36 (SE 0.22) in the experimental group and 2.20 (SE 0.22) in the control group at 8 weeks (P=.03)
Cairo et al [40], 2020	6 months	—	100	• BW and BMI • PA level • Adherence to a healthy diet (27-item "Rate Your Plate" survey) • Presence and severity of fatigue (VAS[ac]-Fatigue) • Depression and anhedonia (PHQ[ad] tool)	• Mean reduced BW 1.8 (SD 4.9) kg in the experimental group (P<.01) and –0.2 (SD 3.7) kg in the control group (P=.70); mean reduced BMI 0.7 (SD 1.8) kg/m^2 in the experimental group (P<.01) and –0.7 (SD 1.4) kg/m^2 in the control group (P=.68) • Mean reduced fatigue score 1.2 (SD 2.4) in the experimental group (P<.001) and 0.65 (SD 2.3) in the control group (P=.03); P=.36 for depression between experimental and control groups • Improved adherence to a plant-based diet: mean change in score –6.2 (SD 5.8) in the experimental group (P<.001) and –2.0 (SD 6.5) in the control group (P=.02)
Fuemmeler et al [38], 2020	8 weeks	—	NR	• Height, BW, BMI, z score, and percentile • Intake of calories and nutrients • PA level • Diet and PA self-efficacy (PACE[ae]) • User satisfaction and narrative engagement	• Mean pretest BMI 22.6 (SD 4.1) kg/m^2 and mean posttest BMI 22.8 (SD 4.1) kg/m^2 in the experimental group (P=.41); mean pretest BMI 22.7 (SD 2.7) kg/m^2 and mean posttest BMI 23.1 (SD 2.6) kg/m^2 in the control group (P=.24) • Mean pretest sweet food intake 8.4% (SD 3.6%) of kcal and mean posttest sweet food intake 13.5% (SD 9%) of kcal in the experimental group (P=.12) and mean pretest sweet food intake 8.8% (SD 6.3%) of kcal and mean posttest sweet food intake 7.5% (SD 4.8%) of kcal in the control group (P=.35; between-group P=.049); mean pretest sugary beverage intake 206.5 (SD 202.1) g and mean posttest sugary beverage intake 156.6 (SD 145.0) g in the experimental group (P=.08) and mean pretest sugary beverage intake 336.8 (SD 367.7) g and mean posttest sugary beverage intake 370.4 (SD 410.9) g in the control group (P=.04; between-group P=.04); mean pretest FV self-efficacy score 4.2 (SD 0.8) and mean posttest FV self-efficacy score 4.3 (SD 0.6) in the experimental group (P=.35) and mean pretest FV self-efficacy score 4.0 (SD 0.8) and mean posttest FV self-efficacy score 4.0 (SD 0.6) in the control group (P=.24; between-group P=.80)
Allicock et al [29], 2021	4 weeks	4 weeks	100	• BMI and WC • FV intake and percentage of energy from fat and fiber • PA level	• Mean change in BMI –0.19 (SD 0.35) kg/m^2 in the experimental group (P=.10) and –0.24 (SD 0.76) kg/m^2 in the control group (P=.76); mean WC change –1.04 (SD 0.95) cm in the experimental group (P=.003) and –0.47 (SD 1.57) cm in the control group (P=.39) • Mean FV change 0.67 (SD 2.35) servings in the experimental group (P=.34) and 0.78 (SD 2.48) servings in the control group (P=.38); mean fast food intake change –1.5 (SD 1.98) servings in the experimental group (P=.008) and –1.11 (SD 1.45) servings in the control group (P=.09)

[a]QALY: quality-adjusted life year.

[b]QoL: quality of life.
[c]EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire.
[d]PG-SGA: Patient-Generated Subjective Global Assessment.
[e]BW: body weight.
[f]Not applicable.
[g]NR: not reported.
[h]SMM: skeletal muscle mass.
[i]FFM: fat-free mass.
[j]BIA: bioimpedance analysis.
[k]FACT-ES: Functional Assessment of Cancer Therapy–Endocrine Symptoms.
[l]NCCN: National Comprehensive Cancer Network.
[m]SMI: skeletal muscle index.
[n]CT: computed tomography.
[o]GHS: global health status.
[p]NLR: neutrophil-to-lymphocyte ratio.
[q]PLR: platelet-to-lymphocyte ratio.
[r]PNI: prognostic nutritional index.
[s]PA: physical activity.
[t]HEI: Healthy Eating Index.
[u]PROMIS: Patient-Reported Outcomes Measurement Information System.
[v]WC: waist circumference.
[w]FACT-G: Functional Assessment of Cancer Therapy–General.
[x]MOCHA: Methodist Hospital Cancer Health Application.
[y]FACT-B: Functional Assessment of Cancer Therapy–Breast.
[z]CASE-Cancer: Communication and Attitudinal Self-Efficacy Scale for Cancer.
[aa]FV: fruits and vegetables.
[ab]EMM: estimated marginal mean.
[ac]VAS: Visual Analog Scale.
[ad]PHQ: Patient Health Questionnaire.
[ae]PACE: Patient-Centered Assessment and Counseling for Exercise.

Risk-of-Bias Assessment

The assessment of risk of bias was conducted for each study. The RoB 2 assessment is shown in Figure 2 [29-35]. In total, 8% (1/13) of the studies had a low risk of bias in all aspects [35]. A total of 15% (2/13) of the studies were reported as double blind [34,35]. Due to uncertainty or unblinded treatment allocation, the quality of 38% (5/13) of the trials was considered low with regard to performance and detection bias [29-33]. There was an unclear risk of selection bias in these 5 trials due to limited information about allocation concealment [29-33] and generation of a randomized sequence [30-32]. Huggins et al [34] reported a low retention rate (<50%), with the use of the multiple imputation approach for handling missing data. The suboutcomes resulting from a web-based intervention were not reported in a breast cancer study investigating the effect of mobile app–based training on QoL [32]. One study did not report the *P* value for the difference in breast cancer well-being after the intervention [30].

The ROBINS assessment is shown in Figure 3 [36-41]. A total of 17% (1/6) of the studies mentioned the frequency of mobile app use (at least 5 days) during the intervention [37], whereas the remaining 83% (5/6) of the studies did not report the intervention status. In total, 67% (4/6) of the studies reported a low retention rate or uncertainty about missing data management [36,37,39,41]. A total of 50% (3/6) of the studies had a high risk of bias in the selection of study participants, which could affect the quality of the intervention and outcomes [36,38,39]. In total, 33% (2/6) of the studies did not provide information on whether there was a deviation from the intended intervention [37,41]. Fuemmeler et al [38] failed to show the changes in weight and height measurements after the intervention. A total of 17% (1/6) of the studies did not provide information about *P* values of weight loss data [37]. In total, 33% (2/6) of the studies reported no information on whether any confounding factors were present [38,41]. A total of 17% (1/6) of the studies had a high risk of bias in outcome measurements that resulted from inappropriate methods of delivering the intervention [37] and measuring outcomes [42].

Figure 2. Risk-of-bias assessment of randomized controlled trials (n=7) using the Risk of Bias 2 tool, with a quality rating of low risk (–), high risk (+), or unclear risk (?).

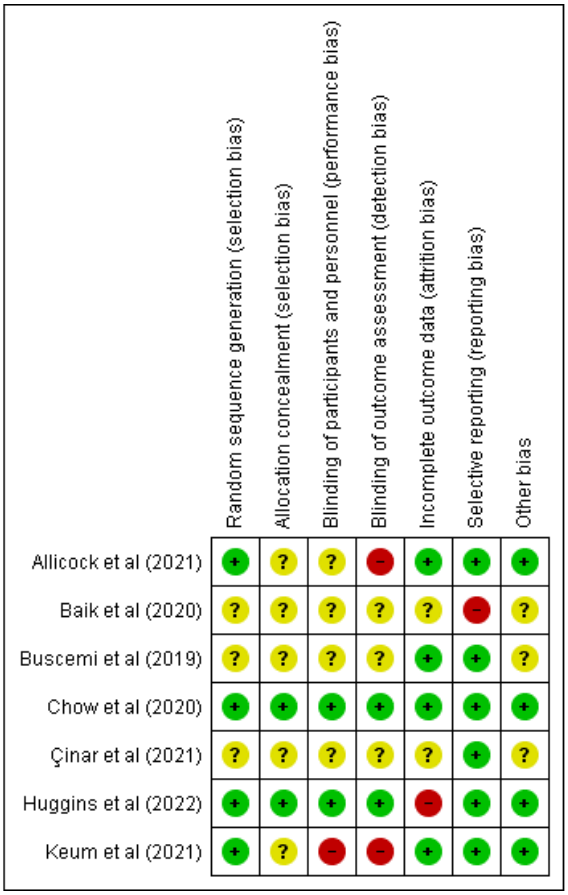
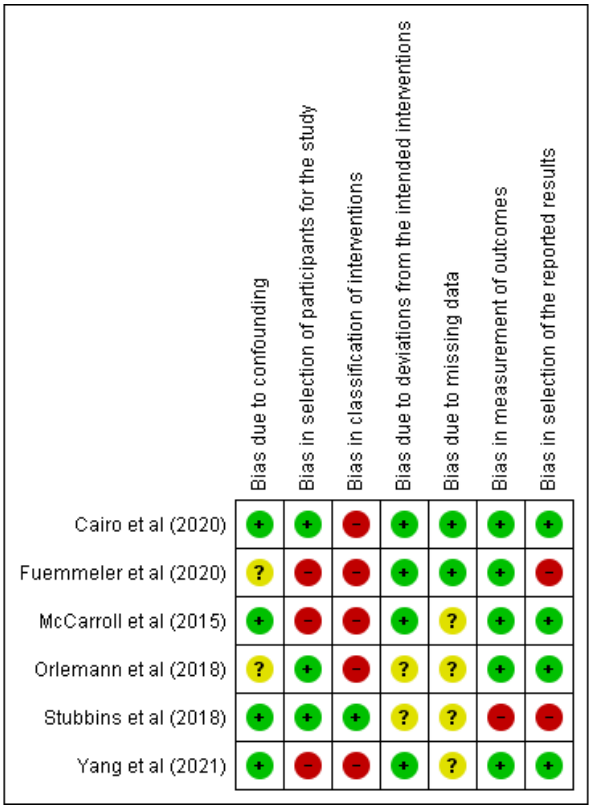


Figure 3. Risk-of-bias assessment using the Risk of Bias in Nonrandomized Studies tool in 6 studies, with a quality rating of low risk (–), high risk (+), or unclear risk (?).



Outcome Measures

A summary of outcome measures and study findings can be found in Table 2.

Anthropometry and Body Composition

Of the 13 studies, 7 (54%) analyzed anthropometry measures, including body weight, BMI [29,34,36-38,40,41], and waist circumference [29,36]. Of these 7 studies, 5 (71%) intended to support weight reduction [29,36-38,40], and 2 (29%) supported weight retention [34,41]. Of the 5 studies supporting weight reduction, 2 (40%) reported significant improvement in weight after the intervention [36,40]. On the other hand, only 50% (1/2) of the studies that supported weight retention reported significant weight gain in patients with cancer who were at risk of malnutrition [41]. Huggins et al [34] reported attenuation of weight loss in patients with upper gastrointestinal cancer who received a symptom-directed nutrition intervention via telephone compared to a mobile app. A total of 29% (2/7) of the studies did not find significant changes in BMI between groups [29,38]. A study showed a decrease in weight among survivors of breast cancer with overweight; however, neither the *P* value nor the significance of the change was stated [37]. A total of 29% (2/7) of the studies reported a significant reduction in waist circumference after the intervention [29,36].

In total, 23% (3/13) of studies aiming to combat cancer-induced malnutrition assessed body composition, namely, skeletal muscle mass, fat-free mass, fat mass, and bone mineral density [33,39,41]. Significant increases in skeletal muscle mass and fat-free mass were reported in app users based on the results of bioimpedance analysis [41]. However, the studies by Keum et al [33] and Yang et al [39] did not show significant results of the skeletal muscle index using computed tomography.

Nutritional Status or Index

Nutritional status was evaluated in 23% (3/13) of studies aiming at weight gain [33,34,39]. According to the Scored PG-SGA, Keum et al [33] reported significant improvements in nutritional status in both the experimental and control groups but with no statistically significant difference between groups. Similarly, a nonsignificant difference in PG-SGA scores in the intervention groups (delivered via telephone or mobile app) compared with the control group was reported by Huggins et al [34]. The PG-SGA score is derived from 7 domains, namely, weight, food intake, nutrition impact symptoms, functional capabilities, presence of catabolic condition, metabolic demand, and physical examination. The scores range from 0 to 53, with higher scores indicating poorer nutritional status [43].

Another study measured the prognostic nutritional index (PNI), neutrophil-to-lymphocyte ratio, and platelet-to-lymphocyte ratio for nutritional status assessment. Only the PNI showed a significant reduction in the experimental group compared to the control group [39]. These 3 indexes were derived from laboratory parameters (PNI: 10 × albumin + 0.005 × absolute lymphocyte count; neutrophil-to-lymphocyte ratio: absolute neutrophil count/absolute lymphocyte count; platelet-to-lymphocyte ratio: platelet/absolute lymphocyte count). Higher readings indicate higher level of inflammation or severity of malnutrition.

Dietary Factors

A total of 62% (8/13) of the studies examined the effect of nutritional mobile apps on dietary outcomes in cancer [29,31,33,35,36,38,40,41]. The common outcome measures were daily nutrient intakes [29,31,33,36,38,41] and level of adherence to dietary recommendations [35,40]. App users reported reduced consumption of high-fat food, including fast food, after the intervention [29,31]. A higher consumption of sugary beverages was observed in non–app users compared to app users, but no significant results were reported for the intake of fruits and vegetables [38]. Keum et al [33] reported higher intake of protein and energy in app users, whereas 33% (2/6) of the studies that measured daily nutrient intake did not report any significant findings [36,41].

The level of adherence to a healthy diet was analyzed in 25% (2/8) of these studies. On the basis of a Rate Your Plate survey, app users reported a significantly improved adherence to a plant-based diet [40]. However, no significant results were reported using the Healthy Eating Index score [35].

QoL and Symptom Burden

The impact of nutritional mobile apps on QoL was evaluated in 46% (6/13) of the studies. In total, 33% (2/6) of these studies measured QoL using the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) [33,34]. A total of 50% (3/6) of the studies used the Functional Assessment of Cancer Therapy–General (FACT-G) [36], the Functional Assessment of Cancer Therapy–Breast (FACT-B) [30], and Functional Assessment of Cancer Therapy–Endocrine Symptoms (FACT-ES) [32]. In total, 17% (1/6) of the studies used the Patient-Reported Outcomes Measurement Information System Global Health–10 to assess QoL [35]. In total, 33% (2/6) of the studies reported significant improvements in QoL based on the EORTC QLQ-C30 [33] and FACT-ES [32] tools. Higher scores were reported for overall perception of QoL and physical, emotional, and functional well-being, whereas lower scores were reported for endocrine symptoms and psychosocial distress. The remaining studies did not report any significant QoL results [34-36,40].

The EORTC QLQ-C30 covers 5 functional domains (physical, emotional, social, role, and cognitive), 9 symptoms (fatigue, nausea and vomiting, pain, dyspnea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties), and a general health perception. The total score ranges from 0 to 100, with higher scores indicating greater symptoms or better functional status [44]. The FACT-G comprises 27 items and 5 Likert rating scales (0-4), similar to the FACT-B and FACT-ES. The FACT-G measures the domains of physical, social, emotional, and functional well-being, whereas the FACT-B and FACT-ES have 11 additional breast cancer–related items and 9 additional endocrine-related items, respectively. The total score of these QoL tools can be >100, with higher scores indicating greater symptoms or better functional status [45].

The Visual Analog Scale–Fatigue and 2-item Patient Health Questionnaire were used in the study by Cairo et al [40]. Although the experimental group reported improved levels of fatigue and depression after the intervention, these changes were

not statistically significant. The Visual Analog Scale–Fatigue comprises 18 items answered on a Likert scale from 0 to 10, with higher scores indicating higher levels of pain or fatigue. On the other hand, the 2-item Patient Health Questionnaire comprises 2 items pertaining to anhedonia and depression ranging from 0 to 6, with higher scores indicating more depressive symptoms [46].

No significant results were found using quality-adjusted life years (QALYs) in the study by Huggins et al [34].

Discussion

Principal Findings

Nutritional mobile apps for populations with cancer have the potential to improve body weight or composition, nutritional status, dietary adherence, and QoL across the continuum of cancer care. The apps offered the basic functions of recording and tracking users' food intake and weight in general. It was unclear whether custom-developed mobile apps were efficacious for nutrition-related health outcomes and QoL in cancer care. However, incorporating commercial mobile apps seemed to be beneficial for improving nutritional care in populations with cancer. This could be due to the implementation of self-monitoring of their progress, a necessary step in delivering quality nutrition care [5]. This review observed that the beneficial effect of stand-alone interventions was comparable with that of multicomponent interventions. However, the results may not be able to discern the magnitude of the difference due to limited data. Overall, the studies included in this review were of low to moderate quality. For RCTs, lack of blinding and biased treatment allocation were the major concerns. Failure to define the intervention status in terms of types, frequency, and timing reported by non-RCTs made the evaluation of nutritional mobile apps challenging.

Comparison to Prior Work

Among the interventions that aimed to support weight loss, almost half (2/5, 40%) reported successful weight control among patients with cancer and overweight. In public health research, the common measures are BMI, waist circumference, waist-to-hip ratio, and body fat percentage [47]. Waist-to-height ratio has also been known to be a good surrogate in predicting the risk of noncommunicable diseases [48,49]. Implementing effective dietary strategies for successful weight loss is highly recommended to reduce the risk of cancer recurrence in long-term survivorship [4]. A review underpinned the beneficial effects of eHealth interventions on weight management in survivors of cancer, with a greater impact if combined with dietary counseling [50].

Our review found reduced intake of fast food [29] and sugary food [38] in app users. When aiming at weight loss, adhering to healthy eating guidelines should be the goal to sustain good health and well-being. A bariatric study highlighted the need to change eating behaviors for sustainable weight management [51]. Self-monitoring weight changes and dietary behavior is a common feature in app-based weight loss programs. The use of mobile app interventions for improved eating behavior and diet quality seems to be promising [52]. There are multiple factors influencing eating habits among school-aged children, particularly role modeling and parenting styles [53]. In addition to app gamification, creating a conducive learning environment in schools and at home could be a way to promote healthy eating habits among children.

The primary concern regarding undernutrition is the lack of energy that the body needs to undergo cancer treatment, which could result in treatment toxicities, longer hospital stays, or reduced QoL [4]. Among interventions that aim to support weight gain in patients with cancer who were malnourished, delivering nutrition support via mobile app platforms may help prevent drastic weight loss and improve skeletal muscle mass and overall nutritional status. However, the findings of this review do not reflect the long-term beneficial effects due to lack of data. Despite the growing development of nutrition apps, tailoring dietary interventions to individuals' needs, nutritional status, cancer type, treatment plan, and comorbid conditions is still an unmet need [54].

Of the 6 studies that focused on QoL, only 2 (33%) showed significant changes in QoL at the treatment phase based on the EORTC QLQ-C30 and FACT-ES [32,33]. A review that focused on app-based interventions to improve nutrition or lifestyle behaviors in patients with breast cancer showed a similar finding during chemotherapy [55]. This could be due to enhanced user engagement by improving self-motivation, health information, social support, and goal setting [56]. The 2 cancer-specific tools used in our review were the EORTC QLQ-C30 [44] and FACT-G [45], which allow for a multidimensional assessments of QoL. These tools yield a comprehensive evaluation of individuals' progress. QALYs, which account for both QoL and survival, have been increasingly used as a standard measure to evaluate disease burden at the population or regional level [57,58]. However, Huggins et al [34] reported no significant results for QALYs in groups that received the intervention via mobile app or telephone compared to controls. Failure to obtain significant results could be due to less participants who continued to use the mobile app after the intervention period.

Strengths and Limitations

This is the first review that has evaluated the impact of app-based dietary interventions in cancer care. The review was based on a systematic search strategy that focused on nutrition interventions delivered via mobile app platforms and on populations with cancer. However, this review has certain limitations. First, only English-language articles were included in the search for this review. Second, the heterogeneity of study designs, interventions, app features, and cancer types was substantial, requiring the results to be interpreted cautiously. Third, the inconsistent measurement and reporting of incomplete data made comparisons difficult across the studies. Finally, this review included pilot studies that comprised small sample sizes (11/13, 85% of the interventions enrolled <70 participants per group), resulting in limited generalizability of the study findings.

Conclusions

Mobile app–based nutrition interventions have a favorable effect on nutritional status and QoL in patients with cancer. In addition, mobile apps that incorporate nutrition interventions could also

be beneficial for survivors after cancer treatment. However, it was unclear whether custom-developed apps were efficacious for improved nutrition-related outcomes and QoL. The continuity of nutritional care in patients with cancer via mobile app platforms could help in achieving a healthy weight by improving their adherence to dietary guidelines. Although most studies yielded favorable outcomes, they were rated as being of low to moderate quality.

Future studies should emphasize randomized controlled designs, larger sample sizes, diet-only mobile apps, greater homogeneity of outcome measures and population characteristics, and high participant engagement and retention within the study.

Acknowledgments

This review would not have been possible without the support of the University of Nottingham for database subscriptions. No funding was granted for this review.

Authors' Contributions

KLSN and MM contributed to conceptualization and methodology. KLSN and MA contributed to data curation and formal analysis. KLSN contributed to project administration, resources, software, and writing (original draft). KLSN, JBYL, and MA contributed to review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist [26].
[DOCX File , 31 KB - cancer_v11i1e50662_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 507 KB - cancer_v11i1e50662_app2.docx]

References

1. de Pinho NB, Martucci RB, Rodrigues VD, D'Almeida CA, Thuler LC, Saunders C, et al. Malnutrition associated with nutrition impact symptoms and localization of the disease: results of a multicentric research on oncological nutrition. Clin Nutr 2019 Jun;38(3):1274-1279. [doi: 10.1016/j.clnu.2018.05.010] [Medline: 29853223]
2. Marshall KM, Loeliger J, Nolte L, Kelaart A, Kiss NK. Prevalence of malnutrition and impact on clinical outcomes in cancer services: a comparison of two time points. Clin Nutr 2019 Apr;38(2):644-651. [doi: 10.1016/j.clnu.2018.04.007] [Medline: 29789167]
3. Muscaritoli M, Lucia S, Farcomeni A, Lorusso V, Saracino V, Barone C, et al. Prevalence of malnutrition in patients at first medical oncology visit: the PreMiO study. Oncotarget 2017 Aug 10;8(45):79884-79896 [FREE Full text] [doi: 10.18632/oncotarget.20168] [Medline: 29108370]
4. Muscaritoli M, Arends J, Bachmann P, Baracos V, Barthelemy N, Bertz H, et al. ESPEN practical guideline: clinical nutrition in cancer. Clin Nutr 2021 May;40(5):2898-2913. [doi: 10.1016/j.clnu.2021.02.005] [Medline: 33946039]
5. Cederholm T, Barazzoni R, Austin P, Ballmer P, Biolo G, Bischoff SC, et al. ESPEN guidelines on definitions and terminology of clinical nutrition. Clin Nutr 2017 Feb;36(1):49-64 [FREE Full text] [doi: 10.1016/j.clnu.2016.09.004] [Medline: 27642056]
6. Arensberg MB, Besecker B, Weldishofer L, Drawert S. Commentary: quality nutrition care is integral to the oncology care model. Support Care Cancer 2021 Dec;29(12):7139-7142 [FREE Full text] [doi: 10.1007/s00520-021-06436-x] [Medline: 34297220]
7. Deftereos I, Yeung JM, Arslan J, Carter VM, Isenring E, Kiss N. Preoperative nutrition intervention in patients undergoing resection for upper gastrointestinal cancer: results from the multi-centre NOURISH point prevalence study. Nutrients 2021 Sep 15;13(9):3205 [FREE Full text] [doi: 10.3390/nu13093205] [Medline: 34579082]
8. Trujillo EB, Dixon SW, Claghorn K, Levin RM, Mills JB, Spees CK. Closing the gap in nutrition care at outpatient cancer centers: ongoing initiatives of the oncology nutrition dietetic practice group. J Acad Nutr Diet 2018 Apr;118(4):749-760. [doi: 10.1016/j.jand.2018.02.010] [Medline: 29576094]
9. Anderson AS, Martin RM, Renehan AG, Cade J, Copson ER, Cross AJ, et al. Cancer survivorship, excess body fatness and weight-loss intervention-where are we in 2020? Br J Cancer 2021 Mar;124(6):1057-1065 [FREE Full text] [doi: 10.1038/s41416-020-01155-2] [Medline: 33235316]

10. Gregory K, Zhao L, Felder TM, Clay-Gilmour A, Eberth JM, Murphy EA, et al. Prevalence of health behaviors among cancer survivors in the United States. J Cancer Surviv 2024 Jun;18(3):1042-1050 [FREE Full text] [doi: 10.1007/s11764-023-01347-8] [Medline: 36933085]
11. Rock CL, Thomson CA, Sullivan KR, Howe CL, Kushi LH, Caan BJ, et al. American Cancer Society nutrition and physical activity guideline for cancer survivors. CA Cancer J Clin 2022 May;72(3):230-262 [FREE Full text] [doi: 10.3322/caac.21719] [Medline: 35294043]
12. Sullivan ES, Rice N, Kingston E, Kelly A, Reynolds JV, Feighan J, et al. A national survey of oncology survivors examining nutrition attitudes, problems and behaviours, and access to dietetic care throughout the cancer journey. Clin Nutr ESPEN 2021 Feb;41:331-339 [FREE Full text] [doi: 10.1016/j.clnesp.2020.10.023] [Medline: 33487286]
13. Mardani A, Farahani MA, Khachian A, Vaismoradi M. Fear of cancer recurrence and coping strategies among prostate cancer survivors: a qualitative study. Curr Oncol 2023 Jul 16;30(7):6720-6733 [FREE Full text] [doi: 10.3390/curroncol30070493] [Medline: 37504353]
14. Park J, Kim J, Shin DW, Shin J, Cho B, Song YM. Factors associated with dietary habit changes in korean stomach cancer survivors after cancer treatment. Nutrients 2023 Jul 24;15(14):3268 [FREE Full text] [doi: 10.3390/nu15143268] [Medline: 37513686]
15. Global strategy on digital health 2020-2025. World Health Organization. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2025-01-12]
16. Whitelaw S, Mamas MA, Topol E, Van Spall HG. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug;2(8):e435-e440 [FREE Full text] [doi: 10.1016/S2589-7500(20)30142-4] [Medline: 32835201]
17. Taha AR, Shehadeh M, Alshehhi A, Altamimi T, Housser E, Simsekler MC, et al. The integration of mHealth technologies in telemedicine during the COVID-19 era: a cross-sectional study. PLoS One 2022 Feb 24;17(2):e0264436 [FREE Full text] [doi: 10.1371/journal.pone.0264436] [Medline: 35202424]
18. Dorsey ER. The new platforms of health care. NPJ Digit Med 2021 Jul 15;4(1):112 [FREE Full text] [doi: 10.1038/s41746-021-00478-5] [Medline: 34267312]
19. Hopstaken JS, Verweij L, van Laarhoven CJ, Blijlevens NM, Stommel MW, Hermens RP. Effect of digital care platforms on quality of care for oncological patients and barriers and facilitators for their implementation: systematic review. J Med Internet Res 2021 Sep 24;23(9):e28869 [FREE Full text] [doi: 10.2196/28869] [Medline: 34559057]
20. Lee YH, Huang LH, Chen SH, Shao JH, Lai CH, Yang NP. Effects of mobile application program (app)-assisted health education on preventive behaviors and cancer literacy among women with cervical intraepithelial neoplasia. Int J Environ Res Public Health 2021 Nov 04;18(21):11603 [FREE Full text] [doi: 10.3390/ijerph182111603] [Medline: 34770117]
21. Jamnadass E, Rai BP, Veneziano D, Tokas T, Rivas JG, Cacciamani G, et al. Do prostate cancer-related mobile phone apps have a role in contemporary prostate cancer management? A systematic review by EAU young academic urologists (YAU) urotechnology group. World J Urol 2020 Oct;38(10):2411-2431 [FREE Full text] [doi: 10.1007/s00345-020-03197-w] [Medline: 32322996]
22. Cruz C, Prado CM, Punja S, Tandon P. Use of digital technologies in the nutritional management of catabolism-prone chronic diseases: a rapid review. Clin Nutr ESPEN 2021 Dec;46:152-166. [doi: 10.1016/j.clnesp.2021.10.020] [Medline: 34857190]
23. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
24. Zhang L, Misir A, Boshuizen H, Ocké M. A systematic review and meta-analysis of validation studies performed on dietary record apps. Adv Nutr 2021 Dec 01;12(6):2321-2332 [FREE Full text] [doi: 10.1093/advances/nmab058] [Medline: 34019624]
25. Ng K, Al-Shagga M, Lim JB, Munisamy M. Feasibility and effect of nutritional mobile applications in improving cancer outcomes: a systematic review. JMIR Preprints Preprint posted online on July 8, 2023 [FREE Full text] [doi: 10.2196/preprints.50662]
26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
27. Sanfilippo F, Tigano S, Palumbo GJ, Astuto M, Murabito P. Importance of inclusion criteria in systematic reviews. Br J Anaesth 2020 Nov;125(5):e398-e399 [FREE Full text] [doi: 10.1016/j.bja.2020.07.010] [Medline: 32739043]
28. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions Version 6.3. London, UK: Cochrane Collaboration; 2022.
29. Allicock M, Kendzor D, Sedory A, Gabriel KP, Swartz MD, Thomas P, et al. A pilot and feasibility mobile health intervention to support healthy behaviors in African American breast cancer survivors. J Racial Ethn Health Disparities 2021 Feb;8(1):157-165. [doi: 10.1007/s40615-020-00767-x] [Medline: 32385847]

30. Baik SH, Oswald LB, Buscemi J, Buitrago D, Iacobelli F, Perez-Tamayo A, et al. Patterns of use of smartphone-based interventions among Latina breast cancer survivors: secondary analysis of a pilot randomized controlled trial. JMIR Cancer 2020 Dec 08;6(2):e17538 [FREE Full text] [doi: 10.2196/17538] [Medline: 33289669]
31. Buscemi J, Oswald LB, Baik SH, Buitrago D, Iacobelli F, Phillips SM, et al. My health smartphone intervention decreases daily fat sources among Latina breast cancer survivors. J Behav Med 2020 Oct;43(5):732-742 [FREE Full text] [doi: 10.1007/s10865-020-00136-3] [Medline: 31970652]
32. Çınar D, Karadakovan A, Erdoğan AP. Effect of mobile phone app-based training on the quality of life for women with breast cancer. Eur J Oncol Nurs 2021 Jun;52:101960. [doi: 10.1016/j.ejon.2021.101960] [Medline: 33882446]
33. Keum J, Chung MJ, Kim Y, Ko H, Sung MJ, Jo JH, et al. Usefulness of smartphone apps for improving nutritional status of pancreatic cancer patients: randomized controlled trial. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e21088 [FREE Full text] [doi: 10.2196/21088] [Medline: 34463630]
34. Huggins CE, Hanna L, Furness K, Silvers MA, Savva J, Frawley H, et al. Effect of early and intensive telephone or electronic nutrition counselling delivered to people with upper gastrointestinal cancer on quality of life: a three-arm randomised controlled trial. Nutrients 2022 Aug 07;14(15):3234 [FREE Full text] [doi: 10.3390/nu14153234] [Medline: 35956410]
35. Chow EJ, Doody DR, Di C, Armenian SH, Baker KS, Bricker JB, et al. Feasibility of a behavioral intervention using mobile health applications to reduce cardiovascular risk factors in cancer survivors: a pilot randomized controlled trial. J Cancer Surviv 2021 Aug;15(4):554-563 [FREE Full text] [doi: 10.1007/s11764-020-00949-w] [Medline: 33037989]
36. McCarroll ML, Armbruster S, Pohle-Krauza RJ, Lyzen AM, Min S, Nash DW, et al. Feasibility of a lifestyle intervention for overweight/obese endometrial and breast cancer survivors using an interactive mobile application. Gynecol Oncol 2015 Jun;137(3):508-515. [doi: 10.1016/j.ygyno.2014.12.025] [Medline: 25681782]
37. Stubbins R, He T, Yu X, Puppala M, Ezeana CF, Chen S, et al. A behavior-modification, clinical-grade mobile application to improve breast cancer survivors' accountability and health outcomes. JCO Clin Cancer Inform 2018 Dec;2:1-11 [FREE Full text] [doi: 10.1200/CCI.18.00054] [Medline: 30652617]
38. Fuemmeler BF, Holzwarth E, Sheng Y, Do EK, Miller CA, Blatt J, et al. Mila Blooms: a mobile phone application and behavioral intervention for promoting physical activity and a healthy diet among adolescent survivors of childhood cancer. Games Health J 2020 Aug;9(4):279-289 [FREE Full text] [doi: 10.1089/g4h.2019.0060] [Medline: 32391734]
39. Yang K, Oh D, Noh JM, Yoon HG, Sun JM, Kim HK, et al. Feasibility of an interactive health coaching mobile app to prevent malnutrition and muscle loss in esophageal cancer patients receiving neoadjuvant concurrent chemoradiotherapy: prospective pilot study. J Med Internet Res 2021 Aug 27;23(8):e28695 [FREE Full text] [doi: 10.2196/28695] [Medline: 34448714]
40. Cairo J, Williams L, Bray L, Goetzke K, Perez AC. Evaluation of a mobile health intervention to improve wellness outcomes for breast cancer survivors. J Patient Cent Res Rev 2020 Oct 23;7(4):313-322 [FREE Full text] [doi: 10.17294/2330-0698.1733] [Medline: 33163551]
41. Orlemann T, Reljic D, Zenker B, Meyer J, Eskofier B, Thiemt J, et al. A novel mobile phone app (OncoFood) to record and optimize the dietary behavior of oncologic patients: pilot study. JMIR Cancer 2018 Nov 20;4(2):e10703 [FREE Full text] [doi: 10.2196/10703] [Medline: 30459139]
42. May CN, Ho AS, Yang Q, McCallum M, Iyengar NM, Comander A, et al. Comparing outcomes of a digital commercial weight loss program in adult cancer survivors and matched controls with overweight or obesity: retrospective analysis. Nutrients 2021 Aug 24;13(9):2908 [FREE Full text] [doi: 10.3390/nu13092908] [Medline: 34578787]
43. da Silva SH, de Oliveira LC, E Silva Lopes MS, Wiegert EV, Motta RS, Ferreira Peres WA. The patient generated-subjective global assessment (PG-SGA) and ECOG performance status are associated with mortality in patients hospitalized with breast cancer. Clin Nutr ESPEN 2023 Feb;53:87-92. [doi: 10.1016/j.clnesp.2022.11.019] [Medline: 36657935]
44. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]
45. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. J Clin Oncol 1993 Mar;11(3):570-579. [doi: 10.1200/JCO.1993.11.3.570] [Medline: 8445433]
46. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007 Nov;22(11):1596-1602 [FREE Full text] [doi: 10.1007/s11606-007-0333-y] [Medline: 17874169]
47. Andreacchi AT, Griffith LE, Guindon GE, Mayhew A, Bassim C, Pigeyre M, et al. Body mass index, waist circumference, waist-to-hip ratio, and body fat in relation to health care use in the Canadian longitudinal study on aging. Int J Obes (Lond) 2021 Mar;45(3):666-676. [doi: 10.1038/s41366-020-00731-z] [Medline: 33432110]
48. Choi JR, Koh SB, Choi E. Waist-to-height ratio index for predicting incidences of hypertension: the ARIRANG study. BMC Public Health 2018 Jun 19;18(1):767 [FREE Full text] [doi: 10.1186/s12889-018-5662-8] [Medline: 29921256]
49. Oumer A, Ale A, Tariku Z, Hamza A, Abera L, Seifu A. Waist-to-hip circumference and waist-to-height ratio could strongly predict glycemic control than body mass index among adult patients with diabetes in Ethiopia: ROC analysis. PLoS One 2022 Nov 09;17(11):e0273786 [FREE Full text] [doi: 10.1371/journal.pone.0273786] [Medline: 36350840]

50. Kupila SK, Joki A, Suojanen LU, Pietiläinen KH. The effectiveness of eHealth interventions for weight loss and weight loss maintenance in adults with overweight or obesity: a systematic review of systematic reviews. Curr Obes Rep 2023 Sep 24;12(3):371-394 [FREE Full text] [doi: 10.1007/s13679-023-00515-2] [Medline: 37354334]
51. Allison KC, Wu J, Spitzer JC, McCuen-Wurst C, Ashare RL, Tewksbury C, et al. Changes in eating behaviors and their relation to weight change 6 and 12 months after bariatric surgery. Obes Surg 2023 Mar;33(3):733-742 [FREE Full text] [doi: 10.1007/s11695-022-06442-w] [Medline: 36690865]
52. Scarry A, Rice J, O'Connor EM, Tierney AC. Usage of mobile applications or mobile health technology to improve diet quality in adults. Nutrients 2022 Jun 12;14(12):2437 [FREE Full text] [doi: 10.3390/nu14122437] [Medline: 35745167]
53. Scaglioni S, De Cosmi V, Ciappolino V, Parazzini F, Brambilla P, Agostoni C. Factors influencing children's eating behaviours. Nutrients 2018 May 31;10(6):706 [FREE Full text] [doi: 10.3390/nu10060706] [Medline: 29857549]
54. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR Mhealth Uhealth 2021 Apr 01;9(6):e20037 [FREE Full text] [doi: 10.2196/20037] [Medline: 34254938]
55. Saevarsdottir SR, Gudmundsdottir SL. Mobile apps and quality of life in patients with breast cancer and survivors: systematic literature review. J Med Internet Res 2023 Jul 26;25:e42852 [FREE Full text] [doi: 10.2196/42852] [Medline: 37494111]
56. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
57. Abdin E, Chong SA, Vaingankar JA, Shafie S, Verma S, Luo N, et al. Impact of mental disorders and chronic physical conditions on quality-adjusted life years in Singapore. Sci Rep 2020 Feb 14;10(1):2695 [FREE Full text] [doi: 10.1038/s41598-020-59604-0] [Medline: 32060390]
58. Briggs AH, Goldstein DA, Kirwin E, Meacock R, Pandya A, Vanness DJ, et al. Estimating (quality-adjusted) life-year losses associated with deaths: with application to COVID-19. Health Econ 2021 Mar;30(3):699-707. [doi: 10.1002/hec.4208] [Medline: 33368853]

Abbreviations

EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire
FACT-B: Functional Assessment of Cancer Therapy–Breast
FACT-ES: Functional Assessment of Cancer Therapy–Endocrine Symptoms
FACT-G: Functional Assessment of Cancer Therapy–General
MeSH: Medical Subject Headings
PG-SGA: Patient-Generated Subjective Global Assessment
PNI: prognostic nutritional index
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY: quality-adjusted life year
QoL: quality of life
RCT: randomized controlled trial
RoB 2: Risk of Bias 2
ROBINS: Risk of Bias in Nonrandomized Studies

Edited by N Cahill; submitted 08.07.23; peer-reviewed by B Nezami, M Khazaee-Pool, LDCL Neri; comments to author 07.03.24; revised version received 07.07.24; accepted 30.10.24; published 05.02.25.

Please cite as:
Ng KLS, Munisamy M, Lim JBY, Alshagga M
The Effect of Nutritional Mobile Apps on Populations With Cancer: Systematic Review
JMIR Cancer 2025;11:e50662
URL: https://cancer.jmir.org/2025/1/e50662
doi:10.2196/50662
PMID:

Review

Investigating the Complexity of Multidimensional Symptom Experiences in Patients With Cancer: Systematic Review of the Network Analysis Approach

Vincent Richard[1,2], MD, PhD; Allison Gilbert[1], MD, PhD; Emanuela Pizzolla[1,3], PhD; Giovanni Briganti[1], MD, PhD

[1]Department of Computational Medicine and Neuropsychiatry, Faculty of Medicine, University of Mons, Mons, Belgium

[2]Department of Medical Oncology, CHU HELORA Hôpital de MONS, site Kennedy, Mons, Belgium

[3]Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy

Corresponding Author:
Giovanni Briganti, MD, PhD
Department of Computational Medicine and Neuropsychiatry
Faculty of Medicine
University of Mons
Avenue du Champ de Mars, 8
Mons, 7000
Belgium
Phone: 32 491254140
Email: giovanni.briganti@umons.ac.be

Abstract

Background: Advances in therapies have significantly improved the outcomes of patients with cancer. However, multidimensional symptoms negatively impact patients' quality of life. Traditional symptom analysis methods fail to capture the dynamic and interactive nature of these symptoms, limiting progress in supportive care. Network analysis (NA) is a promising method to evaluate complex medical situations.

Objective: We performed a systematic review to explore NA's contribution to understanding the complexity of symptom experiences in patients with cancer.

Methods: The research question was as follows: "In patients with cancer (population), what is the contribution of NA (intervention) to understanding the complexity of multidimensional symptom experiences (outcome)?" The keywords "network analysis" AND "symptoms" AND "cancer survivors" OR "cancer patients" were searched in MEDLINE, Embase, Google Scholar, and Scopus between 2010 and 2024. Citations were extracted using Covidence software. Two reviewers independently screened the articles and resolved inclusion disagreements through consensus. Data were synthetized, and results have been narratively described. Bias analysis was performed using the Methodological Index for Non-Randomized Studies tool.

Results: Among 764 articles initially identified, 22 were included. Studies evaluated mixed solid tumors (n=10), digestive tract cancers (n=4), breast cancer (n=3), head and neck cancer (n=2), gliomas (n=2), and mixed solid and hematological cancers (n=1). Twelve studies used general symptom assessment tools, whereas 10 focused on neuropsychological symptoms. Moreover, 1 study evaluated symptoms at diagnosis, 1 evaluated them during curative radiotherapy, 4 evaluated them during the perioperative period, 5 evaluated them during chemotherapy, 4 evaluated them during ongoing cancer therapies, and 7 evaluated them after acute treatments. Among these, 3 evaluated the longitudinal changes in symptom networks across chemotherapy cycles, and 1 evaluated changes during radiotherapy. Three studies investigated the associations between symptoms and biological parameters. Several NA approaches were used: network visualization (n=1), Bayesian network (n=1), pairwise Markov random field and IsingFit method (n=1), unregularized Gaussian graphical model (n=2), regularized partial correlation network (n=6), network visualization and community NA (n=1), network visualization and Walktrap algorithm (n=1), undirected network model with the Fruchterman-Reingold and edge-betweenness approaches (n=4), biased correlation and concise pattern diagram (n=1), extended Bayesian information criterion graphical LASSO method (n=3), cross-lagged panel network (n=1), and unspecified NA (n=3). Psychological symptoms, particularly anxiety, depression, and distress, were frequently identified as central and stably interconnected. Fatigue consistently emerged as a core symptom, closely linked to sleep disturbances, cognitive impairment, and emotional distress. Associations between symptoms and inflammatory biomarkers (eg, interleukin-6, C-reactive protein, and tumor necrosis factor-α) suggest a biological basis for symptom interconnectivity.

Conclusions: NA consistently identified core symptoms, particularly psychological symptoms and fatigue, and associations with inflammatory biomarkers. NA may deepen the understanding of symptom interconnectivity and guide more effective interventions. However, further longitudinal homogeneous studies using standardized methodologies are needed.

(JMIR Cancer 2025;11:e66087) doi:10.2196/66087

KEYWORDS

network analysis; symptoms; cancer patients; systematic review; cancer treatment; symptom management

Introduction

The global burden of cancer is continuously increasing, with Europe accounting for one-fifth of the total cancer cases and cancer deaths [1]. Over the past 2 decades, advances in multidisciplinary management and tailored drug therapies have significantly improved treatment outcomes, offering potential cures or long-term remission and leading to the concept of cancer survivorship [2,3]. However, despite these medical advancements, many patients with cancer continue to experience persistent and complex symptoms resulting from both the disease and its treatments, negatively affecting their quality of life (QoL) for years after diagnosis [4,5]. New treatment opportunities provided by cancer research are often paired with unpleasant side effects, such as those observed with recent advances in immunotherapy, highlighting the need for a deeper understanding of symptom interactions to improve symptom management strategies [6,7].

Traditional approaches to symptom analysis, such as the symptom cluster approach [8,9], have sought to identify groups of co-occurring symptoms that share common mechanisms and clinical outcomes [10-12]. However, the symptom cluster approach has faced criticism due to its reliance on statistical grouping techniques that do not fully capture the dynamic relationships between symptoms and clusters [11,13-15]. Specifically, it lacks the ability to assess direct interactions within or between symptom clusters and does not account for causal relationships between symptoms [13,14]. These limitations have prompted researchers to explore network analysis (NA) as a novel methodological framework for studying symptom complexity [16,17]. NA, originally developed in mathematics and graph theory, has gained traction in psychological and medical research for its ability to estimate complex patterns of relationships and to reveal core features of mental disorders [18,19]. This approach grants a new ontological view on mental diseases, conceiving them as complex systems of components, which are maintained by mutual relationships between them, without the need to identify causal latent variables [17,19,20].

This network-based approach differs fundamentally from traditional models by conceptualizing diseases as interconnected systems rather than relying on predefined diagnostic categories [18].

In cancer research, NA offers a powerful framework for understanding symptom interactions, identifying core symptoms, and refining symptom management strategies. This approach could enable clinicians to develop targeted interventions, prioritizing symptoms that have the highest impact on patients' QoL, which can ultimately enhance patient care [21]. In the study by Kossakowski et al [21], NA was used to analyze data related to health-related QoL in both a healthy population and patients with cancer, showing that maintaining daily routines and work activities could prevent symptom-related vicious cycles. Their findings emphasized the importance of psychosocial interventions in cancer treatment strategies [21].

Beyond symptom management, NA also holds promise for uncovering the underlying biological mechanisms driving symptom progression [22]. By integrating biological markers into symptom networks, this approach could provide new insights into pathophysiological pathways, offering opportunities for more biologically informed therapeutic strategies.

Kosvyra et al [22] explored the application of NA in the study of the biological data of patients with cancer, highlighting a significant gap in multiomics and predictive analyses, which limits the integration of biological mechanisms into symptom network research [22].

Despite promising findings, the application of NA in cancer symptom research remains fragmented, with existing studies often limited by sample heterogeneity, varied methodologies, and a lack of integration with biological data and therapeutic interventions. In this systematic review, we propose to investigate this complex and heterogeneous literature with a precise research question focusing on the contribution of NA in understanding the symptom experience of patients with cancer. The results will be detailed, and we will discuss methodological approaches used in existing studies, including differences in network construction and analysis; identify knowledge gaps; and propose future research directions.

By critically evaluating the existing literature, this review provides the first comprehensive assessment of the role of NA in understanding cancer symptomatology, emphasizing its potential to refine symptom management and enhance patient outcomes.

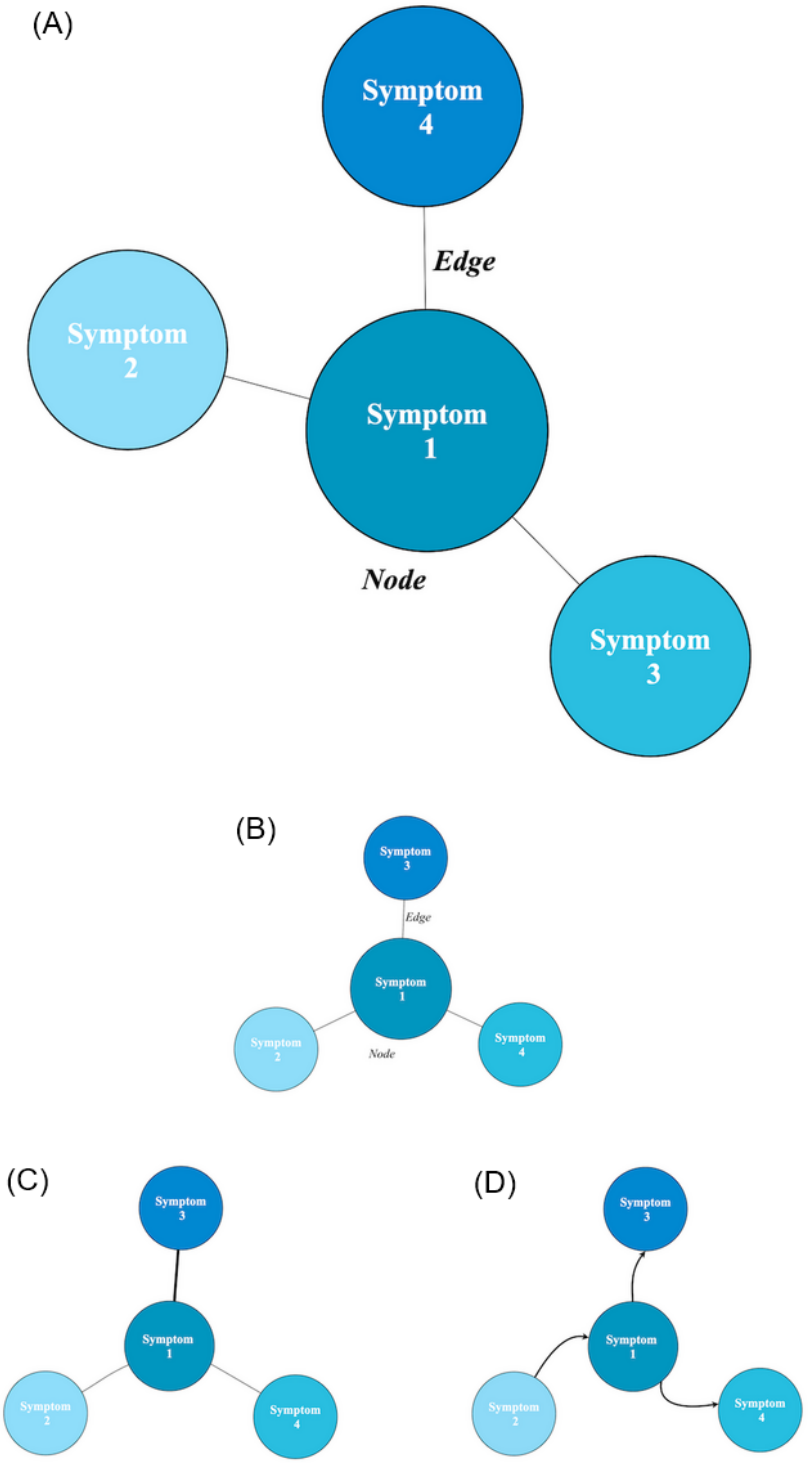
Methods

NA Approach

A network is a set of nodes (variables) and a set of edges (statistical relationships) connecting the nodes [19]. In the medicine field, nodes are symptoms, and a network is a graphic representation of the complex association observable between symptoms. Several types of networks have been developed: *directed networks* (cyclic or acyclic), in which the direction of the edges is determined; *undirected networks*, in which the direction of the edges is unknown [19]; *weighted networks*, in which the weight of the edges is represented by their thickness

and can represent a positive association or negative association; and *unweighted networks*, in which the edges either exist or not, and if they exist, they all have the same importance [18]. The classical structures of networks are represented in Figure 1.

Figure 1. (A) Representation of 3 nodes (symptoms) with their relationships (edges). Networks: (B) unweighted, undirected network; (C) weighted network; and (D) cyclic or acyclic directed network.



NA has to follow a precise methodology: collect the data of interest (from cross-sectional, longitudinal, or panel data studies), construct the network, describe it, and analyze its stability [18,19]. The choice of the NA method influences network structure, the relationships captured, and the assumptions imposed on the data [20].

Once constructed, the structures of the network have to be analyzed in terms of its properties: what is the importance of nodes, is the global structure dense, and are the nodes isolated? These properties are described in terms of centrality (degree, node strength, closeness, betweenness, and clustering) [18]. Finally, the network accuracy has to be evaluated [19].

Pairwise Markov random fields and directed acyclic graphs are the most used methods in the psychopathological sciences [18]. Pairwise Markov random fields (Ising model and Gaussian graphical model) involve undirected models used to represent conditional dependence or independence between pairs of variables and are constructed using local conditional probability distributions [18]. The presence of an edge between 2 nodes indicates that they are conditionally dependent, and the absence of an edge indicates that they are conditionally independent. However, they do not explain model causal relationships. In contrast, directed acyclic graphs represent causal relationships, mapping directed interactions between symptoms without relying on probability distributions [18]. A comparative summary of these models is provided in Table 1.

Table 1. Comparison between pairwise Markov random fields and directed acyclic graphs.

Variable	Pairwise Markov random field	Directed acyclic graph
Graph type	Undirected graph	Directed acyclic graph
Edge interpretation	Encodes conditional dependencies between variables	Represents causal relationships between variables
Edge direction	No direction (edges are bidirectional)	Directed edges (A → B means A influences B)
Conditional independence	An edge's presence or absence represents conditional dependence or independence	Uses d-separation to determine conditional independence
Causality	Does not assume causal relationships	Explicitly models cause-and-effect relationships
Loops/cycles	Can contain cycles	Acyclic (no feedback loops allowed)
Factorization of probability	Factorizes the joint distribution using local conditional distributions	Uses the chain rule to express joint probability based on parent-child relationships
Mathematical representation	Typically modeled using local Markov properties	Follows Bayes' theorem to express probabilities
Common models	Ising model, Gaussian graphical model, and mixed graphical model	Bayesian network and structural equation model (SEM)
Handling of latent variables	Typically does not incorporate latent variables directly	Can explicitly include latent variables
Parameter estimation	Uses maximum likelihood estimation (MLE) or regularization techniques (eg, LASSO)	Parameters estimated using MLE, Bayesian inference, or SEM methods

Research Question and Design

The research question was structured using the specialized PICO (population, intervention, comparator, outcome) framework. The final research question was as follows: "Considering patients with cancer (population), what is the contribution of the NA approach (intervention) to the understanding of the complexity of multidimensional symptom experiences (outcome)?" The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist was used to structure the report (Multimedia Appendix 1).

Search Strategy

We systematically searched the following databases: PubMed (MEDLINE), Embase, Scopus, and Google Scholar. The search strategy was developed by the authors using a combination of medical subject headings, EMTREE thesaurus terms, and free-text keywords informed by an initial scoping review of the literature. No librarian or information specialist was consulted.

The search combined the terms "network analysis," "symptoms," and ("cancer patients" or "cancer survivors") using Boolean operators. For multiword terms, quotation marks were used where appropriate (eg, "network analysis"). Filters were applied to limit results to studies on human adults published in English between 2010 and February 2024. A full description of the search strings and filters applied in each database is available in Multimedia Appendix 2.

References retrieved from the databases were imported into Covidence systematic review software, which automatically identified and removed duplicates. Additional references were identified through manual handsearching of Google Scholar.

Selection Criteria

To be included in the review, the articles had to evaluate symptoms or symptom clusters in adult patients with cancer via an NA approach, either at diagnosis or during acute cancer treatment, long-lasting adjuvant therapy, and follow-up alone. To maintain some disease homogeneity, studies focusing on hematological patients alone were excluded, although those with mixed patient populations, solid tumors, or hematological cancers were admissible. Given that this review focuses on symptoms, articles evaluating QoL, coping strategies, or symptom-targeted interventions alone were excluded. Reviews or meta-analyses were also excluded. Eligible articles had to be written in English. This systematic review was not registered.

Study Selection

The reference management software Covidence was used to export citations from database searches. Two reviewers (VR and AG) independently screened the titles and abstracts, and full-text screening was performed by both reviewers. Disagreements on inclusion were resolved through consensus.

Data Extraction

A predefined extraction form was developed for data extraction. The process was performed by one reviewer (VR) and verified by a second reviewer (AG). Data were synthesized regarding different parameters: design of the study, main purpose of the study, sample size, cancer type, time of symptom assessment,

tools used for symptom assessment and measures, NA methods, and main findings of the NA. The results are narratively described.

Bias Analysis

The methodological quality and risk of bias of the included studies were assessed using the Methodological Index for Non-Randomized Studies (MINORS) tool. This validated instrument was chosen as it is specifically designed to assess the methodological quality of nonrandomized surgical studies, whether comparative or noncomparative, and has been adapted for use in systematic reviews across various medical fields [23].

The MINORS tool evaluates studies across 12 items: 8 items for noncomparative studies and an additional 4 items for comparative studies. Each item is scored as 0 (not reported), 1 (reported but inadequate), or 2 (reported and adequate). For noncomparative studies, the global ideal score is 16, while for comparative studies, it is 24. Two reviewers (EP and GB) independently conducted the bias assessment, with disagreements resolved through discussion until consensus was reached.

The evaluation criteria included clearly stated aims, consecutive patient inclusion, prospective data collection, appropriate endpoints, unbiased outcome assessment, appropriate follow-up period, loss to follow-up analysis, and prospective calculation of study size. For comparative studies, additional criteria included adequate control group selection, contemporary groups, baseline equivalence, and adequate statistical analysis.

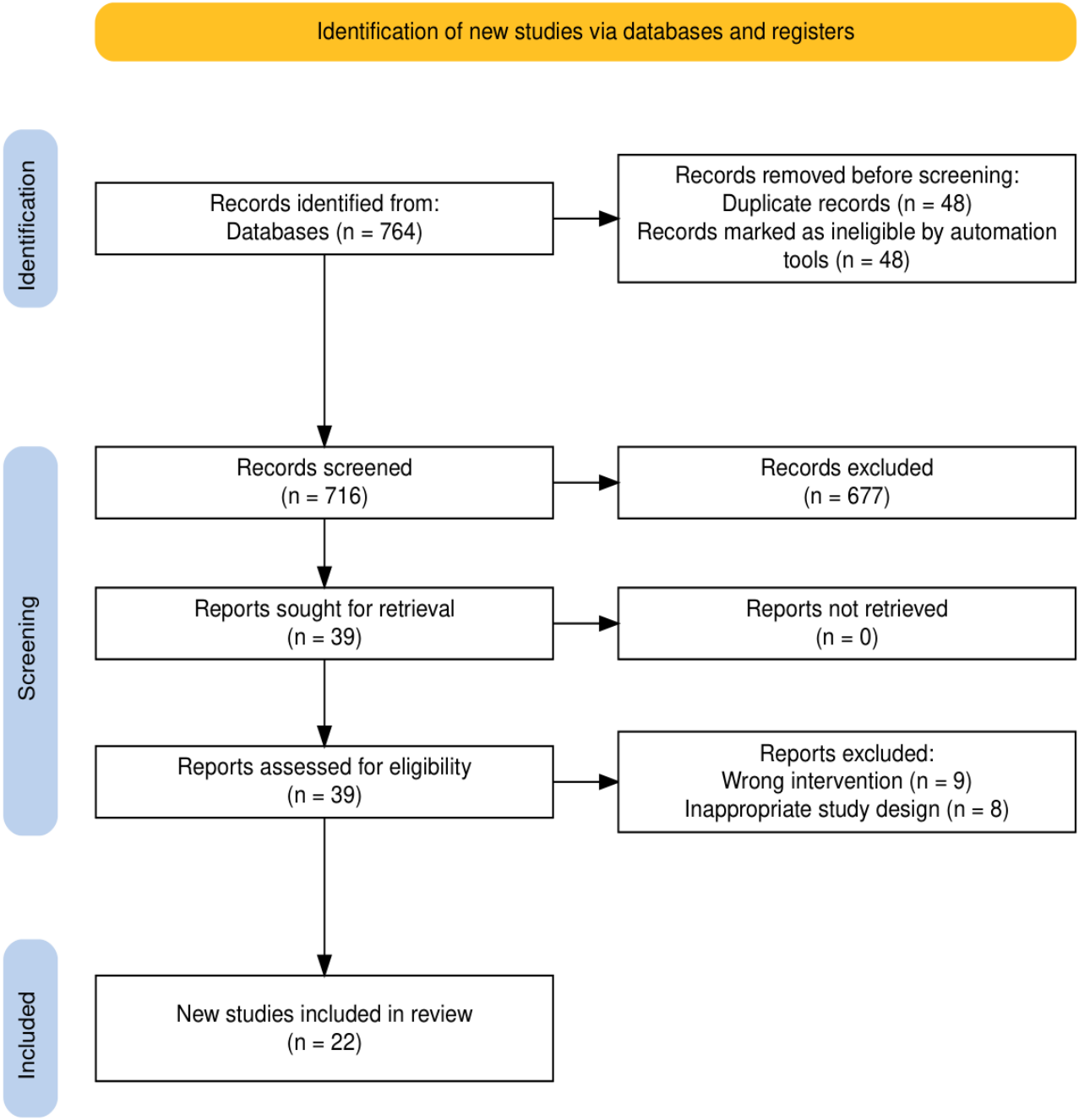
Results

Search Results

A total of 764 articles were initially identified through searches across 4 literature databases. After title and abstract screening, 677 articles were excluded. Of the 39 full-text articles assessed for eligibility, 17 were excluded (9 due to the use of the wrong intervention and 8 due to an inappropriate study design). Ultimately, 22 studies were included in this review, comprising a cumulative total of 20,393 participants.

The complete PRISMA flow diagram is presented in Figure 2. The diagram was generated using the PRISMA Flow Diagram Tool [24].

Figure 2. PRISMA 2020 flow diagram depicting the identification, screening, and inclusion process for studies in the systematic review.



Characteristics of the Included Studies

All characteristics of the selected studies are summarized in the subsections below and detailed in Table 2.

Table 2. Selected studies for the review.

Study and country	Study design	Main purposes	Sample description (including cancer type[a]) and assessment period	Symptom assessment: tools[b] and biomarkers[c]
Bhavnani et al [25], 2010; United States	Secondary analysis of a randomized trial (2007); Longitudinal	Explore symptom co-occurrence and overlap patterns using network analysis. Explore quantitative measures to analyze symptom co-occurrence and overlap (observed patterns).	• N=665 (463 female); Age: 21 years or older; Cancer type: solid tumors 94%, NHK 6%; Period: during CT[d]	Tools: MDASI (18 symptoms)
Xu et al [26], 2018; United States	Secondary analysis of a NIH[e]-funded study (2004-2010); Longitudinal	Investigate how depression, fatigue, and sleep interactions affect cognition and QoL[f] during CT.	• N=74 (74 female); Age: 51.8 (SD 9.5) years; Cancer type: BC; Period: pre-CT, post-CT, and 1 year after	Tools: PSQI (sleep quality), MFSI-SF (fatigue), CES-D (mood), FACT-B (quality of life), FOSQ (functional outcomes of sleepiness), and NP (cognition)
Papachristou et al [27], 2019; United States	Secondary analysis part of a longitudinal study	Evaluate the occurrence, severity, and distress of 38 cancer symptoms. Compare symptom networks based on occurrence, severity, and distress.	• N=1328 (1032 female); Age: 57.2 (SD 12.4) years; Cancer type: BC 40.2%, GI 30.7%, GYN 17.3%, LC 13.2%; Period: during CT	Tools: mMSAS (38 symptoms)
Hartung et al [28], 2019; Germany	Cross-sectional from 2 studies: cross-sectional prospective patients with cancer; survey control general population	Compare depressive symptom severity, frequency, and networks between patients with cancer and the general population.	• Study 1: N=4020 patients with cancer (2050 female); Age: 58 (SD 11) years • Study 2: N=4020 individuals from the general population (2050 female); Age: 55 (SD 15) years • Cancer type: BC 22.54%, PC 15.85%, CR 12.69%, and others 15.47%; Period: 14 months after diagnosis (mean)	Tools: PHQ-9 (depressive symptoms)
Schellekens et al [29], 2020; The Netherlands	Cross-sectional	Examine relationships among symptoms and psychosocial risk or protective factors.	• N=342 (264 female); Age: 51.35 (SD 10.62) years; Cancer type: BC 45.6%, metastatic 36.8%; Period: ongoing treatments.	Tools: CIS-FS (fatigue), CES-D (depressive and anxiety symptoms), HDI (well-being), RSCL (physical symptoms), GSBQ (social withdrawal), ICQ (illness cognition), GAS (goal engagement), and WGS (partner support)
de Rooij et al [30], 2021; The Netherlands	Secondary analysis from the PROFILES registry and NCR[g]; Cross-sectional survey data	Identify symptom clustering across cancer types using network modeling.	• N=1330 (835 female); Age: 61 (SD 15) years; Cancer type: BC 14.29%, CR 14.29%, Ov 14.29%, Thy 14.29%, HK 14.29%, NHK 14.29%, and CLL 14.29%; Period: years after diagnosis	Tools: EORTC-QLQ-C30 (30 symptoms; emotional and cognitive functioning scales)
Rha and Lee [31], 2021; Korea	Secondary data analysis from the SMILE RCT[h]; Longitudinal	Identify stable symptom clusters and their interrelationships across treatment cycles.	• N=249 (184 female); Age: 51.89 (SD 9.75) years; Cancer type: BC 60.3%, GC 33.3%, and LC 6.4%; Period: across CT cycles	Tools: assessment of 20 symptoms, including 12 core symptoms, with a numerical rating scale (0-10)
Shim et al [32], 2021; Korea	Longitudinal	Study physical or psychological symptoms and QoL changes before or after gastric surgery.	• N=256 (92 female); Age: 62.41 (SD 10.72) years; Cancer type: GC; Period: before and 1 week and 3-6 months after surgery	Tools: K-MDASI (Korean version; 13 symptoms), K-HADS (Korean version; depressive and anxiety symptoms), and FACT-Ga (QoL)

Study and country	Study design	Main purposes	Sample description (including cancer type[a]) and assessment period	Symptom assessment: tools[b] and biomarkers[c]
Henneghan et al [33], 2021; United States	Cross-sectional	Visualize symptom-cytokine networks and evaluate centrality in BC survivors.	• N=66 (66 female); Age: 48.44 (SD 8.73) years; Cancer type: BC; Period: after adjuvant CT (6 months-10 years)	Tools: PCI-total (cognition), UCLA Loneliness Scale (loneliness), Perceived Stress Scale (stress), PROMIS (fatigue, anxiety, depression), PSQI (sleep quality), and Epworth Sleepiness Scale (daytime sleepiness); Biomarkers: 13 cytokines (TNF-α, GM-CSF, INF-γ, IL-2, IL-1b, IL-5, IL-7, IL-8, IL-10, IL-13, IL-6, IL-2, and IL-4)
Kalantari et al [34], 2022; United Kingdom	Longitudinal	Analyze changes in symptom clusters across treatment time points.	• N=987 (779 female); Age: 56.9 (SD 12) years; Cancer type: BC 41.3%, GI 29.8%, GYN 17.7%, and LC 11.2%; Period: across 2 CT cycles, 6 time points	Tools: mMSAS (38 symptoms)
Lin et al [35], 2022; United States	Longitudinal	Examine temporal networks of psychoneurological symptoms.	• N=172 (45 female); Age: 59.8 (SD 9.9) years; Cancer type: HNC; Period: 4 times across radiotherapy	Tools: PHQ-8 (depressive symptoms), MFI (fatigue), PSQI (sleep quality), and PRO-CTCAE (cognitive dysfunction and pain)
Santoso et al [36], 2022; The Netherlands	Cross-sectional	Link 5 psychoneurological symptoms with stress biomarkers in newly diagnosed HNC.	• Cohort 1 (complete data): N=264 (55 female); Age: 65 (SD 8.2) years • Cohort 2 (incomplete data): N=475 (135 female); Age: 62 (SD 10.4) years • Cancer type: HNC; Period: at diagnosis and before treatment	Tools: PSQI (sleep quality), HADS (depressive and anxiety symptoms), EORTC-QLQ-H&N35 (EORTC-oral pain-related symptoms), and MFI (fatigue); Biomarkers: cortisol saliva, serum CRP, IL-6, IL-10, and TNF-α
Zhu et al [37], 2022; China	Cross-sectional	Explore network structure and symptom centrality in cancer survivors.	• N=1065 (712 female); Age: 65.00 (SD 11.42) years; Cancer type: BC 29.3%, GI 22.6%, HNK 14.74%, and LC 14.46%; Period: cancer treatments completed (years)	Tools: MDASI (18 symptoms)
Ji et al [38], 2023; China	Cross-sectional	Identify clusters and core symptoms after esophageal cancer surgery.	• N=286 (114 female); Age: 55.5% 65 years or older; Cancer type: early esophageal; Period: early postoperative	Tools: MDASI-GI
Röttgering et al [39], 2023; The Netherlands	Retrospective; Secondary analysis of merged studies	Compare global strength between symptom networks to understand if symptoms are more tightly connected in different subgroups of patients.	• N=256 (95 female); Age: mean 47 years; Cancer type: glioma; Period: pre- and postoperative	Tools: CIS-FS (fatigue), CES-D (depressive symptoms), MOS-cog (cognitive functioning), EORTC-BN-20 (EORTC brain-tumor-related symptoms), and SF-36 (HRQoL[i])
Jing et al [40], 2023; China	Secondary data analysis from a cross-sectional study	Explore symptom networks in patients with BC under endocrine therapy.	• N=613 (613 female); Age: 49.5 (SD 9.4) years; Cancer type: BC; Period: endocrine therapy after acute care	Tools: FACT-ES (19 items)
Li et al [41], 2023; China	Cross-sectional	Study links between symptoms and inflammatory biomarkers in glioma.	• N=203 (102 female); Age: 54.10 (SD 14.1) years; Cancer type: glioma; Period: during treatments	Tools: HAMA-14 (anxiety), HAMD-24 (depressive symptoms), PSQI (sleep quality), MFI (fatigue), and numerical rating scale 0-10 (pain); Biomarkers: IL-1β, IL-6, IL-4, IL-10, CRP, and TNF-α
Wang et al [42], 2023; China	Cross-sectional	Identify core symptom clusters in patients with DC.	• N=202 (58 female); Age: 66.01 (SD 8.97) years; Cancer type: DC; Period: ongoing therapies.	Tools: MDASI-GI

Study and country	Study design	Main purposes	Sample description (including cancer type[a]) and assessment period	Symptom assessment: tools[b] and biomarkers[c]
Teng et al [43], 2024; China	Cross-sectional	Map symptom clusters and central symptoms after CT in patients with LC.	• N=512 (139 female); Age: 65.21 (SD 8.94) years; Cancer type: LC (advanced 68%).; Period: post-CT	Tools: mMSAS (32 symptoms)
Kuang et al [44], 2024; China	Secondary analysis; Cross-sectional	Compare symptom networks by survivorship groups in elderly patients with cancer.	• N=485 (295 female); Age: 72.54 (SD 6.44) years; Cancer type: elderly patients with cancer; Period: after acute treatments	Tools: MDASI (18 symptoms)
Shang et al [45], 2024; China	Prospective	Track predictive interactions between symptoms over time.	• N=230 (94 female); Age: 66.13 (SD 10.80) years; Cancer type: operable CR; Period: pre- and postsurgery	Tools: MDASI-GI
Gong et al [46], 2024; China	Cross-sectional survey	Explore demoralization symptom networks in female patients with cancer.	• N=413 (413 female); Age: 54.01 (SD 10.35) years; Cancer type: BC 63.2%, GC 18.4%, DC 10.7%, and others 7.7%; Period: ongoing therapies	Tools: DS-MV

[a]The following cancer types were identified in the included studies: breast cancer (BC), chronic lymphocytic leukemia (CLL), colorectal cancer (CR), digestive cancer (DC), gastric cancer (GC), gastrointestinal cancer (GI), gynecological cancer (GYN), Hodgkin lymphoma (HK), head and neck cancer (HNC), lung cancer (LC), non-Hodgkin lymphoma (NHK), ovarian cancer (Ov), prostate cancer (PC), and thyroid cancer (Thy).

[b]The following assessment tools were used in the included studies: Center for Epidemiologic Studies Depression Scale (CES-D), Checklist Individual Strength–Fatigue Severity (CIS-FS), Demoralization Scale Mandarin version (DS-MV), European Organization for Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC-QLQ-C30), Functional Assessment of Cancer Therapy–Breast Cancer (FACT-B), Functional Assessment of Cancer Therapy–Endocrine Subscale (FACT-ES), Functional Assessment of Cancer Therapy–Gastric Cancer (FACT-Ga), Functional Outcomes of Sleepiness Questionnaire (FOSQ), Goal Adjustment Scale (GAS), Groningen Social Behavior Questionnaire (GSBQ), Hospital Anxiety and Depression Scale (HADS), Hamilton Anxiety Scale–14 items (HAMA-14), Hamilton Depression Scale–24 items (HAMD-24), health and disease inventory (HDI), illness cognitions questionnaire (ICQ), MD Anderson Symptom Inventory (MDASI), MDASI–gastrointestinal cancer version (MDASI-GI), Multidimensional Fatigue Symptom Inventory–Short Form (MFSI-SF), Modified Memorial Symptom Assessment Scale (mMSAS), Medical Outcomes Study Cognitive Functioning Scale (MOS-cog), neuropsychological test battery composite score (NP), Perceived Cognitive Impairment scale (PCI), Patient Health Questionnaire–8 items (PHQ-8), Patient Health Questionnaire–9 items (PHQ-9), Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE), Patient-Reported Outcomes Measurement Information System (PROMIS), Pittsburgh Sleep Quality Index (PSQI), Rotterdam Symptom Checklist (RSCL), 36-item short-form survey (SF-36), and ways of giving support (WGS).

[c]The following biomarkers were evaluated in the included studies: C-reactive protein (CRP), granulocyte-macrophage colony-stimulating factor (GM-CSF), interleukin (IL), interferon gamma (INF-γ), and tumor necrosis factor alpha (TNF-α).

[d]CT: chemotherapy.

[e]NIH: National Institutes of Health.

[f]QoL: quality of life.

[g]NCR: Netherlands Cancer Registry.

[h]RCT: randomized controlled trial.

[i]HRQoL: health-related quality of life.

Time of Publication

The included studies were published between 2010 and 2024, with 18 out of the 22 studies published between 2021 and February 2024.

Sample Sizes

The sample sizes ranged from 66 patients [33] to 4020 patients [28].

Design of the Study

Of the 22 studies, 7 were based on secondary data analyses and 15 were based on primary data (5 longitudinal and 10 cross-sectional studies).

Types of Cancer

Among the studies, mixed solid tumor populations were evaluated in 10 studies, digestive tract cohorts were assessed in 4 studies, breast cancer was evaluated in 3 studies, and head and neck cancers and gliomas were assessed in 2 studies. One study investigated a mixed cohort of solid and hematological cancers.

Tools for Symptom Assessment and Measures

Symptoms were assessed via 2 classes of validated tools. Twelve studies used general symptom assessment tools, such as the MD Anderson Symptoms Inventory and cancer type versions (n=6), the Modified Memorial Symptom Assessment Scale (n=3), the European Organization for the Research and Treatment of

Cancer Quality of Life Questionnaire (n=1), the Twenty Symptoms List (n=1), and the Functional Assessment of Cancer Therapy-Endocrine Subscale (n=1), whereas 10 studies focused on neuropsychological symptoms via tools such as the Patient Health Questionnaire-9 (n=1), the Perceived Cognitive Impairment scale (n=1), the Hospital Anxiety and Depression Scale (n=1), and 7 other tools.

Time of Symptom Assessment

Considering the timing of symptom assessments, only 1 study evaluated symptoms at diagnosis before any treatment [36], 1 evaluated symptoms during curative radiotherapy [28], 4 evaluated symptoms during the perioperative period [32,38,39,45], 5 evaluated symptoms during chemotherapy [25-27,31,34], 4 evaluated symptoms during ongoing cancer therapies [29,41,42,46], and 7 evaluated symptoms after acute treatments [28,30,33,37,40,43,44]. Among these, 3 studies evaluated the longitudinal changes in symptom networks across chemotherapy cycles [26,31,34] and 1 evaluated changes during radiotherapy sessions [35]. Three studies investigated the associations between symptoms and biological parameters [33,41,43].

NA Approach

Several NA approaches and models were used to perform the studies included in this review. Some studies used different models: network visualization (n=1), Bayesian network (n=1), pairwise Markov random field and IsingFit method (n=1), unregularized Gaussian graphical model (n=2), regularized partial correlation network (n=6), network visualization and community NA (n=1), network visualization and Walktrap algorithm (n=1), undirected network model with the Fruchterman-Reingold approach and edge-betweenness approach (n=4), biased correlation network and concise pattern network diagram (n=1), extended Bayesian information criterion graphical LASSO method (n=3), cross-lagged panel network (n=1), and unspecified NA (n=3).

Main Findings

In the following sections, we delve into the 22 included studies on the basis of the time of symptom assessment: at diagnosis and during or after acute cancer treatment.

Symptom Networks at Diagnosis

Only 1 study [36] evaluated symptom networks at cancer diagnosis (head and neck tumor). The connections between neuropsychological symptom networks and serum cytokines were also investigated. Four nodes had the most important position in the network: fatigue, poor sleep quality, C-reactive protein (CRP), and interleukin (IL)-6.

Symptom Networks During Acute Cancer Treatments

Chemotherapy

Five studies identified and evaluated symptom networks during chemotherapy.

Bhavnani et al [25] used NA to evaluate symptom co-occurrence in oncology, with a cohort of 665 patients with mixed tumors. Inspired by the results of symptom cluster research and its limitations, they used networks to visually analyze how 18 symptoms co-occurred across patients. They described a strongly nested structure of symptom co-occurrence, offering a new approach to the complexity of symptoms in patients with cancer.

Papachristou et al [27] investigated the relationships among 38 common symptoms in a cohort of 1328 patients with cancer, at 1 evaluation time point. They reported that the connections between and among symptoms may differ depending on the symptom dimension used to create the network (occurrence, severity, and distress). They identified a psychological symptom cluster that was stable across all 3 dimensions. They offered perspectives on the use of the network theory to develop new models aiming at improving therapeutic interventions for patients with cancer.

Other authors have reported the need for new interventions in patients with cancer. Rha and Lee [31] investigated clusters and the evolution of symptom networks across chemotherapy cycles in mixed solid tumor populations. They reported stable symptom clusters and evolving networks depending on the evaluation time point and the type of cancer, and the most central symptom identified was fatigue. The authors argued for the development of interventions focusing on central symptoms.

Kalantari et al [34] investigated 38 symptoms in 987 patients with cancer and assessed 4 different cancer types across 2 cycles of chemotherapy. They identified 8 relatively stable symptom clusters.

Xu et al [26] evaluated neuropsychological symptoms and QoL in a newly diagnosed breast cancer population across several chemotherapy cycles. They applied Bayesian network methods to investigate the role of sleep, fatigue, and mood on cognition and QoL across and after chemotherapy. They revealed strong direct and indirect links among symptoms, cognitive performance, and QoL. Sleep quality was directly linked to cognitive performance with late chemotherapy cycles. The authors concluded that a better understanding of the interrelationships among those symptoms, QoL, and cognition could guide the design of further effective interventions [26].

Radiotherapy

Lin et al [35] evaluated psychoneurological symptoms during radiotherapy (4 times) in 172 patients with newly diagnosed head and neck cancer. Depression and fatigue were the 2 core symptoms identified. The network structure was relatively stable over the treatment time. As previously suggested by other authors, Lin et al [35] argued that identifying core symptoms represents an opportunity to decrease other co-occurring symptoms.

Perioperative Period

Four studies evaluated symptom networks during the perioperative period (3 digestive tract tumor populations and 1 glioma population).

Röttgering et al [39] evaluated patterns of associations among depression, cognition, brain tumor–related symptoms, and QoL in a population of 256 diffuse gliomas. They constructed 6 networks based on the presence or absence of 3 disease statuses (surgical, tumor grade, and fatigue). Fatigue severity, depression, and social functioning were nodes highly correlated across the

6 networks. The number of nodes in the nonfatigued network was lower than that in the fatigued network. They suggested the need for integrative symptom management and targeted fatigue as a priority.

Other authors have reported the need to target specific psychological symptoms to reduce other interconnected symptoms [32]. Indeed, Shim et al [32] evaluated associations between cancer-related physical and psychological symptoms and QoL, before and after intent-to-cure surgery in 256 patients with gastric cancer. Distress and sadness were the most central symptoms across all time points. They identified connections between emotional and physical well-being.

Ji et al [38] reported that multiple symptom clusters occurred in a cohort of 286 patients with early esophageal cancer who were surgically treated. Sadness and fatigue were the core symptoms.

Shang et al [45] conducted a prospective study among 230 patients with operable colorectal cancer and evaluated 18 symptoms before and after surgery. They described a stable network with disturbed sleep and distress, which are the most prevalent symptoms to be targeted.

Ongoing Treatments

Several cross-sectional studies have evaluated networks in mixed cancer patient cohorts or mixed treatments (surgery, radiotherapy, chemotherapy, and hormonal treatment).

Among the objectives of NA research, identifying risk and protective factors has been discussed as an interesting approach to help further treatment strategies in patients with cancer. In 342 treated patients with cancer seeking psychological care, Schellekens et al [29] investigated the interconnectedness of fatigue, depression, anxiety, and potential risk and protective factors (physical symptoms, social withdrawal, illness cognition, goal adjustment, and partner support). Depressed mood, loss of enjoyment, and worthlessness were central nodes. Fatigue, anxiety, and depression appear strongly interconnected. Acceptance of illness was centrally embedded in the networks.

Wang et al [42] conducted a study among 202 treated patients with digestive cancer and identified distress, disturbed sleep, poor appetite, and sadness as the most common symptoms. The psychoemotional symptom cluster was the core symptom cluster.

Gong et al [46] explored the core and bridge symptoms of demoralization in 420 treated female patients with cancer. Discouragement, a lack of self-worth, hopelessness, and vulnerability were identified as the core and bridge symptoms.

Symptom Networks After Acute Cancer Treatments

Seven studies investigated symptom networks after acute cancer treatments, ranging from 1 week to several years.

Teng et al [43] identified 4 symptom clusters with a high stability network in 512 patients with advanced lung cancer 1 week after chemotherapy cycles.

Jing et al [40] explored the relationship of endocrine therapy–related symptoms to identify the core symptoms in a population of 613 patients with breast cancer receiving adjuvant hormonal treatment (average duration: 3.6 years). Mood swings and irritability were the most prevalent symptoms, and loss of interest in sex and joint pains were the most severe symptoms. There were no significant differences in network structure or global strength across treatment types (aromatase inhibitors vs selective estrogen receptor modulators) [40].

Concerning depressive symptoms in patients with cancer, Hartung et al [28] compared the frequency and relationships of depressive symptoms between patients with cancer and those in the general population. Depressive symptoms were much more common in patients with cancer but were less closely related to each other. Individual depressive symptom patterns should be considered in individuals rather than in group analyses.

de Rooij et al [30] aimed to explore the full complexity of symptoms. In a study evaluating symptom clusters in 1330 survivors of 7 cancer types, they reported that fatigue was consistently the most central symptom in an identified cluster and should be targeted. They concluded that interrelated symptoms may share the same underlying pathophysiological mechanisms, offering opportunities for new reflections on treatment strategies [30].

Henneghan et al [33] explored symptom networks in 66 patients with breast cancer after adjuvant chemotherapy (6 months to 10 years) and reported that stress, loneliness, depressive symptoms, and fatigue co-occur rather than occur as individual symptoms.

Zhu et al [37] investigated the symptom network of multidimensional symptom experiences and explored centrality indices and density networks in a cohort of 1065 patients with cancer who survived. They demonstrated that fatigue was the most severe symptom and that the density of the “less than 5 years” network was significantly different from that of the longest survivorship network. Distress, sadness, and lack of appetite were the core symptoms.

Kuang et al [44] explored symptom networks in 483 elderly patients with cancer who survived. The most common and severe symptoms were fatigue, disturbed sleep, and difficulty remembering. The density network showed differences between “less than 5 years” and “more than 5 years” survival.

Symptom Networks and Biological Parameters

A study by Santoso et al [36] examined potential connections between psychoneurological symptoms (poor sleep, anxiety, and fatigue) and biomarkers of stress (cortisol slope, CRP, IL-6, IL-10, and tumor necrosis factor-α) in more than 264 patients with newly diagnosed head and neck cancer before treatment. Four nodes had the most important position in the network (fatigue, poor sleep quality, CRP, and IL-6) and may play a role in the interconnections between symptoms and inflammatory conditions.

Henneghan et al [33] investigated different symptoms and 13 cytokines in 66 patients with breast cancer at least 6 months and up to 10 years after adjuvant chemotherapy. Node betweenness was the highest for perceived cognitive impairment and the IL-2 level. Two separate communities of nodes

(symptoms and cytokines) within the network were revealed and connected by several edges. They concluded that perceived cognitive impairment, stress, loneliness, depressive symptoms, and fatigue co-occur and that cytokines may be involved in these biological pathways.

A study by Li et al [41] evaluated the interrelations between neuropsychological symptoms and inflammatory biomarkers in a cohort of 203 patients with glioma. Depression, anxiety, fatigue, IL-6, and tumor necrosis factor-α had higher strength centrality indices and were identified as the most central nodes within the symptom-biomarker networks.

Methodological Quality and Risk of Bias

The methodological quality of the 22 included studies was evaluated using the MINORS tool, and the results are summarized in Table 3. Overall, studies demonstrated moderate to high methodological rigor. Many studies, including the studies by Bhavnani et al [25], Xu et al [26], and Shim et al [32], reported clear aims and appropriate statistical analyses. Several studies, such as the studies by Rha and Lee [31], Kalantari et al [34], and Lin et al [35], employed longitudinal designs, enhancing their capacity to assess symptom dynamics over time. Other studies, such as the studies by Papachristou et al [27] and de Rooij et al [30], used large, heterogeneous samples with comprehensive network models, though some lacked follow-up or sample size reporting. Comparative studies scored well across all 12 MINORS criteria (eg, [28,40,42]). In contrast, noncomparative studies were assessed on the first 8 criteria and showed greater variability (eg, [33,36,46]). No studies were excluded based on their MINORS scores, but rather, the risk of bias assessment informed our interpretation of the findings and provided essential context for understanding methodological strengths and limitations across the reviewed literature.

Table 3. Quality assessment of the included studies.

Study	Category														Key commentary
	CA[a]	CP[b]	PD[c]	AE[d]	UA[e]	FP[f]	Loss <5%[g]	SSC[h]	CG[i]	CoG[j]	GE[k]	SA[l]	NCS[m]	CS[n]	
Bhavnani et al [25], 2010	2	2	1	2	2	0	0	0	2	0	0	2	13	13	Innovative study using network analysis to show symptoms that follow a nested structure rather than distinct clusters; the main limitation was secondary data analysis.
Xu et al [26], 2018	2	1	2	2	2	2	0	0	0	0	0	2	11	11	Sophisticated Bayesian network modeling study; found that sleep quality during chemotherapy was directly linked to cognitive performance in patients with breast cancer.
Papachristou et al [27], 2019	2	1	2	2	2	0	0	1	0	2	2	2	10	14	Large sample study showing that symptom networks differ by the dimension assessed (occurrence).
Hartung et al [28], 2019	2	1	1	2	2	0	0	2	2	2	2	2	10	16	Rigorous study with large matched samples showing that depressive symptoms were more frequent but less interconnected in patients with cancer; suggested that external factors drive symptoms.
Schellekens et al [29], 2020	2	2	2	2	2	0	0	1	0	0	0	2	11	11	Robust preregistered analysis of distressed patients with cancer; identified depressed mood.
de Rooij et al [30], 2021	2	1	0	2	2	0	0	2	1	2	1	2	10	15	Strong methodological study examining symptom clusters across 7 cancer types; identified fatigue as a consistently central symptom; limited by cross-sectional design.
Rha and Lee [31], 2021	2	1	2	2	2	2	0	1	0	0	0	2	12	12	Strong longitudinal analysis identified 3 stable symptom clusters across chemotherapy cycles, with fatigue as the most central symptom across all time points.
Shim et al [32], 2021	2	2	2	2	2	2	0	0	0	0	0	2	12	12	Prospective study highlighting the central role of distress and sadness across perioperative time points; psychological symptoms served as bridges connecting symptoms to quality of life.
Henneghan et al [33], 2021	2	0	2	2	2	0	0	0	0	0	0	2	8	8	Innovative exploratory study examining symptom-cytokine networks in breast cancer survivors; identified IL-2 and cognitive impairment as central; limited by a small sample size.
Kalantari et al [34], 2022	2	1	2	2	2	2	0	0	0	2	1	2	11	14	Rigorous longitudinal study identifying 8 symptom clusters across chemotherapy cycles; demonstrated the stability of core symptoms over time despite changing severity.
Lin et al [35], 2022	2	1	2	2	2	2	0	1	0	2	1	2	12	15	Strong longitudinal study examining symptom networks in patients with head and neck cancer; identified depression and fatigue as core symptoms across time points.
Santoso et al [36], 2022	2	2	2	2	2	0	0	0	0	0	0	2	10	10	Large sample study found poor sleep.
Zhu et al [37], 2023	2	1	2	2	2	0	0	0	0	0	0	2	9	9	Large sample study found distress, sadness, and lack of appetite as core symptoms in cancer survivors; network density was higher in survivors <5 years vs >5 years.

Study	Category														Key commentary
	CA[a]	CP[b]	PD[c]	AE[d]	UA[e]	FP[f]	Loss <5%[g]	SSC[h]	CG[i]	CoG[j]	GE[k]	SA[l]	NCS[m]	CS[n]	
Ji et al [38], 2023	2	1	2	2	2	0	0	0	0	0	0	2	9	9	Identified symptom clusters in early recovery after esophageal cancer surgery, with sadness and fatigue as core symptoms; good sample size but convenience sampling.
Röttgering et al [39], 2024	2	1	1	2	2	0	0	0	0	2	1	2	8	11	Found fatigue.
Jing et al [40], 2023	2	1	1	2	2	0	0	1	0	2	1	2	9	14	Strong analysis of endocrine therapy–related symptoms in patients with breast cancer; identified emotional symptoms as central regardless of treatment type.
Li et al [41], 2023	2	1	2	2	2	0	0	0	0	0	0	2	9	9	Examined symptom-biomarker interconnections in patients with glioma; found depression.
Wang et al [42], 2025	2	1	2	2	2	0	0	2	0	0	0	2	11	11	Well-designed study that identified a psychoemotional symptom cluster as core in patients with digestive cancer; distress had the highest centrality and the strongest bridge effect.
Teng et al [43], 2024	2	1	2	2	2	0	0	0	0	0	0	2	9	9	Large sample study that identified the sickness behavior symptom cluster as most central in postchemotherapy patients with lung cancer; feeling irritable was a core symptom.
Kuang et al [44], 2024	2	1	1	2	2	0	0	0	0	2	1	2	8	11	Examined symptom networks in older adults with cancer; found vomiting.
Shang et al [45], 2024	2	1	2	2	2	2	0	0	0	0	0	2	11	11	Innovative longitudinal study using cross-lagged panel network analysis; identified disturbed sleep during admission as a predictor of subsequent symptoms in patients with colorectal cancer.
Gong et al [46], 2024	2	1	2	2	2	0	0	0	0	0	0	2	9	9	Identified key demoralization symptoms in Chinese female patients with cancer; strengths included prospective data collection; limited by convenience sampling.

[a]CA: clear aim.

[b]CP: consecutive patients.

[c]PD: prospective data.

[d]AE: appropriate end points.

[e]UA: unbiased assessment.

[f]FP: follow-up period.

[g]Loss <5%: loss to follow-up <5%.

[h]SSC: sample size calculation.

[i]CG: control group.

[j]CoG: contemporary groups.

[k]GE: group equivalence.

[l]SA: statistical analysis.

[m]NCS: noncomparative score (/16).

[n]CS: comparative score (/24).

Discussion

NA and Current Knowledge

Patients with cancer experience a highly interconnected network of co-occurring symptoms, which arise from both the disease itself and its treatment, significantly affecting QoL [4,5]. Traditional symptom analysis methods have primarily relied on symptom clustering approaches that fail to capture the dynamic interplay and mutual reinforcement between symptoms [13,14,16]. This review highlights the growing application of NA in oncology, demonstrating its potential to redefine symptom

management by shifting the focus from treating individual symptoms or clusters of symptoms to identifying core symptoms that exert a broader influence on the overall network. As Papachristou et al [27] suggested, the network theory could offer interesting perspectives for understanding and focusing on specific symptoms to implement new therapeutic interventions, subsequently improving the management of patients with cancer.

Across the 22 included studies, NA was applied at different cancer treatment stages to identify key symptom interconnections and potential intervention targets. One of the most consistent findings was the prominent role of psychological symptoms, particularly anxiety, depression, and distress, which formed stable and interconnected networks, especially during chemotherapy and long-term survivorship [27,37,40]. For instance, Papachristou et al [27] found that psychological symptoms, such as anxiety, depression, and distress, tend to form stable networks during chemotherapy, influencing overall symptom burden. The stability of these networks suggests that psychological symptoms play a central role in shaping cancer symptomatology, potentially exacerbating physical symptoms through stress-related mechanisms. In support of this assumption, previous studies that did not apply NA have suggested that psychological disorders can substantially worsen physical symptoms in patients with cancer [47,48]. A clear example is the study by Renna et al [47] showing that breast cancer survivors with a distress disorder may be particularly at risk for more physical symptoms and treatment, reducing their QoL and increasing the recurrence risk.

Another recurrent finding in our work was the identification of fatigue as a central symptom across all treatment phases, with strong interconnections to sleep disturbance, cognitive impairment, and emotional distress [30,31,35]. Prominent studies, such as those by Rha and Lee [31] and Lin et al [35], reported that fatigue and depression consistently emerge as core symptoms, suggesting that targeting these symptoms may alleviate multiple co-occurring symptoms. The widespread influence of fatigue highlights the importance of psychophysiological symptoms in cancer symptom monitoring, reinforcing the need for targeted interventions that address not only fatigue itself but also its cascading effects on other symptoms.

In addition, while most studies were purely descriptive, a subset integrated biological markers into NA models, revealing significant associations among fatigue, depression, sleep disturbances, and inflammatory biomarkers such as IL-6, CRP, and tumor necrosis factor-α [33,36,41]. These findings suggest a possible biological underpinning of symptom clustering, aligning with existing evidence that inflammatory pathways may contribute to cancer-related fatigue and neuropsychological symptoms [49,50]. However, the mechanistic links between inflammation and symptom networks remain unclear, warranting further investigation.

Collectively, these findings support the hypothesis that targeting core symptoms, particularly fatigue and psychological distress, may provide a more effective therapeutic approach than treating symptoms in isolation. Despite the promising insights provided by NA, the studies reviewed were primarily descriptive, limiting their immediate clinical applicability. Further studies, particularly longitudinal studies and interventional trials, are necessary to determine whether NA-informed symptom management strategies can improve patient outcomes and facilitate the integration of network-based approaches into clinical practice.

Implications for Clinical Practice

From the clinician's point of view, we strongly believe that the NA approach could lead to innovations in interventions for patients with cancer. We thus argue for more longitudinal design studies investigating homogeneous patient cohorts. Consensus is required on tools to measure symptoms, with preference for polyvalent assessment (somatic and psychological symptoms).

With respect to network types, we suggest the use of Bayesian networks as the primary approach, considering the potential implementation of their outcomes in artificial intelligence datasets and consequently their use in clinical settings, especially in health-risk prediction and the evaluation of specific therapeutic interventions.

Finally, we believe that more clinician involvement (medical oncology, radiotherapy, oncological surgery, supportive care, and palliative care) in this area of research is highly necessary.

Limitations and Perspectives for This Research Area

While this review offers a comprehensive synthesis of the current literature on the use of NA in cancer symptomatology, with most articles published during the last 3 years [30-46], some limitations must be acknowledged. First, there was considerable heterogeneity among the included studies in terms of cancer types, patient populations, sample sizes, symptom assessment tools, and network modeling techniques. This variability limited the ability to make direct comparisons across studies and precluded meta-analytic synthesis. Additionally, the majority of studies employed cross-sectional and exploratory designs, which, although valuable for hypothesis generation, limit the capacity to infer causality or observe symptom network evolution over time. Only a small number of studies incorporated biological markers [33,36,41], and none examined the impact of NA-informed therapeutic interventions, which constrains the applicability of the current findings to clinical practice.

In addition, although many included studies examined psychological symptoms, such as anxiety and depression, relatively few explicitly assessed neurocognitive functioning, despite its well-documented vulnerability to cancer treatments [51]. For example, while some studies incorporated cognitive performance indicators (eg, [26,39]), a more systematic and targeted exploration of cancer-related cognitive impairment within network models remains lacking. This represents an important research gap, as cancer-related cognitive impairment is increasingly recognized as a major component of cancer survivorship [51].

These limitations reflect broader gaps in the field and point to important directions for future research. First, there is a pressing need for longitudinal studies that can track changes in symptom

networks across different treatment phases and survivorship. Such designs would enable researchers to identify critical time points at which symptom interconnectivity shifts, potentially informing more precise and timely interventions. Moreover, future research should move beyond descriptive modeling to include interventional studies, particularly randomized controlled trials designed to test whether targeting core symptoms identified through NA leads to measurable improvements in symptom burden and QoL. For example, evidence from non-NA–based trials has shown beneficial effects of physical activity and mind-body interventions (eg, yoga and mindfulness) on neuropsychological symptoms in cancer populations [52,53]. Incorporating such interventions into future network-informed studies could provide valuable insights into how these therapies affect symptom interconnectivity.

Another key area for development is the integration of biological and physiological data into NA frameworks. The limited but promising evidence linking symptom networks with inflammatory markers (eg, IL-6, CRP, and tumor necrosis factor-α) suggests that incorporating physiological data, including genomic, proteomic, and metabolomic variables, could shed light on the underlying mechanisms driving symptom interactions [33,36,41]. This could, in turn, support the development of biologically informed, personalized treatment strategies.

Furthermore, to ensure greater consistency and comparability across future studies, standardization of methodological approaches is essential. This includes the use of validated and comprehensive symptom assessment tools, consistent time points for symptom evaluation, and transparent reporting of network construction and statistical parameters. The development of consensus guidelines for conducting and reporting NA in oncology would represent a valuable step toward building a more cohesive and interpretable body of research.

In addition to study design variability, methodological considerations inherent to NA approaches must also be acknowledged. Several included studies used different NA techniques, such as Gaussian graphical models or Bayesian networks, with varying assumptions, sparsity constraints, and estimation procedures. The sensitivity and specificity of these models in capturing symptom interconnections depend heavily on data quality, sample size, and statistical regularization methods [20]. Furthermore, the stability of centrality measures and the reproducibility of network structures were not systematically evaluated across studies, limiting conclusions about the robustness and generalizability of findings [54]. Greater methodological standardization and reporting transparency are needed to ensure the validity of symptom networks in oncology.

In summary, while current studies provide compelling evidence for the potential of NA to enhance symptom understanding in cancer care, addressing methodological limitations and expanding the scope of research are essential next steps.

Conclusions

Cancer is a complex disease that causes significant disruption to both biological systems and overall health, leading to complex, interrelated symptom experiences in patients with cancer. This review highlights the growing application of NA as a valuable tool for understanding the complexity of cancer-related symptoms.

Across the included studies, NA consistently identified core symptoms, particularly psychological symptoms and fatigue, that appear central to patients' experiences across treatment stages. These findings suggest that focusing on core symptom interconnectivity may offer more effective avenues for symptom management than traditional approaches targeting isolated symptoms.

While current research offers compelling evidence for the application of NA in cancer symptomatology, several methodological limitations persist. Future studies should focus on longitudinal designs that track symptom networks across different phases of cancer treatment and survivorship. Further research should also explore interventional approaches to determine whether NA-informed strategies can improve symptom management and enhance QoL. Integrating biological and physiological data into NA frameworks holds promise for developing personalized, biologically informed treatment strategies.

Standardization of methodological approaches, including validated symptom assessment tools, and transparent reporting of network construction are essential to strengthen the consistency and comparability of findings across studies. Ultimately, network-informed research can contribute to a deeper understanding of cancer symptom interconnectivity and lead to more effective and targeted interventions, improving outcomes for patients with cancer.

Acknowledgments

The authors would like to thank the staff and colleagues at the Department of Computational Medicine and Neuropsychiatry, Faculty of Medicine, University of Mons, for their support and collaboration throughout the development of this study. We also acknowledge the contributions of all researchers whose work formed the basis of this systematic review. Generative OpenAI (ChatGPT) was used under the authors' supervision for language editing and grammatical improvement only, without any intellectual contribution. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

VR contributed to conceptualization, data curation, formal analysis, writing–original draft, and writing–review and editing. AG contributed to conceptualization, data curation, formal analysis, methodology, writing–original draft, and writing–review and editing. EP contributed to visualization and writing–review and editing. GB contributed to supervision, writing–review and editing, and validation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[PDF File (Adobe PDF File), 129 KB - cancer_v11i1e66087_app1.pdf]

Multimedia Appendix 2
Detailed search strategies.
[DOCX File , 21 KB - cancer_v11i1e66087_app2.docx]

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024 Apr 04;74(3):229-263 [FREE Full text] [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Mayer DK, Nasso SF, Earp JA. Defining cancer survivors, their needs, and perspectives on survivorship health care in the USA. The Lancet Oncology 2017 Jan;18(1):e11-e18. [doi: 10.1016/s1470-2045(16)30573-3]
3. Institute of Medicine and National Research Council. From Cancer Patient to Cancer Survivor: Lost in Transition. Washington, DC: The National Academies Press; 2006.
4. Tevaarwerk A, Denlinger CS, Sanft T, Ansbaugh SM, Armenian S, Baker KS, et al. Survivorship, Version 1.2021. J Natl Compr Canc Netw 2021 Jun 30;19(6):676-685 [FREE Full text] [doi: 10.6004/jnccn.2021.0028] [Medline: 34214969]
5. Joshy G, Thandrayen J, Koczwara B, Butow P, Laidsaar-Powell R, Rankin N, et al. Disability, psychological distress and quality of life in relation to cancer diagnosis and cancer type: population-based Australian study of 22,505 cancer survivors and 244,000 people without cancer. BMC Med 2020 Dec 01;18(1):372 [FREE Full text] [doi: 10.1186/s12916-020-01830-4] [Medline: 33256726]
6. Lagergren P, Schandl A, Aaronson NK, Adami H, de Lorenzo F, Denis L, European Academy of Cancer Sciences. Cancer survivorship: an integral part of Europe's research agenda. Mol Oncol 2019 Mar 08;13(3):624-635 [FREE Full text] [doi: 10.1002/1878-0261.12428] [Medline: 30552794]
7. Multinational Association of Supportive Care in Cancer (MASCC). URL: https://mascc.org/ [accessed 2025-06-16]
8. Dodd MJ, Miaskowski C, Lee KA. Occurrence of symptom clusters. J Natl Cancer Inst Monogr 2004 Jul 01;2004(32):76-78. [doi: 10.1093/jncimonographs/lgh008] [Medline: 15263044]
9. Miaskowski C, Barsevick A, Berger A, Casagrande R, Grady PA, Jacobsen P, et al. Advancing symptom science through symptom cluster research: expert panel proceedings and recommendations. J Natl Cancer Inst 2017 Apr 24;109(4):djw253 [FREE Full text] [doi: 10.1093/jnci/djw253] [Medline: 28119347]
10. Miaskowski C, Dodd M, Lee K. Symptom clusters: the new frontier in symptom management research. J Natl Cancer Inst Monogr 2004(32):17-21. [doi: 10.1093/jncimonographs/lgh023] [Medline: 15263036]
11. Harris CS, Kober KM, Conley YP, Dhruva AA, Hammer MJ, Miaskowski CA. Symptom clusters in patients receiving chemotherapy: A systematic review. BMJ Support Palliat Care 2022 Mar 17;12(1):10-21 [FREE Full text] [doi: 10.1136/bmjspcare-2021-003325] [Medline: 34921000]
12. So WKW, Law BMH, Ng MSN, He X, Chan DNS, Chan CWH, et al. Symptom clusters experienced by breast cancer patients at various treatment stages: A systematic review. Cancer Med 2021 Apr 21;10(8):2531-2565 [FREE Full text] [doi: 10.1002/cam4.3794] [Medline: 33749151]
13. Kim H, Abraham I, Malone P. Analytical methods and issues for symptom cluster research in oncology. Curr Opin Support Palliat Care 2013 Mar;7(1):45-53. [doi: 10.1097/SPC.0b013e32835bf28b] [Medline: 23196378]
14. Chen E, Nguyen J, Khan L, Zhang L, Cramarossa G, Tsao M, et al. Symptom clusters in patients with advanced cancer: a reanalysis comparing different statistical methods. J Pain Symptom Manage 2012 Jul;44(1):23-32 [FREE Full text] [doi: 10.1016/j.jpainsymman.2011.07.011] [Medline: 22658252]
15. Aktas A. Cancer symptom clusters: current concepts and controversies. Curr Opin Support Palliat Care 2013 Mar;7(1):38-44. [doi: 10.1097/SPC.0b013e32835def5b] [Medline: 23287418]
16. Zhu Z, Xing W, Hu Y, Wu B, So WK. Paradigm shift: Moving from symptom clusters to symptom networks. Asia Pac J Oncol Nurs 2022 Jan;9(1):5-6 [FREE Full text] [doi: 10.1016/j.apjon.2021.12.001] [Medline: 35528791]

17. Borsboom D. A network theory of mental disorders. World Psychiatry 2017 Feb 26;16(1):5-13 [FREE Full text] [doi: 10.1002/wps.20375] [Medline: 28127906]
18. Briganti G, Scutari M, Epskamp S, Borsboom D, Hoekstra RHA, Golino HF, et al. Network analysis: An overview for mental health research. Int J Methods Psychiatr Res 2024 Dec 14;33(4):e2034. [doi: 10.1002/mpr.2034] [Medline: 39543824]
19. Hevey D. Network analysis: a brief overview and tutorial. Health Psychol Behav Med 2018 Sep 25;6(1):301-328 [FREE Full text] [doi: 10.1080/21642850.2018.1521283] [Medline: 34040834]
20. Borsboom D, Deserno MK, Rhemtulla M, Epskamp S, Fried EI, McNally RJ, et al. Network analysis of multivariate data in psychological science. Nat Rev Methods Primers 2021 Aug 19;1(1):58. [doi: 10.1038/s43586-021-00055-w]
21. Kossakowski JJ, Epskamp S, Kieffer JM, van Borkulo CD, Rhemtulla M, Borsboom D. The application of a network approach to Health-Related Quality of Life (HRQoL): introducing a new method for assessing HRQoL in healthy adults and cancer patients. Qual Life Res 2016 Apr 14;25(4):781-792 [FREE Full text] [doi: 10.1007/s11136-015-1127-z] [Medline: 26370099]
22. Kosvyra A, Ntzioni E, Chouvarda I. Network analysis with biological data of cancer patients: A scoping review. J Biomed Inform 2021 Aug;120:103873 [FREE Full text] [doi: 10.1016/j.jbi.2021.103873] [Medline: 34298154]
23. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg 2003 Sep;73(9):712-716. [doi: 10.1046/j.1445-2197.2003.02748.x] [Medline: 12956787]
24. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. : An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. Campbell Syst Rev 2022 Jun 27;18(2):e1230 [FREE Full text] [doi: 10.1002/cl2.1230] [Medline: 36911350]
25. Bhavnani SK, Bellala G, Ganesan A, Krishna R, Saxman P, Scott C, et al. The nested structure of cancer symptoms. Implications for analyzing co-occurrence and managing symptoms. Methods Inf Med 2010;49(6):581-591 [FREE Full text] [doi: 10.3414/ME09-01-0083] [Medline: 21085743]
26. Xu S, Thompson W, Ancoli-Israel S, Liu L, Palmer B, Natarajan L. Cognition, quality-of-life, and symptom clusters in breast cancer: Using Bayesian networks to elucidate complex relationships. Psychooncology 2018 Mar 21;27(3):802-809 [FREE Full text] [doi: 10.1002/pon.4571] [Medline: 29055062]
27. Papachristou N, Barnaghi P, Cooper B, Kober KM, Maguire R, Paul SM, et al. Network analysis of the multidimensional symptom experience of oncology. Sci Rep 2019 Feb 19;9(1):2258 [FREE Full text] [doi: 10.1038/s41598-018-36973-1] [Medline: 30783135]
28. Hartung TJ, Fried EI, Mehnert A, Hinz A, Vehling S. Frequency and network analysis of depressive symptoms in patients with cancer compared to the general population. J Affect Disord 2019 Sep 01;256:295-301. [doi: 10.1016/j.jad.2019.06.009] [Medline: 31200167]
29. Schellekens MPJ, Wolvers MDJ, Schroevers MJ, Bootsma TI, Cramer AOJ, van der Lee ML. Exploring the interconnectedness of fatigue, depression, anxiety and potential risk and protective factors in cancer patients: a network approach. J Behav Med 2020 Aug 22;43(4):553-563 [FREE Full text] [doi: 10.1007/s10865-019-00084-7] [Medline: 31435892]
30. de Rooij BH, Oerlemans S, van Deun K, Mols F, de Ligt KM, Husson O, et al. Symptom clusters in 1330 survivors of 7 cancer types from the PROFILES registry: A network analysis. Cancer 2021 Dec 15;127(24):4665-4674 [FREE Full text] [doi: 10.1002/cncr.33852] [Medline: 34387856]
31. Rha SY, Lee J. Stable symptom clusters and evolving symptom networks in relation to chemotherapy cycles. J Pain Symptom Manage 2021 Mar;61(3):544-554 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.08.008] [Medline: 32828931]
32. Shim E, Ha H, Suh Y, Kong S, Lee H, Yang H, et al. Network analyses of associations between cancer-related physical and psychological symptoms and quality of life in gastric cancer patients. Psychooncology 2021 Jun 13;30(6):946-953. [doi: 10.1002/pon.5681] [Medline: 33760355]
33. Henneghan A, Wright ML, Bourne G, Sales AC. A cross-sectional exploration of cytokine-symptom networks in breast cancer survivors using network analysis. Can J Nurs Res 2021 Sep;53(3):303-315 [FREE Full text] [doi: 10.1177/0844562120927535] [Medline: 32482100]
34. Kalantari E, Kouchaki S, Miaskowski C, Kober K, Barnaghi P. Network analysis to identify symptoms clusters and temporal interconnections in oncology patients. Sci Rep 2022 Oct 12;12(1):17052 [FREE Full text] [doi: 10.1038/s41598-022-21140-4] [Medline: 36224203]
35. Lin Y, Bruner DW, Paul S, Miller AH, Saba NF, Higgins KA, et al. A network analysis of self-reported psychoneurological symptoms in patients with head and neck cancer undergoing intensity-modulated radiotherapy. Cancer 2022 Oct;128(20):3734-3743 [FREE Full text] [doi: 10.1002/cncr.34424] [Medline: 35969226]
36. Santoso AMM, Jansen F, Peeters CFW, Baatenburg de Jong RJ, Brakenhoff RH, Langendijk JA, et al. Psychoneurological symptoms and biomarkers of stress and inflammation in newly diagnosed head and neck cancer patients: a network analysis. Curr Oncol 2022 Sep 28;29(10):7109-7121 [FREE Full text] [doi: 10.3390/curroncol29100559] [Medline: 36290836]
37. Zhu Z, Sun Y, Kuang Y, Yuan X, Gu H, Zhu J, et al. Contemporaneous symptom networks of multidimensional symptom experiences in cancer survivors: A network analysis. Cancer Med 2023 Jan;12(1):663-673 [FREE Full text] [doi: 10.1002/cam4.4904] [Medline: 35651298]

38. Ji Y, Zhang Y, Shi Z, Wang P, Chen T, Fan Z, et al. The symptom cluster and core symptoms of early recovery in patients after esophageal cancer surgery: A network analysis. ResearchGate. 2023. URL: https://www.researchgate.net/publication/376373425_The_symptom_cluster_and_core_symptoms_of_early_recovery_in_patients_after_esophageal_cancer_surgery_A_network_analysis [accessed 2025-06-16]
39. Röttgering JG, Varkevisser TMCK, Gorter M, Belgers V, De Witt Hamer PC, Reijneveld JC, et al. Symptom networks in glioma patients: understanding the multidimensionality of symptoms and quality of life. J Cancer Surviv 2024 Jun 16;18(3):1032-1041 [FREE Full text] [doi: 10.1007/s11764-023-01355-8] [Medline: 36922442]
40. Jing F, Zhu Z, Qiu J, Tang L, Xu L, Xing W, et al. Contemporaneous symptom networks and correlates during endocrine therapy among breast cancer patients: A network analysis. Front Oncol 2023 Mar 31;13:1081786 [FREE Full text] [doi: 10.3389/fonc.2023.1081786] [Medline: 37064124]
41. Li H, Shi X, Li J, Zhang X, Li F. Psychoneurological symptoms and inflammatory markers in patients with glioma in China: a network analysis. Support Care Cancer 2023 Jul 03;31(7):435. [doi: 10.1007/s00520-023-07873-6] [Medline: 37395813]
42. Wang K, Diao M, Yang Z, Salvador JT, Zhang Y. Identification of core symptom cluster in patients with digestive cancer. Cancer Nurs 2023 Oct 26;48(1):E55-E63. [doi: 10.1097/ncc.0000000000001280]
43. Teng L, Zhou Z, Yang Y, Sun J, Dong Y, Zhu M, et al. Identifying central symptom clusters and correlates in patients with lung cancer post-chemotherapy: A network analysis. Asia Pac J Oncol Nurs 2024 Apr;11(4):100383 [FREE Full text] [doi: 10.1016/j.apjon.2024.100383] [Medline: 38495643]
44. Kuang Y, Jing F, Sun Y, Zhu Z, Xing W. Symptom networks in older adults with cancer: A network analysis. J Geriatr Oncol 2024 Apr;15(3):101718. [doi: 10.1016/j.jgo.2024.101718] [Medline: 38340638]
45. Shang B, Bian Z, Luo C, Lv F, Wu J, Lv S, et al. Exploring the dynamics of perioperative symptom networks in colorectal cancer patients: a cross-lagged panel network analysis. Support Care Cancer 2023 Dec 27;32(1):62. [doi: 10.1007/s00520-023-08288-z] [Medline: 38150034]
46. Gong Y, Shang B, Tan J, Luo C, Bian Z, Wu X, et al. Core and bridge symptoms of demoralization in Chinese female cancer patients: a network analysis. Front Psychiatry 2024 Feb 5;15:1273411 [FREE Full text] [doi: 10.3389/fpsyt.2024.1273411] [Medline: 38374974]
47. Renna ME, Shrout MR, Madison AA, Lustberg M, Povoski SP, Agnese DM, et al. Distress disorder histories relate to greater physical symptoms among breast cancer patients and survivors: findings across the cancer trajectory. Int J Behav Med 2023 Aug 13;30(4):463-472 [FREE Full text] [doi: 10.1007/s12529-022-10115-4] [Medline: 35831698]
48. Grassi L, Caruso R, Riba M, Lloyd-Williams M, Kissane D, Rodin G, ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org. Anxiety and depression in adult cancer patients: ESMO Clinical Practice Guideline. ESMO Open 2023 Apr;8(2):101155 [FREE Full text] [doi: 10.1016/j.esmoop.2023.101155] [Medline: 37087199]
49. Bower JE. Cancer-related fatigue--mechanisms, risk factors, and treatments. Nat Rev Clin Oncol 2014 Oct 12;11(10):597-609 [FREE Full text] [doi: 10.1038/nrclinonc.2014.127] [Medline: 25113839]
50. Saligan LN, Olson K, Filler K, Larkin D, Cramp F, Yennurajalingam S, Multinational Association of Supportive Care in Cancer Fatigue Study Group-Biomarker Working Group. The biology of cancer-related fatigue: a review of the literature. Support Care Cancer 2015 Aug 15;23(8):2461-2478 [FREE Full text] [doi: 10.1007/s00520-015-2763-0] [Medline: 25975676]
51. Whittaker AL, George RP, O'Malley L. Prevalence of cognitive impairment following chemotherapy treatment for breast cancer: a systematic review and meta-analysis. Sci Rep 2022 Feb 08;12(1):2135. [doi: 10.1038/s41598-022-05682-1] [Medline: 35136066]
52. Campbell KL, Zadravec K, Bland KA, Chesley E, Wolf F, Janelsins MC. The effect of exercise on cancer-related cognitive impairment and applications for physical therapy: systematic review of randomized controlled trials. Phys Ther 2020 Mar 10;100(3):523-542 [FREE Full text] [doi: 10.1093/ptj/pzz090] [Medline: 32065236]
53. Janelsins MC, Peppone LJ, Heckler CE, Kesler SR, Sprod LK, Atkins J, et al. YOCAS©® yoga reduces self-reported memory difficulty in cancer survivors in a nationwide randomized clinical trial. Integr Cancer Ther 2016 Jul 27;15(3):263-271. [doi: 10.1177/1534735415617021]
54. Forbes MK, Wright AGC, Markon KE, Krueger RF. Quantifying the reliability and replicability of psychopathology network characteristics. Multivariate Behav Res 2021 May 29;56(2):224-242 [FREE Full text] [doi: 10.1080/00273171.2019.1616526] [Medline: 31140875]

Abbreviations

CRP: C-reactive protein
IL: interleukin
MINORS: Methodological Index for Non-Randomized Studies
NA: network analysis
PICO: population, intervention, comparator, outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life

Edited by N Cahill; submitted 03.09.24; peer-reviewed by YY Kristian, J Drott; comments to author 20.12.24; revised version received 01.04.25; accepted 14.05.25; published 09.07.25.

Please cite as:
Richard V, Gilbert A, Pizzolla E, Briganti G
Investigating the Complexity of Multidimensional Symptom Experiences in Patients With Cancer: Systematic Review of the Network Analysis Approach
JMIR Cancer 2025;11:e66087
URL: https://cancer.jmir.org/2025/1/e66087
doi:10.2196/66087
PMID:

Chatbot for the Return of Positive Genetic Screening Results for Hereditary Cancer Syndromes: Prompt Engineering Project

Emma Coen[1], PhD; Guilherme Del Fiol[2], MD, PhD; Kimberly A Kaphingst[3], ScD; Emerson Borsato[2], PhD; Jackilen Shannon[4], PhD; Hadley Smith[5], PhD; Aaron Masino[1], PhD; Caitlin G Allen[6,7], MPH, PhD

[1]School of Computing, Clemson University, 105 Sikes Hall, Clemson, SC, United States

[2]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, United States

[3]Department of Communication, University of Utah, Salt Lake City, UT, United States

[4]Cancer Population Sciences, Div. Oncological Science, Oregon Health & Science University, Portland, OR, United States

[5]Department of Population Medicine, Harvard University, Cambridge, MA, United States

[6]Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, United States

[7]Department of Implementation Science, School of Medicine, Wake Forest University, 525 Vine St, Winston Salem, NC, United States

Corresponding Author:
Caitlin G Allen, MPH, PhD
Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, United States

Abstract

The increasing demand for population-wide genomic screening and the limited availability of genetic counseling resources have created a pressing need for innovative service delivery models. Chatbots powered by large language models (LLMs) have shown potential in genomic services, particularly in pretest counseling, but their application in returning positive population-wide genomic screening results remains underexplored. Leveraging advanced LLMs like GPT-4 offers an opportunity to address this gap by delivering accurate, contextual, and user-centered communication to individuals receiving positive genetic test results. This project aimed to design, implement, and evaluate a chatbot integrated with GPT-4, tailored to support the return of positive genomic screening results in the context of South Carolina's In Our DNA SC program. This initiative offers free genetic screening to 100,000 individuals, with over 33,000 results returned and numerous positive findings for conditions such as Lynch syndrome, hereditary breast and ovarian cancer syndrome, and familial hypercholesterolemia. A 3-step prompt engineering process using retrieval-augmented generation and few-shot techniques was used to create the chatbot. Training materials included patient frequently asked questions, genetic counseling scripts, and patient-derived queries. The chatbot underwent iterative refinement based on 13 training questions, while performance was evaluated through expert ratings on responses to 2 hypothetical patient scenarios. The 2 scenarios were intended to represent common but distinct patient profiles in terms of gender, race, ethnicity, age, and background knowledge. Domain experts rated the chatbot using a 5-point Likert scale across 8 predefined criteria: tone, clarity, program accuracy, domain accuracy, robustness, efficiency, boundaries, and usability. The chatbot achieved an average score of 3.86 (SD 0.89) across all evaluation metrics. The highest-rated criteria were tone (mean 4.25, SD 0.71) and usability (mean 4.25, SD 0.58), reflecting the chatbot's ability to communicate effectively and provide a seamless user experience. Boundary management (mean 4.0, SD 0.76) and efficiency (mean 3.88, SD 1.08) also scored well, while clarity and robustness received ratings of 3.81 (SD 1.05) and 3.81 (SD 0.66), respectively. Domain accuracy was rated 3.63 (SD 0.96), indicating satisfactory performance in delivering genetic information, whereas program accuracy received the lowest score of 3.25 (SD 1.39), highlighting the need for improvements in delivering program-specific details. This project demonstrates the feasibility of using LLM-powered chatbots to support the return of positive genomic screening results. The chatbot effectively handled open-ended patient queries, maintained conversational boundaries, and delivered user-friendly responses. However, enhancements in program-specific accuracy are essential to maximize its utility. Future research will explore hybrid chatbot designs that combine the strengths of LLMs with rule-based components to improve scalability, accuracy, and accessibility in genomic service delivery. The findings underscore the potential of generative artificial intelligence tools to address resource limitations and improve the accessibility of genomic health care services.

(JMIR Cancer 2025;11:e65848) doi:10.2196/65848

KEYWORDS

prompt engineering; few-shot learning; retrieval-augmented generation; population screening program; cancer; genetics; screening; syndrome; genomic; counseling; large language model; LLM; engineering; chatbot; prompt; RAG; mobile phone

Introduction

The increased demand for genomic testing, resulting growth in patient volume, and limited access to providers with genomic expertise has necessitated new, innovative genetic service delivery models [1-6]. Prior research has demonstrated the feasibility and acceptability of incorporating technologies such as chatbots to support common communication that occurs throughout the genomic service delivery process [7-10]. Chatbots are a highly accessible and scalable platform that allows for simulated conversations. Accessible via the web through a hyperlink or downloadable app, chatbots can be used on a smartphone, tablet, or computer. The use of chatbots has also been shown to improve access to services and support health equity by providing personalized health education, being available in multiple languages, and offering continuous access to information [11-15].

The integration of chatbots into routine and ancillary tasks such as pretest counseling education, informed consent, delivery of negative results, and cascade testing have been shown to be feasible and effective in supporting genomic service delivery [8,16]. For example, chatbots have been used to collect family health history, provide pretest support, communicate with family members about results, and obtain consent for genomic research [8,17-19]. Prior results from the BRIDGE (Broadening the Reach, Impact, and Delivery of Genetic Services) trial showed equivalence between a technology-based chatbot approach and standard of care in the completion of pretest genetics education and completion of genetic testing among unaffected primary care patients meeting criteria for cancer genetic evaluation [20]. Additional research in other health service delivery contexts has found that patients using chatbots reported a better understanding of their condition or procedure, being more prepared for upcoming appointments, and feeling more informed when making health care decisions [21-29].

To date, the integration of chatbot technology into genomic service delivery has yet to focus on the return of positive genetic test results directly to patients. Currently, the return of positive results has been carried out largely through direct communication, due to the complex and sensitive nature of the information, the potential psychological impact of learning about genetic predisposition, and the need to ensure understanding of the results and their implications. However, nonchatbot technology-based solutions, such as digital patient portals, are available to communicate with patients about these results and have been shown to be highly acceptable and preferred in genomics research [8,10,16,30-35]. Furthermore, a large-scale study across 3 academic medical centers found that individuals preferred laboratory test results to be delivered immediately digitally [30].

Prior qualitative data have indicated that patients are favorable toward receiving results via chatbots, as they are convenient and allow for the opportunity to contemplate information and ask questions [8]. Digital health communication approaches, such as chatbots, may be especially appropriate for the disclosure of population-wide genomic screening (PGS) results. PGS is often conducted on a large scale, targeting asymptomatic individuals as part of public health initiatives. As a result, the communication typically emphasizes general risk awareness, with initial results disclosure indicating increased risk rather than confirming a diagnosis. The Consent and Disclosure of Recommendations workgroup funded by the National Cancer Institute's Clinical Genome Resource (ClinGen) recommends considering factors such as test complexity, testing situation complexity, implications of genetic diagnosis to the patient and family, evidence of potential adverse psychological impact, and availability of high-quality and patient-friendly materials when deciding on the level of interaction with the patient [36,37]. Since PGS is typically completed through research and consent from participants and individuals are receiving results for well-defined hereditary conditions, the necessary level of initial communication about positive PGS results is lower than more complex, clinical results.

While high levels of acceptability, usability, and understanding of chatbots have been found in prior research, the majority of chatbots developed to date are rule-based, meaning that they operate on a set of predefined navigation paths with predefined scripted options and responses [8,9,19]. This approach allows for reliability and consistency in managing response options. However, user testing of rule-based chatbots has also revealed a need for chatbots that allow users to ask open-ended questions and receive responses in real time [8,9,19]. More recently, the release of large language models (LLMs) such as ChatGPT offers an opportunity to direct open-ended questions to LLMs to better support the return of positive genetic testing results, as open-ended questions allow for more nuanced and personalized responses. However, it is critical to test such systems to ensure that patients would receive accurate and clear information. Indeed, creating a hybrid chatbot with both rule-based and LLM components can offer a versatile and streamlined user experience by ensuring that key information is covered in the rule-based components of the chatbot and allowing for the LLM component to support complex, open-ended queries that are not covered in the scripted content. The objectives of the present project were to (1) prompt engineer an LLM-based chatbot focused on answering questions about the return of positive PGS results, and (2) conduct an intrinsic evaluation of the prompt engineering approach based on hypothetical cases and expert raters. This viewpoint paper offers insight into the application of specific prompt engineering methods to create patient-facing chatbots in the hereditary cancer diagnostic process.

Methods

Project Setting

We trained this chatbot using prompt engineering for the context of answering questions about the return of PGS results for an ongoing PGS program being delivered at the Medical University of South Carolina (MUSC). The PGS program was established in November 2021 with a focus on providing free genetic screening to 100,000 individuals in South Carolina. At the time of analysis, the program has recruited 59,352 individuals, returned 33,142 results, and identified 132 individuals with Lynch syndrome, 265 individuals with hereditary breast and

ovarian cancer syndrome, and 191 individuals with familial hypercholesterolemia.

Prompt Engineering Approach for Open-Ended Content

Overview

LLM models have been applied to improve accuracy and standardization for a variety of biomedical tasks including medical guidelines retrieval, diagnostics, medical reporting, and medical education [38-40]. The LLM selected depends on the task at hand, with a variety of LLMs developed for specific medical tasks and specialties [41]. Commonly used LLMs include ChatGPT, Perplexity AI, Claude AI, and Google Bard [42]. Developing generative artificial intelligence (AI) standards emphasizes the need to design generative AI tools responsibly for user mental models and build trust while allowing for generative variability, cocreation, and imperfection [43]. Meeting these standards requires effective prompt engineering, which includes the process of developing the text that instructs the LLM to complete a given task [44].

We used a 3-step prompt using the retrieval-augmented generation (RAG) technique which integrates retrieval-based methods with generative models, enabling the generation of contextually informed responses by retrieving relevant knowledge from a large corpus and incorporating it into the output generation process. RAG has been shown to improve LLM model performances by incorporating external information as a domain-specific knowledge base [45,46]. This project used OpenAI's GPT Version 4-Turbo-Preview model, as new research has indicated GPT version 4 performs significantly better at answering genetics questions than version 3.5 [43,44,47]. OpenAI's Playground was used for prompt engineering and testing. GPT4 was trained to respond about a variety of topics including providing examples of the impact of positive results, screening recommendations, and family history and cascade testing resources, and providing details regarding genetic counseling and specific PGS programs. Boundaries were also provided to ensure GPT4 responses remained within the intended scope of the chatbot.

Step 1: Provide Content and Context to GPT4

We used the RAG technique for prompt development. The RAG approach consisted of providing supplementary materials that were uploaded through OpenAI's Playground "File Search" function which allows GPT4 to access the additional information in real time when responding to users' questions. The additional files uploaded were: (1) detailed descriptions and frequently asked questions from the MUSC's PGS website; (2) MUSC Genetic Counseling Scripts: standard scripts used by genetic counselors at MUSC, providing insights into professional communication and common queries; and (3) Genome Medical Genetic Counseling Scripts: scripts from Genome Medical to offer additional perspectives. These documents expanded the model's knowledge base to increase the detail, consistency, and accuracy of responses. The team observed an improvement in the chatbot's replies after including these documents based on the established evaluation criteria.

Step 2: Establish a Bank of Commonly Asked Questions

To train and test the LLM, a bank of commonly asked questions was developed. This bank of questions was derived from patient quality improvement interviews and expert input. This step ensures that the model is trained on a wide array of realistic and relevant scenarios, enabling it to provide accurate and helpful responses. The list of 27 questions was randomly divided into 13 training questions and 14 evaluation questions (Multimedia Appendix 1).

Step 3: Develop and Refine Prompts

The core of prompt engineering involves creating and refining prompts that train the AI model to elicit the most accurate and appropriate responses. The prompt development process used OpenAI GPT assistants to develop an initial draft prompt. The prompt aimed to not only inform the chatbot about the situational context and content to be discussed but also about the writing style and limitations it should adhere to. We completed iterative testing by inputting the prompt as the instructions for the AI assistant and running the 13 training questions through the messaging feature. Adjustments were made to the initial prompt until the chatbot answers were deemed accurate, clear, and appropriate by our internal team. This process is subject to the bias of the team. However, the team was careful to evaluate the chatbot responses strictly based on the evaluation criteria and quality of responses to the test questions. The prompt indicated to the LLM that patient cases would be provided as input.

Prompt Engineering Evaluation

Overview

After completing the prompt engineering of our LLM chatbot, we conducted an intrinsic evaluation based on 2 hypothetical cases that were presented to domain experts in clinical genomics. The evaluation consisted of 2 steps described below.

Step 1: Establish the Prompt Evaluation Criteria

Previous literature has indicated relevant criteria to consider for chatbots in health communication [48]. Considering this previous work, we established relevant evaluation criteria tailored to this project through discussion and consensus among the team (Table 1). Based on 8 criteria, an evaluation instrument was developed in REDCap (Research Electronic Data Capture; Vanderbilt University) consisting of the 8 criteria, their definitions, and the ability to rate each criterion using a 5-point Likert scale from 1=Very Poor to 5=Excellent. Because prompt engineering in this context is a relatively new field, these criteria were optimized as much as possible with limited precedent.

Table . Evaluation criteria.

Criteria	Quality definition
Tone	The ability of the chatbot to express information in a way that is appropriate for the type of information being delivered
Clarity	The ability of the chatbot to communicate information clearly and in a way that avoids ambiguity or confusion
Program accuracy	The ability of the chatbot to provide correct information about the PGS[a] program
Domain accuracy	The ability of the chatbot to provide correct information about the genetic test results and care implications
Robustness	Ability to handle ambiguous queries or incomplete information
Efficiency	Ability to provide answers that are direct, concise, and complete
Boundaries	Ability to avoid answering questions that are unrelated to the topic
Usability	Ease of interfacing with the chatbot

[a]PGS: population-wide genomic screening.

Step 2: Development of Case Scenarios and Expert Ratings

We developed 2 hypothetical scenarios focused on returning results to individuals who participated in PGS. We used scenarios to allow the expert raters to view the chatbot from the perspective of a specific hypothetical patient (Textbox 1).

Textbox 1. Case scenarios.

Case 1: Anna

- A 34-year-old White woman named Anna participated in the In Our DNA SC population-wide genomic screening program in November. She had no personal or family history of cancer. She saw a digital advertisement and decided to participate in the program out of curiosity about her health. She had previously completed a direct-to-consumer genetic screen through 23andMe but did not receive any concerning results. A few months passed after she provided her sample. Her results showed that she has a BRCA1 mutation. Anna was shocked to find this news about her BRCA1 mutation and is concerned about the next steps in her care.

Case 2: Josh

- A 56-year-old Black man named Josh previously had colorectal cancer. His sister participated in In Our DNA SC based on a recommendation from her doctor. Josh's sister convinced him that it would be beneficial for him to participate. He reluctantly agreed to participate but did not see much of a point since he had already been treated for colorectal cancer. Josh received results showing that he was positive for the MLH1 gene, which means he has Lynch syndrome. Josh was confused by these results. He did not understand how these results would be important to him since he already had cancer.

Results

Final Prompt and Example Responses

The final prompt used for the LLM is included in Multimedia Appendix 2. Instructions for the prompt included a description of the type of patients the chatbot would be interfacing with (eg, patients who have done genetic testing as part of MUSC's population genomics program and are receiving their results digitally), the types of questions the chatbot may receive (eg, patient questions may range from understanding their specific probability of developing cancer to inquiries about insurance coverage), and boundaries (eg, do not provide any kind of medical advice […] if the patient asks questions outside of your boundaries, reply with "I am unable to answer questions unrelated to genetic testing or helping you get to your first appointment with a genetic counselor."). Other elements of the prompt included the style of the response and literacy level (eg, you will be replying to adults, but use a fifth-grade level vocabulary. Provide a clear, direct, concise response, preferably in bullet point format). Examples of responses provided by the chatbot are included in Table 2.

Table . Sample questions and LLM[a] responses.

Question	LLM response
What is the cost of genetic counseling visits and the screening?	I'm here to help, but remember, I can't give you specific costs because they can vary a lot depending on several factors. Here's what you should know about the costs of genetic counseling visits and screenings: • Insurance coverage: Many health insurance plans cover genetic counseling and screening, especially if you have a higher risk for certain conditions. However, coverage can vary, so it's important to check with your insurance provider. • Genetic counseling costs: The cost of a genetic counseling session can vary based on the location, the facility, and whether the counselor is part of a hospital or a private clinic. Some genetic counselors may work with you on payment options if you are paying out of pocket. • Screening costs: Like counseling, the cost for genetic screenings can vary widely. It depends on the type of screening and whether it's covered by insurance. • Financial assistance programs: Some hospitals, clinics, and organizations offer financial assistance or programs that might help cover the cost of genetic testing and counseling for those who qualify. It's a good idea to: • Contact your insurance company to find out what they cover and if you need a referral. • Ask the genetic counseling service about costs and payment options when you schedule your appointment.

What is the recommended treatment or preventative care?	For individuals with an MLH1 mutation, recommended treatment and preventive care focus on lowering the risk of cancer and catching any signs of cancer early. While I can't give personal medical advice, here are some general recommendations:
	• Regular screenings:
	  • Colonoscopy: Starting at a younger age and happening more often than for the general population, usually every 1 to 2 years.
	  • Physical exams: Including checking for changes in your body that might indicate cancer.
	• Lifestyle changes: Eating a healthy diet.
	  • Staying active.
	  • Avoiding smoking.
	• Family planning discussions: Since this mutation can be passed to your children, talking with a genetic counselor about family planning options is helpful.
	• Consider surgery: In some cases, to prevent cancer, doctors might suggest surgery to remove certain parts of the body that are at high risk of developing cancer. This is a personal decision and requires a detailed discussion with your health care team.
	• Stay informed: New recommendations and treatments can become available. Keep in regular contact with your health care team.
	
	A genetic counselor can provide personalized information based on your specific situation and help you understand your options for prevention and early detection. It's important to discuss the best plan for you with your health care provider or genetic counselor.
	
	Remember, these are general recommendations, and the best approach for you should be discussed with your health care provider. If you need help scheduling an appointment with a genetic counselor at the MUSC[b] Hollings Cancer Center Hereditary Cancer Clinic, you can visit their website or call 843-792-0745 for more personalized support.

Question	LLM response
How can I stay up to date on the latest treatments and science about my pathogenic variant?	Staying informed about the latest treatments and science related to your MLH1 mutation is a great way to actively participate in your health care. Here are some ways to keep updated: • Follow reputable sources: Websites of major cancer centers, genetic organizations, and health institutions often share updates about new research and treatments. Some reliable sources include the National Cancer Institute, American Cancer Society, and the Genetic Alliance. • Attend appointments: Regular check-ins with your health care team, including your genetic counselor, can keep you informed on any new recommendations or treatments relevant to your situation. • Join support groups: There are support groups and communities for individuals with hereditary cancer risk. These can be great places to share information and experiences. • Subscribe to newsletters: Many organizations offer newsletters that summarize the latest research and developments in cancer treatment and genetics. • Educational events: Workshops, webinars, and conferences designed for patients and families can provide valuable information and updates. • Ask questions: Whenever you have an appointment with your health care provider or genetic counselor, ask if there are any new findings or treatments relevant to your mutation. Remember, while it's great to stay informed, always discuss any information or potential treatments you come across with your health care provider before making any changes to your health care plan. They can help interpret how new information may apply to your specific situation. If you're looking for specific resources or have more questions, feel free to ask.

[a]LLM: large language model.

[b]MUSC: Medical University of South Carolina.

Prompt Engineering Evaluation

The expert ratings of chatbot responses are provided in Table 3. Overall, the chatbot received an average score of 3.86 (SD 0.89) across all criteria and cases. The ability to express information (tone) and ease of interface (usability) received the highest scores (4.25, SD 0.68 and 4.25, SD 0.58, respectively). Boundary was rated as 4.0 (SD 0.73) followed by efficiency (3.88, SD 1.09). Clarity and robustness received scores of 3.81 (SD 1.05) and 3.81 (SD 0.66), respectively, followed by domain accuracy (3.63, SD 0.96). The lowest-rated domain was program accuracy (3.25, SD 1.39).

Table . Expert ratings for each case and combined.

Quality	Quality definition	Case 1: Anna			Case 2: Josh			Combined		
		Median (IQR)	Mean (SD)	Range	Median (IQR)	Mean (SD)	Range	Median (IQR)	Mean (SD)	Range
Tone	Ability of chatbot to express information in a way that is appropriate for the type of information being delivered	4 (4-5)	4.25 (0.71)	3 - 5	4 (4-5)	4.25 (0.71)	3 - 5	4 (4-5)	4.25 (0.68)	3 - 5
Clarity	Ability of chatbot to communicate information clearly and in a way that avoids ambiguity or confusion	4 (3-5)	3.88 (1.1)	2 - 5	4 (3-4.5)	3.75 (1.0)	2 - 5	4 (3-5)	3.81 (1.05)	2 - 5
Program accuracy	Ability of chatbot to provide correct information about the In Our DNA SC program	3.5 (2-4.5)	3.25 (1.58)	1 - 5	3.5 (2.5-4)	3.25 (1.28)	1 - 5	3.5 (2.5-4)	3.25 (1.39)	1 - 5
Domain accuracy	Ability of chatbot to provide correct information about the genetic test results and care implications	4 (4-4)	3.88 (0.83)	2 - 5	4 (3-4)	3.38 (1.06)	1 - 4	4 (3.5-4)	3.63 (0.96)	1 - 5
Robustness	Ability to handle ambiguous queries or incomplete information	4 (3-4)	3.75 (0.71)	3 - 5	4 (3.5-4)	3.88 (0.64)	3 - 5	4 (3-4)	3.81 (0.66)	3 - 5
Efficiency	Ability to provide answers that are direct, concise, and complete	4 (3-5)	4 (1.07)	3 - 5	3.5 (3-5)	3.75 (1.16)	2 - 5	3.5 (3-5)	3.88 (1.09)	2 - 5

Quality	Quality definition	Case 1: Anna			Case 2: Josh			Combined		
		Median (IQR)	Mean (SD)	Range	Median (IQR)	Mean (SD)	Range	Median (IQR)	Mean (SD)	Range
Boundaries	Ability to avoid answering questions that are unrelated to the topic	4 (3.5-4.5)	4 (0.76)	3 - 5	4 (3.5-4.5)	4 (0.76)	3 - 5	4 (3.5-4.5)	4 (0.73)	3 - 5
Usability	Ease of interfacing with the chatbot	4 (4-5)	4.38 (0.52)	4 - 5	4 (4-4.5)	4.13 (0.64)	3 - 5	4 (4-5)	4.25 (0.58)	3 - 5
Average scores	—[a]	3.92 (3-5)	3.94 (0.92)	1 - 5	3.80 (3-4)	3.88 (0.91)	1 - 5	3.88 (3-5)	3.86 (0.89)	1 - 5

[a]Not applicable.

We provided the 2 case scenarios, the test questions, and answers the chatbot had provided to those questions and were asked to rate the quality of the chatbot responses based on the designated criteria listed in Table 1. The experts independently evaluated, scored, and submitted their scores to the team. The 2 scenarios were selected to represent 2 common patient profiles that differed in age, race, gender, and background. The evaluators were aware that the responses were generated by an LLM. Eight experts completed the evaluation of the LLM output for the 2 hypothetical scenarios (Konstantinos N. Lazaridis, Libby Malphrus, Samantha Norman, Ravi Sharaf, JS, HS, Sarah English, and Anne Madeo). Experts included: 2 clinician-researchers with expertise in genomics, one genetic counselor, 3 program managers working with genomic screening programs, and 2 PhD-trained researchers with expertise in genomics. Experts were recruited based on their domain-specific knowledge and experience to provide a comprehensive evaluation of the chatbot. Descriptive statistics were calculated, including median and mean scores for each evaluation criterion.

Discussion

Principal Findings

We completed prompt engineering and intrinsic evaluation of the LLM component of a chatbot designed to facilitate the return of positive PGS results. Through the RAG technique, we successfully developed a prompt tailored for this application. Eight experts performed an intrinsic evaluation, which assessed the chatbot's responses to 14 questions across 8 distinct domains in 2 hypothetical case scenarios. The chatbot achieved an overall average score of 3.88 across all domains, with the highest ratings in the tone domain and the lowest in program accuracy. These findings will inform further refinement of the prompt and integration of the LLM with the existing rule-based system, ultimately leading to the development of a hybrid chatbot to support the return of genomic screening results. As indicated by the range of scores, there was some disagreement among raters regarding the chatbot's performance.

Comparison to Prior Work

Prior studies have indicated that individuals are favorable toward the use of chatbots for patient follow-up and genetic test results disclosure, with a preference to include open-ended response options [8]. However, to date, few chatbots have incorporated LLMs to answer open-ended responses to questions about genetic testing in real time [8,19]. LLM responses must be carefully engineered to ensure confidence in the accuracy and reliability of responses, as well as the ability to handle ambiguous questions [49]. Our prompt engineering process resulted in a chatbot that performed well in the criteria of boundaries (ability to avoid answering questions that are unrelated to the topic), domain accuracy (ability of chatbot to provide correct information about the genetic test result and care implications), and robustness (ability to handle ambiguous queries or incomplete information). Another project focused on generative AI solutions for personalized pharmacogenomics recently identified similar trends. Prior research indicated found that the accuracy (the degree to which the responses align with guidelines) of their chatbot was rated at the 75th percentile and relevance (similar to our criteria of boundaries) was rated at the 78th percentile for patient-facing responses delivered by their chatbot [50]. These significant differences in performance metrics for these domains across responses provided by ChatGPT 3.5 and their pharmacogenomics-specific AI assistant (71st percentile vs 75th percentile for accuracy and 68th percentile vs 78th percentile for relevancy) indicate the value in prompt engineering for specific use cases. Challenges exist in ensuring domain accuracy and boundaries, such as limitations in LLM's context retrieval and ability to process specialized biomedical and genomic data [51,52].

The combination of high domain accuracy and boundaries is essential for managing sensitive health information and mitigates concerns about chatbots offering misinformation and medical advice beyond the scope of the chatbot. As the LLM is further refined, it will be important to document all steps of the prompt engineering and be clear and transparent about the prompt engineering process used to develop the model in order to instill trust in the quality of responses and reduce the risk of

misinformation [49]. It will also be critical to involve patient stakeholders in the future evaluation process. Other approaches to prompt development and evaluation include the involvement of experts (genetic counselors, oncologists) to help identify unintentional sources of bias and decide on high-quality data sources that can be used to train the model [53]. Furthermore, given that the evaluation process included only a limited set of test questions, the inclusion of a more comprehensive question set could provide additional insight into the chatbot's performance and ensure its ability to manage a greater set of user interactions. For example, our testing included 14 questions, whereas other projects have included over 30 questions [50]. In particular, future studies should incorporate adversarial examples in both engineering and testing, especially to more comprehensively test the model accuracy and boundaries [53].

In addition to domain accuracy and boundaries, it is critical to ensure open-ended, LLM-generated responses are delivered in a tone that instills trust and engagement with the individual. Expert ratings indicated that the chatbot had good quality tone (ability to express information in a way that is appropriate for the type of information being delivered), usability (ease of interfacing with the chatbot), efficiency (ability to answer in a way that is direct, concise, and complete), and clarity (ability to communicate information clearly and in a way that avoids confusion) in both case scenarios. Prior research assessed a similar domain of language and bias (clarity and neutrality of responses, ensuring the context is understandable and devoid of bias), which was rated highly (87th percentile) [50].

Lessons Learned

Our prompt engineering approach incorporated multiple techniques to develop an LLM chatbot that was well-rated across several quality domains. Several valuable lessons were learned. We used RAG as our approach to prompt development, but other techniques such as few-shot, supervised fine-tuning, and reinforcement learning from human feedback could be used to further adjust the model's responses [45]. In addition, we focus on a use case of returning positive results for PGS, as PGS results return is among the least complex types of results being disclosed and could benefit from incorporating automation. Limitations of the project include our small sample size for the intrinsic evaluation of the chatbot responses and the lack of patients reviewing the responses. The reviewers are subject to bias when considering the perspective of the hypothetical scenarios which does limit the reliability of their scores.

Future Directions

At this phase of the project, our goal was to develop the initial prompt and assess the feasibility of the prompt to respond to questions about the return of results. Thus, we did not include patients but will include patient perspectives and ratings of the quality of responses in future refinement of the LLM. Patients may identify areas for improvement that are not apparent to expert reviewers. Further, we only evaluate the script produced by the LLM component of the chatbot across 2 use cases. Additional use cases should be assessed (eg, other genes) to identify whether one prompt can be used or whether multiple prompts need to be developed for specific open-ended components of a hybrid chatbot. Finally, our assessment is only focused on the LLM component of the chatbot. Our future work will integrate the LLM component with the rule-based script, allowing us to assess different hybrid approaches. For example, we could address whether open-response options should be available as part of each component of the chatbot, which may require specific prompts for each component, or whether the open-response LLM component is generic.

While the final prompt delivered relatively high-quality responses in an appropriate tone, it is important to note that we did not assess perceptions of the quality of delivery among patients. Many chatbots have been designed to support mental health and behavior change modifications and are explicitly focused on building relationships and natural language experience for genomics-focused chatbots, and this is an important aspect of communication that will need to be evaluated before implementing a similar chatbot [54,55]. Furthermore, we tested the responses for hypothetical scenarios returning Lynch syndrome pathogenic variant (MLH1) and hereditary breast and ovarian cancer syndrome (BRCA) results. There may be a need to further refine and test response quality and tone across specific genes, as each has unique implications and may require distinct prompts. User testing among patients will also help address potential adaptations needed to ensure culturally appropriate responses [56].

Our long-term goal is to incorporate the LLM component of the chatbot described here with an existing rule-based chatbot called Genetic Risk Assessment for Cancer Education. This hybrid approach could be ideal for the return of positive PGS results, as it integrates scripted content that is critical for results disclosure with patient preference for open-ended response options. The combined approach can address the limitations of purely rule-based or purely LLM-driven systems to combine consistency and accuracy with conversational fluidity and content comprehensiveness. Some information may be more suitable for rule-based or scripted content. For example, in our intrinsic evaluation, the LLM chatbot received poor scores for program accuracy (ability of chatbot to provide correct information about the genomic screening program). The program accuracy referred to the ability of the chatbot to provide factually correct information about the specific program that patients were engaged in through this testing process. Although provided materials about the specific program were included as part of prompt engineering, experts rated this lowest among the domains they evaluated. This may indicate that additional contextual knowledge is required to sufficiently explain the complexities of individual programs. This type of information does not require personalization and may be most suited for prescripted, educational content, whereas the LLM components are most suitable for complex and open-ended questions and more nuanced interactions [49]. This additional personalization may make education more accessible and streamlined for patients seeking genetic care, potentially increasing their participation. As a result, improving the program accuracy score is an important future research topic.

One hybrid approach could incorporate a scripted component that provides a predetermined set of information, followed by an LLM component that is engineered specifically to support open-ended questions about a certain domain (Multimedia

Appendix 3). This may include key domains of: overview of the PGS program, returning positive results, screening recommendations, impact on family, and next steps.

Another hybrid approach could vary when the LLM or rule-based components are used throughout the chatbot. For example, the return of results process involves 3 main stages: engagement, activation, and addressing information needs. In the engagement stage, the rule-based component of the chatbot would provide an overview of the PGS program, inform the individual of their positive results, and educate the individual about what this means for their long-term care. The activation phase could also use rule-based content and guide individuals through a core set of scripted information to encourage the next steps. In the subsequent open-ended content, participants' information needs could be addressed by allowing them to ask additional questions about topics they choose, which could be answered through the LLM. This hybrid approach has benefits and drawbacks [57]. While the increased efficiency of resources and centralized communication are benefits of implementing the technology, the technology can introduce new types of errors, have biases of their own, and be perceived as less personable.

Conclusions

This project demonstrated the initial feasibility of prompt engineering for the LLM component of a chatbot designed to return positive genomic screening results, with high expert ratings across most of the evaluation criteria. These preliminary findings will be used to further develop a hybrid chatbot that integrates the rule-based and LLM components to enhance the delivery of results by providing essential information with the flexibility of managing a range of patient queries. This increased efficiency has the potential to save health care systems financial and time resources. Additionally, hybrid AI tools such as these offer the potential to support patients' decision-making and improve their education and health behaviors. Further refinements of the prompt are needed, as well as broad user-testing that involves individuals with various genomic conditions and cultural preferences, and testing of the best integration of LLM and rule-based components of the chatbot. This new approach to conveying positive genetic screening results has promise and can help address the limitations of the current genomic workforce that would be needed for the return of all positive results in a population genomic screening context.

Acknowledgments

This study was supported by the National Cancer Institute (grant U24CA274582).

Data Availability

All data generated or analyzed in this study are presented in the main manuscript.

Authors' Contributions

EC contributed to methodology, formal analysis, original draft writing, and project administration. GDF was responsible for conceptualization, supervision, review and editing of the manuscript, and funding acquisition. KAK contributed to conceptualization, methodology, supervision, and review and editing. EB, JS, HS, and AM contributed to review and editing. CGA was involved in conceptualization, supervision, investigation, original draft writing, and funding acquisition.

Conflicts of Interest

HS received consulting income from Illumina, Inc, unrelated to this work.

Multimedia Appendix 1
Training and test questions.
[DOCX File, 21 KB - cancer_v11i1e65848_app1.docx]

Multimedia Appendix 2
Prompt content.
[DOCX File, 7 KB - cancer_v11i1e65848_app2.docx]

Multimedia Appendix 3
Description of PGS chat content. PGS: population-wide genomic screening.
[DOCX File, 15 KB - cancer_v11i1e65848_app3.docx]

References

1. Foss KS, O'Daniel JM, Berg JS, et al. The rise of population genomic screening: characteristics of current programs and the need for evidence regarding optimal implementation. J Pers Med 2022 Apr 26;12(5):692. [doi: 10.3390/jpm12050692] [Medline: 35629115]

2. Murray MF, Giovanni MA, Doyle DL, et al. DNA-based screening and population health: a points to consider statement for programs and sponsoring organizations from the American College of Medical Genetics and Genomics (ACMG). Genet Med 2021 Jun;23(6):989-995. [doi: 10.1038/s41436-020-01082-w] [Medline: 33727704]
3. Khoury MJ, Dotson WD. From genes to public health: are we ready for DNA-based population screening? Genet Med 2021 Jun;23(6):996-998. [doi: 10.1038/s41436-021-01141-w] [Medline: 33790422]
4. Hoskovec JM, Bennett RL, Carey ME, et al. Projecting the supply and demand for certified genetic counselors: a workforce study. J Genet Couns 2018 Feb;27(1):16-20. [doi: 10.1007/s10897-017-0158-8] [Medline: 29052810]
5. Attard CA, Carmany EP, Trepanier AM. Genetic counselor workflow study: the times are they a-changin'? J Genet Couns 2019 Feb;28(1):130-140. [doi: 10.1002/jgc4.1041] [Medline: 30629774]
6. Maiese DR, Keehn A, Lyon M, Flannery D, Watson M, Working Groups of the National Coordinating Center for Seven Regional Genetics Service Collaboratives. Current conditions in medical genetics practice. Genet Med 2019 Aug;21(8):1874-1877. [doi: 10.1038/s41436-018-0417-6] [Medline: 30686822]
7. Siglen E, Vetti HH, Lunde ABF, et al. Ask Rosa—the making of a digital genetic conversation tool, a chatbot, about hereditary breast and ovarian cancer. Patient Educ Couns 2022 Jun;105(6):1488-1494. [doi: 10.1016/j.pec.2021.09.027] [Medline: 34649750]
8. Schmidlen T, Schwartz M, DiLoreto K, Kirchner HL, Sturm AC. Patient assessment of chatbots for the scalable delivery of genetic counseling. J Genet Couns 2019 Dec;28(6):1166-1177. [doi: 10.1002/jgc4.1169] [Medline: 31549758]
9. Schmidlen T, Jones CL, Campbell-Salome G, McCormick CZ, Vanenkevort E, Sturm AC. Use of a chatbot to increase uptake of cascade genetic testing. J Genet Couns 2022 Oct;31(5):1219-1230. [doi: 10.1002/jgc4.1592] [Medline: 35616645]
10. Chaix B, Bibault JE, Pienkowski A, et al. When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer 2019 May 2;5(1):e12856. [doi: 10.2196/12856] [Medline: 31045505]
11. Suther S, Kiros GE. Barriers to the use of genetic testing: a study of racial and ethnic disparities. Genet Med 2009 Sep;11(9):655-662. [doi: 10.1097/GIM.0b013e3181ab22aa] [Medline: 19752639]
12. Hall MJ, Olopade OI. Disparities in genetic testing: thinking outside the BRCA box. J Clin Oncol 2006 May 10;24(14):2197-2203. [doi: 10.1200/JCO.2006.05.5889] [Medline: 16682739]
13. Jones T, McCarthy AM, Kim Y, Armstrong K. Predictors of BRCA1/2 genetic testing among Black women with breast cancer: a population-based study. Cancer Med 2017 Jul;6(7):1787-1798. [doi: 10.1002/cam4.1120] [Medline: 28627138]
14. Reiter PL, Katz ML. Racial/ethnic differences in knowledge, attitudes, and beliefs about COVID-19 among adults in the United States. Front Public Health 2021;9:653498. [doi: 10.3389/fpubh.2021.653498] [Medline: 34046389]
15. Nadarzynski T, Knights N, Husbands D, et al. Achieving health equity through conversational AI: a roadmap for design and implementation of inclusive chatbots in healthcare. PLOS Digital Health 2024 May;3(5):e0000492. [doi: 10.1371/journal.pdig.0000492] [Medline: 38696359]
16. Heald B, Keel E, Marquard J, et al. Using chatbots to screen for heritable cancer syndromes in patients undergoing routine colonoscopy. J Med Genet 2021 Dec;58(12):807-814. [doi: 10.1136/jmedgenet-2020-107294] [Medline: 33168571]
17. Ponathil A, Ozkan F, Welch B, Bertrand J, Chalil Madathil K. Family health history collected by virtual conversational agents: an empirical study to investigate the efficacy of this approach. J Genet Couns 2020 Dec;29(6):1081-1092. [doi: 10.1002/jgc4.1239] [Medline: 32125052]
18. Ireland D, Bradford D, Szepe E, et al. Introducing Edna: a trainee chatbot designed to support communication about additional (secondary) genomic findings. Patient Educ Couns 2021 Apr;104(4):739-749. [doi: 10.1016/j.pec.2020.11.007] [Medline: 33234441]
19. Chavez-Yenter D, Kimball KE, Kohlmann W, et al. Patient interactions with an automated conversational agent delivering pretest genetics education: descriptive study. J Med Internet Res 2021 Nov 18;23(11):e29447. [doi: 10.2196/29447] [Medline: 34792472]
20. Kaphingst KA, Kohlmann WK, Lorenz Chambers R, et al. Uptake of Cancer Genetic Services for Chatbot vs Standard-of-Care Delivery Models: The BRIDGE Randomized Clinical Trial. JAMA network open 2024 Sep 3;7(9):e2432143. [doi: 10.1001/jamanetworkopen.2024.32143] [Medline: 39250153]
21. Milne-Ives M, de Cock C, Lim E, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346. [doi: 10.2196/20346] [Medline: 33090118]
22. Tudor Car L, Dhinagaran DA, Kyaw BM, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 7;22(8):e17158. [doi: 10.2196/17158] [Medline: 32763886]
23. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106. [doi: 10.2196/12106] [Medline: 30470676]
24. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19. [doi: 10.2196/mental.7785] [Medline: 28588005]
25. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64. [doi: 10.2196/mental.9782] [Medline: 30545815]

26. Harper R, Nicholl PN, McTear MF, Wallace JG, Black LA, Kearney PM. Automated phone capture of diabetes patients readings with consultant monitoring via the web. Presented at: 2008 15th Annual IEEE International Conference on Engineering of Computer Based Systems (ECBS); Mar 31 to Apr 4, 2008; Belfast, Northern Ireland. [doi: 10.1109/ECBS.2008.31]
27. Levin E, Levin A. Evaluation of spoken dialogue technology for real-time health data collection. J Med Internet Res 2006 Dec 11;8(4):e30. [doi: 10.2196/jmir.8.4.e30] [Medline: 17213048]
28. Rhee H, Allen J, Mammen J, Swift M. Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study. Patient Prefer Adherence 2014;8:63-72. [doi: 10.2147/PPA.S53504] [Medline: 24470755]
29. Wang A, Qian Z, Briggs L, Cole AP, Reis LO, Trinh QD. The use of chatbots in oncological care: a narrative review. Int J Gen Med 2023;16:1591-1602. [doi: 10.2147/IJGM.S408208] [Medline: 37152273]
30. Steitz BD, Turer RW, Lin CT, et al. Perspectives of patients about immediate access to test results through an online patient portal. JAMA Netw Open 2023 Mar 1;6(3):e233572. [doi: 10.1001/jamanetworkopen.2023.3572] [Medline: 36939703]
31. LaRocque JR, Davis CL, Tan TP, D'Amico FJ, Merenstein DJ. Patient preferences for receiving reports of test results. J Am Board Fam Med 2015 Nov 1;28(6):759-766. [doi: 10.3122/jabfm.2015.06.150030]
32. Shultz SK, Wu R, Matelski JJ, Lu X, Cram P. Patient preferences for test result notification. J Gen Intern Med 2015 Nov;30(11):1651-1656. [doi: 10.1007/s11606-015-3344-0] [Medline: 25944020]
33. Leekha S, Thomas KG, Chaudhry R, Thomas MR. Patient preferences for and satisfaction with methods of communicating test results in a primary care practice. Jt Comm J Qual Patient Saf 2009 Oct;35(10):497-501. [doi: 10.1016/s1553-7250(09)35068-0] [Medline: 19886088]
34. Grimes GC, Reis MD, Budati G, Gupta M, Forjuoh SN. Patient preferences and physician practices for laboratory test results notification. J Am Board Fam Med 2009;22(6):670-676. [doi: 10.3122/jabfm.2009.06.090078] [Medline: 19897696]
35. Bibault JE, Chaix B, Guillemassé A, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. J Med Internet Res 2019 Nov 27;21(11):e15787. [doi: 10.2196/15787] [Medline: 31774408]
36. Ormond KE, Hallquist MLG, Buchanan AH, et al. Developing a conceptual, reproducible, rubric-based approach to consent and result disclosure for genetic testing by clinicians with minimal genetics background. Genet Med 2019 Mar;21(3):727-735. [doi: 10.1038/s41436-018-0093-6] [Medline: 29976988]
37. Faucett WA, Peay H, Coughlin CR. Genetic testing: consent and result disclosure for primary care providers. Med Clin North Am 2019 Nov;103(6):967-976. [doi: 10.1016/j.mcna.2019.07.001] [Medline: 31582007]
38. Lozano A, Fleming SL, Chiang CC, Shah N. Clinfo.ai: an open-source retrieval-augmented large language model system for answering medical questions using scientific literature. Pac Symp Biocomput 2024;29:8-23. [Medline: 38160266]
39. Zakka C, Shad R, Chaurasia A, et al. Almanac—retrieval-augmented language models for clinical medicine. NEJM AI 2024 Feb;1(2). [doi: 10.1056/aioa2300068] [Medline: 38343631]
40. Sacoransky E, Kwan BYM, Soboleski D. ChatGPT and assistive AI in structured radiology reporting: a systematic review. Curr Probl Diagn Radiol 2024;53(6):728-737. [doi: 10.1067/j.cpradiol.2024.07.007] [Medline: 39004580]
41. Ray PP. Timely need for navigating the potential and downsides of LLMs in healthcare and biomedicine. Brief Bioinform 2024 Mar 27;25(3):bbae214. [doi: 10.1093/bib/bbae214] [Medline: 38725154]
42. Uppalapati VK, Nag DS. A comparative analysis of AI models in complex medical decision-making scenarios: evaluating ChatGPT, Claude AI, Bard, and Perplexity. Cureus 2024 Jan;16(1):e52485. [doi: 10.7759/cureus.52485] [Medline: 38371109]
43. Lahat A, Sharif K, Zoabi N, et al. Assessing generative pretrained transformers (GPT) in clinical decision-making: comparative analysis of GPT-3.5 and GPT-4. J Med Internet Res 2024 Jun 27;26:e54571. [doi: 10.2196/54571] [Medline: 38935937]
44. Taloni A, Borselli M, Scarsi V, et al. Comparative performance of humans versus GPT-4.0 and GPT-3.5 in the self-assessment program of American Academy of Ophthalmology. Sci Rep 2023 Oct 29;13(1):18562. [doi: 10.1038/s41598-023-45837-2] [Medline: 37899405]
45. Giuffrè M, Kresevic S, Pugliese N, You K, Shung DL. Optimizing large language models in digestive disease: strategies and challenges to improve clinical outcomes. Liver Int 2024 Sep;44(9):2114-2124. [doi: 10.1111/liv.15974] [Medline: 38819632]
46. Zhou Q, Liu C, Duan Y, et al. GastroBot: a Chinese gastrointestinal disease chatbot based on the retrieval-augmented generation. Front Med 2024;11:1392555. [doi: 10.3389/fmed.2024.1392555]
47. McGrath SP, Kozel BA, Gracefo S, Sutherland N, Danford CJ, Walton N. A comparative evaluation of ChatGPT 3.5 and ChatGPT 4 in responses to selected genetics questions. J Am Med Inform Assoc 2024 Oct 1;31(10):2271-2283. [doi: 10.1093/jamia/ocae128] [Medline: 38872284]
48. Denecke K, Abd-Alrazaq A, Househ M, Warren J. Evaluation metrics for health chatbots: a Delphi study. Methods Inf Med 2021 Dec;60(5-06):171-179. [doi: 10.1055/s-0041-1736664] [Medline: 34719011]
49. Sorin V, Barash Y, Konen E, Klang E. Large language models for oncological applications. J Cancer Res Clin Oncol 2023 Sep;149(11):9505-9508. [doi: 10.1007/s00432-023-04824-w] [Medline: 37160626]
50. Murugan M, Yuan B, Venner E, et al. Empowering personalized pharmacogenomics with generative AI solutions. J Am Med Inform Assoc 2024 May 20;31(6):1356-1366. [doi: 10.1093/jamia/ocae039] [Medline: 38447590]

51. Jin Q, Yang Y, Chen Q, Lu Z. GeneGPT: augmenting large language models with domain tools for improved access to biomedical information. Bioinformatics 2024 Feb 1;40(2):btae075. [doi: 10.1093/bioinformatics/btae075] [Medline: 38341654]
52. Mahbub M, Srinivasan S, Begoli E, Peterson GD. BioADAPT-MRC: adversarial learning-based domain adaptation improves biomedical machine reading comprehension task. Bioinformatics 2022 Sep 15;38(18):4369-4379. [doi: 10.1093/bioinformatics/btac508] [Medline: 35876792]
53. Sorin V, Barash Y, Konen E, Klang E. Deep-learning natural language processing for oncological applications. Lancet Oncol 2020 Dec;21(12):1553-1556. [doi: 10.1016/S1470-2045(20)30615-X] [Medline: 33271088]
54. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. J Med Internet Res 2020;22(9):e22845. [doi: 10.2196/22845]
55. Wilson L, Marasoiu M. The development and use of chatbots in public health: scoping review. JMIR Hum Factors 2022 Oct 5;9(4):e35882. [doi: 10.2196/35882] [Medline: 36197708]
56. Hu G, Liu L, Xu D. On the responsible use of chatbots in bioinformatics. Genomics Proteomics Bioinf 2024 May 9;22(1):qzae002. [doi: 10.1093/gpbjnl/qzae002] [Medline: 38862428]
57. Kerdvibulvech C, Chang CC. A new study of integration between social robotic systems and the metaverse for dealing with healthcare in the post-COVID-19 situations. In: Lecture Notes in Computer Science: Springer; 2022:392-401. [doi: 10.1007/978-3-031-24670-8_35]

Abbreviations

AI: artificial intelligence
BRIDGE: Broadening the Reach, Impact, and Delivery of Genetic Services
LLM: large language model
MUSC: Medical University of South Carolina
PGS: population-wide genomic screening
RAG: retrieval-augmented generation
REDCap: Research Electronic Data Capture

Edited by N Cahill; submitted 27.08.24; peer-reviewed by L Guo, T Siriborvornratanakul, Y Qiu; revised version received 17.03.25; accepted 21.03.25; published 10.06.25.

Please cite as:
Coen E, Del Fiol G, Kaphingst KA, Borsato E, Shannon J, Smith H, Masino A, Allen CG
Chatbot for the Return of Positive Genetic Screening Results for Hereditary Cancer Syndromes: Prompt Engineering Project
JMIR Cancer 2025;11:e65848
URL: https://cancer.jmir.org/2025/1/e65848
doi:10.2196/65848

Leveraging Artificial Intelligence for Digital Symptom Management in Oncology: The Development of CRCWeb

Darren Liu[1,2,3*], MS; Yufen Lin[2,4*], PhD; Runze Yan[2,3], PhD; Zhiyuan Wang[5], BE; Delgersuren Bold[2,3], MS; Xiao Hu[2,3,6], PhD

[1]Department of Computer Science, Laney Graduate School, Emory University, Atlanta, GA, United States

[2]Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Rd NE, Atlanta, GA, United States

[3]Center for Data Science, Emory University, Atlanta, GA, United States

[4]Winship Cancer Institute, Emory University, Atlanta, GA, United States

[5]Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA, United States

[6]The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, United States

[*]these authors contributed equally

Corresponding Author:
Xiao Hu, PhD
Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Rd NE, Atlanta, GA, United States

Abstract

Digital health interventions offer promise for scalable and accessible health care, but access is still limited by some participatory challenges, especially for disadvantaged families facing limited health literacy, language barriers, low income, or living in marginalized areas. These issues are particularly pronounced for patients with colorectal cancer (CRC), who often experience distressing symptoms and struggle with educational materials due to complex jargon, fatigue, or reading level mismatches. To address these issues, we developed and assessed the feasibility of a digital health platform, CRCWeb, to improve the accessibility of educational resources on symptom management for disadvantaged patients with CRC and their caregivers facing limited health literacy or low income. CRCWeb was developed through a stakeholder-centered participatory design approach. Two-phase semistructured interviews with patients, caregivers, and oncology experts informed the iterative design process. From the interviews, we developed the following 5 key design principles: user-friendly navigation, multimedia integration, concise and clear content, enhanced accessibility for individuals with vision and reading disabilities, and scalability for future content expansion. Initial feedback from iterative stakeholder engagements confirmed high user satisfaction, with participants rating CRCWeb an average of 3.98 out of 5 on the postintervention survey. Additionally, using generative artificial intelligence tools, including large language models like ChatGPT and multimedia generation tools such as Pictory, complex health care guidelines were transformed into concise, easily comprehensible multimedia content, and made accessible through CRCWeb. User engagement was notably higher among disadvantaged participants with limited health literacy or low income, who logged into the platform 2.52 times more frequently than nondisadvantaged participants. The structured development approach of CRCWeb demonstrates that generative artificial intelligence–powered multimedia interventions can effectively address health care accessibility barriers faced by disadvantaged patients with CRC and caregivers with limited health literacy or low income. This structured approach highlights how digital innovations can enhance health care.

International Registered Report Identifier (IRRID): RR2-10.2196/48499

(JMIR Cancer 2025;11:e68516) doi:10.2196/68516

KEYWORDS

colorectal cancer; health disparity; health equity; generative artificial intelligence; large language model; software engineering; artificial intelligence

Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related deaths in the United States [1]. For patients with CRC, educational materials play a critical role in understanding their diagnosis, exploring treatment options, self-managing their side effects and symptoms, and navigating posttreatment care, ultimately empowering them to make informed decisions about their health [2-4]. However, most of these materials are primarily text-based, which poses significant accessibility challenges for disadvantaged populations, such as individuals with low income or limited health literacy [5].

Caregivers, typically family members who provide emotional and physical support, frequently experience similar distressing symptoms, making it equally challenging for them to engage with complex medical content [6,7]. For disadvantaged populations with limited health literacy or low income, these barriers are even more pronounced, which reduces their ability to access and understand crucial health information [8]. Many educational materials are only available in English, further excluding non-English-speaking individuals from receiving critical guidance on symptom management and supportive care [9]. These factors combine to create a critical gap in the ability of patients with CRC and their caregivers to effectively manage symptoms and make informed decisions about their treatment, highlighting the urgent need for more accessible solutions.

Recent advancements in generative artificial intelligence (GenAI) offer a transformative solution to these challenges. GenAI can convert traditional text-based educational materials into multimedia formats at a fraction of the cost and time, making it an efficient and scalable option [10,11]. In this work, we introduce CRCWeb, a novel GenAI-driven digital health mobile platform designed to provide accessible, tailored symptom management resources to patients with CRC and their caregivers. Powered by state-of-the-art GenAI models like ChatGPT [12], CRCWeb transforms dense health care texts into digestible multimedia formats, including videos and audio, concise and easy-to-understand health knowledge, practical activity prompts, and quizzes. This approach makes essential health knowledge on symptom management and cancer care more accessible to individuals with CRC facing low literacy and other accessibility challenges [13]. By reducing the cognitive load required to process health information, CRCWeb empowers patients and caregivers with easy-to-understand materials, improving their abilities to manage symptoms and adhere to treatment plans and guidelines. Therefore, the purpose of this viewpoint is to describe the development of CRCWeb, which leverages GenAI for digital symptom management in patients with CRC and their caregivers, and present preliminary data to support its feasibility and potential to reduce barriers to accessing health information and improve user engagement, satisfaction, and symptom management.

Contextual Background

We used a stakeholder-centered participatory design approach to develop CRCWeb at the Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, Georgia, United States. The motivation for this project was to develop and test a symptom management tool tailored to the needs of disadvantaged patients with cancer, particularly those with limited health literacy or low income. The project was led by a multidisciplinary research team with expertise in oncology, nursing, AI, technology, psychology, behavioral science, and clinical trial design. The development and evaluation of CRCWeb took place between November 2022 and May 2025.

Designing CRCWeb

Overview

CRCWeb was developed through an iterative stakeholder-centered participatory design approach [14,15]. We conducted 2-phase, semistructured interviews using an interview guide developed from the existing literature that included open-ended questions and probes to elicit their specific needs, challenges, and expectations for a technology-based intervention tool for symptom management during chemotherapy treatment. Each interview lasted between 30 and 45 minutes and was conducted in either a private conference room at the clinic or virtually via Zoom (Zoom Video Communications, Inc), depending on the participant's preference. All interviews were audio-recorded and transcribed, and field notes were taken to document nonverbal cues. Content analysis was used to analyze the qualitative data through Dedoose (SocioCultural Research Consultants) and was completed in 4 steps: data preparation, writing memos, coding, and categorizing and connecting [16].

Semistructured Interviews: Phase 1

In Phase I, 11 patients with CRC, 8 caregivers, and 4 oncologists were asked about their perspectives and suggestions for a technology-based intervention to manage symptoms. We gathered their feedback before the platform's development. This early involvement of key users allowed us to ensure that the platform would be designed to address their specific needs, challenges, and expectations from the outset. During the interview, the participants were guided by the questions in Table 1. We started by discussing their needs for a technology-based intervention tool during chemotherapy (Question 1) and their experiences with existing technological tools for symptom management (Question 2), allowing us to identify key gaps in current digital solutions. Providers were also asked to offer suggestions for the development of this tool (Question 3).

Table . Two-phase semistructured interview questions designed for iterative stakeholder-centered participatory design[a].

Phase	Examples of interview questions
I	1. Can you describe your need for a technology-based intervention tool to manage symptoms during chemotherapy?
	2. Have you used any technological tools for symptom management? If so, which ones have you used, and what is your experience with them?
	3. Do you have any suggestions on developing a technology-based intervention program for patients and caregivers? What are they?
II	4. What are your likes and dislikes regarding the CRCWeb intervention components (eg, family involvement, symptom management, and coping strategies), delivery methods (eg, doses and intervals), and formats (eg, video, audio, and evaluations)?
	5. How easy or difficult was it for you to navigate and understand this technology tool?
	6. What challenges or barriers have you encountered when accessing cancer care and symptom management?
	7. In your opinion, how can we achieve health equity in cancer care?
	8. What are the facilitators and challenges related to implementing a technology-based intervention for patients and caregivers?

[a]Questions 1 to 3 were used in phase I to collect participants' needs for CRCWeb prior to development, while questions 4 to 8 were used in phase II to iteratively gather feedback from participants to propose new designs and functionalities.

Semistructured Interviews: Phase 2

In Phase II, we expanded the participant pool beyond the original 23 from Phase I to include a more diverse group: 5 additional patients and 5 caregivers, as well as a palliative care physician, a nurse practitioner, and a clinical leader. These participants were iteratively asked to provide feedback on CRCWeb's proposed content, including family involvement, symptom management, and coping strategies, as well as the delivery methods and formats (Question 4) and their user experiences (Question 5). Additionally, the interviews addressed topics such as barriers to accessing cancer care and symptom management (Question 6), participants' perspectives on achieving health equity (Question 7), and the facilitators and challenges associated with implementing the intervention (Question 8). After each iteration, we proposed new designs and functionalities based on participants' feedback. Given that over half of the participants that we interviewed came from disadvantaged backgrounds with limited health literacy or low income and were unfamiliar with technology, the iterative design process allowed them to actively engage with CRCWeb during its early development stages, ensuring that the approach aligned with their needs and preferences. These participants provided valuable insights into the specific needs that CRCWeb aimed to address, particularly the need for improved accessibility to health care resources and information.

Ethical Considerations

The study protocol (STUDY00004750) was approved by the institutional review board at Emory University. Written consent was obtained from participants. All participants were informed of the voluntary nature of their participation and their right to withdraw at any time without consequence. All research data were anonymized to maintain confidentiality. Study materials were securely stored and accessible only to authorized research team members. Participants received up to $60 compensation for their involvement. The study was conducted in accordance with the US Common Rule (45 CFR 46).

Design Principles

The design principles outlined in Table 2 were identified from our 2-phase semistructured interview transcripts using content analysis. To reduce the learning curve of using a new app, particularly for disadvantaged populations with limited health literacy or low income, we applied design principle 1, ensuring that CRCWeb includes intuitive and user-friendly navigation features. To increase participant engagement, we implemented design principle 2 to include multimedia components to enrich the content [17,18]. Design principle 3 ensures that the learning modules are concise, containing only essential information. This design principle minimizes learning time and makes the content easier for disadvantaged individuals with limited health literacy to comprehend. To assist individuals with vision impairments and reading disabilities, we designed the vision principle, optimizing CRCWeb to enhance accessibility for these users. Design principle 5 was implemented to ensure that the system is scalable, allowing for the inclusion of additional educational materials in various formats in the future without requiring changes to the system architecture. These qualitative data are invaluable in identifying key features that would drive the development of CRCWeb.

Table . Five design principles were outlined from the 2-phase semistructured interviews with patients and caregivers[a].

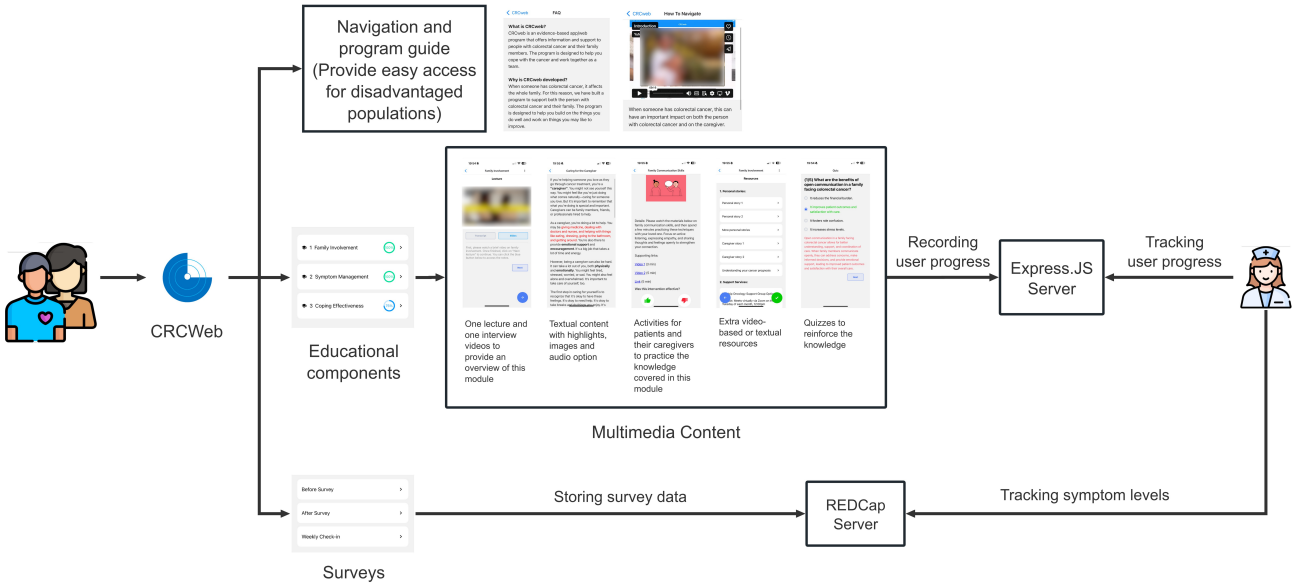
Design principle	Explanation
User-friendly principle	Our platform should feature intuitive and user-friendly navigation.
Multimedia principle	Our platform should feature multimedia components.
Concise principle	Extraneous material should be excluded to keep the content short and easy to understand.
Vision principle	Our app should feature functions that help people with limited vision to access educational content.
Scalability principle	Our platform should be scalable to include more topics and content formats in the future.

[a]These open-ended questions aim to address accessibility challenges for disadvantaged populations facing limited health literacy or low income by incorporating intuitive navigation, multimedia components, concise and easy-to-understand content, and functionalities designed to assist individuals with limited vision.

Development of CRCWeb

The development of CRCWeb strictly followed the design principles outlined in the previous section. As shown in Figure 1, CRCWeb contains 3 main components: navigation and program guide, educational materials, and surveys. The navigation and program guide simplify access while the educational components offer engaging, multimedia content tailored to different learning needs. The survey feature tracks symptom levels and user progress, with data securely stored on the REDCap (Research Electronic Data Capture; Vanderbilt University) server. Then, we discussed in detail how CRCWeb's design increased accessibility by focusing on these 3 key components.

Figure 1. System architecture of CRCWeb, highlighting its 3 core components: navigation and program guide, educational components, and surveys. The educational components feature 5 multimedia sections: lectures, content, activities, resources, and quizzes. User progress is securely stored on our server, ensuring confidentiality while enabling administrators to monitor and track advancement. REDCap: Research Electronic Data Capture.



Functionalities of CRCWeb

As shown in Figure 2, CRCWeb's main interface includes 4 navigation tabs at the bottom of the screen: Home, Content, Survey, and Account. This streamlined layout is crucial for enabling users to easily access the app's core functionalities without confusion. By providing a consistent and simple navigation system, CRCWeb ensures that users, especially those with limited technological proficiency, can effortlessly switch between essential features, aligning with design principle 1. As shown in Figure 2A, users can also quickly navigate between sections using a button located in the top right corner. Additionally, the large blue buttons at the bottom of the screen, as displayed in Figure 2C, allow users to move to the next or previous sections within a module. When a user reaches the final section, the right button turns green, indicating that they have completed the current module. To further assist users in tracking their progress, as demonstrated in Figure 2B, a green check mark appears next to the title of a section when it is finished, providing a visual reminder of completion. The Home tab includes essential resources such as "How to Navigate?" and "FAQ" sections, designed to guide users through the platform and answer common questions. These features are critical for users who may not be familiar with technology, ensuring that everyone can easily access and use CRCWeb's resources. Users can also send direct messages to the research team for additional support by tapping the message icon in the top right corner. This feature is particularly useful for

disadvantaged users with limited access to advanced technologies who may require additional help to navigate the platform or understand the content. As shown Figure 3B, the Content tab grants users access to educational modules, complete with a progress bar that visually tracks their completion status. This progress bar is essential in keeping users motivated and aware of their learning journey. The Survey tab, illustrated in Figure 2C, enables users to complete pre- and postintervention surveys, as well as weekly symptom check-ins. This structure allows CRCWeb to deliver notifications that adjust the learning experience based on user progress and survey responses. Finally, the Account tab, shown in Figure 2D, enables users to manage personal information and settings and pair their account with a caregiver. This paired account feature fosters collaboration, enabling patients and caregivers to share learning progress and better coordinate care efforts.

One of the key tools supporting this interactivity is the smart content tagging system, which enhances the readability and accessibility of text-based materials. The content tagging system allows critical information to be highlighted using different font weights, colors, and multimedia elements. For example, key points can be bolded or marked in red or blue to guide users' attention, as shown in Figure 3. In addition, users can choose to increase font size in settings. Driven by design principles 2 and 4, this system ensures that important information is easy to spot, particularly for users with low literacy or cognitive challenges. Additionally, for users with reading difficulties or visual impairments, we integrated a text-to-speech option that allows users to listen to the content instead of reading it. This feature is available throughout the app, ensuring that all users—regardless of their abilities—can engage with the educational materials. The system is designed to strip out tags before converting the text to speech, preventing unnecessary audio distractions. To reinforce learning, each module concludes with a quiz, as shown in Figure 4D. The quizzes provide immediate feedback, informing users whether their answers are correct, followed by explanations to deepen their understanding. Users are encouraged to achieve at least 80% accuracy before moving on to the next module. Notifications are sent to remind users to retake quizzes if they do not meet this threshold, ensuring that learning is reinforced and that users fully comprehend the material before progressing.

To ensure CRCWeb can be accessed on a wide variety of devices, we built the platform using the React Native framework [19]. This allows the platform to be deployed as an iOS app, Android app, and web app from a single code base, ensuring maximum accessibility across platforms. Since most disadvantaged populations with limited health literacy or low income are using older versions of Android devices, we made additional efforts to optimize CRCWeb's performance in the Android environment to ensure a smooth, bug-free experience. Following an agile software development process [20], we continuously refined the platform based on feedback from internal testers and semistructured interviews with patients and caregivers, resulting in an iterative and user-driven design. The backend server of CRCWeb is built using the Express.js framework [21] and object-relational mapping [22] to accommodate any relational database, allowing flexibility in terms of database management while ensuring robust performance and scalability.

Figure 2. System architecture of CRCWeb, highlighting its 3 core components: navigation and program guide, educational components, and surveys. The educational components feature 5 multimedia sections: lectures, content, activities, resources, and quizzes. User progress is securely stored on our server, ensuring confidentiality while enabling administrators to monitor and track advancement.

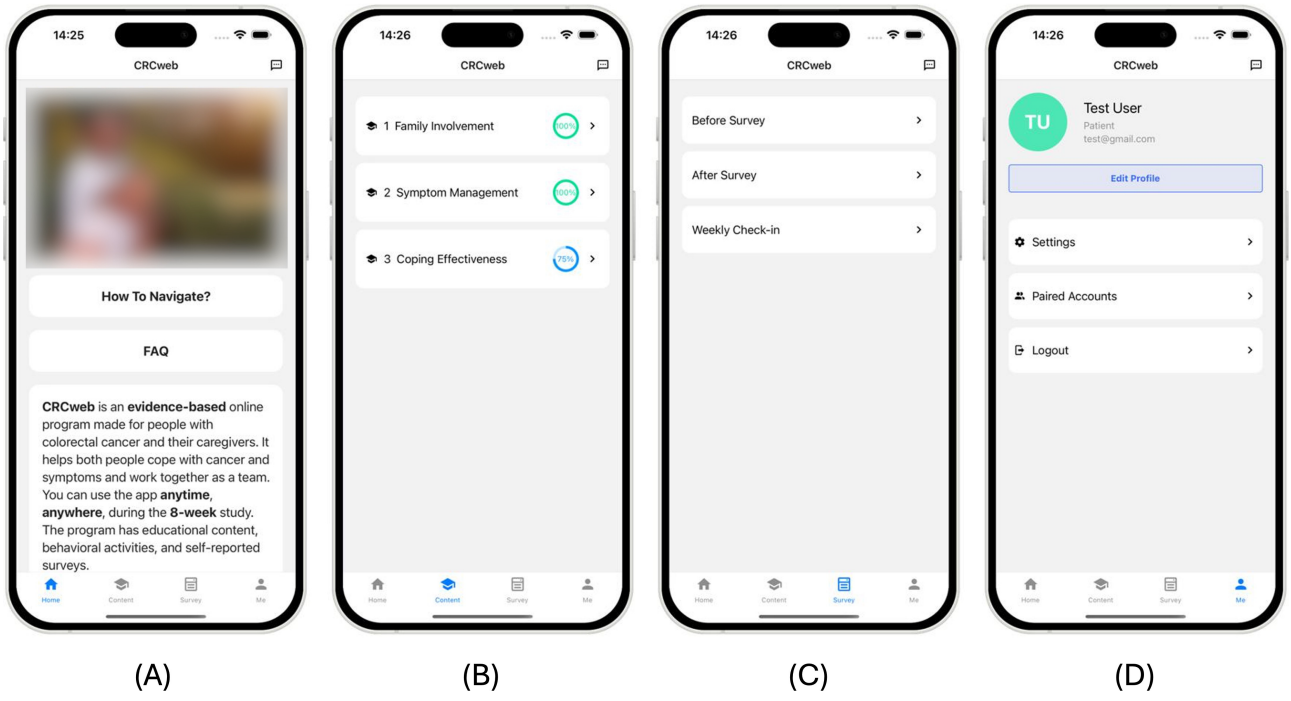
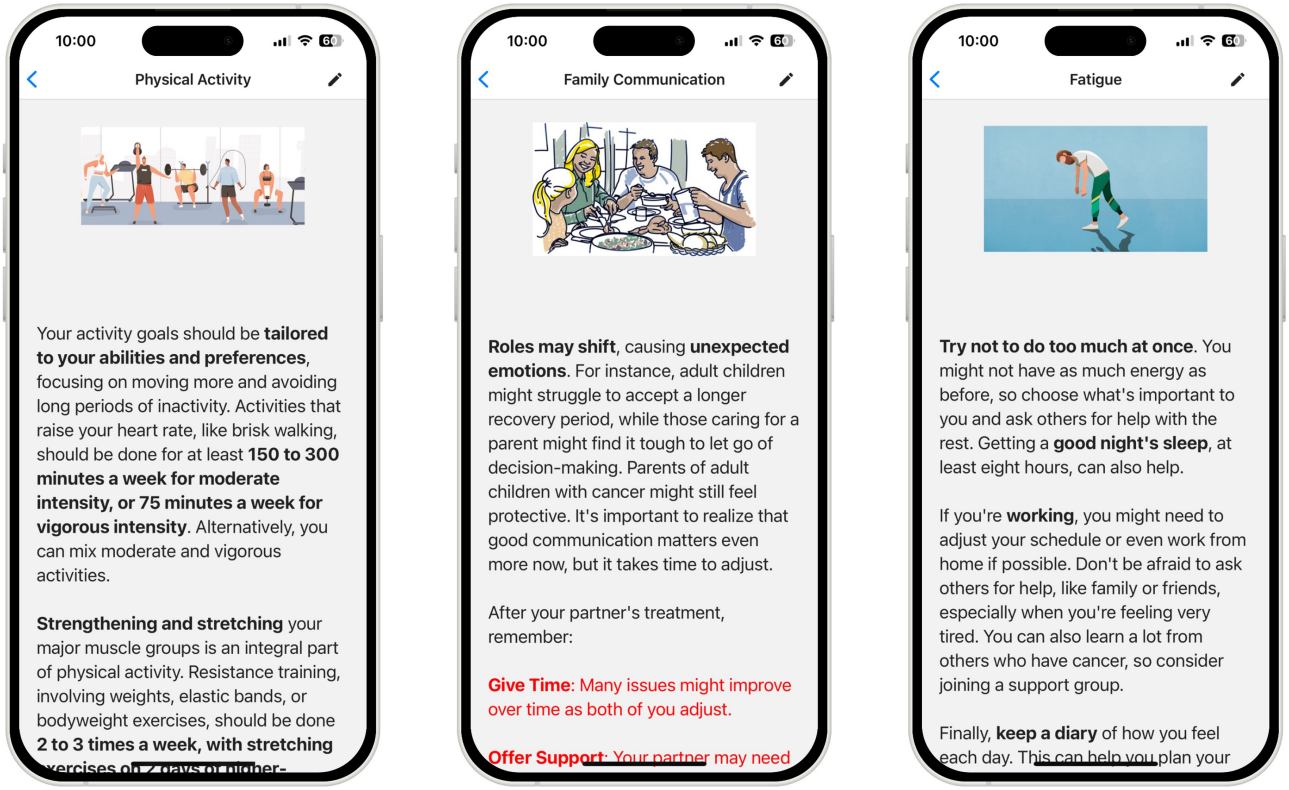
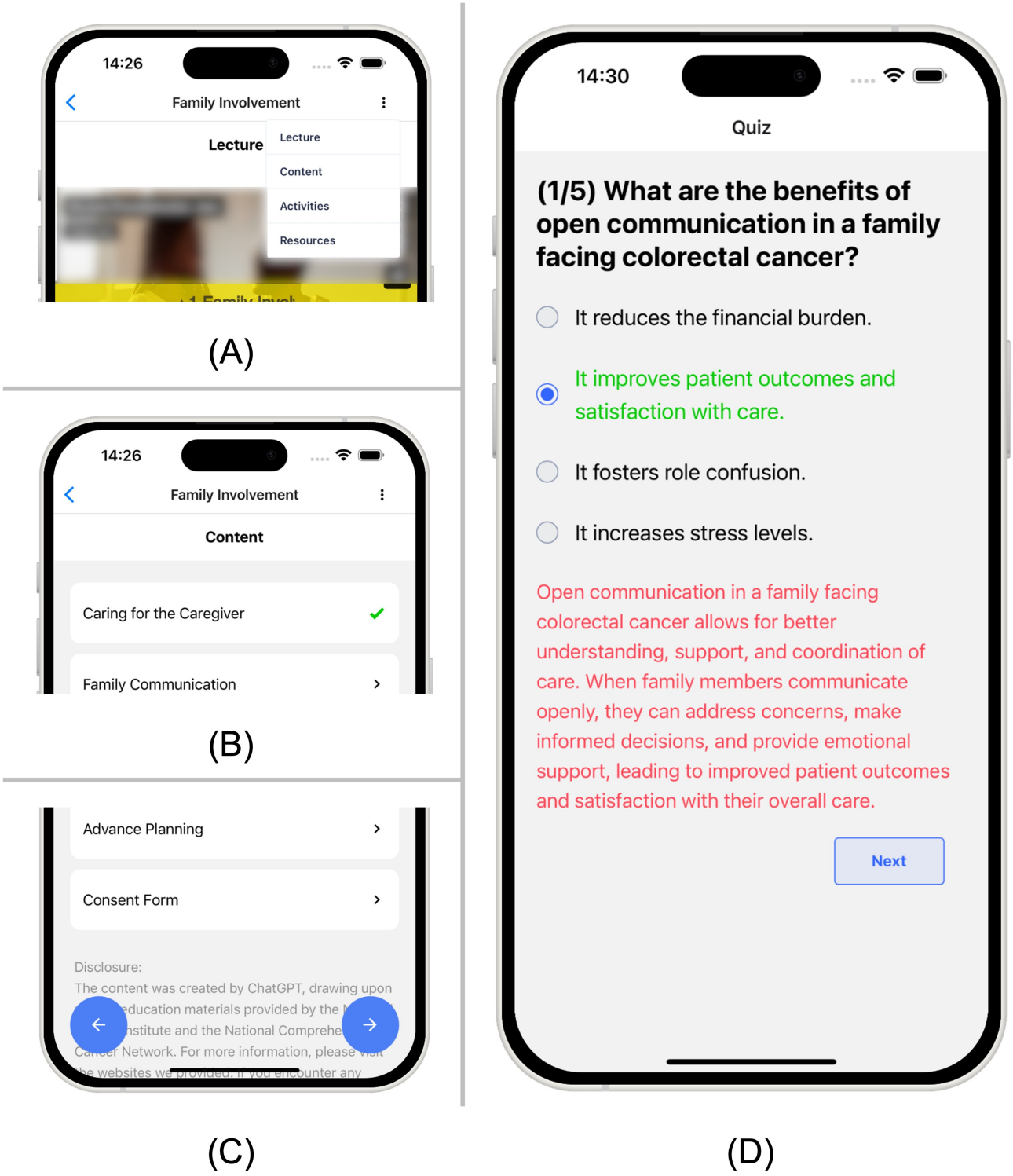


Figure 3. Examples of educational topics in CRCWeb: (A) Physical Activity, offering practical exercise recommendations; (B) Family Communication, guiding effective caregiver-patient discussions; and (C) Fatigue, providing practical strategies for managing energy levels during treatment.



(A) (B) (C)

Figure 4. The user-friendly navigation features of CRCWeb: (A) a drop-down menu to quickly navigate through different sections; (B) a green check mark to indicate when a section has been reviewed; (C) large buttons for navigating between sections; and (D) quizzes with immediate feedback to inform users if they answer correctly and providing a detailed explanation.



Educational Content

Tailoring Educational Materials for Symptom Management Using GenAI

CRCWeb's educational content is organized around 3 core modules—Family Involvement, Symptom Management, and Coping Effectiveness—all of which are informed by Stress Coping Theory [23] and Family Systems Theory [24]. These theories emphasize the critical role that psychological resilience and family dynamics play in how individuals manage chronic conditions like CRC. Specifically, Stress Coping Theory highlights how emotional and cognitive responses influence symptom management, while Family Systems Theory indicates the importance of involving family members in the care process. By integrating these theoretical frameworks, CRCWeb was designed to improve both patient and caregiver outcomes and engagement, particularly for disadvantaged populations with

limited health literacy or low income. Therefore, we hypothesized that enhanced family involvement and targeted psychosocial education alleviated the symptom burden for patients with CRC and their caregivers. This tool aimed to provide the knowledge, strategies, and emotional support necessary to manage symptoms effectively, ultimately improving both the patients' and caregivers' quality of life.

Each module is structured into 5 key sections: lectures, content, activities, resources, and quizzes. As illustrated in Figure 1, the educational materials were tailored into 4 distinct formats and distributed across different sections of CRCWeb: videos in the lectures section, textual content in the content section, practical activities in the activity section, extra-textual and video-based resources in the resources section, and quizzes in the quiz section. These educational materials were developed in alignment with guidelines from the National Cancer Institute [25] and the National Comprehensive Cancer Network [26], ensuring that the content is evidence-based and authoritative. As shown in Figure 3, the topics address a comprehensive range of practical, emotional, and physical challenges faced by patients with CRC and their caregivers. These topics are crafted to not only educate but also empower patients and caregivers, encouraging them to take an active role in managing cancer treatment and improving overall well-being.

To improve both engagement and accessibility, we integrated multimedia content throughout all modules, in alignment with design principle 2. Drawing from the semistructured interviews, we developed 5 distinct types of content—videos, slides, textual content, practical activities, extra resources, and quizzes—tailored to fit within the 5 sections of the program. Each lecture section includes 2 short videos: a primary lecture video that provides a comprehensive overview of the module's content and a supplemental video featuring interviews with patients with CRC who share their lived experiences. These videos are complemented by slides that offer additional visual summaries, making the information accessible to users with different learning preferences. Each content section features concise textual information, complemented by images and highlighted key points. The activity section offers practical activities with detailed instructions, while the resources section provides additional more detailed resources. The quiz section at the end of each module includes 5 quizzes to reinforce learning.

In the activities section, users engage in 9 digital exercises (3 activities per module) designed to reinforce what they learned in the lectures and content sections. These activities are tailored to daily life routines, such as symptom tracking and communication exercises between patients and caregivers. The activities also include a rating feature that allows users to provide feedback on their experience through thumbs-up or down ratings and optional comments. This feedback loop is a crucial part of CRCWeb's design, as it enables continuous refinement of activities based on real-time user input, ensuring that the content remains relevant and effective.

The resources section offers optional supplementary materials for users who wish to delve deeper into specific topics. These include additional videos, articles, and external links to trusted cancer resources. In response to user feedback, we curated this section to ensure that it provides meaningful yet nonoverwhelming options for further exploration. For example, users can access interviews with others managing similar symptoms, providing both practical tips and emotional support through shared experiences.

Each module concludes with a 5-question multiple-choice quiz, generated by ChatGPT and reviewed by our expert panel. These quizzes serve as a reinforcement tool, helping participants solidify their understanding of the key concepts covered in each module. Immediate feedback is provided for each question, with detailed explanations to clarify any misunderstandings and further support the learning process.

This comprehensive design, grounded in theory and informed by direct user feedback, ensures that the educational materials are not only accessible but also actionable, empowering patients with CRC and caregivers to take an active role in symptom management and care.

Generating Accessible Multimedia Materials Using GenAI

To enable the accessibility and engagement of educational materials, CRCWeb incorporates a framework for using GenAI to develop multimedia content, as shown in Figure 5. This framework begins by using the PyPDF2 [27] package to extract textual information from the original PDF documents. ChatGPT then generates concise summaries of the extracted text for each module. In line with design principle 4, these summaries are further processed by Pictory to create lecture videos tailored to individuals with limited vision and reading disabilities.

Additionally, the extracted text is processed by ChatGPT to produce various content formats tailored to the needs of disadvantaged populations with limited health literacy or low income. For instance, concise, low-reading-level text is generated to teach symptom management, while practical activities and quizzes are created to support hands-on learning and retention. These tailored formats, in adherence to design principles 2 and 3, are designed to make the educational content more accessible and actionable. The majority of the educational materials are created by GenAI-powered tools and subsequently reviewed by a team of oncology experts to ensure accuracy and relevance before being made available to patients and caregivers.

In the lecture section, we use ChatGPT to distill the core content of each module into a brief, cohesive summary that highlights the key topics. This summary is then processed by Pictory, a leading GenAI-powered video generation tool, which swiftly transforms text into engaging video content. These videos provide an alternative to text, allowing users to watch or listen to the material, reducing cognitive load and increasing engagement. As shown in Figure 5, Pictory creates relevant videos by incorporating the transcript and automatically highlighting key terms to reinforce understanding. The videos also include relaxing background music and a human voiceover to ensure a smooth, engaging viewing experience. For users who prefer text, we provide video transcripts that offer the same content in written form.

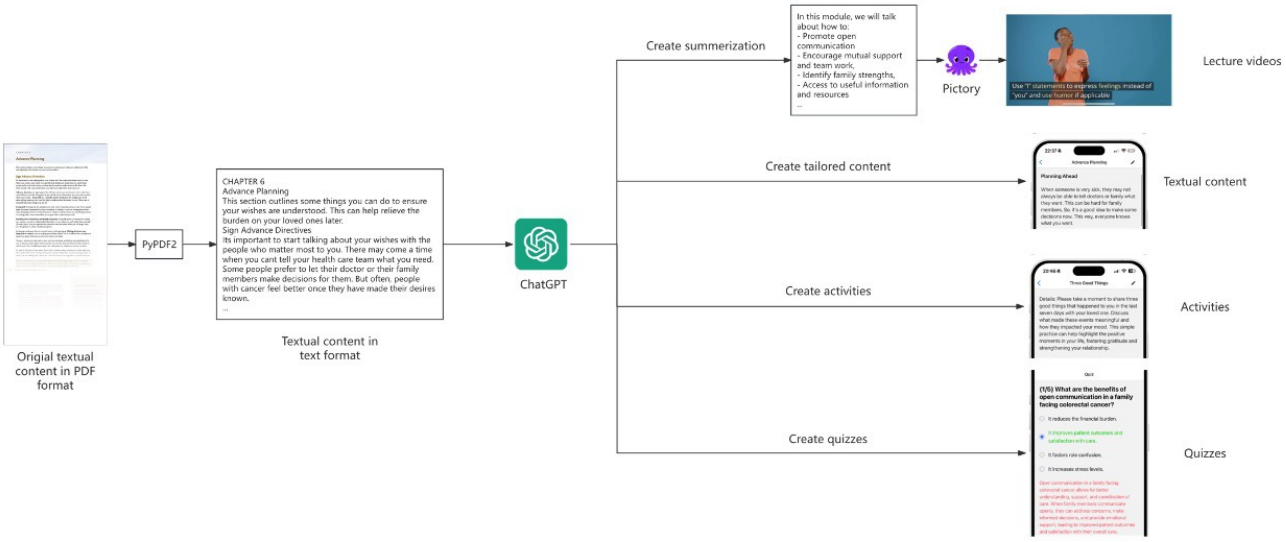
In the content section, CRCWeb presents text-based materials adapted from national guidelines for each topic. To ensure these materials are concise and easy to understand, as per design principle 3, ChatGPT is tasked with summarizing guideline documents into no more than 250 words with a Flesch-Kincaid Grade level of 6, making the content accessible to users of all literacy levels. A smart tagging system highlights key points using predefined tags, which are then reviewed and refined by our oncology experts to ensure clarity and relevance. Links to the original, full-length documents are included at the end of each section for users who wish to explore the source materials in more detail. Additionally, a text-to-speech feature powered by the React Native TTS package [28] provides an auditory option for users with vision impairments or reading difficulties. To prevent the tags from being read aloud, we use regular expressions to remove them from the audio transcript, ensuring a smooth listening experience.

In the activity section, we use ChatGPT to generate simple, practical activities that patients and caregivers can complete together. These activities are designed to reinforce the key concepts covered in the module and are structured to be easy to implement in daily routines. User feedback on these activities is collected through a rating system, which allows us to continuously improve their relevance and usability.

For the quiz section, ChatGPT generates 10 multiple-choice questions for each module, each accompanied by a detailed explanation of the correct answer. From these, we select the 5 most appropriate questions, refining them to match the challenge level needed to reinforce the material without overwhelming users. As shown in Figure 5, these quizzes provide immediate feedback with an explanation, helping participants reinforce the knowledge they learned from the module.

By integrating GenAI with expert review and accessibility features such as video, text, and audio options, CRCWeb ensures that its educational materials are engaging, user-friendly, and tailored to the needs of both patients with CRC and caregivers.

Figure 5. The framework was designed to leverage GenAI in creating multimedia content. First, the original PDF materials are converted into text using the Python package PyPDF2. The extracted text is then tailored by ChatGPT for 4 specific purposes: summarization of the topic, shortening the content to a lower reading level, creating practical activities, and generating quizzes. The topic summary is then processed by Pictory, which converts it into a video alongside a pleasant human reading voice. GenAI: generative artificial intelligence.



Performance Evaluation of Large Language Models in Tailoring Educational Content

We compared 3 models from the GPT family (GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo) to generate tailored content for disadvantaged patients with CRC and their caregivers with limited health literacy using predetermined prompts. The GPT-generated content was then evaluated by oncology experts and applied to CRCWeb.

To promote the accessibility and comprehension of educational materials for disadvantaged patients with CRC and their caregivers with limited health literacy, we structured prompts to have large language models (LLMs) produce content at a low reading level (Flesch-Kincaid Grade level of 6), maintain a word limit of 250, and provide Spanish and Chinese translations for each topic.

Our primary sources for content generation were education materials from the National Cancer Institute and the National Comprehensive Cancer Network guidelines [26]. We carefully selected 30 distinct topics that encompass a broad spectrum of content, including fatigue, depression, anxiety, pain, cognitive impairment, nutrition, and more. A subset of the topics was implemented in CRCWeb.

As reported in our prior work [29], the GPT family of models exhibited outstanding capability in tailoring educational materials for disadvantaged patients with CRC and their caregivers with limited health literacy or low income, only with deviations from the designated reading level.

Pilot Study

We conducted an 8-week single-arm pre-post pilot clinical trial among patients and their caregivers in 2 cancer clinics as our

test users to guide the development of CRCWeb. Among all 40 enrolled participants (20 patients and their caregivers), 22 of them came from disadvantaged backgrounds (ie, income ≤250% Federal Poverty Level, Medicaid, and uninsured), while the remaining 18 were from nondisadvantaged backgrounds, including individuals with higher incomes and adequate insurance coverage. The participants were instructed to download and navigate through CRCWeb by a clinical coordinator.

Participant Satisfaction With CRCWeb

The average satisfaction score was 3.980 out of 5 (1=strongly disagree, 2=disagree, 3=partially agree, 4=agree, and 5=strongly agree), with minimal variation (0.004), indicating that participants found the platform helpful on average.

When comparing satisfaction levels between the nondisadvantaged and disadvantaged groups, the average scores were nearly identical: 3.985 for the nondisadvantaged group and 3.971 for the disadvantaged group, resulting in a minimal difference of 0.014.

A nonparametric Two One-Sided Test using the Mann-Whitney *U* test was applied. The results, presented in Table S3 in Multimedia Appendix 1, revealed that for Questions 1 and 2, which assessed overall satisfaction and perceived helpfulness of CRCWeb, the distributions between the 2 groups are equivalent (both $P<.05$). For Questions 3 through 6—addressing the relevance of content, clarity, understanding of personal situations, and skills for symptom management—both Test 1 and Test 2 were also significant ($P<.05$). This indicates that both groups rated CRCWeb similarly on these dimensions, with no notable differences in how they perceived the app's relevance, clarity, or utility in managing symptoms. This demonstrates strong equivalence in the experiences of both groups in these areas. Finally, for Question 7, which focused on whether CRCWeb provided practical suggestions for everyday life, only Test 2 was significant ($P=.001$), while Test 1 was not ($P=.05$). This suggests that the disadvantaged group seems to have found CRCWeb more useful for providing practical, everyday suggestions compared to the nondisadvantaged group. This observation further reinforces the usability of CRCWeb for disadvantaged populations, as they perceived the platform to provide more practical support than their nondisadvantaged counterparts.

User Engagement and Login Frequency

To assess user engagement, we tracked attendance records, adherence rate, and login frequency across the study period. A total of 40 participants were enrolled in the intervention and the retention rate was 75%. Among the 40 participants, 87.5% completed all 3 modules and logged into CRCWeb at least 3 times. There was a clear difference in engagement between the disadvantaged and nondisadvantaged groups. The disadvantaged group logged in 209 times in total, compared to 83 logins from the nondisadvantaged group, making the disadvantaged group 2.52 times more engaged. A Mann-Whitney *U* test was performed to verify whether this difference was statistically significant, and the results confirmed that the disadvantaged group had significantly higher login frequencies ($P=.047$).

Lessons Learned and Future Directions

This viewpoint presents that by transforming textual data into multimedia components and tailoring educational content to the needs of low-health-literacy populations, CRCWeb addressed the significant health disparities that exist for disadvantaged groups with limited health literacy or low income [30]. Our assessment suggests that CRCWeb significantly enhanced user engagement for disadvantaged groups with limited health literacy or low income and achieved high levels of user satisfaction. This result is particularly noteworthy, as prior research has often highlighted lower engagement among disadvantaged populations in digital health interventions [31]. While our results demonstrated CRCWeb's effectiveness in delivering accessible educational content, several limitations should be addressed in future directions.

First, language translations were not included in this pilot study as we only recruited English-speaking participants. Prior research indicates that LLM performance varies by language, performing better in high-resource languages like German, French, and Spanish, but less effectively in lower-resource languages such as Kannada and Occitan [32,33]. To make CRCWeb more inclusive for non-English speakers, future work will evaluate auto-translation features and incorporate LLM-based translation for additional languages. This will allow CRCWeb to serve a broader, multilingual population.

Second, while commercial GenAI-powered tools such as ChatGPT and Pictory help streamline content creation, they come with risks such as potential pricing increases and service suspensions. To mitigate these risks, future work will explore developing our own GenAI models based on open-source frameworks [34-36]. This will give CRCWeb greater control over its content generation processes and ensure long-term scalability and cost-effectiveness.

Finally, in terms of user engagement, while login frequency is an important metric, it does not fully capture the complexity of user interactions with CRCWeb. To improve measurement accuracy, future iterations will include native apps for each platform, enabling more detailed tracking of user behaviors such as screen time and in-app navigation. Although this approach requires greater development effort, it will provide more inclusive metrics of user engagement.

Conclusions

Improving the accessibility of educational content on symptom management is essential to empowering patients with CRC and their caregivers, enabling them to more effectively manage symptoms throughout treatment. This is particularly vital for disadvantaged populations with limited health literacy or low income, who often lack access to national guidelines or frequent hospital-based care. CRCWeb addresses this challenge by leveraging GenAI-powered tools to transform overwhelming health care guidance into accessible multimedia formats, specifically tailored for patients with CRC and their caregivers. With features like low-reading-level text, engaging videos, and user-friendly navigation, CRCWeb ensures that patients and

caregivers can better understand and manage their symptoms. Designed with a stakeholder-centered approach, the platform prioritizes the needs of its users, making it a valuable tool for improving health outcomes. Moreover, the scalable design of CRCWeb demonstrates its potential to be adapted for broader disadvantaged populations with limited health literacy or low income, extending its impact beyond patients with CRC and caregivers to enhance health care accessibility for diverse groups.

Acknowledgments

This work was supported by the Oncology Nursing Foundation Research Grant (2022 RE03). The authors also appreciate the support from Emory University School of Nursing, Winship Cancer Institute, and Grady Memorial Hospital. The front end of CRCWeb is available under an MIT license [37].

Data Availability

The datasets generated and analyzed during the current clinical trial are not publicly available due to their containing private sensitive patient information but are available from the corresponding author upon reasonable request.

Authors' Contributions

DL designed and developed the mobile app, analyzed and interpreted the data, and wrote the manuscript. YL was responsible for conducting the clinical trial, recruiting participants, collecting the data, and writing the manuscript. RY assisted with the technical aspects of the clinical trial and contributed to writing the manuscript. ZW also provided support with manuscript writing. DB developed the foundation of the backend server and cloud services. XH supervised the entire project and provided valuable insights and feedback throughout the process. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File, 16 KB - cancer_v11i1e68516_app1.docx]

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74(1):12-49. [doi: 10.3322/caac.21820] [Medline: 38230766]
2. Chan RJ, Milch VE, Crawford-Williams F, et al. Patient navigation across the cancer care continuum: an overview of systematic reviews and emerging literature. CA Cancer J Clin 2023;73(6):565-589. [doi: 10.3322/caac.21788] [Medline: 37358040]
3. Lin Y, Bailey DE Jr, Docherty SL, et al. Distinct profiles of multiple co-occurring symptoms in patients with gastrointestinal cancers receiving chemotherapy. Support Care Cancer 2021 Aug;29(8):4461-4471. [doi: 10.1007/s00520-020-05946-4] [Medline: 33454824]
4. Lin Y, Bailey DE Jr, Docherty SL, et al. Distinct sleep disturbance profiles in patients with gastrointestinal cancers receiving chemotherapy. Cancer Nurs 2022;45(2):E417-E427. [doi: 10.1097/NCC.0000000000000975] [Medline: 34183521]
5. Saeed SA, Masters RM. Disparities in health care and the digital divide. Curr Psychiatry Rep 2021 Jul 23;23(9):61. [doi: 10.1007/s11920-021-01274-4] [Medline: 34297202]
6. Schulz R, Sherwood PR. Physical and mental health effects of family caregiving. Am J Nurs 2008 Sep;108(9 Suppl):23-27. [doi: 10.1097/01.NAJ.0000336406.45248.4c] [Medline: 18797217]
7. Lin Y, Porter LS, Chee W, et al. A web-based dyadic intervention to manage psychoneurological symptoms for patients with colorectal cancer and their caregivers: protocol for a mixed methods study. JMIR Res Protoc 2023 Jun 28;12(1):e48499. [doi: 10.2196/48499] [Medline: 37379055]
8. Knighton AJ, Brunisholz KD, Savitz ST. Detecting risk of low health literacy in disadvantaged populations using area-based measures. EGEMS (Wash DC) 2017 Dec 15;5(3):7. [doi: 10.5334/egems.191] [Medline: 29930971]
9. Al Shamsi H, Almutairi AG, Al Mashrafi S, Al Kalbani T. Implications of language barriers for healthcare: a systematic review. Oman Med J 2020 Mar;35(2):e122. [doi: 10.5001/omj.2020.40] [Medline: 32411417]
10. Moulaei K, Yadegari A, Baharestani M, Farzanbakhsh S, Sabet B, Reza Afrash M. Generative artificial intelligence in healthcare: a scoping review on benefits, challenges and applications. Int J Med Inform 2024 Aug;188:105474. [doi: 10.1016/j.ijmedinf.2024.105474] [Medline: 38733640]
11. Li J, Luo X, Cao Q, Lin Y, Xu Y, Li Q. Communication needs of cancer patients and/or caregivers: a critical literature review. J Oncol 2020;2020(1):7432849. [doi: 10.1155/2020/7432849] [Medline: 32454826]

12. Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, et al. GPT-4 technical report. arXiv. Preprint posted online on Mar 15, 2023. [doi: 10.48550/arXiv.2303.08774]
13. Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet Things Cyber-Phys Syst 2023;3:121-154. [doi: 10.1016/j.iotcps.2023.04.003]
14. Lu AKM, Chung YL, Wang YH, Lin SK, Hsieh JL. Developing and evaluating a campus-based health management app for nursing students: a pilot study on usability and user experience. Digital Health 2025;11:20552076251330577. [doi: 10.1177/20552076251330577] [Medline: 40162172]
15. Jimenez J, Del Rio A, Berman AN, Grande M. Personalizing digital health: adapting health technology systems to meet the needs of different older populations. Healthcare (Basel) 2023 Jul 26;11(15):2140. [doi: 10.3390/healthcare11152140] [Medline: 37570380]
16. Drisko JW, Maschi T. Content Analysis: Oxford University Press; 2016. [doi: 10.1093/acprof:oso/9780190215491.001.0001]
17. Buch SV, Treschow FP, Svendsen JB, Worm BS. Video- or text-based e-learning when teaching clinical procedures? A randomized controlled trial. Adv Med Educ Pract 2014;5:257-262. [doi: 10.2147/AMEP.S62473] [Medline: 25152638]
18. Mayer RE. Multimedia learning. Psychol Learn Motiv 2002:85-139. [doi: 10.1016/S0079-7421(02)80005-6]
20. React Native. 2024. URL: https://reactnative.dev [accessed 2024-09-13]
20. Abrahamsson P, Salo O, Ronkainen J, Warsta J. Agile software development methods: review and analysis. arXiv. Preprint posted online on Sep 25, 2017. [doi: 10.48550/arXiv.1709.08439]
21. Archer R. ExpressJs: Web App Development With NodeJs Framework: CreateSpace Independent Publishing Platform; 2016.
22. Ireland C, Bowers D, Newton M, Waugh K. Understanding object-relational mapping: a framework based approach. Int J Adv Softw Eng 2009;2(2) [FREE Full text]
23. Krohne HW. Stress and coping theories. Int Encycl Soc Behav Sci 2002. [doi: 10.1016/B0-08-043076-7/03817-1]
24. Rothbaum F, Rosen K, Ujiie T, Uchida N. Family systems theory, attachment theory, and culture. Fam Process 2002;41(3):328-350. [doi: 10.1111/j.1545-5300.2002.41305.x] [Medline: 12395563]
25. Patient education publications. National Cancer Institute. 2024. URL: https://www.cancer.gov/publications/patient-education [accessed 2025-05-30]
26. Guidelines for patients. NCCN. 2024. URL: https://www.nccn.org/patientresources/patient-resources/guidelines-for-patients [accessed 2025-05-30]
27. Fenniak M. PyPDF2. pypi. 2022. URL: https://pypi.org/project/PyPDF2 [accessed 2025-05-30]
28. React-native-TTS. NPM. 2024. URL: https://www.npmjs.com/package/react-native-tts [accessed 2025-05-30]
29. Liu D, Hu X, Xiao C, et al. Evaluation of large language models in tailoring educational content for cancer survivors and their caregivers: quality analysis. JMIR Cancer 2025 Apr 7;11:e67914. [doi: 10.2196/67914] [Medline: 40192716]
30. Epari A, Kim K, Xiao C, et al. Perceptions and needs for a technology-based dyadic intervention on symptom management among patients with colorectal cancer and their caregivers: a qualitative study. Cancer Nurs 2024 Nov 29. [doi: 10.1097/NCC.0000000000001429] [Medline: 39625810]
31. Radu I, Scheermesser M, Spiess MR, Schulze C, Händler-Schuster D, Pehlke-Milde J. Digital health for migrants, ethnic and cultural minorities and the role of participatory development: a scoping review. Int J Environ Res Public Health 2023 Oct 23;20(20):6962. [doi: 10.3390/ijerph20206962] [Medline: 37887700]
32. Zan C, Ding L, Shen L, Zhen Y, Liu W, Tao D. Building accurate translation-tailored llms with language aware instruction tuning. arXiv. Preprint posted online on Jun 26, 2024. [Medline: 38979485]
33. Li Z, Shi Y, Liu Z, Yang F, Liu N, Du M. Quantifying multilingual performance of large language models across languages. arXiv. Preprint posted online on Apr 17, 2024. [doi: 10.48550/arXiv.2404.11553]
34. Team G, Mesnard T, Hardin C, Dadashi R, Bhupatiraju S, Pathak S, et al. Gemma: open models based on gemini research and technology. arXiv. Preprint posted online on Mar 13, 2024. [doi: 10.48550/arXiv.2403.08295]
35. Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, et al. Llama: open and efficient foundation language models. arXiv. Preprint posted online on Feb 27, 2023. [doi: 10.48550/arXiv.2302.13971]
36. Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, et al. Llama 2: open foundation and fine-tuned chat models. arXiv. Preprint posted online on Jul 18, 2023. [doi: 10.48550/arXiv.2307.09288]
37. Liu D. CRCWeb. GitHub. 2025. URL: https://github.com/ImNotDarren/CRCWeb [accessed 2025-05-26]

Abbreviations

CRC: colorectal cancer
GenAI: generative artificial intelligence
LLM: large language model
REDCap: Research Electronic Data Capture

Edited by N Cahill; submitted 09.11.24; peer-reviewed by C Wang, CN Hang, L Zhu; revised version received 03.05.25; accepted 05.05.25; published 16.06.25.

Please cite as:
Liu D, Lin Y, Yan R, Wang Z, Bold D, Hu X
Leveraging Artificial Intelligence for Digital Symptom Management in Oncology: The Development of CRCWeb
JMIR Cancer 2025;11:e68516
URL: https://cancer.jmir.org/2025/1/e68516
doi:10.2196/68516

Process Re-Engineering and Data Integration Using Fast Healthcare Interoperability Resources for the Multidisciplinary Treatment of Lung Cancer

Ching-Hsiung Lin[1], MD; Bing-Yen Wang[2], MD; Sheng-Hao Lin[1], MD; Pei Hsuan Shih[3], BS; Chin-Jing Lee[4], PhD; Yung Ting Huang[5], BS; Shih Chieh Chen[6], MS; Mei-Lien Pan[7], PhD

[1]Division of Chest Medicine, Department of Internal Medicine, Changhua Christian Hospital, Changhua, Taiwan

[2]Division of Thoracic Surgery, Department of Surgery, Changhua Christian Hospital, Changhua, Taiwan

[3]Department of Information Systems, Changhua Christian Hospital, Changhua, Taiwan

[4]Project Management Office for Intelligence Healthcare, Kaohsiung Medical University Chung-Ho Memorial Hospital, Kaohsiung, Taiwan

[5]Roche Diagnostics Ltd Taiwan, Taipei, Taiwan

[6]Roche Information Solution, APAC, Taipei, Taiwan

[7]Institute of Hospital and Health Care Administration, National Yang Ming Chiao Tung University, No. 155, Section 2, Linong Street Beitou District, Taipei, Taiwan

Corresponding Author:
Mei-Lien Pan, PhD
Institute of Hospital and Health Care Administration, National Yang Ming Chiao Tung University, No. 155, Section 2, Linong Street Beitou District, Taipei, Taiwan

Abstract

Multidisciplinary team (MDT) meetings play a critical role in cancer care by fostering collaboration between different health care professionals to develop optimal treatment recommendations. However, meeting scheduling and coordination rely heavily on manual work, making information-sharing and integration challenging. This results in incomplete information, affecting decision-making efficiency and impacting the progress of MDT. This project aimed to optimize and digitize the MDT workflow by interviewing the members of an MDT and implementing an integrated information platform using the Fast Healthcare Interoperability Resources (FHIR) standard. MDT process re-engineering was conducted at a central Taiwan medical center. To digitize the workflow, our hospital adopted the NAVIFY Tumor Board (NTB), a cloud-based platform integrating medical data using international standards, including Logical Object Identifiers, Names, and Codes, Systemized Nomenclature of Medicine–Clinical Terms, M-code, and FHIR. We improved our hospital's information system using application programming interfaces to consolidate data from various systems, excluding sensitive cases. Using FHIR, we aggregated, analyzed, and converted the data for seamless integration. Using a user experience design, we gained insights into the lung cancer MDT's processes and needs. We conducted 2 phases: pre- and post-NTB integration. Ethnographic observations and stakeholder interviews revealed pain points. The affinity diagram method categorizes the pain points during the discussion process, leading to efficient solutions. We divided the observation period into 2 phases: before and after integrating the NTB with the hospital information system. In phase 1, there were 83 steps across the 6 MDT activities, leading to inefficiencies and potential delays in patient care. In phase 2, we streamlined the tumor board process into 33 steps by introducing new functions and optimizing the data entry for pathologists. We converted the related medical data to the FHIR format using 6 FHIR resources and improved our hospital information system by developing functions and application programming interfaces to interoperate among various systems; consolidating data from different sources, excluding sensitive cases; and enhancing overall system efficiency. The MDT workflow reduced steps by 60% (50/83), lowering the coordinated activity time from 30 to 5 minutes. Improved efficiency boosted productivity and coordination in each case of manager feedback. This study optimized and digitized the workflow of MDT meetings, significantly enhancing the efficiency and accuracy of the tumor board process to benefit both medical professionals and patients. Based on FHIR, we integrated the data scattered across different information systems in our hospital and established a system interoperability interface that conformed to the standard. While digitizing the work of MDT meetings, we also promoted the optimization and transformation of related information systems and improved their service quality. We recommend additional research to assess the usability of a tumor board platform.

(JMIR Cancer 2025;11:e53887) doi:10.2196/53887

KEYWORDS

multidisciplinary team meetings; process re-engineering; multidisciplinary cancer care; Fast Healthcare Interoperability Resources; tumor board; multidisciplinary team; cancer; lung cancer; treatment; lung; health care professionals; health care; MDT; digitize; API; hospital; information system; HIS; medical data; platform; data integration; information and communication technology; ICT; decision support; eHealth; digital tools; clinic; patient care; application programming interface; hospital information system

Introduction

Cancer care is a complex process that requires collaboration among health care professionals to develop the best treatment strategies. A multidisciplinary team (MDT) approach, in which experts from various specialties come together to discuss and share knowledge on cancer cases, provides an effective platform for delivering comprehensive and personalized care [1]. The MDT typically includes oncologists, surgeons, radiation oncologists, oncology nurses, clinical psychologists, social workers, dietitians, pharmacists, and physical therapists. Each member contributes their expertise to ensure the patient receives the most appropriate treatment and support. The team conducts a thorough assessment of the patient's condition, considering factors such as cancer type, stage, physical health, psychological well-being, and social circumstances. From this evaluation, a tailored treatment plan is developed, addressing all aspects of patient care. This collaborative process fosters a patient-centered approach that not only aims to improve survival rates but also enhances the quality of life by addressing the diverse needs of the patient.

The MDT approach is particularly valuable in cancer care as it ensures continuous communication and coordination among health care providers, leading to more effective treatment and improved overall outcomes. MDTs enhance treatment efficiency and patient care by bringing together health care professionals from various disciplines to collaborate on treatment plans. This approach facilitates shared decision-making and provides comprehensive care by addressing the social, psychological, nutritional, and physical needs of patients with cancer [2]. By integrating the expertise of diverse professionals, the MDT approach enhances the comprehensiveness and precision of care, aiding patients in navigating the complexities of cancer treatment [3]. However, to fully realize the benefits of this approach, it is essential to overcome systemic barriers, attitudinal challenges, and knowledge gaps through multilevel interventions [4].

According to the requirements of the Taiwan National Cancer Diagnosis and Treatment Quality Certification, MDT meetings must be held regularly every year to discuss new diagnostic cases. However, the treatment course for patients with cancer is long, and patients are often cared for by different clinical departments. As a result, cancer clinical data are scattered across different information systems (eg, outpatient, inpatient, and personal cancer management) and cover various diverse data elements, including patient demographic data, laboratory reports, and medications. Convening an MDT of cancer care meetings requires coordinating the schedules of various specialized teams and collecting and consolidating data from different information systems. Medical staff work tirelessly, which can be challenging. It is necessary to collect and integrate patient data before the meeting and follow up on the comments afterward. Effective MDT decision-making requires having access to relevant information, giving structured case presentations, exercising leadership skills, and organizing an effective meeting infrastructure [5-7]. Meeting tools can help boost MDT meeting efficiency [8-10]. Moreover, considerations regarding the clinical workflows of end users and existing information systems (eg, electronic health records [EHR]) are increasing. It is challenging to integrate information tools into the current system to meet end user needs and expectations and to effectively improve MDT practices [11]. Therefore, a thorough understanding of the MDT decision-making process, information flow, and routine workflow of key stakeholders is required.

In recent years, hospital information systems (HISs) have vigorously developed, with various clinical departments increasingly using information systems to support their clinical work [12-15]. Although most of the work in hospitals has been digitized, when handling cancer cases, especially in MDT meetings, it remains necessary to communicate through multiple emails and telephone calls to complete treatment proposals. Repetitive manual operations are required when preparing and conducting MDT meetings. To prepare meeting materials, team members collect case- and patient-care-related information such as literature, guidelines, and clinical trial information from different information systems. The collected materials are scattered and stored separately by each team member, and the meeting minutes are stored in the cancer center in the form of paper documents, making it difficult for clinicians to read meeting discussion materials and minutes before making clinical decisions. MDTs involve many people, specialties, data, and knowledge, and there is often a lack of standard execution procedures for holding meetings and preparing data. Moreover, the varying team members in their respective professional groups can lead to complex dynamics regarding authority and responsibility [4]. Addressing these challenges requires effective management strategies to ensure the seamless operation of MDTs and the provision of high-quality care to patients. Few studies discuss the integration of data and workflow in oncology care. The rapid expansion of medical knowledge, especially in oncology, has led to information overload and requires well-designed digital tools to manage and use this data effectively. The digital tool has streamlined the preparation and conduct of multidisciplinary tumor board meetings, with potential future applications in virtual meetings and patient engagement [16]. To build a patient-oriented cancer precision medicine platform that integrates all the information for cancer care, clinical decision support is necessary.

Using international standards is a common practice for interoperability and data integration across systems. The commonly used international medical information exchange standard is Fast Healthcare Interoperability Resources (FHIR) [17]. FHIR is designed to enable quick and efficient health data exchange, including clinical and administrative data. It has a

strong focus on implementation and also strengthens health data interoperability. FHIR solutions are built from a set of modular components called “resources” that can be easily assembled into working systems. Existing medical data can also be exchanged with other information systems through FHIR resources. By adhering to the standard, all medical information, including EHRs, medical images, and laboratory results, can be transformed into a consistent and easily interpretable format. Its implementation means that health care data is converted and integrated, thus ensuring efficient communication among various information systems. The FHIR standard streamlines data conversion processes, making it easier for health care providers to rapidly access and interpret patient information. This leads to faster decision-making and better patient management in MDT meetings.

We aimed to optimize and digitize the workflow of MDT meetings in cancer care. This viewpoint article described the implementation of an integrated information platform using the FHIR standard to enhance the efficiency and accuracy of the tumor board process. By interviewing MDT members and re-engineering, the MDT process, we aimed to address the challenges of manual work, information-sharing, and coordination in MDT meetings. We hypothesized that digitizing the MDT workflow and integrating data using FHIR would significantly improve the efficiency, accuracy, and overall quality of cancer care provided by MDTs.

Setting and Project Overview

Overview

This project was conducted at a major medical center located in central Taiwan. The hospital is renowned for its comprehensive cancer care services, which include diagnosis, treatment, and follow-up care for various types of cancer. The hospital has a dedicated oncology center that provides multidisciplinary care to patients with cancer, ensuring that they receive the best possible treatment and support throughout their cancer treatment journey.

This hospital is a large medical facility with over 1800 beds and a wide range of specialized departments. It serves a diverse patient population from central Taiwan and beyond, offering advanced medical care and cutting-edge treatments. The oncology center at the hospital is equipped with state-of-the-art technology and staffed by a team of highly skilled health care professionals, including oncologists, radiologists, surgeons, specialized oncology nurses, clinical psychologists, social workers, dietitians, pharmacists, and physical therapists.

The team involved in this project includes members from various departments within the hospital. The MDT members include pulmonologists, pathologists, radiologists, radiation oncologists, and case coordinators. To optimize and digitize the workflow of MDT meetings in cancer care, we have developed an integrated information platform based on the FHIR standard to enhance the efficiency and accuracy of the tumor board process. A flow diagram is shown below (Figure 1).

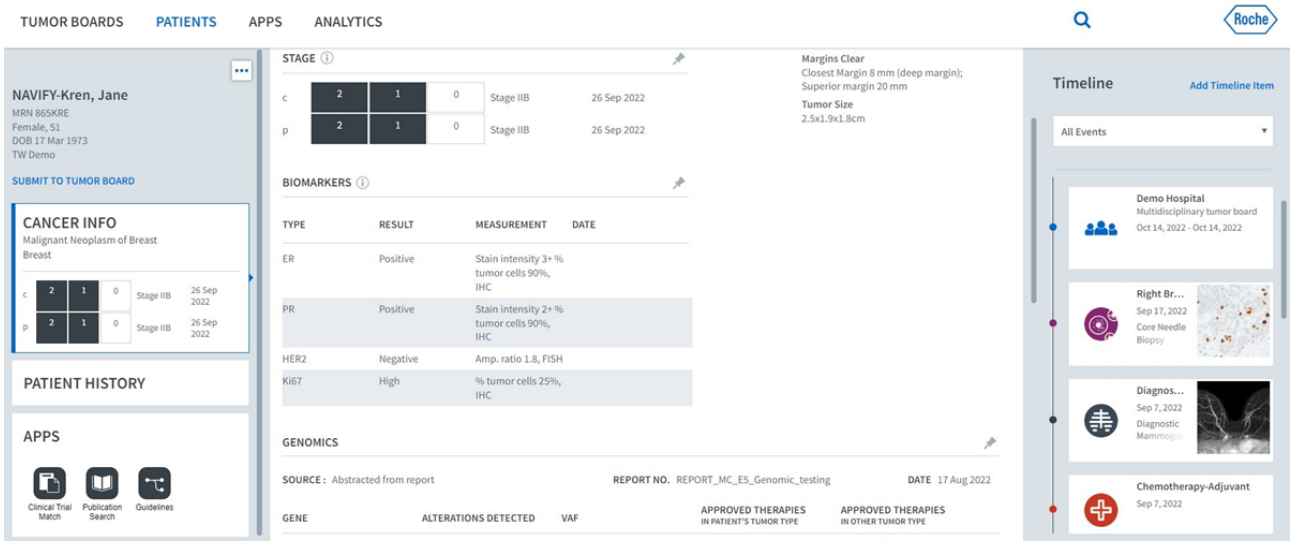
Figure 1. The research flow diagram. EHR: electronic health record; FHIR: Fast Healthcare Interoperability Resources.

Implementation of the Tumor Board Platform

In this project, the tumor board platform refers to a digital tool designed to streamline and enhance the workflow of multidisciplinary tumor board meetings. These platforms integrate various types of medical data, such as EHRs, medical images, laboratory results, and pathology reports, into a single, accessible system. This integration allows health care professionals, including oncologists, radiologists, surgeons, and other specialists, to collaboratively analyze, discuss, and develop personalized treatment plans for patients with cancer. The tumor board process involves regular meetings where an MDT of health care providers with different specialties come together to discuss cancer cases. During these meetings, the team reviews patient information, shares knowledge, and formulates comprehensive treatment plans tailored to the individual needs of each patient. The goal of the tumor board process is to ensure that patients receive the best possible care through coordinated and collaborative decision-making. To digitize the working process, an information system (NAVIFY Tumor Board [NTB]; Roche Molecular Systems), was adopted in our hospital. The NTB is a cloud-based workflow platform that integrates all relevant medical data to facilitate tumor board workflows [18]. This software as a tumor board platform uses various international standard terminologies (eg, Logical Object Identifiers, Names, and Codes, Systemized Nomenclature of Medicine–Clinical Terms, and M-code) and has a core database built using the FHIR standard to help prepare, present, and document the workflow of multidisciplinary tumor board meetings. In the early stages of system construction, we imported patient medical data into the information platform. The inclusion criteria for the patient data consolidated into the platform involve patients meeting the qualifications for multidisciplinary care. These include patients with stage III lung cancer, complex cases, patients with special events, and cases recommended for discussion by attending physicians. We excluded sensitive cases before integrating the medical data. Sensitive cases refer to patient histories of domestic violence and medical disputes, as defined by our hospital.

We also define standardized processes and operating procedures for various roles and provide system training courses to facilitate the use of this system. One of the key values that MDT members find in NTB is its ability to integrate data. In the past, patient data was scattered across different systems, requiring members of MDTs to log into multiple systems to review a single case. However, with NTB, they only need to log into the NTB platform to complete case reviews. In addition to data integration, NTB's interface visually presents patient examinations and treatment schedules through a "timeline," which helps MDTs understand the progression of a patient's condition more easily. Whether through data integration or the timeline feature, effective data exchange between systems is achieved using the FHIR standard. Moreover, the platform directly facilitates collaborative case editing among MDTs, allowing team members to prepare their respective data from different locations and times, thereby enriching the completeness of patient information (Figure 2).

Figure 2. User interface of NAVIFY Tumor Board.



We aggregated and analyzed the data requirements for interfacing with each information system. Additionally, we mapped and converted all of the data to align with the FHIR format. This comprehensive process involved the mapping of diagnosis and treatment information to the FHIR standard, ensuring integration with the tumor board information system. Moreover, to facilitate seamless data exchange, we designed an interoperability data model that adheres to internationally recognized FHIR standards. This approach enhances data compatibility and promotes efficient communication across diverse health care systems.

Our primary objective was to conduct process re-engineering in a multidisciplinary lung cancer team to address the difficulties and challenges faced by the teams in their workflow. To achieve this goal, we applied the user experience design approach to gain in-depth insights into the processes and requirements of lung cancer MDT.

This project uses "step reduction" as an indicator to assess workflow rather than "time reduction." The primary reason for this choice is that the complexity of cancer cases varies widely in clinical settings, ranging from standard treatments to complex cases. The preparation time for different cases varies

accordingly. Hence, “step reduction” is chosen as a metric for evaluating workflow processes. Additionally, qualitative observation is used for two reasons: (1) Participant perspectives: Qualitative methods capture direct participant involvement in workflows, allowing insights into personnel perspectives and experiences, thereby providing insights into human factors that influence efficiency and effectiveness. (2) Holistic observation: Qualitative methods offer a comprehensive view of workflow processes, considering various elements and their interactions, which is crucial for identifying potential areas for improvement [19,20].

We divided the work of process re-engineering into two phases: (1) before the use of the NTB and (2) after the integration of the NTB and the HISs. Through observations of the tasks of multidisciplinary decision-making and meetings, we recorded the workflow and working steps from the MDT and interviewed the members of the multidisciplinary cancer team members to understand the related work and information needs of their work. We participated in 2 observation meetings covering 3 cases in each meeting. Interview participants included: 1 supervisor from the oncology center, 1 case manager, 1 resident physician, 1 radiologist physician, 1 pathologist physician, 1 lung MDT leader, and 1 chest nurse practitioner. Participants involved in defining pain points for discussion included: 1 supervisor from the oncology center, 1 case manager, 1 resident physician, 1 radiologist physician, 1 pathologist physician, 1 lung MDT leader, and 1 chest nurse practitioner.

Ethical Considerations

This project was approved by the Institutional Review Board of the Changhua Christian Hospital (No. 200816). Ethical approval was obtained on December 30, 2021. The requirement for informed consent was waived by the Institutional Review Board because the research involved minimal risk to the participants and could not be practicably carried out without the waiver. All data collected were anonymized to ensure the privacy of the participants. No identifiable personal information was used in the analysis or reporting of the results. No compensation was provided to the participants as the study involved minimal risk and did not require active participation.

Process Re-Engineering Phase 1

Overview

In phase 1, we used AEIOU (Activities, Environments, Interactions, Objects, and Users), ethnographic observation, and stakeholder interviews to gather pain points and requirements. Through on-site observations, we gained a preliminary understanding of the team members, their discussion processes, and the content involved. Subsequently, we conducted stakeholder interviews to gather more information from the lung cancer MDT members. These methods enabled us to gain deeper insights into users’ needs and pain points.

The MDT workflow was divided into 6 activities: patient collection, coordination, preparation, meetings, documentation, and follow-up.

From phase 1 to phase 2, the number of steps in the workflow decreased from 83 steps to 33 steps. The coordinate activity saw the largest reduction, dropping from 35 to 12 steps (Figure 3). Phase 1 consisted of 83 steps across the 6 activities, most of which were carried out manually. We performed various tasks, including patient data collection, documentation, and follow-up, manually. During the coordination activity, MDT members had to repeatedly query and retrieve data from various HIS and EMR systems. These data were then compiled into Microsoft Word (Microsoft Corp) files containing diagnoses, radiology image reports, and pathological summaries, mostly in text format. During meetings, MDT members use their individual files and different information systems to present case details, which could lead to duplicate or inconsistent content among the different files. After meetings, case managers compile the meeting minutes in text form and store them at the tumor center. If physicians need to review the meeting content during follow-up, they must request access from the tumor center. This process is time-consuming and inefficient, leading to potential delays in patient care.

Figure 3. Workflow of MDT: (**A**) phase 1: before the use of the NTB and (**B**) phase 2: after optimizing workflow and integrating the information system. MDT: multidisciplinary team; NTB: NAVIFY Tumor Board.

Collect 4 steps	Coordinate 35 steps	Prepare 11 steps	Conduct 16 steps	Document 11 steps	Follow-up 6 steps
-Patient collection: 4	-Administration: 5 -Data query: 13 -Notify: 5 -Records summary: 6 -Prepare materials: 5 -Discussion: 1	-Administration: 4 -Records summary: 2 -Prepare materials: 2 -Meeting room: 3	-Data query: 5 -Prepare materials: 4 -Discussion: 4 -System switching: 3	-Administration: 7 -Notify: 4	-Administration: 1 -Data query: 1 -Notify: 2 -Recommendation: 2

(A) Phase 1

Collect 1 step	Coordinate 12 steps	Prepare 3 steps	Conduct 7 steps	Document 4 steps	Follow-up 6 steps
-Patient collection: 1	-Administration: 1 -Data query: 6 -Notify: 5	-Administration: 1 -Records summary: 1 -System switching: 1	-Data query: 4 -Prepare materials: 3	-Administration: 2 -Notify: 2	-Data query: 1 -Notify: 1 -Records summary: 2 -Recommendation: 2

(B) Phase 2

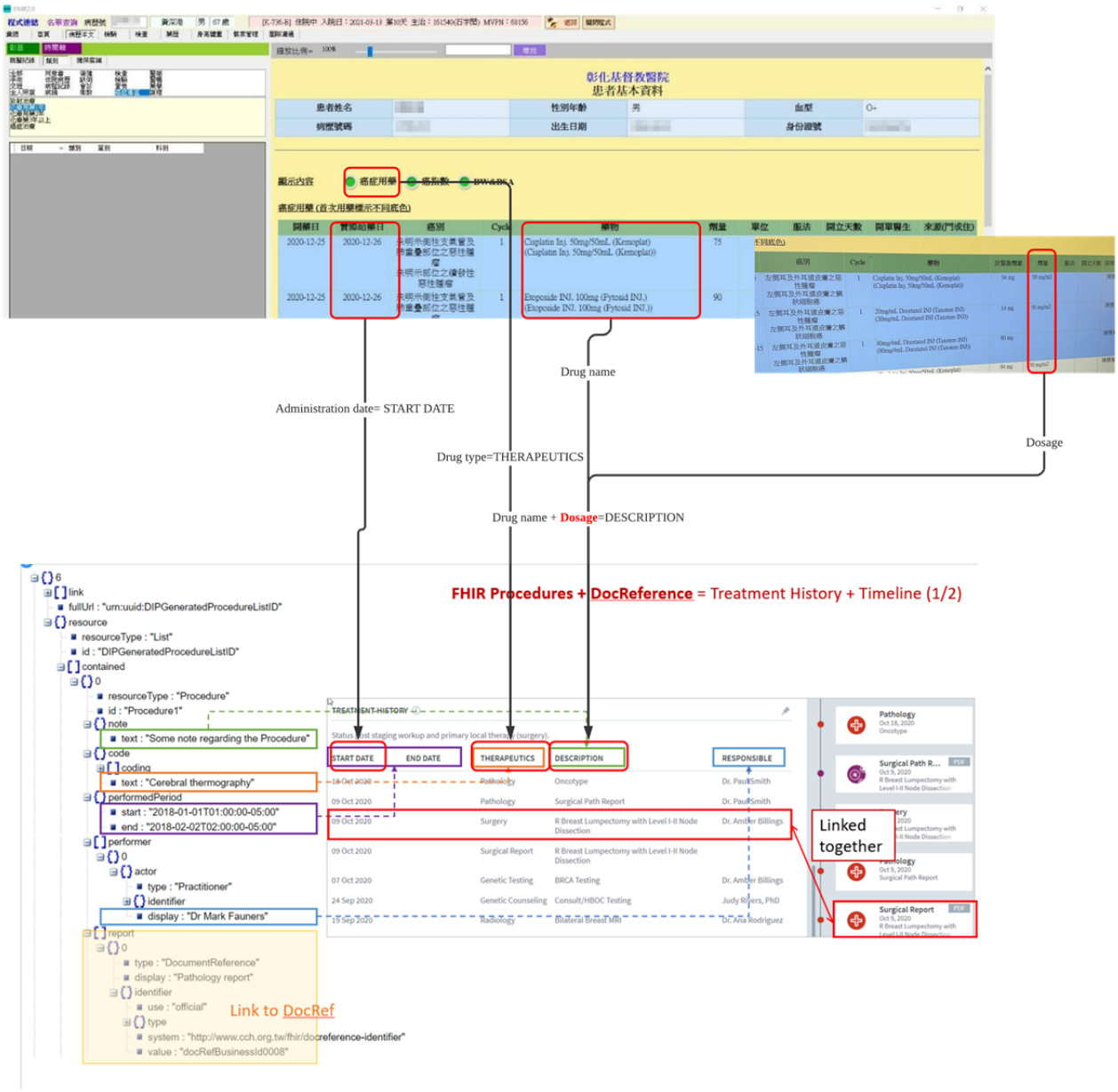
FHIR Adoption

For data integration, we conducted a thorough review of the medical charts and meticulously analyzed patient records and relevant information. To ensure accurate and effective data mapping, we engaged in extensive discussions with medical professionals and experts. Through these collaborative sessions, we identified the key data elements and attributes in the medical charts and carefully aligned them with the appropriate FHIR resources, which encompassed six major categories.

1. Patient: contains essential patient information (eg, name, date of birth, sex, medical record number, and other relevant data).
2. Condition: provides details of the clinical conditions, problems, diagnoses, or events that require attention, including key information (eg, the date of initial diagnosis).
3. Body structure: records information about anatomical structures (eg, location and laterality data).
4. Procedures: encompasses actions that are currently or have been performed on the patient (eg, surgical dates, preoperative and postoperative diagnoses, operating physicians, surgical images [JPEG format], medical orders, and radiation therapy information).
5. Diagnostic report: contains data on patient diagnostic results (eg, radiology images and other relevant findings).
6. Observation: comprises data from patient examinations (eg, genetic testing reports and other observational data).

We transformed medical data into a format that adheres to FHIR standards. This comprehensive process allowed us to seamlessly integrate medical data into the FHIR standard, thereby enabling smooth interoperability and data exchange across various health care systems. We mapped the data from the clinical information system to the corresponding fields compliant with the FHIR standard. We presented this comparison systematically through the system interface, allowing clinical physicians to review and confirm the accuracy of data transformation. An example of clinical data transformation into FHIR is shown in Figure 4.

Figure 4. Example of chemotherapy treatment history clinical data transformation into FHIR. FHIR: Fast Healthcare Interoperability Resources.



Process Re-Engineering Phase 2

Overview

In phase 2, we used the affinity diagram method to define pain points. First, we mapped the complete pre-, during-, and postdiscussion processes of lung cancer MDT. Then, using the affinity diagram method, we categorized pain points at various stages of the discussion. This approach helped us gain a clearer understanding of the process of pain points and design effective solutions more efficiently.

Integration of the Platform With EHR

We also improved our hospital's information system and modified some of its functions. We developed several functions and application programming interfaces (APIs) to interoperate these systems and consolidated the data from several information systems within our hospital, including patient demographics, inpatient and outpatient order data, the cancer registry, the American Joint Committee on Cancer stage data, surgical data, biomarkers, radiology, picture archiving and communication systems, pathology reports, and cancer treatment plans.

Figure 5 displays a sample code of the APIs developed to enable the system interface to meet the specified requirements. These APIs were instrumental in integrating 10 information systems into the NTB, thereby facilitating process digitization and automation. This seamless integration, coupled with process optimization, allowed MDT members to collaborate effectively on the platform, streamlining the coordination and preparation of the meeting content. The platform's data integration capabilities enable the simultaneous presentation of diagnoses, findings, images, and imaging reports during meetings. Participants can access the complete meeting materials, reports, and images together, ensuring an efficient and comprehensive discussion. Moreover, the platform automatically saves screenshots and image data during meeting minutes, thereby simplifying the documentation process. Physicians can

conveniently access the meeting content directly from the NTB during follow-up, thus enhancing accessibility and continuity of care. The successful development and implementation of these APIs significantly improved the efficiency and effectiveness of the MDT workflow at the NTB.

Figure 5. Sample code of the APIs. API: application programming interface.

```
class CancerInfoController : ApiController
{
    Dictionary<string, string> _connstrings;
    string CancerInfo(int chart)
    {
        string result = "";
        OracleConnection conn = new OracleConnection(_connstrings);
        OracleCommand cmdSel = null;
          OracleCommand cmdUpd = null;
        try
          {
            conn.Open();
            string selSQL = BuildSql(chart);
            cmdSel = new OracleCommand(selSQL, conn);
            cmdSel.CommandType = CommandType.Text;
            cmdSel.Parameters.Add("chart", OracleDbType.Decimal).Value = chart;
            DataTable dtCancerInfo = ExecuteQuery(cmdSel);
            result = ConvertToJson(dtCancerInfo);
        }
        catch (Exception ex)
          {
            result = HandleException(ex);
        }
        finally
          {
                string updSQL = BuildSql(chart);
            cmdUpd = new OracleCommand(updSQL, conn);
                cmdUpd.CommandType = CommandType.Text;
            cmdUpd.Parameters.Add("chart", OracleDbType.Decimal).Value = chart;
                cmdUpd.ExecuteNonQuery();
                conn.Close();
        }
        return result;
    }
}
```

In phase 2, we successfully reduced the number of steps in the tumor board process to 33 across 6 activities (Figure 3). To achieve this, we implemented several improvements to our HIS, specifically for the tumor board process. First, we introduced 2 new functions in the HIS to streamline the tumor board process. The “patient collection” function allowed MDT team members to easily request a tumor board workflow digitally, simplifying the initiation process. Additionally, we developed a new function tailored for pathologists. Through in-depth user interviews, we determined the system operation requirements and reorganized the pathology report entry process. In the original system, pathologists had to enter a template code, and the corresponding gross findings and descriptive content appeared in text fields. To improve the data structure and efficiency, we converted the gross findings and descriptions into a fixed structure and extracted vital information for inclusion in the primary diagnosis feature. We carefully discussed these templates with key users and pathologists to

ensure their accuracy and relevance. To minimize errors, we designed a check function for the proposed system. In cases where templates could not be converted into structured inputs, we extracted and converted the content of the remark field into structured information. Pathologists need only input keywords or key numerical values, and the system automatically converts the data into an edited narrative report. Furthermore, we provided a comment field for physicians to offer supplementary explanations when needed. These enhancements significantly improved the efficiency and accuracy of the tumor board process, benefiting both medical professionals and patients.

The pain points identified and categorized were as follows: phase 1: before the use of the NTB, there were a total of 48 pain points, which were reduced to 12 after optimizing workflow and integrating the information system (Table 1). For the detailed pain point with the MDT workflow, please refer to the Multimedia Appendix 1.

Table . Total pain points were identified in 6 activities.

Pain points	Collect	Coordinate	Prepare	Conduct	Document	Follow-up	Subtotal
Phase 1, n	4	22	5	8	6	3	48
Phase 2, n	0	4	0	4	1	3	12

Despite having the same number of activities in the MDT workflow, a significant reduction of 60% (50/83) in the number of steps was achieved in the tumor board process. This streamlining effort effectively optimized the efficiency and effectiveness of the tumor board process. Due to these improvements, the case managers reported a remarkable decrease in the average time spent on coordinated activities, from 30 to 5 minutes. Feedback from the case managers highlighted the considerable time-saving benefits generated by the enhanced workflow, leading to increased productivity and smoother coordination within the MDT.

Lessons Learned

Overview

We encountered several issues and challenges during the implementation of the FHIR standard. First, there is a significant scarcity of personnel proficient in both medical data content and the FHIR standard. Second, the data collected by existing information systems must meet the basic requirements of the FHIR standard format and be structured data. Throughout the project, we reviewed medical charts and engaged in discussions with medical and informatics experts to map our data into FHIR resources. Additionally, FHIR education training courses were organized to ensure that both clinical and information staff in our hospital could learn this medical information standard. Furthermore, the data fields of the HIS system were adjusted to meet the requirements of FHIR conversion.

Future Work and Implications for Practice

By implementing a framework and applying multispecialty discussions in cancer care, this study is expected to help medical information teams refer to available FHIR resources and provide a standard interface in an efficient, low-cost manner that does not affect daily operations. The interface integrates the complete diagnosis and treatment experience of cancer (eg, diagnosis, treatment, outpatient and inpatient consultations, previous discussions, and other imaging information) and can be used for in-hospital treatment, teaching, and research. FHIR resources, data models, and related systems used for MDT discussions on cancer can also be referenced by those who wish to construct FHIR standards.

Other MDTs engaged in cancer care can use our workflow optimization experience as a reference. With the growth in cancer cases and the development of precision medicine, genetic diagnosis, and digital medicine, the demand for digital assistance platforms for cancer care has gradually increased. At the same time, medical and health data are expected to increase at a rapid rate. These developments present additional challenges for clinical teams. Our optimized workflow can be used as a reference by other hospitals that wish to digitize their MDTs.

The development of a structured medical record entry system in our hospital was accelerated while digitizing the MDT meeting workflow. The essence of any successful structured medical record entry system lies in its ability to standardize or make data collection uniform across patients through an easy reporting system while allowing improved decision support, real-time quality assessment, and opportunities for patient-oriented clinical research [21]. The MDT meeting workflow provides incentives and application scenarios, prompting users to participate in the system design and development. However, with the NTB platform, MDTs experienced a profound improvement in their clinical discussions. The ease of accessing patient data, imaging results, pathology reports, and treatment histories allows for a more holistic understanding of each case. This comprehensive approach facilitates in-depth discussions and fosters collaboration in devising tailored treatment plans for patients. Furthermore, the platform streamlines operational processes, reduces the administrative burden, and saves time. Improved workflow and efficient decision-making contribute to enhanced patient outcomes and overall operational efficiency.

In addition to the NTB platform, the adoption of the FHIR international standard has revolutionized data exchange among cancer-related systems. By providing a consistent and standardized interface, FHIR enhances system openness and interoperability, allowing different systems to seamlessly communicate and share information. This standardized approach empowers health care providers to integrate patient data from various sources, including EHRs, imaging systems, laboratory results, and treatment records. Consequently, clinicians have a more comprehensive and real-time view of a patient's health status, leading to better-informed clinical decisions and improved patient outcomes.

The comprehensive rollout plan to other cancer teams can optimize the MTD workflow within the hospital, enhancing clinical work efficiency. We will modify the HISs to add a new function that enables physicians to read the NTB meeting outcomes through outpatient information systems, facilitating a closed-loop data use process. Moreover, additional data sources such as ultrasound reports and medical histories will be integrated to enrich the data sources.

This platform can be extended to other clinical settings. Attending physicians have found that the NTB platform is not only useful for multidisciplinary discussions, but also that the integrated data on NTB is suitable for team assessments and discussions before patient surgeries. Our educational department has also noted that storing comprehensive case discussion data in NTB can serve as gold cases for training post graduate year doctors, establishing a repository of gold cases without medical record numbers. Additionally, it provides case materials for teaching faculty and restricts access to post graduate year doctors only, achieving educational objectives. Furthermore, in response to Taiwan's cancer next-generation sequencing (NGS) reimbursement policy, establishing a Molecular Tumor Board to discuss NGS cases is an upcoming initiative for deeper application.

We also observed that, although the new NTB platform provided integrated data that could save data search time, it increased the working time to prepare meeting materials. Professional software tools often have a learning curve, and their initial use may require additional time. Although it might take a bit more time to learn how to use the software, as users become more proficient, their data preparation speed increases, and more time is saved. Additionally, an integrated platform can provide users with more comprehensive patient records, allowing them to gain insights from consolidated data rather than simply copying and pasting information from various systems, as was the custom in the past. This increased time usage may enhance the quality of health care deliveries. The NTB also has many new functions (eg, data annotation), which can make presentation materials more appealing; however, this takes more time. During follow-up, although they could access the NTB to view the meeting content, physicians looked forward to an efficient way to view the meeting minutes. The NTB platform is being continuously optimized and integrated with other information systems, in which medical professionals can view the meeting minutes directly. Inviting various clinical teams to use the platform to improve decision-making support is the next step.

After data integration, information can be reused in a format compliant with the FHIR standard. In Taiwan, NGS cases can be uploaded to national-level biobanks in FHIR format, as many disease and case notification data requirements also mandate the FHIR format. Using data for future clinical research will be easier, especially for studies involving cross-institutional or international clinical databases. Our research demonstrates the integration of full medical report data in meetings, allowing team members to review the reports together. If the data are insufficient, team members can directly enter the patient's timeline function to view other data without a cross-system query. A previous study [22] referred to information technology as a solution to achieve real-time data collection and imaging, which may improve patient-centered care coordination. In this study, we not only accomplished information integration but also optimized the workflow for tumor boards.

Conclusions

The use of information technology in MDT meetings has become common; however, the full potential of information systems for data collection, integration, and collaboration remains underused despite its immense value to health professionals. In this project, we attempted to optimize and digitize the workflow of MDT meetings. By leveraging the international data exchange standard, FHIR, we successfully integrated data from various information systems within our hospital, establishing a system interoperability interface compliant with the FHIR standard.

During the digitization process, we not only optimized and transformed related information systems but also enhanced the overall service quality of our hospital's information system. This digital transformation has facilitated physicians' use of medical record data for research by implementing a structured medical record entry interface, thereby improving the accessibility and availability of medical records.

In addition to the lung MDT, we encouraged other cancer care teams to adopt the new process and integrated platform. Currently, 4 cancer groups are using NAVIFY, and we anticipate that by the end of 2023, all cancer groups within our hospital will be on board, amounting to a total of 10 teams. Overall, we emphasize the importance of efficient processes that use standardized and leveraged technology to optimize the tumor board process and enhance cancer care delivery. We will share our experience of information systems and process improvement with other hospitals and health care professionals, and encourage further research to assess the usability of tumor board meetings for multidisciplinary care teams. We believe that sharing knowledge and experience will drive advances in health care and improve patient outcomes.

The most significant impact of an optimized workflow is its support for timely, data-driven decisions. By integrating fragmented processes and data, the oncology center can more effectively manage the operations of each tumor board workflow for different cancer types. This increases efficiency in the preparation of meeting materials and enables standardization of meetings.

Acknowledgments

The authors would like to express their appreciation to the lung cancer MDT members and participants from the Cancer Center, AI Center, and Information Systems Department in Changhua Christian Hospital. This study was supported in part by Roche Diagnostics Ltd.

Authors' Contributions

CHL, BYW, SHL, and CJL initiated the conception. CHL, BYW, SHL, CJL, YTH, and MLP designed this study. PHS, CJL, YTH, SCC, and MLP performed experiments. MLP wrote this paper and prepared the figures. CHL and SCC revised this paper. All the authors reviewed this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information.
[DOCX File, 479 KB - cancer_v11i1e53887_app1.docx]

References

1. Patkar V, Acosta D, Davidson T, Jones A, Fox J, Keshtgar M. Cancer multidisciplinary team meetings: evidence, challenges, and the role of clinical decision support technology. Int J Breast Cancer 2011;2011:831605. [doi: 10.4061/2011/831605] [Medline: 22295234]
2. Scott B. Multidisciplinary team approach in cancer care: a review of the latest advancements. EMJ Oncol 2021;9(9):2-13 [FREE Full text]
3. Taberna M, Moncayo FG, Jané-Salas E, et al. The multidisciplinary team (MDT) approach and quality of care. Front Oncol 2020;10:85. [doi: 10.3389/fonc.2020.00085] [Medline: 32266126]
4. Das IP, Baker M, Altice C, Castro KM, Brandys B, Mitchell SA. Outcomes of multidisciplinary treatment planning in US cancer care settings. Cancer 2018 Sep 15;124(18):3656-3667. [doi: 10.1002/cncr.31394] [Medline: 30216477]
5. Lamb BW, Brown KF, Nagpal K, Vincent C, Green JSA, Sevdalis N. Quality of care management decisions by multidisciplinary cancer teams: a systematic review. Ann Surg Oncol 2011 Aug;18(8):2116-2125. [doi: 10.1245/s10434-011-1675-6] [Medline: 21442345]
6. Taylor C, Atkins L, Richardson A, Tarrant R, Ramirez AJ. Measuring the quality of MDT working: an observational approach. BMC Cancer 2012 May 29;12:202. [doi: 10.1186/1471-2407-12-202] [Medline: 22642614]
7. Croke JM, El-Sayed S. Multidisciplinary management of cancer patients: chasing a shadow or real value? An overview of the literature. Curr Oncol 2012 Aug;19(4):e232-e238. [doi: 10.3747/co.19.944] [Medline: 22876151]
8. Hammer RD, Fowler D, Sheets LR, Siadimas A, Guo C, Prime MS. Digital tumor board solutions have significant impact on case preparation. JCO Clin Cancer Inform 2020 Aug;4:757-768. [doi: 10.1200/CCI.20.00029] [Medline: 32816529]
9. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital “advanced use” divide. J Am Med Inform Assoc 2017 Nov 1;24(6):1142-1148. [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
10. Simpson L. A measurement and collection of AONN+ metrics within an automated tumor board workflow. J Oncol Navig Surviv 2019;10(2) [FREE Full text]
11. Clarke S, Wilson ML, Terhaar M. Using dashboard technology and clinical decision support systems to improve heart team efficiency and accuracy: review of the literature. Stud Health Technol Inform 2016;225:364-366. [doi: 10.3233/978-1-61499-658-3-364] [Medline: 27332223]
12. Pender A, Garcia-Murillas I, Rana S, et al. Efficient genotyping of KRAS mutant non-small cell lung cancer using a multiplexed droplet digital PCR approach. PLoS ONE 2015;10(9):e0139074. [doi: 10.1371/journal.pone.0139074] [Medline: 26413866]
13. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk Manag Healthc Policy 2011;4:47-55. [doi: 10.2147/RMHP.S12985] [Medline: 22312227]
14. Booth RG. Educating the future eHealth professional nurse. Int J Nurs Educ Scholarsh 2006;3:Article 13. [doi: 10.2202/1548-923X.1187] [Medline: 16646940]
15. Rosenberg L. Are healthcare leaders ready for the real revolution? J Behav Health Serv Res 2012 Jul;39(3):215-219. [doi: 10.1007/s11414-012-9285-z] [Medline: 22736047]
16. Hammer RD, Prime MS. A clinician’s perspective on co-developing and co-implementing a digital tumor board solution. Health Informatics J 2020 Sep;26(3):2213-2221. [doi: 10.1177/1460458219899841] [Medline: 31969041]
17. HL7 FHIR. 2023. URL: https://hl7.org/fhir/ [accessed 2025-04-24]
18. Krupinski EA, Comas M, Gallego LG, GISMAR Group. A new software platform to improve multidisciplinary tumor board workflows and user satisfaction: a pilot study. J Pathol Inform 2018;9:26. [doi: 10.4103/jpi.jpi_16_18] [Medline: 30167341]
19. Creswell JW. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 4th edition: SAGE Publications, Inc; 2018.
20. Yin RK. Case Study Research and Applications: Design and Methods, 6th edition: SAGE Publications, Inc; 2018.

21. Bleeker SE, Derksen-Lubsen G, van Ginneken AM, van der Lei J, Moll HA. Structured data entry for narrative data in a broad specialty: patient history and physical examination in pediatrics. BMC Med Inform Decis Mak 2006 Jul 13;6:29. [doi: 10.1186/1472-6947-6-29] [Medline: 16839414]
22. Janssen A, Robinson T, Brunner M, Harnett P, Museth KE, Shaw T. Multidisciplinary teams and ICT: a qualitative study exploring the use of technology and its impact on multidisciplinary team meetings. BMC Health Serv Res 2018 Jun 13;18(1):444. [doi: 10.1186/s12913-018-3242-3] [Medline: 29898716]

Abbreviations

AEIOU: Activities, Environments, Interactions, Objects, and Users
API: application programming interface
EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
HIS: hospital information system
MDT: multidisciplinary team
NGS: next-generation sequencing
NTB: NAVIFY Tumor Board

Edited by N Cahill; submitted 29.10.23; peer-reviewed by G Grieve, S Lampridis; revised version received 24.03.25; accepted 24.03.25; published 05.05.25.

Please cite as:
Lin CH, Wang BY, Lin SH, Shih PH, Lee CJ, Huang YT, Chen SC, Pan ML
Process Re-Engineering and Data Integration Using Fast Healthcare Interoperability Resources for the Multidisciplinary Treatment of Lung Cancer
JMIR Cancer 2025;11:e53887
URL: https://cancer.jmir.org/2025/1/e53887
doi:10.2196/53887

Adapting a Self-Guided eHealth Intervention Into a Tailored Therapist-Guided eHealth Intervention for Survivors of Colorectal Cancer

Johanne Dam Lyhne[1], MD; Allan 'Ben' Smith[2], PhD; Tina Birgitte Wisbech Carstensen[3,4], PhD; Lisa Beatty[5,6], PhD; Adeola Bamgboje-Ayodele[7,8], PhD; Britt Klein[9,10], PsyD; Lars Henrik Jensen[1], MD, PhD; Lisbeth Frostholm[3,4], Prof Dr

[1]Department of Clinical Oncology, University Hospital of Southern Denmark, Beriderbakken 4, Vejle, Denmark

[10]Biopsychosocial and eHealth Research & Innovation Hub, Federation University Australia, Ballarat, Australia

[2]Daffodil Centre, A joint venture between Cancer Council NSW and University of Sydney, Sydney, Australia

[3]Clinic for Functional Disorders, Aarhus University Hospital, Aarhus, Denmark

[4]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

[5]Flinders University Institute of Mental Health & Wellbeing, College of Education, Psychology & Social Work, Flinders University, Adelaide, Australia

[6]Psycho-Oncology Cooperative Research Group, Sydney, Australia

[7]Ingham Institute for Applied Medical Research, South West Sydney Clinical Campuses, UNSW Medicine & Health, University of New South Wales, Liverpool, New South Wales, Australia

[8]Biomedical Informatics and Digital Health, School of Medical Sciences, Charles Perkins Centre, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[9]Health Innovation & Transformation Centre, Federation University Australia, Ballarat, Australia

Corresponding Author:
Johanne Dam Lyhne, MD
Department of Clinical Oncology, University Hospital of Southern Denmark, Beriderbakken 4, Vejle, Denmark

Abstract

Therapist-guided eHealth interventions have been shown to engage users more effectively and achieve better outcomes than self-guided interventions when addressing psychological symptoms. Building on this evidence, this viewpoint aimed to describe the adaptation of iConquerFear, a self-guided eHealth intervention targeting fear of cancer recurrence, into a therapist-guided version (TG-iConquerFear) tailored specifically for survivors of colorectal cancer (CRC). The goal was to optimize patient outcomes while minimizing the need for extensive resources. The adaptation process followed the Information System research framework, which facilitated a systematic integration of knowledge and iterative testing. Drawing on insights from the original iConquerFear development, as well as feedback from end users, oncologists, and therapists, we began by identifying areas for improvement. These insights formed the foundation for the first design cycle. Initial internal testing revealed the need for several adjustments to enhance the intervention. While the core concept of iConquerFear remained unchanged, we made significant modifications to improve access by optimizing the platform for mobile devices, to support adherence by expanding the exercises, and to equip therapists with tools such as reflective questions and a monitoring control panel. External field testing with 5 survivors of CRC provided further validation. Participants reported a high level of acceptability, and their feedback guided additional minor points to consider incorporating in future versions. This study illustrates how a self-guided eHealth intervention can be successfully adapted into a therapist-guided format for fear of cancer recurrence, tailored to meet the needs of survivors of CRC. The described approach serves as a valuable framework for integrating therapist guidance into similar interventions, ensuring their relevance and effectiveness for targeted populations.

Trial Registration: ClinicalTrials.gov NCT04287218; https://clinicaltrials.gov/study/NCT04287218

International Registered Report Identifier (IRRID): RR2-10.1186/s12885-020-06731-6

(JMIR Cancer 2025;11:e63486) doi:10.2196/63486

KEYWORDS

fear of cancer recurrence; therapist-guided; self-guided; online intervention; colorectal cancer; digital health; psychosocial intervention; survivorship; eHealth; adaptation; survivors; oncologists; therapists; acceptability; mobile phone

Introduction

eHealth interventions, defined as programs that provide information and support for physical or mental health problems via digital platforms [1], can overcome barriers to accessing support including travelling distance to site of intervention, time constraints, disease burden, financial issues, perceived stigma, and mobility or logistics constraints due to pandemics such as COVID-19 [2]. eHealth interventions in diverse cancer settings address challenges related to scalability and cost-effectiveness with effects comparable to traditional face-to-face therapy [3-7]. However, the initial development, evaluation, and implementation of effective eHealth interventions is a complex and multidisciplinary process [8], which requires substantive financial and human resources [9].

eHealth interventions may fill an important gap in psychosocial cancer care especially, by augmenting limited available services [10]. However, the efficacy of self-guided psychological eHealth interventions can be limited by low uptake and engagement [11]. Furthermore, these interventions might increase disparities in health care, as those with digital skills and more resources will be more likely to engage [12], while those with late effects such as peripheral neuropathy and fatigue might face challenges in using required devices (eg, computers, keyboards, or mice) [13]. Promisingly, meta-analytic evidence shows that adding guidance to interventions yields greater efficacy when treating anxiety, distress, fatigue [14], and fear of cancer recurrence (FCR) [7] in people living with cancer compared with nonguided interventions. However, guidance comes with greater costs and reduced scalability due to the use of health care personnel, infrastructure, and safety measures [15-19].

Leveraging existing self-guided eHealth interventions is one way of reducing resources required to design an entirely new guided eHealth service. Several psycho-oncological interventions have undergone successful adaptations, for example FindingMyWay [20] from Australia and Fear Of Recurrence Therapy [21] from Canada. FindingMyWay has undergone 2 adaptations: first, into a UK-version needing a contextual adaptation [22] to reflect the UK health care system and terminology, and second, to a metastatic-breast cancer specific version (Finding My Way-Advanced [23]). Fear Of Recurrence Therapy was adapted to family caregivers and to an eHealth format [24].

Intervention models that facilitate greater access to FCR treatment have been identified as a top international research priority [25]. To address this priority, Smith et al [26] adapted the effective face-to-face therapist-delivered ConquerFear [27] FCR treatment to an eHealth self-guided format (iConquerFear). However, feasibility testing of iConquerFear revealed that some individuals needed guidance and more relatable content for optimal engagement and benefit [28], as reported for other psychological symptoms [15,29]. To address these two key recommendations, this current study aimed to adapt iConquerFear into a tailored asynchronous therapist-guided eHealth version (TG-iConquerFear).

While guidance can be provided either in real time (synchronous) or as delayed (asynchronous), a systematic review by Cox et al [30] has shown the superior convenience, flexibility, and limited interruptions of daily routines with asynchronous guidance in telehealth interventions. As recommendations are mixed [31,32] the choice of asynchronous guidance will be evaluated.

To make intervention content more personal and relatable, a further aim of this study was to tailor iConquerFear specifically for survivors of colorectal cancer (CRC). CRC ranks as the third most frequently diagnosed cancer worldwide, with its prevalence steadily rising due to prolonged survivorship [33]. The prevalence and characteristics of FCR experienced by individuals affected by CRC have been described in detail [34-40]. However, no intervention customized to address FCR in survivors of CRC has been developed [7,41].

In summary, this paper describes the process of adapting iConquerFear into TG-iConquerFear targeting survivors of CRC. We report using the Information System research framework [42] to integrate recommended improvements from the original Australian development study [26] and pilot study [28] with end user feedback from field testing with oversight by a multidisciplinary research team as a template for other researchers seeking to make similar adaptations.

Methods

Intervention Content

The iConquerFear is a metacognitive intervention consisting of 5 modules. The content of each module is outlined in Table 1 [26].

Table . iConquerFear content and features. Features common across all modules include the following: web-based questionnaires, interactive exercises, downloadable hand-outs, progress graphs, email and SMS reminders, and safety plan.

Module	Content and features
1. Introduction and goal setting	• Introduction to FCR[a] and treatment model (survivor and clinician videos) • Values clarification and goal setting (interactive card sort exercise)
2. Attention training	• Introduction to attention training (survivor and clinician videos) • Attention training practice (audio, monitoring and feedback, and reminders)
3. Detached mindfulness	• Introduction to detached mindfulness (survivor and clinician videos) • Detached mindfulness practice (audio, monitoring and feedback, and reminders)
4. Learning to live well and manage worry	• Psychoeducation about appropriate threat monitoring (clinician video) • Compliance with follow-up and self-examination (assessment with feedback) • Challenging unhelpful metacognition (assessment with feedback) • Worry management techniques (textual overview and downloadable PDF)
5. Treatment summery and relapse plan	• Reflection on change in FCR during treatment (assessment with feedback) • Consolidation of new strategies for managing FCR through relapse prevention (personalized feedback and downloadable action plan)

[a]FCR: fear of cancer recurrence.

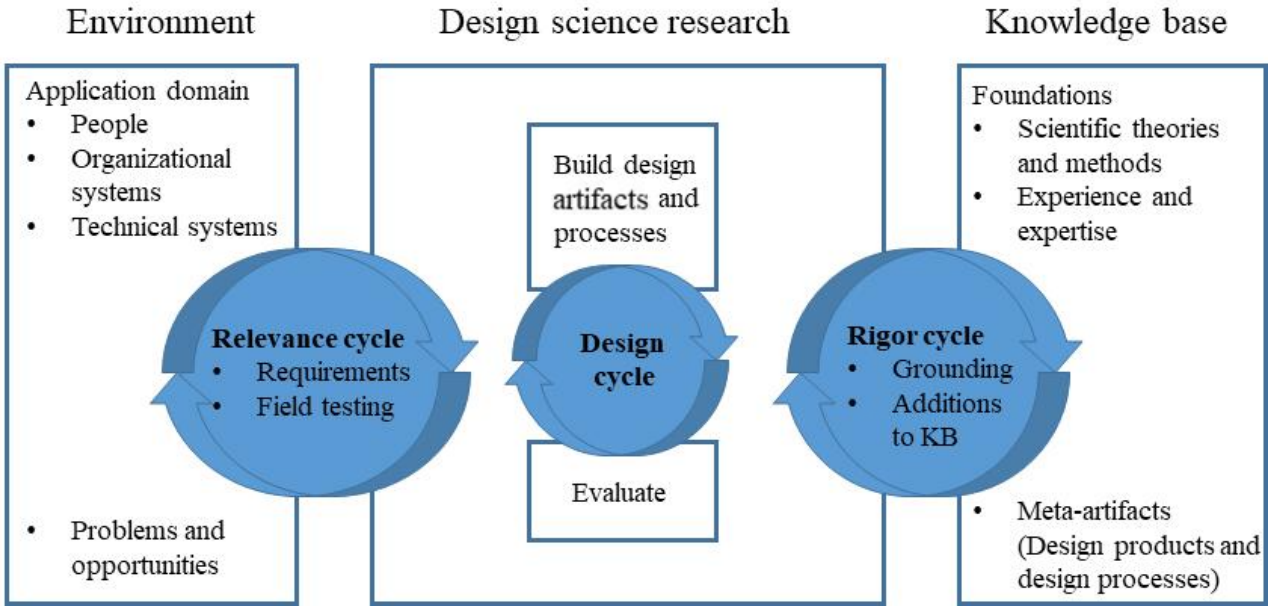
Language and Cultural Adaptation

The adaptation of iConquerFear to TG-iConquerFear was preceded by a draft translation and cultural adaptation of all written material from iConquerFear by a professional translator with experience in the psychiatric setting (Multimedia Appendix 1). This Danish draft was then built into an existing web-based treatment platform made available by the Department for Functional Disorders at Aarhus University Hospital in Denmark. The design of the Danish platform was comparable to that of iConquerFear. When needed, new technical features were programmed to reflect the intervention treatment content of iConquerFear.

Framework

A participatory design approach, guided by the Information System research framework [42], was used in the adaptation process. The Information System research framework urges end users' inclusion and active engagement in designing and evaluating information systems. The framework consists of 3 overarching user participatory design cycles: The relevance cycle determines end user requirements; the design cycle involves prototype development and evaluation; the rigor cycle focuses on assessing "past knowledge" from the knowledge base (KB) and underpinning theories (Figure 1) [42].

Figure 1. The Information System research framework. KB: knowledge base.



Adaptation Process

At each stage of adaptation, the intervention was reviewed by a multidisciplinary research team (8 members), comprised of: an oncologist specializing in CRC, a health psychologist with expertise in developing guided online interventions, one of the iConquerFear developers, 2 psychologists specializing in anxiety-related online therapy (the therapists), a health researcher with expertise in psycho-oncology, and 2 survivors of CRC.

Relevance Cycle

First, information related to the adapted intervention's relevance were gathered from the environment. Three consumer representatives (community partners from the Danish Bowel Cancer Association and the Association for Late Effects After Cancer), and 4 people with lived experience (survivors of CRC) were consulted for problem definition and justification of the interventions' relevance, including need and potential challenges. Within the same two 1-hour focus groups, the main elements from the iConquerFear intervention were presented to explore views on whether the online platform would be relevant in addressing survivors of CRC needs and how to promote engagement. The challenges raised in these interviews were validated with existing literature on online intervention engagement. The environment was revisited twice during internal and external field testing.

Rigor Cycle

Existing knowledge from the KB consisted mainly of expertise and experience from the development [26] and piloting of iConquerFear [28], which was rigorously developed based on the ConquerFear therapy manual [43] and iterative user feedback. This knowledge was complemented by feedback on the iConquerFear program from the TG-iConquerFear research team. Research team members evaluated iConquerFear content to assess suitability for survivors of CRC and the therapist-guided format. Feedback was collected via email or shared during live meetings. A selected group of research team members reviewed all aspects of the modules, including exercises, text, images, graphics, examples, and videos. Group discussions were held to assess their relevance to survivors of CRC experiencing elevated FCR. The unique needs and challenges specific for survivors of CRC were also deliberated, playing a crucial role in making intervention content more personal and relatable. The core therapeutic concept of iConquerFear, including general structure, therapeutic principles, and key goals remained unaltered in TG-iConquerFear.

Design Cycle

Insights added through the relevance cycle and rigor cycle contributed to the design science research consisting of implementation of all knowledge and suggestions into iConquerFear. The loop in the design cycle was between members of the research team and 3 software designers. Multiple iterations of the design cycle were conducted before internally field-testing TG-iConquerFear version 1 in the environment as part of a further relevance cycle. The 3 cycles allowed for an iterative process resulting in a second version, which was field-tested externally.

Field Testing

During both internal and external field testing, the intervention was used as intended based on the original manual from PoCoG (Psycho-Oncology Co-Operative Research Group) in Australia, with therapist guidance provided over a span of 10 weeks. The amount of therapist-guidance was measured by assessing the quality, quantity and length of messages exchanged between therapist and participant in each module. Every participant had a designated personal therapist and messages were answered within 3 working days. As the main focus of the guidance was to increase engagement and adherence, the substance of the guidance was not predetermined or restricted, and no limits were established concerning frequency of communication. It was planned that there would be communication between participant and therapist at least once a week. Both therapist and participant could initiate communication. Specific tasks for the therapist included welcoming the participant, addressing

inquiries regarding content and tools, providing feedback on exercises within the program, and motivating engagement and adherence.

Internal Field Testing

The internal field testing was performed within the research team. The 2 survivors of CRC were test-participants and the 2 therapists, who have extensive experience in online therapy, provided the guidance. Both the survivors of CRC and the therapists gave written feedback on various feasibility measures, including usability, adherence, acceptability, and safety after each module and at the end of the intervention.

External Field Testing

For the external field testing, interested volunteers from the bowel cancer association were informed about the project by telephone by the primary investigator and subsequently screened for eligibility via an online questionnaire sent directly to their digital citizen mailbox (used for communication between citizens and public authorities in Denmark). Volunteers were eligible if they were aged 18 years or older, had completed surgery for CRC with curative intent, were without sign of recurrence, and gave electronic informed consent. A minimum Fear of Cancer Recurrence Inventory–Short Form score of 13, indicating moderate or higher FCR levels [44], was also required. Volunteers were excluded if they self-reported clinical levels of depression, psychotic illness, or abuse of alcohol or drugs. Volunteer participants were given a unique link to the online platform of TG-iConquerFear. The guidance was performed by the same 2 therapists from the research team. Participants and therapists were encouraged to provide written feedback to the primary investigator on usefulness and suggestions for improvement after completing each module, and at the end of the intervention. This study was approved by the Regional Research Ethics Committee of Southern Denmark (S—20190061).

Results

TG-iConquerFear Version 1

During the initial relevance cycle, group interviews with representatives from the environment (consumer representatives and survivors of CRC) justified adapting iConquerFear to better engage prospective participants. Identified challenges included concerns about the complexity of pages overloaded with text (“You need to get to the point faster”), relevance of certain elements (such as sunscreen and breast palpation; “that just annoys me”), videos featuring only women (“seems like something only women have?”), and assessing further support if needed. “… Being on your own” was a major concern, and given that a large portion of end users are expected to be used, the focus group participants favored asynchronous communication. Suggestions for adaptations from the KB were either general or focused on 2 key recommendations: enhancing personal and relatable content, and adding guidance. Suggestions for adaptations from the research team centered on how to integrate therapist guidance, drawing from experience with eHealth interventions targeting health anxiety, and how to tailor hand-outs to address the specific needs of survivors of CRC, particularly regarding follow-up and late effects. These suggestions served as guidance for the design cycle, during which the adapted TG-iConquerFear version 1 was crafted. The specific adaptations made in response to these suggestions are outlined in Figure 2.

TG-iConquerFear version 1 then progressed to the relevance cycle, where it underwent internal field testing in the environment.

Figure 2. Specific adaptations made in response to suggestions from the environment and the knowledge base. CRC: colorectal cancer; FCR: fear of cancer recurrence.

Adaptions made in response to general suggestions:

- A forewarning about the difficulty of the attention training exercise was implemented.
- Clear links to additional support were added.
- The "How to" navigation video was replaced with a short descriptive text.
- Tailored automatic feedback was delivered more smoothly with fewer clicks.
- The text versus visual content balance was tipped toward more images.
- Measuring FCR was reduced from 5 times to twice during the 10 weeks of intervention.
- Emphasis was directed towards refining exercises and tools.
- Sections featuring nonessential descriptive text were succinctly edited.

Adaption made to enhance "personal and relatable" content:

- A male consumer consultant was added (65 years old).
- The mood of people appearing on the images was more representative for survivors with severe FCR.
- Pictures representing sex and age of CRCS were chosen.
- All content unrelated to CRC, such as advises on sunscreen use and instructions for breast self-examinations, was revised.
- Examples were made relevant and manageable to survivors of colorectal cancer.
- Existing handouts were modified to target Danish survivors of colorectal cancer (especially the handout on cancer follow-up).
- New handouts concerning late effects of cancer treatment and the Danish Colorectal Cancer Screening Program were formulated.

Adaptions made to add therapist guidance:

- A recognizable message-logo in the upper right corner on the dashboard was added.
- A securely encrypted message system for written communication that could gather and save all messages up to 6 months after intervention was added.
- Information pages were updated to inform about the collaboration with the therapist and how to be in contact.
- All "save" buttons in relation to exercises were changed to "send" buttons.

Internal Field Testing

The outcomes from the internal field testing are presented in Table 2. The feedback was presented to the research team, who determined that a second design cycle was necessary. Adjustments were made to TG-iConquerFear version 1 (last column in Table 2), leading to TG-iConquerFear version 2.

Table . Outcomes from internal field testing and adjustments.

Feasibility measure and feedback from end users		Adjustments
Usability		
	"It has proven to be a barrier for me that the system cannot be accessed via a smartphone. In an already busy daily life, I repeatedly find that I do not sit down at my PC." [Female participant, aged 51 years]	The intervention was adjusted to enable easier access on smart phone and iPad.
	"It takes really long time to figure out how far the participants are within the program, and if they had made any new exercises since last therapist log-in." [Female therapist, aged 38 years]	A therapist-monitored control panel with information on all active participants within treatment, including registration of activity for each participant, and expected date for completion of the program was added.
Usability of the asynchronous communication		
	"In many ways, it's easier to sit here and write with you than if we were face to face - I like that I have the opportunity to write something, consider, rephrase, etc - or just write freely in a flood of 'unloading,' depending on what I need on that day." [Female participant, aged 51 years]	No adjustments
	"For me as a therapist, asynchronous or delayed communication between me and the patient, means that I don't have to schedule specific agreed-upon sessions with patients… Additionally, I can take my time to consider the responses I give to patients. Sometimes, in face-to-face interactions, it can be challenging to find the right words or consider the right way to challenge the patient in the moment. Asynchronous communication with the patients makes my work less stressful and more flexible." [Female therapist, aged 48 years]	No adjustments
Adherence		
	"We need a clear indication from the participants on whether or not they have been working with the tools and find them useful." [Female therapist, aged 38 years]	In each module, two reflective questions with tailored feedback were added to monitor if the participants had engaged with the exercises of the previous module, and if and how the exercises had been helpful.
Acceptability		
	"The value-based module is too short for 14 days of training, and it is not clear how the participant should work with specific goals that are in line with their values. This results in a rather quick completion of the module and the participants state rather vague and abstract goals without a clear indication of when reaching the goal." [Female therapist, aged 48 years]	A lighthouse metaphor was added to explain values, see Multimedia Appendix 2, for the full formulation.
	"I find especially the first exercises (ref: the value clarification exercise) inadequately explained. For academics with a background in social sciences, the explanations are not so difficult to interpret, but they probably make up the smallest part of the target audience, which is likely more composed of older individuals from all social strata. Among them, there may be many who have difficulty benefiting from this module if they are not further guided." [Male participant, aged 67 years]	Information on the value-clarification exercise was made more concrete so that the program itself was able to guide the participant through the exercises leading the therapist to stand back and support and evaluate.
Safety		

Feasibility measure and feedback from end users		Adjustments
	"I hesitated to fully engage in this digital treatment. I needed my therapist to ensure confidentiality." [Female participant, aged 51 years]	A paragraph on confidentiality was added in the introduction.

TG-iConquerFear Version 2

Following the adjustments from the first relevance cycle, the second design cycle aimed to refine TG-iConquerFear version 1 further. Key modifications, as detailed in Table 2, were implemented to address user feedback and enhance the intervention's feasibility. This second version underwent external field testing within the environment.

External Field Testing

The external field testing took place from February to July 2022 at the Clinic for Functional Disorders at Aarhus University Hospital in Denmark. It involved 5 volunteers from the Danish Bowel Cancer Association. All 5 were survivors of colon cancer and most were female (n=4), married (n=4), and employed (n=4). Mean age was 54 (SD 10.2; range 42-71) years and average time since diagnosis were 2.3 (SD 1.6) years (8 months to 5 years).

Four test-participants completed all 5 modules, while 1 test-participant did not complete module 2. The amount of therapist guidance averaged 15.2 (range 12 - 23) messages per participant, equating to 2.7 (range 2 - 3.8) messages per module. The mean length of a message was 196.8 (range 12 - 745) words. The messages addressed topics such as welcoming participants, elaborating on goal setting (module 1), integrating attention training and detached mindfulness into daily routines (modules 2 and 3), and supporting participants in challenging unhelpful metacognitions. No messages were prompted by misunderstandings of the intervention content or technical issues. Four out of 5 test-participants accessed the intervention via smart phone or tablet. No test-participant reported troubles with smart phone access, and no comments related to the description of the values-clarification exercise, of which both were adjusted after the internal field testing. Outcomes are presented in Textbox 1.

Textbox 1. Outcomes from external field testing.

Positive feedback from test-participants

- "… I am slowly returning to life where cancer doesn't occupy as much space. Not because it's forgotten, but more integrated. The tools in your investigation were really useful." [Male, aged 71 years]
- "I use detached mindfulness regularly. Especially if I'm not occupied with something else at the moment, my mind tends to circle around cancer, and I find that the method helps to stop that." [Female, aged 43 years]
- "I am incredibly grateful that I was allowed to participate in this program. Even with the bumps along the way, it has been a huge help. In many ways, it's easier to sit here and write with you than if we were face to face - I like that I have the opportunity to write something, consider, rephrase, etc - or just write freely in a flood of 'unloading,' depending on what I need on that day... Gold!" [Female, aged 51 years]
- "I feel so lucky being part of this program. While I sit here and write, I hear the birds singing. And enjoy it. I live my life fully. I do not postpone things. I know there are late effects, but I will not have them ruin my mood…" [Female, aged 55 years]

Constructive feedback from test-participants

- "It could also be good if a notification could be sent in e-boks when there is a new message from my therapist." [Female, aged 49 years]
- "… There should be some form of notification when there is new activity on the platform." [Male, aged 71 years]
- "It could be good if one could see how far they are in the program - if the round circles on the front page could possibly change color when a module is completed. I personally know how difficult it is to remember such things when there are several days in between." [Female, aged 49 years]
- "I have copied my action plan step by step so that I can refer to it if I 'forget' it at times. Is it possible to have it compiled for printing at the end of the program? - good to hang up where it is visible in everyday life." [Female, aged 55 years]
- "…so that one can click into what concerns oneself - sometimes it can be quite depressing to read through a long list of late effects - it's not certain that one has them all oneself, but one is reminded of one's vulnerability. Possibly bubbles on a page under 'the box' with 'fatigue,' 'sleep problems,' etc…" [Female, aged 51 years]

The results were presented to the research team in April 2023. The research team deemed TG-iConquerFear version 2 as satisfactory. The minor suggestions from the external field testing were not implemented due to time and resource limitations but will be considered for future updates.

Discussion

Principal Findings

We adapted a self-guided eHealth intervention for FCR into a tailored therapist-guided eHealth intervention aimed at survivors of CRC guided by the Information System research framework. The process was overseen by a multidisciplinary research team including 2 survivors of CRC. Based on knowledge and

experience in this team, both minor and major adaptations were made. The addition of an embedded message system facilitating therapist-participant communication was the primary change in the iConquerFear program. However, during field testing it became evident that additional content (eg, the Lighthouse Metaphor or reflective questions) was necessary to optimize adherence and facilitate guidance. Consequently, TG-iConquerFear is more resource demanding than iConquerFear, and we will evaluate the efficacy and cost-effectiveness of these choices, which may limit scalability (ClinicalTrials.gov NCT04287218).

The smooth adaptation of iConquerFear into TG-iConquerFear was greatly facilitated by the extensive research led by the Australian PoCoG evaluating ConquerFear and iConquerFear [26-28,45]. This prior work served as a valuable foundation, allowing us to expedite the launch of the randomized controlled trial, as comprehensive testing of intervention content had already been conducted. Below, we discuss considerations related to the two key requests that emerged from pilot testing of iConquerFear [26] which we addressed in the first round of adaptation to TG-iConquerFear version 1.

Considerations on How to Enhance "Personal and Relatable" Properties

Some iConquerFear pilot test participants [26] indicated that certain content within the program felt impersonal or unrelatable. Two main directions were considered to promote engagement by tailoring the intervention [46]: targeting a specific gender (men) or targeting a single cancer type (CRC). Choosing a gender-based approach simplifies the selection of colors, images, patients featured in videos, and examples, aiming for enhanced engagement among participants of a specific gender. However, several modules include content pertaining to specific cancer types, for instance, living in alignment with one's values, where focusing on a specific cancer type such as CRC allows for framing personal examples closely tied to everyday life such as "what if my ostomy leaks?" Similarly, cancer type plays a defining role in shaping the cancer follow-up program, advisory elements, and the identification of alarm symptoms to be monitored. These elements do not apply universally to a specific gender, and one could argue that cancer type is a more defining factor in determining the specific needs reported by cancer survivors, rather than gender alone [47]. For instance, a female survivors of CRC is likely to have far more in common with a male survivor of CRC than a female survivor of breast cancer. By considering the cancer type and adjusting relevant advisory elements, the intervention was customized to meet the unique needs of individuals based on their specific cancer experiences. This tailored approach balances scalability and engagement by ensuring that the content remains relevant, relatable, and meaningful to participants, hopefully enhancing the engagement and effectiveness of the TG-iConquerFear intervention. No end user feedback was received regarding impersonal or unrelatable content after adaptation.

A third option involves the creation of distinct intervention versions with personalized content based on demographic factors such as age, gender, cancer type, or needs, thereby presenting only pertinent content to individual participants although this approach risks relevant content being excluded due to individual differences in patient needs and how they report them. However, the software used for our intervention was incapable of accommodating such customization. Furthermore, implementing this approach would have necessitated a substantial allocation of time and resources beyond our available capacity. Nonetheless, this avenue remains an intriguing prospect for future research endeavors, warranting further exploration and consideration.

Consideration on How to Add Therapist Guidance

Some participants in the iConquerFear pilot expressed a desire for personal contact with a researcher or clinician [26], which was seen as a potential way to motivate engagement with and potential benefit from the 8 - 10 week duration of the program. This is in line with a recent meta-review [46]. The potential benefit of adding guidance was also noted by researchers from the cancer recurrence self-help training trial, who found no effect of CBT-based online self-help training [48]. The choice of asynchronous communication was supported by end users and therapists, and the amount of guidance was consistent through the intervention. Reflective questions were incorporated into TG-iConquerFear version 1 at the therapists' request after each module to monitor adherence and perceived usefulness. Participants received automated tailored feedback momentarily, in accordance with recommendations from the literature [46,49], allowing them to seamlessly continue with the intervention, and these reflective questions facilitated a dialogue between participant and therapist regarding the usefulness of the module content, and strategies to enhance participant outcomes.

Other Strategies to Enhance Adherence

Easy Access

Difficulties in integrating the intervention into daily life activities, as exemplified by an internal test-participant, have also been reported in the literature as a barrier for adoption and adherence when targeting fatigue [50]. This issue is closely related to the barrier of perceived time burden associated with participating in digital interventions [46]. Given that only minor changes were required to adapt the TG-iConquerFear version 1 to fit a smart-phone layout, this aspect was prioritized.

Promoting Competence

In TG-iConquerFear version 1, the value-based model was inadequately explained, leading to a lack of self-competence reported by an internal test-participant. Competence can be promoted by encouraging users to set graded goals with smaller achievable steps, thereby increasing confidence through experiences of success as described in the person-based approach to intervention development [8]. Adjustments were made to clearly outline how to work with goals and values to fully use the module's potential. Additionally, tailored feedback was provided to congratulate success in goal achievement and offer remedial advice if goals were not attained [8].

Reminders

Technology-based reminders (eg, prompts) have the potential to promote engagement with digital interventions [1,46], particularly when participants choose to receive them [51]. Since

2 external test-participants explicitly suggested reminders, this could be considered for future updates. However, it should be noted that no studies in these reviews specifically targeted FCR, and the effect of distributing reminders, which could potentially trigger FCR, in this patient population remains unknown.

The TG-iConquerFear version 2 was deemed ready despite further suggestions for improvement, and in May 2023, a randomized controlled trial investigating TG-iConquerFear versus augmented treatment as usual was launched [52]. The suggestions, though noted, were seen as relatively minor, and addressing them would have required a significant allocation of resources disproportionate to their perceived impact. Given that each cycle typically generates some feedback, the multidisciplinary research team convened to make a decision regarding the intervention's readiness for the randomized controlled trial.

Strengths and Limitations

The process of adaptation was guided by an existing framework for information system research. The framework allows for multiple iterative processes, which were needed in this study. It is a strength of this study that the KB comprised insights from multiple sources. The therapists performing the guidance were experienced in delivering anxiety-related online therapy, and the software designers facilitating the adaptation process in the design cycle had expertise in health care software development. The field testing participants were relatively newly diagnosed and the females were younger than the average female patient with colon cancer. Changes due to their comments may have slightly shifted the focus of the final version of the intervention to address a somewhat younger audience. However, as FCR is more prevalent in younger than older survivors [53], and the incidence of CRC in younger patients continues to increase [54], this may be an advantage. Furthermore, we only included 5 participants in the external field testing, but all 5 participants completed all or almost all modules. This study aimed to test the adaptation of an already feasible and pilot tested intervention, which is why the small sample size was accepted. No patients with rectal cancer were included in the test group. We have no reasons to believe that their feedback would be any different as needs and late effects are comparable [55,56].

Clinical Implications

This study demonstrates a process to adapt a self-guided eHealth intervention into a tailored therapist-guided eHealth intervention, which could help efficiently address survivorship concerns such as FCR. Guided eHealth interventions are effective supplements to face-to-face intervention and could be a valuable step in a stepped-care model, where self-guided interventions might be the first step for survivors with mild symptoms [57,58]. Guided interventions require some level of involvement of health personnel, but can significantly increase the accessibility and reach of psychological interventions while promoting engagement and efficacy [14]. Determining the optimal dose of guidance for the severity of symptoms is needed [29].

Future Directions

The quality and content of therapist guidance will be assessed, alongside the investigation of TG-iConquerFear's efficacy in a larger, more diverse population across Denmark [52]. Additionally, further evaluation of TG-iConquerFear's performance across various devices, including smartphones or iPhone, tablets or iPads, and computers, will help refine its usability and ensure compatibility across platforms, thereby increasing scalability. Future studies should investigate the usability of eHealth interventions such as TG-iConquerFear for patients experiencing late effects such as fatigue and peripheral neuropathy, particularly in relation to content load and the touchscreen-based exercises. Finally, linking TG-iConquerFear with health apps or smartwatches could provide valuable insights into the usage of the intervention combined with, for example, participant activity levels and heart rate, allowing for an evaluation of the interplay between intervention content and bio-physiological and behavioral parameters.

Conclusion

It is possible to successfully adapt a self-guided eHealth intervention for people with FCR into a tailored therapist-guided intervention. This paper provides an overview of the process and lists considerations based on experience. The described procedure can be used in similar settings where the wish is to incorporate guidance in an existing self-guided eHealth intervention.

Acknowledgments

This study has been possible to conduct due to major support from the Department of Functional Disorders at Aarhus University Hospital. Our thanks to programmers Philip Riley, Nicolaj Knudsen, and Tobias Christensen and thanks to therapists Mie Wiklund, MSc, and Tina Carstensen. Our thanks to Harald Hammershøi and OPEN, Odense Patient Data Explorative Network, Odense University Hospital, to Alan Hevner, PhD, for copyright permission, and a special thanks to the 7 volunteer participants from The Danish Bowel Cancer Association. This study was supported by the Danish Cancer Society (grant R303-A17152), Trygfonden (grant 146250), Dagmar Marshalls Foundation, The Region of Southern Denmark, and Lillebaelt Hospital.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

In the preparation of this paper, ChatGPT (OpenAI) was used to refine and enhance the clarity of the English language. The artificial intelligence (AI) was specifically used for linguistic and grammatical improvements. No AI-generated content was used

to generate original research ideas, interpret data, or contribute substantively to the intellectual content of this study. The final paper was critically reviewed and approved by all authors to ensure accuracy and alignment with the intended research findings.

Authors' Contributions

JDL, AS, LHJ, and LF contributed to conceptualization and funding acquisition. JDL, AS, TBWC, LHJ, and LF participated in data curation and formal analysis. JDL, TBWC, and LF conducted the research and investigation process. JDL and LF were responsible for project administration. LF was responsible for software development. LF and LHJ worked on supervision and validation. JDL, AS, and LF handled visualization. JDL and AS wrote the first draft. All authors contributed to the design of the methodology, participated in the provision of study materials, and provided a critical review of the final draft before submission and during the publication stages.

Conflicts of Interest

AS, AB and BK hold the license with Blue Note for iConquerFear. The rest of the authors have no conflicts of interest.

Multimedia Appendix 1
All primary changes.
[DOCX File, 104 KB - cancer_v11i1e63486_app1.docx]

Multimedia Appendix 2
The lighthouse metaphor and exercise.
[DOCX File, 324 KB - cancer_v11i1e63486_app2.docx]

References

1. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 8;18(1):e6. [doi: 10.2196/jmir.4790] [Medline: 26747176]
2. Anthony Jnr B. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. J Med Syst 2020 Jun 15;44(7):132. [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
3. Compen F, Bisseling E, Schellekens M, et al. Face-to-face and internet-based mindfulness-based cognitive therapy compared with treatment as usual in reducing psychological distress in patients with cancer: a multicenter randomized controlled trial. J Clin Oncol 2018 Aug 10;36(23):2413-2421. [doi: 10.1200/JCO.2017.76.5669] [Medline: 29953304]
4. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
5. Wagner B, Horn AB, Maercker A. Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. J Affect Disord 2014 Jan;152-154:113-121. [doi: 10.1016/j.jad.2013.06.032] [Medline: 23886401]
6. Luo C, Sanger N, Singhal N, et al. A comparison of electronically-delivered and face to face cognitive behavioural therapies in depressive disorders: a systematic review and meta-analysis. EClinicalMedicine 2020 Jul;24:100442. [doi: 10.1016/j.eclinm.2020.100442] [Medline: 32775969]
7. Kang N, Yu ES. Is digital intervention for fear of cancer recurrence beneficial to cancer patients?: A systematic review and meta-analysis. Psychooncology 2023 Sep;32(9):1348-1358. [doi: 10.1002/pon.6199] [Medline: 37571918]
8. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30. [doi: 10.2196/jmir.4055] [Medline: 25639757]
9. WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. 2019 URL: https://fi-admin.bvsalud.org/login/?next=/bibliographic/new-source%3Fdocument_type%3DNm [accessed 2025-02-21]
10. Leykin Y, Thekdi SM, Shumay DM, Muñoz RF, Riba M, Dunn LB. Internet interventions for improving psychological well-being in psycho-oncology: review and recommendations. Psychooncology 2012 Sep;21(9):1016-1025. [doi: 10.1002/pon.1993] [Medline: 21608075]
11. Gan DZQ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3:764079. [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
12. Lee EW, McCloud RF, Viswanath K. Designing effective eHealth interventions for underserved groups: five lessons from a decade of eHealth intervention design and deployment. J Med Internet Res 2022 Jan 7;24(1):e25419. [doi: 10.2196/25419] [Medline: 34994700]

13. Mowlem FD, Sanderson B, Platko JV, Byrom B. Optimizing electronic capture of patient-reported outcome measures in oncology clinical trials: lessons learned from a qualitative study. J Comp Eff Res 2020 Dec;9(17):1195-1204. [doi: 10.2217/cer-2020-0143] [Medline: 33274651]
14. Akdemir A, Smith AB, Wu VS, et al. Guided versus non-guided digital psychological interventions for cancer patients: a systematic review and meta-analysis of engagement and efficacy. Psychooncology 2024 Jan;33(1):e6290. [doi: 10.1002/pon.6290] [Medline: 38282223]
15. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
16. Musiat P, Johnson C, Atkinson M, Wilksch S, Wade T. Impact of guidance on intervention adherence in computerised interventions for mental health problems: a meta-analysis. Psychol Med 2022 Jan;52(2):229-240. [doi: 10.1017/S0033291721004621] [Medline: 34802474]
17. Winter N, Russell L, Ugalde A, White V, Livingston P. Engagement strategies to improve adherence and retention in web-based mindfulness programs: systematic review. J Med Internet Res 2022 Jan 12;24(1):e30026. [doi: 10.2196/30026] [Medline: 35019851]
18. Xie LF, Itzkovitz A, Roy-Fleming A, Da Costa D, Brazeau AS. Understanding self-guided web-based educational interventions for patients with chronic health conditions: systematic review of intervention features and adherence. J Med Internet Res 2020 Aug 13;22(8):e18355. [doi: 10.2196/18355] [Medline: 32788152]
19. Wagner LI, Tooze JA, Hall DL, et al. Targeted eHealth intervention to reduce breast cancer survivors' fear of recurrence: results from the fortitude randomized trial. J Natl Cancer Inst 2021 Nov 2;113(11):1495-1505. [doi: 10.1093/jnci/djab100] [Medline: 34057469]
20. Beatty L, Kemp E, Coll JR, et al. Finding My Way: results of a multicentre RCT evaluating a web-based self-guided psychosocial intervention for newly diagnosed cancer survivors. Support Care Cancer 2019 Jul;27(7):2533-2544. [doi: 10.1007/s00520-018-4526-1] [Medline: 30411239]
21. Maheu C, Lebel S, Bernstein LJ, et al. Fear of Cancer Recurrence Therapy (FORT): a randomized controlled trial. Health Psychol 2023 Mar;42(3):182-194. [doi: 10.1037/hea0001253] [Medline: 36862474]
22. Hulbert-Williams NJ, Leslie M, Hulbert-Williams L, et al. The Finding My Way UK clinical trial: adaptation report and protocol for a replication randomized controlled efficacy trial of a web-based psychological program to support cancer survivors. JMIR Res Protoc 2021 Sep 20;10(9):e31976. [doi: 10.2196/31976] [Medline: 34542420]
23. Beatty L, Koczwara B, Butow P, et al. Development and usability testing of a web-based psychosocial intervention for women living with metastatic breast cancer: Finding My Way-Advanced. J Cancer Surviv 2021 Jun;15(3):403-409. [doi: 10.1007/s11764-021-01019-5] [Medline: 33723741]
24. Lamarche J, Cusson A, Nissim R, et al. It's time to address fear of cancer recurrence in family caregivers: usability study of an virtual version of the Family Caregiver-Fear Of Recurrence Therapy (FC-FORT). Front Digit Health 2023;5:1129536. [doi: 10.3389/fdgth.2023.1129536] [Medline: 37671170]
25. Shaw J, Kamphuis H, Sharpe L, et al. Setting an international research agenda for fear of cancer recurrence: an online delphi consensus study. Front Psychol 2021;12:596682. [doi: 10.3389/fpsyg.2021.596682] [Medline: 33692719]
26. Smith AB, Bamgboje-Ayodele A, Butow P, et al. Development and usability evaluation of an online self-management intervention for fear of cancer recurrence (iConquerFear). Psychooncology 2020 Jan;29(1):98-106. [doi: 10.1002/pon.5218] [Medline: 31483911]
27. Butow PN, Turner J, Gilchrist J, et al. Randomized trial of ConquerFear: a novel, theoretically based psychosocial intervention for fear of cancer recurrence. J Clin Oncol 2017 Dec 20;35(36):4066-4077. [doi: 10.1200/JCO.2017.73.1257] [Medline: 29095681]
28. Smith A, Bamgboje-Ayodele A, Jegathees S, et al. Feasibility and preliminary efficacy of iConquerFear: a self-guided digital intervention for fear of cancer recurrence. J Cancer Surviv 2024 Apr;18(2):425-438. [doi: 10.1007/s11764-022-01233-9] [Medline: 35876964]
29. Andrews B, Klein B, Van Nguyen H, Corboy D, McLaren S, Watson S. Efficacy of a digital mental health biopsychosocial transdiagnostic intervention with or without therapist assistance for adults with anxiety and depression: adaptive randomized controlled trial. J Med Internet Res 2023 Jun 12;25:e45135. [doi: 10.2196/45135] [Medline: 37307046]
30. Cox A, Lucas G, Marcu A, et al. Cancer survivors' experience with telehealth: a systematic review and thematic synthesis. J Med Internet Res 2017 Jan 9;19(1):e11. [doi: 10.2196/jmir.6575] [Medline: 28069561]
31. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen L. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May 1;4(3):666-684. [doi: 10.1177/193229681000400323] [Medline: 20513335]
32. Ownsworth T, Chan RJ, Jones S, Robertson J, Pinkham MB. Use of telehealth platforms for delivering supportive care to adults with primary brain tumors and their family caregivers: a systematic review. Psychooncology 2021 Jan;30(1):16-26. [doi: 10.1002/pon.5549] [Medline: 32915517]
33. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424. [doi: 10.3322/caac.21492] [Medline: 30207593]

34. Lim CYS, Laidsaar-Powell RC, Young JM, Kao SCH, Zhang Y, Butow P. Colorectal cancer survivorship: a systematic review and thematic synthesis of qualitative research. Eur J Cancer Care (Engl) 2021 Jul;30(4):e13421. [doi: 10.1111/ecc.13421] [Medline: 33733545]
35. Custers JAE, Gielissen MFM, Janssen SHV, de Wilt JHW, Prins JB. Fear of cancer recurrence in colorectal cancer survivors. Support Care Cancer 2016 Feb;24(2):555-562. [doi: 10.1007/s00520-015-2808-4] [Medline: 26108170]
36. Luigjes-Huizer YL, Tauber NM, Humphris G, et al. What is the prevalence of fear of cancer recurrence in cancer survivors and patients? A systematic review and individual participant data meta-analysis. Psychooncology 2022 Jun;31(6):879-892. [doi: 10.1002/pon.5921] [Medline: 35388525]
37. Lyhne JD, Smith AB, Timm S, et al. Validity and screening capacity of the FCR-1r for fear of cancer recurrence in long-term colorectal cancer survivors. Support Care Cancer 2023 Nov 11;31(12):690. [doi: 10.1007/s00520-023-08159-7] [Medline: 37950072]
38. Nahm SH, Blinman P, Butler S, Tan SYC, Vardy J. Factors associated with fear of cancer recurrence in breast and colorectal cancer survivors: a cross-sectional study of cancer survivors. Asia Pac J Clin Oncol 2021 Jun;17(3):222-229. [doi: 10.1111/ajco.13434] [Medline: 33079491]
39. Jakobsen IH, Jeppesen MM, Simard S, Thaysen HV, Laurberg S, Juul T. Initial validation of the Danish version of the Fear of Cancer Recurrence Inventory (FCRI) in colorectal cancer patients. J Cancer Surviv 2018 Dec;12(6):723-732. [doi: 10.1007/s11764-018-0709-5] [Medline: 30128857]
40. Krok-Schoen JL, Naughton MJ, Bernardo BM, Young GS, Paskett ED. Fear of recurrence among older breast, ovarian, endometrial, and colorectal cancer survivors: findings from the WHI LILAC study. Psychooncology 2018 Jul;27(7):1810-1815. [doi: 10.1002/pon.4731] [Medline: 29644766]
41. Tauber NM, O'Toole MS, Dinkel A, et al. Effect of psychological intervention on fear of cancer recurrence: a systematic review and meta-analysis. J Clin Oncol 2019 Nov 1;37(31):2899-2915. [doi: 10.1200/JCO.19.00572] [Medline: 31532725]
42. Hevner AR. A three cycle view of design science research. Scand J Inf Sys 2007;19(2):4 [FREE Full text]
43. Butow PN, Bell ML, Smith AB, et al. Conquer fear: protocol of a randomised controlled trial of a psychological intervention to reduce fear of cancer recurrence. BMC Cancer 2013 Apr 23;13:201. [doi: 10.1186/1471-2407-13-201] [Medline: 23617696]
44. Simard S, Savard J. Screening and comorbidity of clinical levels of fear of cancer recurrence. J Cancer Surviv 2015 Sep;9(3):481-491. [doi: 10.1007/s11764-015-0424-4] [Medline: 25603948]
45. Butow P, Sharpe L, Thewes B, Turner J, Gilchrist J, Beith J. Fear of cancer recurrence: a practical guide for clinicians. Oncology (Williston Park) 2018 Jan 15;32(1):32-38. [Medline: 29447419]
46. Leslie M, Beatty L, Hulbert-Williams L, et al. Web-based psychological interventions for people living with and beyond cancer: meta-review of what works and what does not for maximizing recruitment, engagement, and efficacy. JMIR Cancer 2022 Jul 8;8(3):e36255. [doi: 10.2196/36255] [Medline: 35802418]
47. Sanson-Fisher R, Girgis A, Boyes A. The unmet supportive care needs of patients with cancer. Cancer 2000 Jan 1;88(1):226-237. [doi: 10.1002/(SICI)1097-0142(20000101)88:1<>1.0.CO;2-G] [Medline: 10618627]
48. van Helmondt SJ, van der Lee ML, van Woezik RAM, Lodder P, de Vries J. No effect of CBT-based online self-help training to reduce fear of cancer recurrence: first results of the CAREST multicenter randomized controlled trial. Psychooncology 2020 Jan;29(1):86-97. [doi: 10.1002/pon.5233] [Medline: 31595627]
49. Morrison L, Moss-Morris R, Michie S, Yardley L. Optimizing engagement with internet-based health behaviour change interventions: comparison of self-assessment with and without tailored feedback using a mixed methods approach. Br J Health Psychol 2014 Nov;19(4):839-855. [doi: 10.1111/bjhp.12083] [Medline: 24308806]
50. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: a systematic review and meta-analysis. Psychooncology 2017 Sep;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
51. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86. [doi: 10.2196/jmir.2583] [Medline: 23598614]
52. Lyhne JD, Frostholm L, Fink P, Jensen LH. Study protocol: a randomized controlled trial comparing the efficacy of therapist guided internet-delivered cognitive therapy (TG-iConquerFear) with augmented treatment as usual in reducing fear of cancer recurrence in Danish colorectal cancer survivors. BMC Cancer 2020 Mar 16;20(1):223. [doi: 10.1186/s12885-020-06731-6] [Medline: 32178640]
53. Lim E, Humphris G. The relationship between fears of cancer recurrence and patient age: a systematic review and meta-analysis. Cancer Rep (Hoboken) 2020 Jun;3(3):e1235. [doi: 10.1002/cnr2.1235] [Medline: 32671982]
54. Siegel RL, Torre LA, Soerjomataram I, et al. Global patterns and trends in colorectal cancer incidence in young adults. Gut 2019 Dec;68(12):2179-2185. [doi: 10.1136/gutjnl-2019-319511] [Medline: 31488504]
55. Juul T, Bräuner AB, Drewes AM, et al. Systematic screening for late sequelae after colorectal cancer-a feasibility study. Colorectal Dis 2021 Feb;23(2):345-355. [doi: 10.1111/codi.15519] [Medline: 33420746]

56. Mekhael M, Larsen HM, Lauritzen MB, et al. Bowel dysfunction following pelvic organ cancer: a prospective study on the treatment effect in nurse-led late sequelae clinics. Acta Oncol 2023 Jan;62(1):70-79. [doi: 10.1080/0284186X.2023.2168214] [Medline: 36757368]
57. Pradhan P, Sharpe L, Menzies RE. Towards a stepped care model for managing fear of cancer recurrence or progression in cancer survivors. Cancer Manag Res 2021;13:8953-8965. [doi: 10.2147/CMAR.S294114] [Medline: 34880676]
58. Tran MJ, Jefford M, Smith B, et al. Study protocol for the evaluation of Fear-Less: a stepped-care program for fear of cancer recurrence in survivors with early-stage disease. Pilot Feasibility Stud 2022 Aug 10;8(1):177. [doi: 10.1186/s40814-022-01123-y] [Medline: 35948952]

ABBREVIATIONS

CRC: colorectal cancer
FCR: fear of cancer recurrence
KB: knowledge base
PoCoG: Psycho-Oncology Co-Operative Research Group
TG-iConquerFear: Therapist Guided iConquerFear

Edited by N Cahill; submitted 21.06.24; peer-reviewed by D Bracken-Clarke, N Evbarunegbe; revised version received 22.01.25; accepted 30.01.25; published 05.03.25.

Please cite as:
Lyhne JD, Smith A', Carstensen TBW, Beatty L, Bamgboje-Ayodele A, Klein B, Jensen LH, Frostholm L
Adapting a Self-Guided eHealth Intervention Into a Tailored Therapist-Guided eHealth Intervention for Survivors of Colorectal Cancer
JMIR Cancer 2025;11:e63486
URL: https://cancer.jmir.org/2025/1/e63486
doi:10.2196/63486

Viewpoint

Developing Effective Frameworks for Large Language Model–Based Medical Chatbots: Insights From Radiotherapy Education With ChatGPT

James C L Chow[1,2*], PhD; Kay Li[3*], PhD

[1]Department of Medical Physics, Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada
[2]Department of Radiation Oncology, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[3]Department of English, Faculty of Arts and Science, University of Toronto, Toronto, ON, Canada
[*]all authors contributed equally

Corresponding Author:
James C L Chow, PhD
Department of Medical Physics
Princess Margaret Cancer Centre
University Health Network
Rm 7-606, 7/F, 700 University Ave
Toronto, ON, M5G 1X6
Canada
Phone: 1 416 946 4501
Fax: 1 416 946 6566
Email: james.chow@uhn.ca

Abstract

This Viewpoint proposes a robust framework for developing a medical chatbot dedicated to radiotherapy education, emphasizing accuracy, reliability, privacy, ethics, and future innovations. By analyzing existing research, the framework evaluates chatbot performance and identifies challenges such as content accuracy, bias, and system integration. The findings highlight opportunities for advancements in natural language processing, personalized learning, and immersive technologies. When designed with a focus on ethical standards and reliability, large language model–based chatbots could significantly impact radiotherapy education and health care delivery, positioning them as valuable tools for future developments in medical education globally.

(JMIR Cancer 2025;11:e66633) doi:10.2196/66633

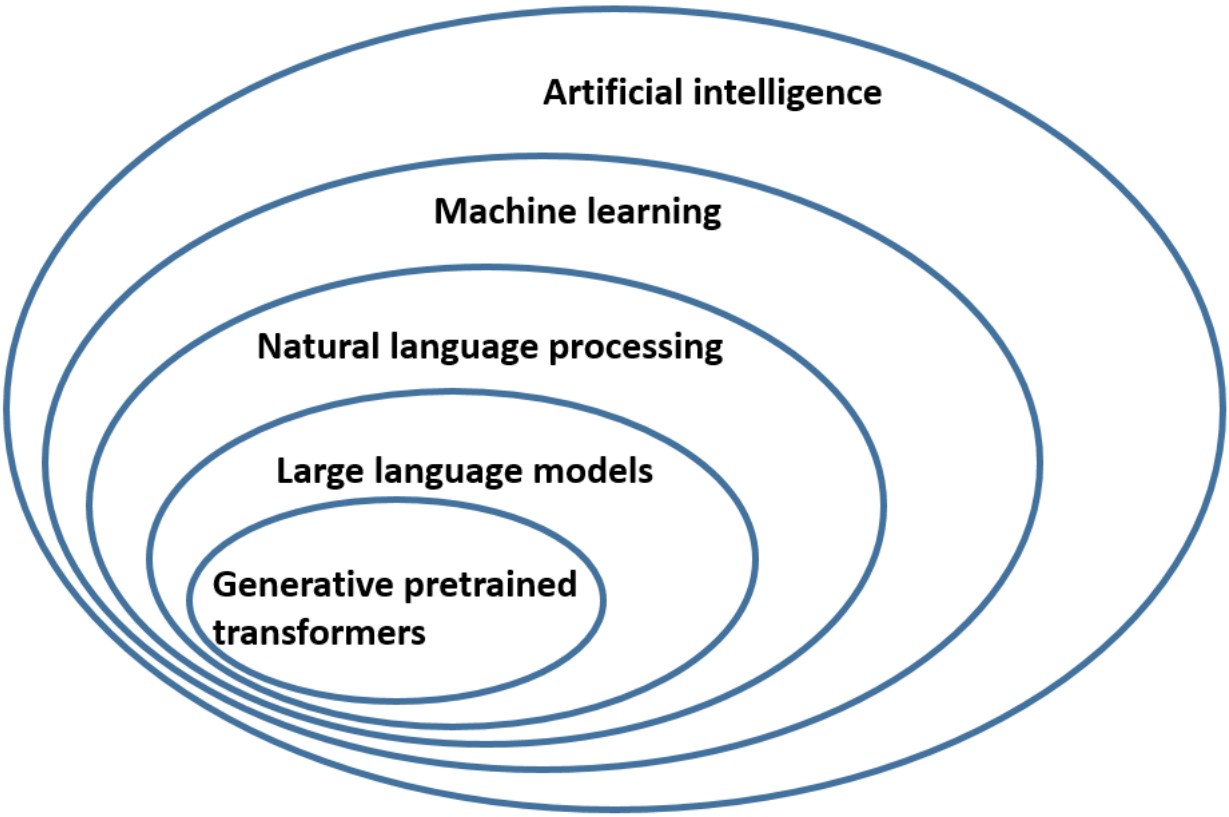
KEYWORDS

artificial intelligence; AI; AI in medical education; radiotherapy chatbot; large language models; LLMs; medical chatbots; health care AI; ethical AI in health care; personalized learning; natural language processing; NLP; radiotherapy education; AI-driven learning tools

Introduction

The integration of artificial intelligence (AI) in health care has rapidly evolved, with large language models (LLMs) such as ChatGPT at the forefront of this change [1-3]. These models, trained on vast datasets, have shown remarkable potential in various domains, including health care, by understanding and generating human-like text [4]. Unlike traditional rule-based chatbots, LLM-based systems can engage in complex conversations, answer medical queries, assist in symptom checking, and provide personalized health advice [5]. This capability is largely due to their training on extensive medical literature and patient data [6], which allow them to access and synthesize information in ways that were previously unimaginable [7]. The relationship between AI, machine learning, natural language processing (NLP), LLMs, and generative pretrained transformers is shown in Figure 1.

Figure 1. Relationships between key concepts in artificial intelligence, including machine learning, natural language processing, large language models, and generative pretrained transformers.



The deployment of LLM-based chatbots in health care is driven by the growing demand for accessible information and the need to alleviate the burden on health care systems [8]. These chatbots operate 24-7, providing immediate responses to patient inquiries; enhancing patient education; and supporting telemedicine, mental health, and chronic disease management [9-13]. However, their integration also presents challenges, such as ensuring the accuracy of information, avoiding biased responses, and safeguarding patient privacy [14].

Radiotherapy is a specialized field within oncology that demands precise and accurate information for effective treatment planning and education [15]. Unlike general medical topics, it involves complex concepts such as dose distribution, radiation physics, and the biological effects of ionizing radiation. Inaccurate information can lead to severe consequences, including inadequate treatment or harm to patients, making reliable dissemination critical [16].

Radiotherapy education serves 3 primary groups: patients, radiation staff, and the public. Patients require clear and empathetic guidance to understand their treatment, manage expectations, and make informed decisions, improving adherence and outcomes [17]. For health care professionals, education focuses on advanced training in treatment planning, safety protocols, and technological advancements to ensure high standards of care [18]. Public education, meanwhile, aims to demystify radiotherapy, reduce stigma, and promote awareness of its benefits, fostering trust in cancer treatment systems [19]. Providing accurate and ethically shared information across these groups is vital to ensure safety, informed decision-making, and the effective application of radiotherapy.

This Viewpoint explores a framework for integrating LLM-based chatbots in radiotherapy education, designed to ensure accuracy, reliability, and ethical integrity. The framework incorporates 3 key components: a controlled and curated database, a robust quality control system, and continuous monitoring mechanisms [20]. The curated database includes only verified and relevant medical literature and patient data, ensuring that the chatbot's responses are accurate and up to date. The quality control system involves regular audits, database updates, and algorithms to identify and correct inaccuracies or biases. Continuous monitoring tracks the chatbot's performance, enabling real-time adjustments and improvements [20]. By addressing challenges such as misinformation and patient privacy, this framework establishes a trustworthy and effective tool in radiotherapy education. Furthermore, it sets a new standard for AI-driven tools in specialized health care education by focusing on accuracy, reliability, and ethical integrity [20,21].

Ultimately, the ethical design and use of LLM-based medical chatbots involve collective efforts from policy makers and governments to maximize benefits while mitigating risks such as data privacy breaches and algorithmic bias [22].

Current State of Medical Chatbots

Development and Implementation of LLM-Based Chatbots

The development and implementation of LLM-based chatbots in health care have marked a significant evolution from earlier chatbot technologies. Unlike traditional rule-based systems that operated within a limited scope, LLM-based chatbots such as those powered by OpenAI's ChatGPT have demonstrated the ability to understand and generate natural language with remarkable fluency and contextual awareness [23]. This advancement has led to a broad range of applications in health care, where these chatbots are being used to support patient interaction, provide medical information, assist with diagnosis, and streamline administrative tasks. A review of current LLM-based medical chatbots reveals a diversity of functionalities tailored to various health care needs [24]. One prominent example is ChatGPT itself, which has been adapted for use in several health care settings. ChatGPT can engage in detailed conversations with patients, answering questions about symptoms, explaining medical conditions, and providing guidance on the next steps in care. Its ability to understand nuanced queries and generate contextually appropriate responses has made it a valuable tool for preliminary consultations, particularly in telemedicine where immediate access to health care professionals may be limited [25]. Another notable use case is the deployment of LLM-based chatbots for chronic disease management [13]. These chatbots are designed to assist patients in managing conditions such as diabetes, hypertension, and asthma by providing personalized advice, reminding them to take medication, and offering lifestyle recommendations based on their medical history and real-time data inputs; for example, some LLM-based chatbots integrate with wearable devices to monitor patient vitals and deliver timely interventions, improving adherence to treatment plans and potentially reducing the need for emergency care [26]. In addition to patient-facing applications, LLM-based chatbots have been implemented to support health care professionals. These chatbots can function as web-based assistants, helping clinicians with tasks such as documentation, coding, and accessing up-to-date medical literature; for instance, some systems are designed to summarize patient records, extract relevant clinical information, and even suggest differential diagnoses, thereby enhancing the efficiency of clinical workflows [27]. In research settings, LLM-based chatbots are used to assist in data analysis, generate research hypotheses, and streamline the process of literature review by quickly summarizing large volumes of academic papers.

The flexibility and scalability of LLM-based chatbots have also enabled their adoption in mental health support. Chatbots such as Woebot [28] and Wysa [29] use LLMs to engage in therapeutic conversations, offering cognitive behavioral therapy techniques, mood tracking, and crisis intervention. These tools provide users with immediate access to mental health support, which is particularly valuable in regions with limited access to mental health professionals or during times when in-person therapy is not feasible. While the functionalities and use cases of LLM-based chatbots in health care are vast, the implementation of these tools is not without challenges. Ensuring the accuracy and reliability of the information provided by these chatbots is crucial, especially in scenarios where they are used for diagnosis or treatment recommendations. Moreover, the ethical implications of using AI in patient care, particularly in terms of data privacy and informed consent, require careful consideration [30]. Despite these challenges, the ongoing development and refinement of LLM-based chatbots continue to push the boundaries of what is possible in health care, offering promising avenues for improving patient care, enhancing health care delivery, and supporting the work of health care professionals.

Strengths and Limitations

The implementation of LLM-based chatbots in health care, particularly in specialized fields such as radiotherapy, offers both significant strengths and notable limitations. As these advanced AI-driven tools become more integrated into medical practice, it is crucial to understand their advantages and challenges to optimize their use while mitigating potential risks.

Strengths

One of the primary strengths of LLM-based chatbots is their ability to process and generate natural language with a high degree of fluency and context sensitivity. This capability allows them to engage in detailed and nuanced conversations with both patients and health care professionals, making complex medical information more accessible and understandable. In radiotherapy, where patients often face intricate treatment regimens and technical jargon, an LLM-based chatbot can break down complex concepts into simpler terms, enhancing patient comprehension and engagement. This improved communication can lead to better patient adherence to treatment plans and a more informed patient population, which is essential for the success of radiotherapy [17]. Another significant advantage of LLM-based chatbots is their scalability and availability. These chatbots can provide consistent, round-the-clock support, making them particularly valuable in settings where access to health care professionals may be limited. In radiotherapy, where timely information is critical, a chatbot can offer immediate answers to patient queries, provide pretreatment education, and even guide patients through posttreatment care [31]. This constant availability helps reduce the burden on health care providers, allowing them to focus on more complex cases and personalized care. Furthermore, LLM-based chatbots can be regularly updated with the latest medical research and guidelines, ensuring that the information they provide remains current. In a field such as radiotherapy, where treatment protocols and technologies are continuously evolving, this ability to quickly incorporate new knowledge is a significant strength. It enables the chatbot to serve as a reliable resource for both patients and health care professionals, supporting evidence-based practice and reducing the likelihood of outdated or incorrect information being disseminated [32].

Limitations

Despite their strengths, LLM-based chatbots also present several limitations, particularly in specialized medical fields such as radiotherapy. One of the most pressing challenges is ensuring the accuracy and reliability of the information provided by these

chatbots [20]. While LLMs can generate text that seems plausible and coherent, they may occasionally produce incorrect or misleading information, especially when faced with highly specialized or uncommon queries. In radiotherapy, where precise details about treatment options, dosimetry, and potential side effects are crucial, even small inaccuracies can have serious consequences for patient care. Another limitation is the potential for LLM-based chatbots to oversimplify complex medical information. While simplifying language is essential for patient comprehension, there is a risk that crucial nuances may be lost, leading to misunderstandings or incomplete knowledge. In radiotherapy, where patients need to fully understand their treatment options, the risks, and the benefits, this oversimplification could impact their ability to make informed decisions [3]. Ethical considerations also pose significant challenges for the implementation of LLM-based chatbots in health care. Issues related to patient privacy, data security, and the potential for bias in the AI's responses are critical concerns. In radiotherapy, where patients are dealing with life-altering decisions and sensitive health information, ensuring that the chatbot operates within strict ethical guidelines is essential. There is also the challenge of maintaining transparency in how these chatbots operate and ensuring that patients are fully aware that they are interacting with an AI, not a human health care provider. Moreover, the reliance on LLM-based chatbots could potentially lead to an overdependence on AI at the expense of human interaction. In specialized fields such as radiotherapy, the human touch is often crucial for providing emotional support and addressing the psychosocial aspects of care. While chatbots can provide information and support, they cannot replace the empathy and personalized care that human health care providers offer [33].

Summary

Table 1 summarizes the key advantages and challenges associated with the deployment of LLM-based chatbots in health care, particularly in the context of radiotherapy, where precision, ethical considerations, and human interaction are critical. It is seen that while LLM-based chatbots offer significant strengths, including enhanced communication, scalability, and up-to-date information, they also come with limitations that must be carefully managed. In specialized fields such as radiotherapy, where there is a need for compassion in addition to accurate information as the users may be in stressful situations, the deployment of these chatbots requires a well–thought-out framework that maximizes their benefits while addressing their inherent challenges.

Table 1. Strengths and limitations of large language model–based chatbots in specialized medical fields such as radiotherapy.

Aspect	Strengths	Limitations
Natural language processing	• Fluent, context-aware communication • Simplifies complex medical information for better patient understanding	• Risk of generating incorrect or misleading information • Potential oversimplification of complex concepts, leading to misunderstandings
Availability and scalability	• 24-7 availability for patient and health care provider support • Reduces the burden on health care professionals by handling routine queries	• Overdependence on AI[a] could reduce human interaction, which is vital in emotionally supportive care, particularly in radiotherapy
Up-to-date information	• Can be regularly updated with the latest research and guidelines, ensuring that the information provided is current	• Challenges in maintaining the accuracy of specialized information, especially as new research emerges and medical knowledge evolves
Patient engagement	• Enhances patient education and engagement through personalized, accessible information • Supports informed decision-making in complex treatments	• Risk of inadequate personalization, potentially leading to generic advice that does not fully meet individual patient needs
Ethical considerations	• Can be programmed to follow ethical guidelines, ensuring responsible dissemination of medical information	• Issues related to patient privacy, data security, and AI bias • Potential lack of transparency in AI operations, affecting trust
Efficiency in health care	• Improves efficiency by assisting in documentation, coding, and access to medical literature for health care providers	• Dependency on chatbot accuracy for clinical tasks might introduce errors if the chatbot provides incorrect or incomplete information

[a]AI: artificial intelligence.

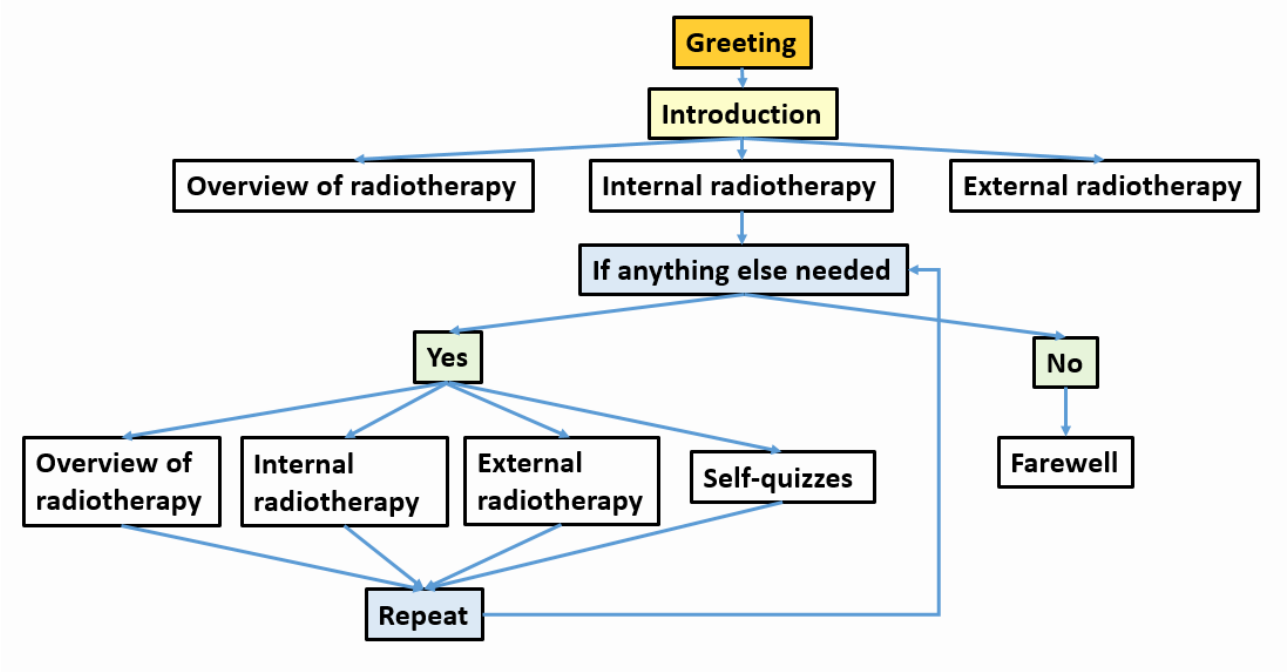
Impact on Patient Education and Health Care Delivery

Influence on Patient Education

LLM-based chatbots play a significant role in enhancing patient education by making complex medical information more accessible. In specialized fields such as radiotherapy, where patients must understand intricate treatment protocols and potential side effects, chatbots provide a valuable resource for simplifying and clarifying these concepts. By translating technical jargon into clear language, chatbots help patients grasp the fundamentals of their treatment, which can be crucial for their engagement and adherence. Moreover, chatbots can offer tailored educational content based on individual patient profiles; for example, a chatbot might provide specific information about

the type of radiotherapy a patient is receiving, potential side effects relevant to their case, and strategies for managing these side effects [31,34]. This personalized approach ensures that patients receive relevant information that directly pertains to their situation, enhancing their understanding and preparedness for their treatment journey. Figure 2 shows a typical dialogue flowchart for a chatbot in radiotherapy education.

Figure 2. Typical dialogue flowchart for a chatbot used in radiotherapy education.



Impact on Decision-Making

In addition to enhancing patient education, LLM-based chatbots significantly influence patient decision-making. By providing timely and accurate information, these tools empower patients to make informed choices about their treatment options; for instance, a chatbot can help patients compare different radiotherapy techniques, discuss the potential benefits and risks of each option, and address any concerns they may have. This support is especially valuable in radiotherapy, where patients often face complex decisions about their care [32]. The chatbot's ability to offer evidence-based information and answer questions in real time helps reduce uncertainty and anxiety, allowing patients to engage more effectively in their treatment planning. Furthermore, by facilitating a deeper understanding of their condition and treatment options, chatbots contribute to a more collaborative decision-making process between patients and their health care providers.

Enhancement of Health Care Delivery

The implementation of LLM-based chatbots also has a notable impact on the overall delivery of health care. These tools streamline administrative tasks, such as appointment scheduling, medication reminders, and follow-up care, thus freeing up health care professionals to focus on more complex patient interactions [35]. In radiotherapy, where treatment regimens are often intensive and require careful management, chatbots can help monitor patient progress, manage treatment schedules, and provide reminders for follow-up appointments or adherence to prescribed protocols. In addition, chatbots can assist in triaging patient queries, directing them to appropriate resources or health care professionals based on the nature of their concerns [36]. This efficient handling of inquiries helps ensure that patients receive timely and relevant support, which can improve their overall experience and satisfaction with their care.

Challenges in Deploying LLM-Based Medical Chatbots

Overview

Despite the promising advancements and benefits of LLM-based medical chatbots, their deployment in health care settings is accompanied by several significant challenges. These challenges span various critical areas, including the accuracy and reliability of the information provided, privacy and data security concerns, ethical considerations, and the potential for misinformation. Addressing these issues is essential to ensure that chatbots contribute positively to health care delivery while mitigating risks.

Accuracy and Reliability

Ensuring the accuracy and reliability of medical information provided by LLM-based chatbots is a critical challenge that affects their effectiveness and safety in health care settings. The complexity of medical knowledge and the rapid evolution of treatment protocols present significant hurdles in maintaining the precision of information delivered by these AI tools. One major challenge is the inherent nature of LLMs, which are trained on vast datasets that include a wide range of information sources. While this breadth of data allows chatbots to generate responses to a variety of queries, it also means that the information they provide can be inconsistent or outdated [37]. In the context of radiotherapy, where treatment guidelines and protocols are continually updated based on the latest research, ensuring that a chatbot reflects the most current and accurate

information is crucial. Failure to do so can result in the dissemination of outdated or incorrect treatment recommendations, potentially compromising patient safety [20].

In addition, LLM-based chatbots rely on pattern recognition and probabilistic inference rather than deep, context-specific understanding [38]. This approach can sometimes lead to the generation of plausible but inaccurate answers, especially when dealing with rare or complex medical scenarios; for example, a chatbot might provide general information about radiotherapy but struggle with specifics related to individual patient conditions or less common treatment modalities [4]. This limitation highlights the need for robust validation processes to ensure that the chatbot's responses are not only accurate but also relevant to the user's specific context. Another aspect of accuracy and reliability involves the management of the chatbot's knowledge base [39]. Regular updates and reviews are necessary to keep the information current and accurate. This process requires a continuous feedback loop from health care professionals who can assess and correct any inaccuracies or gaps in the information provided by the chatbot. Implementing such a system involves collaboration between AI developers and medical experts to ensure that the chatbot's responses align with the latest evidence-based practices. Moreover, there is a challenge in balancing the chatbot's ability to generate detailed responses with the need to provide clear and concise information [40]. Overly complex or technical explanations can be difficult for patients to understand, while overly simplified answers may omit critical details. Striking the right balance is essential to provide information that is both accurate and accessible to users with varying levels of medical knowledge.

Privacy and Data Security

Privacy and data security are paramount concerns when deploying LLM-based chatbots in health care, given the sensitive nature of patient information handled by these tools. Ensuring the protection of these data against unauthorized access and breaches is essential to maintaining patient trust and complying with regulatory requirements [19]. One of the primary challenges in safeguarding patient data involves ensuring compliance with privacy regulations such as the Health Insurance Portability and Accountability Act in the United States [41] or the General Data Protection Regulation in Europe [42]. These regulations mandate stringent measures for the collection, storage, and handling of personal health information. LLM-based chatbots must be designed to adhere to these regulations, incorporating robust data encryption methods, secure storage solutions, and strict access controls to protect sensitive patient information from breaches and unauthorized access. Data transmission security is another critical aspect. When chatbots interact with patients, the data exchanged—such as medical history, treatment details, and personal identifiers—must be encrypted to prevent interception by malicious actors. Implementing secure communication protocols, such as transport layer security, helps ensure that data transmitted between the chatbot and the user are protected from eavesdropping and tampering [43]. In addition, managing user consent and data use transparency is crucial for maintaining privacy. Patients must be informed about what data are being collected, how the data will be used, and their rights regarding their information. Clear consent mechanisms should be integrated into the chatbot's design, allowing users to provide informed consent before any data collection occurs. This transparency helps build trust and ensures that patients are aware of how their data are being handled.

Another challenge is ensuring that data anonymization and deidentification practices are effectively applied. Even when data are stored or processed for purposes such as improving chatbot performance or training new models, it is essential to anonymize the data to prevent the identification of individual patients. Proper anonymization techniques reduce the risk of sensitive information being exposed in case of a data breach. Moreover, continuous monitoring and auditing of the chatbot's security systems are necessary to identify and address potential vulnerabilities. Regular security assessments, including penetration testing and vulnerability scanning, help ensure that the chatbot's infrastructure remains resilient against emerging threats. Promptly addressing any identified weaknesses is crucial for maintaining the overall security of patient data.

Ethical Considerations

Overview

The deployment of LLM-based chatbots in health care raises several ethical dilemmas that must be carefully addressed to ensure the responsible and fair use of AI. Key ethical concerns include AI decision-making processes, transparency, and accountability, each of which plays a critical role in maintaining trust and integrity in health care interactions [44]. To address these concerns, transparency can be enhanced by developing clear documentation that explains the chatbot's decision-making processes, including data sources and logic. Bias detection and mitigation tools, such as fairness-aware algorithms, should be used to identify and correct biases in real time. Regular audits by independent third parties can ensure compliance with ethical standards.

User feedback mechanisms are essential for identifying and addressing ethical issues. Users should be able to report concerns or errors, which should be promptly reviewed and acted upon. In addition, clear guidelines for human oversight must be established, defining the roles of health care professionals in monitoring chatbot interactions. By incorporating these detailed suggestions, processes, and tools, we can better address ethical issues and promote the responsible use of AI in health care [18].

Long-Term Societal Impacts

The widespread adoption of ChatGPT and similar technologies could significantly reshape the job market, potentially leading to both job displacement and the creation of new opportunities, necessitating reskilling and upskilling programs to mitigate negative impacts. Moreover, AI technologies might exacerbate economic inequalities if their benefits are not equitably distributed, highlighting the need for policies that promote inclusive growth. The integration of AI into daily life could alter social interactions, reducing human-to-human contact and impacting social skills and relationships, thus requiring a balance between AI use and genuine human connection. These changes underscore the importance of robust ethical governance frameworks involving diverse stakeholders to ensure that AI development aligns with societal values and ethical principles,

addressing emerging challenges and opportunities through continuous dialogue and adaptive regulation.

AI Decision-Making

One of the central ethical issues is the decision-making capability of AI systems. Unlike human health care providers who can apply clinical judgment and empathy, LLM-based chatbots operate based on algorithms and data patterns [45]. This raises questions about the extent to which these systems can make ethical decisions, especially in complex or nuanced medical scenarios; for instance, while chatbots can provide general information and support, they may lack the ability to fully understand the context of a patient's situation or the ethical implications of certain recommendations [46]. Ensuring that AI-driven tools align with ethical guidelines and medical standards is essential to avoid potentially harmful outcomes.

Transparency

Transparency in AI decision-making is another critical ethical consideration. Patients and health care professionals need to understand how chatbots generate their responses and make recommendations. This includes clarity about the data sources and algorithms that drive the chatbot's behavior. Without transparency, there is a risk that users might overestimate the chatbot's capabilities or misinterpret its advice. Providing clear information about the chatbot's operational mechanisms helps build trust and allows users to make informed decisions about how to use the tool effectively [47]. Furthermore, transparency extends to the disclosure of limitations and potential biases inherent in the AI system. Chatbots are trained on datasets that may reflect existing biases or incomplete information, which can affect the fairness and objectivity of their responses [48]. Openly communicating these limitations and actively working to mitigate bias are crucial for maintaining ethical standards and ensuring that the chatbot serves all users equitably.

Accountability

Accountability is a significant ethical issue concerning the use of LLM-based chatbots. Determining who is responsible for the chatbot's recommendations and decisions is essential, particularly in cases where incorrect or harmful information is provided. Clear lines of accountability must be established, involving both the developers who create and maintain the chatbot and the health care providers who deploy it. This includes ensuring that there are mechanisms for addressing errors or issues that arise from the chatbot's interactions with users [49]. In addition, it is important to have protocols in place for managing situations where the chatbot's advice might lead to adverse outcomes [50]. This includes providing a way for users to report issues and seek redress, as well as ensuring that there are processes for continuous improvement based on feedback and incident analysis. Effective oversight and governance structures help ensure that ethical standards are upheld and that the chatbot contributes positively to patient care.

Potential for Misinformation

Overview

The potential for misinformation is a critical concern in the deployment of LLM-based medical chatbots. Given their reliance on vast datasets and sophisticated algorithms, these chatbots can inadvertently spread incorrect or misleading information, which poses significant risks to patient health and safety. As deep learning pioneer Geoffrey Hinton has noted, neural networks can share what they learn instantly [22]; therefore, any erroneous messages can spread instantaneously, making unwanted wide impact.

Sources of Misinformation

One key source of misinformation is the quality and accuracy of the training data used to develop the chatbot. LLMs are trained on diverse datasets that include information from various sources, some of which may be outdated, biased, or inaccurate. If a chatbot's training data contain erroneous or misleading content, there is a risk that the chatbot will replicate and disseminate these inaccuracies in its responses. In fields such as radiotherapy, where precise and current information is crucial, the presence of outdated or incorrect data can lead to harmful consequences for patients relying on the chatbot for guidance [51]. Another potential source of misinformation is the chatbot's ability to generalize information [52]. While LLM-based chatbots can handle a wide range of queries, they may not always be adept at distinguishing between general knowledge and specific, context-sensitive details. This limitation can lead to the generation of responses that are technically correct but fail to address the nuances of individual patient situations [53]; for example, a chatbot might provide general advice on radiation safety but miss specific recommendations tailored to a patient's unique treatment plan.

Risks and Consequences

The spread of incorrect or misleading information by medical chatbots can have serious repercussions for patient care. Patients may make health-related decisions based on inaccurate information, which can result in inappropriate or ineffective treatments. In radiotherapy, where treatment decisions are complex and require careful consideration of numerous factors, relying on incorrect information can lead to suboptimal care or adverse outcomes. Moreover, misinformation can erode trust in health care technology. If patients or health care providers encounter inaccuracies or inconsistencies in the information provided by a chatbot, their confidence in the tool's reliability and usefulness may be undermined. This loss of trust can diminish the chatbot's effectiveness and impact its acceptance and integration into health care practices.

Mitigation Strategies

To mitigate the risks associated with misinformation, several strategies can be used [54]. Regular updates and maintenance of the chatbot's knowledge base are essential to ensure that it reflects the most current and accurate information. Collaboration with medical experts and continuous validation of the chatbot's responses help identify and correct inaccuracies. Implementing mechanisms for user feedback and reporting can also provide valuable insights into potential issues and facilitate prompt

resolution. Furthermore, incorporating a system of checks and balances, such as providing disclaimers that emphasize the chatbot's limitations and the need for professional medical consultation, can help manage user expectations. Ensuring that the chatbot directs users to seek advice from qualified health care professionals when necessary can prevent reliance on potentially flawed or incomplete information [55].

Proposed Framework for a Resilient Medical Chatbot in Radiotherapy Education

Database Management and Control

Effective database management is crucial for developing a resilient medical chatbot in radiotherapy education. The integrity of the chatbot's responses depends heavily on the accuracy and timeliness of the information it accesses. To achieve this, a robust strategy for building and maintaining a controlled database is essential [56]. The first step in this strategy involves curating a comprehensive and authoritative database from verified medical sources, such as peer-reviewed journals, clinical guidelines, and expert consensus statements. This curated database should be dynamic, with automated systems in place for continuous updates. The proposed framework incorporates an AI-driven mechanism for real-time monitoring of new publications and guidelines, ensuring that the database remains current and reducing the risk of disseminating outdated or inaccurate information [57]. To implement this, automated updates could leverage web scraping tools or application programming interface integrations to collect and validate data against trusted sources. The database should use meta-tagging and hierarchical organization for efficient data management and retrieval. Meta-tagging can be achieved using NLP algorithms to assign contextual keywords, improving the chatbot's ability to interpret and respond to user queries accurately. A hierarchical data structure, such as a tree- or graph-based model, can prioritize data by relevance and reliability, ensuring that the chatbot provides the most appropriate responses [58,59]. In addition, periodic audits of the database should be conducted to verify its accuracy and adherence to updated guidelines.

Maintaining control over the database is equally important. This involves setting strict protocols for data entry and modification, with access restricted to authorized personnel who are well versed in radiotherapy and medical education. Regular audits of the database content should be conducted to identify and rectify any inconsistencies or errors [57]. In addition, using version control systems can help track changes, allowing for the restoration of previous database states if needed, which further enhances the reliability of the chatbot. To ensure that the database remains resilient against emerging challenges such as misinformation or biased data, a layered review process should be integrated. This involves cross-referencing new data entries with multiple sources and using machine learning algorithms to detect and flag anomalies or conflicting information [60]. By implementing these strategies, the database management and control framework will serve as the foundation for a resilient medical chatbot capable of providing accurate, reliable, and up-to-date information in radiotherapy education.

Quality Control System

A robust quality control system is vital for ensuring the reliability and effectiveness of a medical chatbot in radiotherapy education. The system must be designed to maintain the highest standards of accuracy, relevance, and trustworthiness in the chatbot's content, which requires regular reviews, timely updates, and stringent validation processes [61]. The cornerstone of this quality control system is the implementation of a multitiered review process. This begins with the periodic assessment of the chatbot's content by a panel of experts in radiotherapy and medical education. These experts should evaluate the chatbot's responses for accuracy, clarity, and consistency with current clinical practices. Regularly scheduled reviews ensure that the content remains aligned with the latest advancements in radiotherapy and adheres to evolving educational standards [4].

In addition to expert reviews, the quality control system should include automated checks and balances. Machine learning algorithms can be deployed to continuously monitor the chatbot's interactions, identifying patterns of errors or discrepancies in the responses [62]. These algorithms can flag potential issues for further human review, ensuring that errors are caught and corrected promptly. Automated processes should also include regular updates to the chatbot's knowledge base, triggered by new research findings, clinical guidelines, or changes in medical protocols. Another critical aspect of the quality control system is the validation of the chatbot's content before it goes live. This can be achieved through rigorous testing with simulated user interactions, covering a wide range of scenarios that the chatbot is likely to encounter. Feedback loops from these tests should be analyzed to refine the chatbot's algorithms and content delivery mechanisms, ensuring that it provides accurate and contextually appropriate responses [63]. Finally, the quality control system must be adaptable, with mechanisms in place for continuous improvement. This includes incorporating user feedback to identify areas where the chatbot's performance may be lacking and making necessary adjustments. Moreover, the system should be capable of responding to unforeseen challenges, such as the propagation of misinformation, and adapting to emerging trends in radiotherapy by swiftly updating the chatbot's content and protocols [17].

Monitoring and Feedback Loop

Establishing an effective monitoring and feedback loop is essential for the continuous improvement and reliability of a medical chatbot in radiotherapy education. This system ensures that the chatbot consistently meets user needs, adapts to new information, and addresses any issues that arise during its interactions. The foundation of this system is a comprehensive monitoring mechanism designed to track all chatbot interactions in real time. By recording user queries, responses, and outcomes, this monitoring system can identify patterns in user behavior and detect potential issues, such as inaccurate answers, misinterpretations, or gaps in the chatbot's knowledge base [64]. Advanced analytics tools can be integrated to automatically flag problematic interactions for further review by the development team, enabling swift corrective actions. In addition to tracking interactions, the monitoring system should

incorporate a robust feedback loop that allows users to provide direct input on the chatbot's performance [65]. This feedback can be collected through postinteraction surveys, ratings, or optional comment fields, giving users the opportunity to highlight areas where the chatbot excels or falls short. Aggregating and analyzing this feedback offers valuable insights into the chatbot's strengths and weaknesses, guiding future updates and enhancements.

To maximize the effectiveness of the feedback loop, it is crucial to establish a process for prioritizing and addressing the issues identified. A triage system can be implemented to categorize user feedback based on severity and frequency, ensuring that the most critical issues are addressed promptly [66]; for example, if multiple users report the same error or misunderstanding, this issue would be flagged as a high priority for immediate resolution. Less critical feedback, such as suggestions for improved phrasing or additional features, can be scheduled for consideration in future updates. Moreover, the feedback loop should be designed to foster continuous learning and adaptation. Regularly scheduled reviews of the chatbot's performance, informed by user feedback and monitoring data, should be conducted to identify trends and areas for improvement. This iterative process allows the chatbot to evolve over time, enhancing its ability to provide accurate, reliable, and contextually relevant information in radiotherapy education.

Ethical and Legal Safeguards

Implementing robust ethical guidelines and legal safeguards is paramount in developing a medical chatbot for radiotherapy education. These measures are critical to addressing concerns related to privacy, transparency, and accountability, ensuring that the chatbot operates within the highest ethical and legal standards. To begin with, the chatbot must be designed with stringent privacy protections. Given the sensitive nature of medical information, the chatbot should comply with all relevant data protection laws, such as the General Data Protection Regulation in Europe or the Health Insurance Portability and Accountability Act in the United States [19]. This involves implementing secure data encryption methods, anonymizing user interactions, and ensuring that any personal or health-related data collected during interactions are stored securely and used only for their intended purpose [67]. Access to these data should be strictly controlled, with clear protocols for who can view, modify, or delete the data, ensuring that user privacy is respected at all times. Transparency is another critical aspect of ethical chatbot design. Users must be fully informed about how the chatbot functions, including the sources of its information, the limitations of its advice, and the nature of its data collection practices. This can be achieved by providing clear and accessible information within the chatbot interface, such as disclaimers before interactions, a detailed frequently asked questions section, or links to privacy policies. In addition, the chatbot should be designed to clearly distinguish between general information and personalized advice, ensuring that users understand the context and limitations of the information they receive [68].

Accountability mechanisms are also essential to uphold ethical standards. The development and deployment of the chatbot should include a clear governance structure, where responsibilities for content accuracy, data management, and user interactions are well defined. Regular audits of the chatbot's performance and adherence to ethical guidelines should be conducted, with the results made available to relevant stakeholders [69]. Furthermore, there should be a clear process for users to report any ethical concerns or grievances, and these issues should be addressed promptly and transparently. To further enhance ethical and legal safeguards, it is important to establish an oversight committee consisting of experts in ethics, law, and medical education [70]. This committee would be responsible for reviewing the chatbot's operations, ensuring that it remains aligned with ethical principles and legal requirements. The committee should also be tasked with evaluating the impact of the chatbot on users, particularly in terms of potential biases, misinformation, or unintended consequences, and recommending corrective actions as needed. The proposed framework is summarized in Table 2.

Table 2. Summary of the proposed framework for a resilient medical chatbot in radiotherapy education. The key components, their essential elements, and their respective purposes in ensuring the chatbot's accuracy, reliability, and ethical operation are outlined.

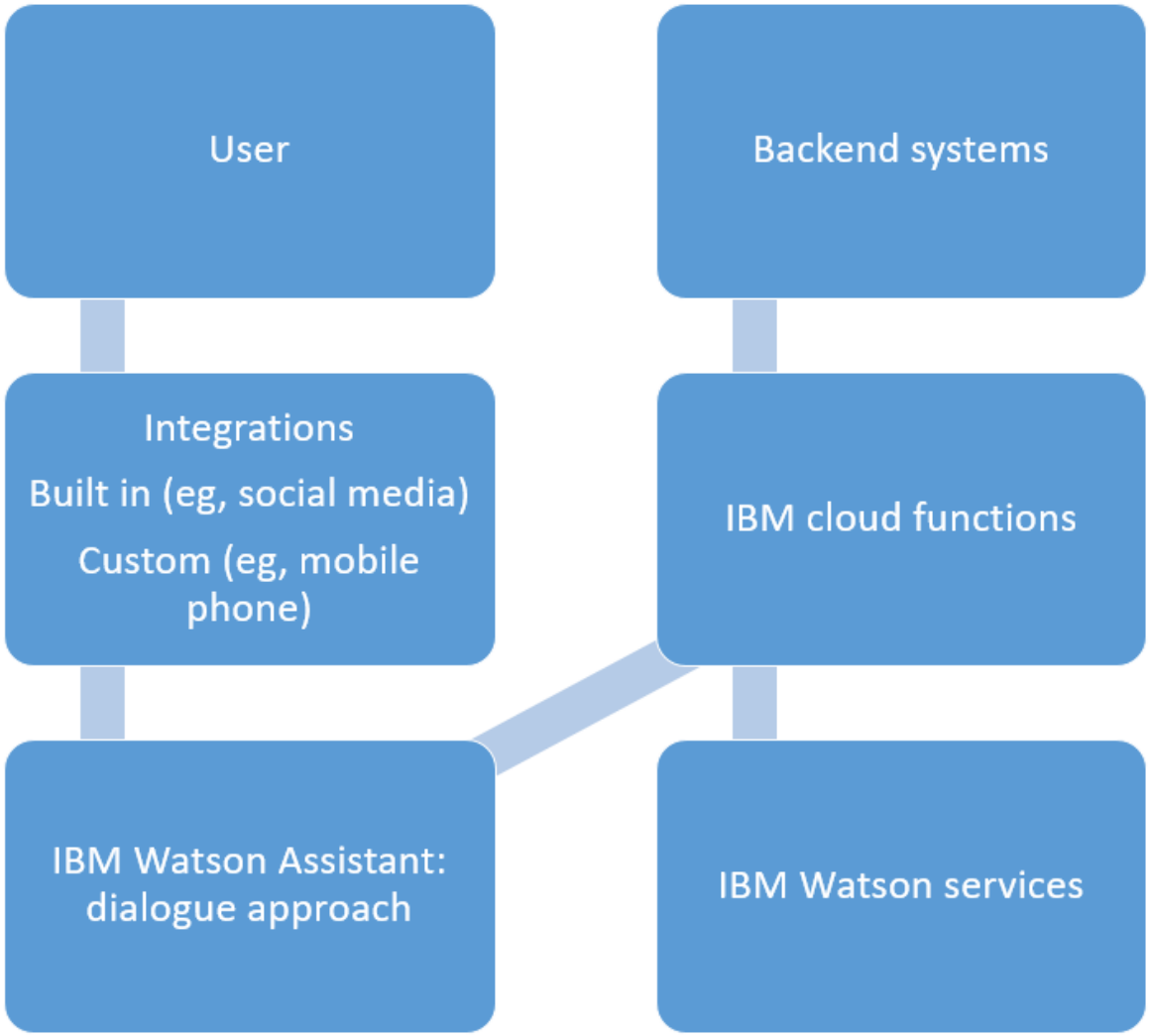
Component	Key elements	Purpose
Database management and control	• Curated and dynamic database • Automated updates • Hierarchical data structure	Ensures that the chatbot has access to accurate, up-to-date, and relevant information
Quality control system	• Multitiered expert review • Automated checks • Content validation	Maintains high standards of accuracy, relevance, and reliability in chatbot responses
Monitoring and feedback loop	• Real-time interaction tracking • User feedback integration • Continuous improvement	Tracks chatbot performance, identifies issues, and incorporates user feedback for ongoing enhancement
Ethical and legal safeguards	• Privacy protections • Transparency measures • Accountability mechanisms	Addresses concerns related to user privacy, ensures transparency, and upholds accountability

Case Study: Application of the Framework in Radiotherapy Education

Outlining the Process of Developing and Testing the Chatbot Using the Proposed Framework

The development of the radiotherapy education chatbot using the proposed framework commenced with a comprehensive needs assessment. This step involved engaging key stakeholders, including radiotherapy educators, clinical oncologists, medical physicists, and students, to identify the specific educational challenges and gaps that the chatbot aimed to address [32]. The focus was on creating a tool that could effectively supplement traditional education by providing accurate, accessible, and contextually relevant information on complex radiotherapy topics. Once the needs were clearly defined, the next phase involved designing the chatbot's conversational architecture. The framework emphasized the integration of AI-driven NLP capabilities with a structured knowledge base specific to radiotherapy [71]. This phase required meticulous curation and validation of content, ensuring that the information was not only accurate but also aligned with the latest guidelines and research in the field. A collaborative approach was adopted, involving experts from various disciplines to ensure that the chatbot's responses would be both scientifically robust and pedagogically sound [72]. The chatbot was then built using a hybrid approach that combined rule-based algorithms with machine learning models [73]. The rule-based components ensured that the chatbot could handle critical educational scenarios with precision, while the machine learning aspects allowed for more dynamic and flexible interactions. The integration with the IBM Watson Assistant cloud platform facilitated the deployment of advanced NLP algorithms, enabling the chatbot to understand and respond to user queries effectively [32]. Figure 3 shows the architectural diagram of a chatbot developed using the IBM Watson Assistant platform [74].

Figure 3. Architectural diagram showing how the chatbot is connected to the IBM Watson Assistant cloud platform.



Testing the chatbot was an iterative process. Initial prototypes were subjected to rigorous testing in simulated environments, where various scenarios were presented to evaluate the chatbot's performance in real time [75]. The testing focused on several key metrics: the accuracy of information, user engagement, response time, and the ability to handle unexpected or complex

queries. Feedback from educators and students was crucial during this phase, providing insights into the chatbot's usability and effectiveness in an educational setting. To ensure the reliability and safety of the chatbot, the testing also included a comprehensive evaluation of ethical considerations. The framework's built-in mechanisms for privacy protection, data security, and bias mitigation were scrutinized to prevent potential harm. The chatbot was tested for its adherence to ethical standards through a Delphi study involving international experts, ensuring that it met global expectations for AI-driven educational tools [76].

The final phase involved pilot-testing the chatbot in real educational settings, specifically within radiotherapy courses. These studies aimed to evaluate both the chatbot's technical performance and its educational impact. Observations from the pilot phase revealed notable improvements in student comprehension and engagement; for example, the chatbot, called RT Bot, was deployed on a website accessible by invitation and introduced during workshops and conferences to test its functionality. Participants were invited to complete a survey evaluating the chatbot's functionality and content. This survey, conducted with 60 participants, aimed to gather statistical insights and feedback for improving the chatbot in the future. The results, summarized in Table 3 [32], highlight varying levels of user satisfaction with RT Bot during its testing phase. The assessment used a scale ranging from 1 (lowest satisfaction) to 5 (highest satisfaction). The data indicate that a substantial 70% (42/60) of the respondents rated the content's helpfulness, understandability, and reading duration as above average.

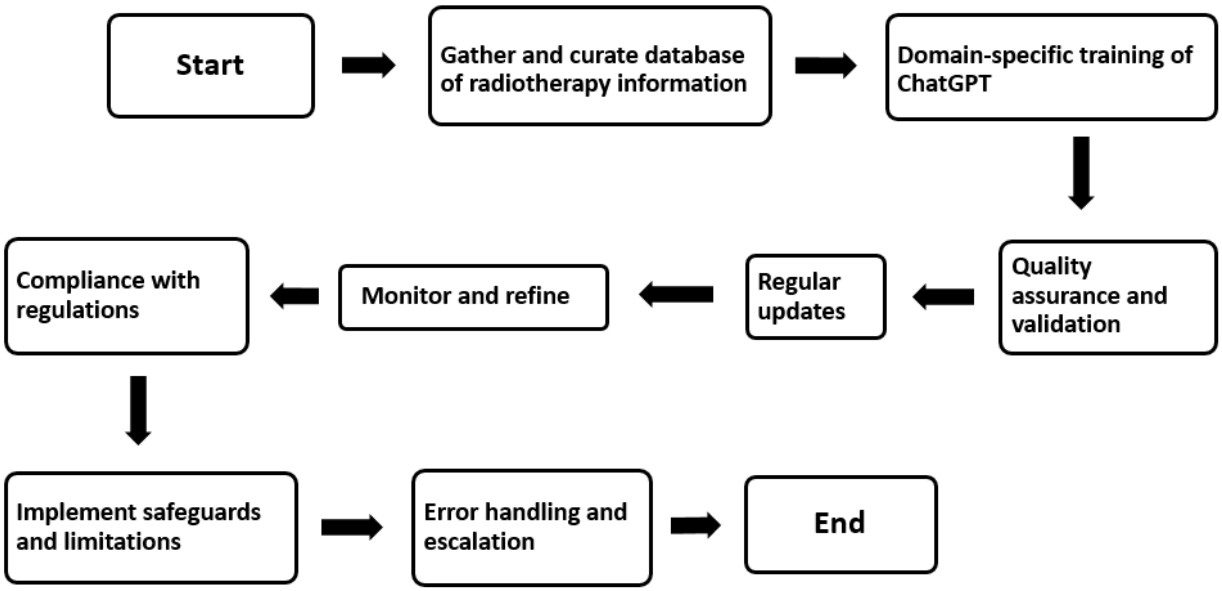
Table 3. Users' satisfaction ratings regarding the contents of RT Bot (N=60) [32].

Degree of satisfaction (ranging from 1=lowest to 5=highest)	Helpfulness of content, n (%)	Understandability of content, n (%)	Reading duration, n (%)
1	3 (5)	0 (0)	0 (0)
2	0 (0)	9 (15)	3 (5)
3	15 (25)	9 (15)	15 (25)
4	24 (40)	24 (40)	18 (30)
5	18 (30)	18 (30)	24 (40)

Specific feedback highlighted areas for improvement, such as the need for more intuitive methods to restart conversations and tailored hints for users with diverse backgrounds when answering radiation safety questions. These insights enabled developers to refine the chatbot's adaptability and responsiveness. Data collected from the pilot tests demonstrated a measurable positive effect on learning outcomes, as students achieved higher quiz scores and expressed increased confidence in applying radiation safety principles. The feedback loop established during this phase ensured that the chatbot evolved into a robust, user-centered educational tool, ready to support radiotherapy training effectively [34,77,78].

Figure 4 illustrates the workflow for creating and maintaining an LLM-based medical chatbot powered by ChatGPT, tailored for radiotherapy. The chatbot relies on a curated database of verified information and undergoes domain-specific training to understand radiotherapy nuances. However, several limitations exist. The quality of training data and the effectiveness of cross-checking mechanisms are crucial; any errors or biases can lead to inaccurate responses. Regular updates are necessary to keep the database current, but this requires significant resources. User feedback helps refine the system, although the feedback can be subjective and challenging to manage. Error-handling mechanisms allow for query escalation, but not all issues may be resolved, potentially affecting user trust. Despite these challenges, the iterative process aims to provide accurate and up-to-date information within the chatbot's expertise.

Figure 4. Flowchart providing a sequential representation of the steps involved in creating, maintaining, and improving a medical chatbot based on the large language model ChatGPT, with a focus on radiotherapy, ensuring accuracy, compliance, and continuous refinement.



This process is crucial because it ensures the reliability, accuracy, and ethical integrity of the medical chatbot in radiotherapy. By curating a precise and verified database, fine-tuning the AI model, and implementing rigorous validation processes, the system can provide trustworthy information. The continuous refinement loop, fueled by user feedback and updates, ensures that the chatbot stays current with evolving practices and advancements in radiotherapy. This reliability and adherence to medical standards instills confidence in users, guiding them with accurate information while emphasizing the chatbot's limitations and the necessity of seeking professional medical advice. Ultimately, this meticulous process not only educates and assists users but also upholds the ethical responsibility of providing reliable health care information through AI-driven technologies.

Performance Evaluation

Evaluating the chatbot's performance was a critical step to ensure that it met the high standards required for educational tools in radiotherapy. The evaluation focused on 3 main areas: accuracy, user satisfaction, and educational impact.

Accuracy

The accuracy of the chatbot was assessed through a multitiered validation process. First, a set of standardized queries covering a wide range of radiotherapy topics was developed. These queries were designed to test the chatbot's ability to provide correct and precise information. Subject matter experts in radiotherapy independently evaluated the chatbot's responses to these queries, comparing them against established medical literature and clinical guidelines [79]. The chatbot's performance was quantified using accuracy metrics such as precision and recall, ensuring that it consistently delivered reliable information [80]. In addition, the chatbot's ability to update and incorporate the latest research findings was tested, emphasizing the need for dynamic content management to maintain long-term accuracy [81].

User Satisfaction

User satisfaction was evaluated through extensive user testing, involving a diverse group of end users, including students, educators, and health care professionals. Surveys and feedback forms were used to gather qualitative and quantitative data on users' experiences with the chatbot. Key aspects such as ease of use, response time, clarity of information, and overall user experience were measured. The feedback loop was integral to refining the chatbot, addressing any issues related to user interaction, and ensuring that the chatbot met the expectations of its target audience. In addition to surveys, usability testing sessions were conducted where users interacted with the chatbot in controlled environments, providing real-time feedback on their experiences. This iterative process helped in identifying and rectifying any usability challenges, enhancing the overall user satisfaction [79].

Educational Impact

The educational impact of the chatbot was evaluated by analyzing its effectiveness as a learning tool in radiotherapy education. This involved conducting controlled studies where the performance of students who used the chatbot as a supplementary learning resource was compared to that of students who did not. Various educational metrics, such as knowledge retention, comprehension, and application of concepts, were measured through pre- and postintervention assessments [82]. The chatbot's impact on these metrics was statistically analyzed to determine its effectiveness in enhancing learning outcomes [77]. Furthermore, longitudinal studies were conducted to assess the sustained educational benefits of the chatbot, ensuring that its use had a lasting positive impact on students' understanding of radiotherapy concepts [83]. In addition to these formal evaluations, the chatbot's performance was assessed in real-world educational settings through pilot studies. These studies provided insights into how the chatbot influenced classroom dynamics, student engagement, and the

overall learning environment [84-86]. Feedback from these settings was crucial in understanding the practical implications of integrating the chatbot into existing curricula and further validated its educational impact.

Lessons Learned and Future Directions

As the radiotherapy education chatbot continues to evolve, there are numerous opportunities for innovation that could significantly enhance its capabilities. Future research and technological advancements hold the potential to expand the chatbot's functionality, improve user experiences, and further integrate it into the educational landscape.

Advanced NLP and Understanding

One of the most promising areas for innovation lies in the continued development of advanced NLP algorithms. Future iterations of the chatbot could benefit from more sophisticated NLP techniques, such as transformer-based models and deep learning architectures, which can better understand and generate human-like responses [87]. These advancements would enable the chatbot to handle more complex queries, recognize nuanced language patterns, and provide more contextually accurate answers. Enhanced NLP capabilities could also improve the chatbot's ability to engage in multiturn conversations, allowing for deeper and more meaningful interactions with users.

Personalized Learning Experiences

Another key area for innovation is the development of personalized learning experiences. Leveraging data-driven insights, the chatbot could be equipped with adaptive learning algorithms that tailor content and responses to individual users' needs, preferences, and learning paces [88]. By analyzing user interactions, the chatbot could identify areas where a student might be struggling and provide targeted educational resources or alternative explanations to enhance understanding. This personalized approach would make the chatbot a more effective tool for diverse learning styles and could significantly improve educational outcomes [89].

Integration With Virtual and Augmented Reality

The integration of the chatbot with virtual and augmented reality (VR/AR) technologies presents exciting possibilities for creating immersive learning environments. By combining the chatbot's informational capabilities with VR/AR platforms, users could engage in interactive simulations of radiotherapy procedures, enhancing their practical understanding of complex concepts [90,91]; for example, students could use the chatbot to guide them through virtual radiotherapy sessions, where they can visualize and manipulate different treatment parameters in a 3D space. This fusion of AI-driven education with immersive technologies could revolutionize how radiotherapy is taught and learned.

Enhanced Multilingual and Cross-Cultural Capabilities

As radiotherapy education becomes increasingly globalized, there is a growing need for the chatbot to support multilingual and cross-cultural communication. Future developments could focus on expanding the chatbot's language capabilities, enabling it to provide accurate and contextually relevant information in multiple languages. This would not only make the chatbot more accessible to non–English-speaking users but also allow it to adapt its responses to different cultural contexts, ensuring that the information is both accurate and culturally sensitive [92]. Such advancements would make the chatbot a valuable educational tool in diverse global settings.

Ethical AI and Bias Mitigation

As AI continues to advance, addressing ethical concerns and mitigating biases in the chatbot's responses will be paramount. Future research could explore more robust methods for ensuring that the chatbot's algorithms remain unbiased and that its responses adhere to ethical guidelines [93]. This might include developing advanced algorithms for detecting and correcting biases in real time, as well as enhancing the transparency and explainability of the chatbot's decision-making processes. By prioritizing ethical AI, the chatbot could set new standards for responsible AI use in education, fostering trust and reliability among users [20].

Integration With Broader Educational Ecosystems

Finally, future innovations could focus on integrating the chatbot with broader educational ecosystems, including learning management systems and other digital educational tools [94,95]. By doing so, the chatbot could become a seamless part of the educational workflow, providing continuous support to students and educators across various platforms. This integration would enable more comprehensive data collection and analysis, allowing educators to monitor student progress in real time and make data-informed decisions to enhance teaching strategies. The chatbot could also facilitate collaborative learning by connecting users with peers, mentors, or experts, creating a more interactive and supportive educational community.

Adaptability to Other Medical Fields

The framework developed for the radiotherapy education chatbot demonstrates significant potential for adaptation to other medical fields. By leveraging advanced NLP algorithms and personalized learning experiences, the chatbot can be tailored to provide specialized educational content across various medical disciplines, such as cardiology, neurology, and oncology. The integration of VR/AR technologies can further enhance learning by offering immersive simulations of medical procedures relevant to each field. Furthermore, the chatbot's multilingual and cross-cultural capabilities ensure that it can deliver accurate and contextually appropriate information to a diverse global audience. By addressing ethical AI considerations and integrating with broader educational ecosystems [18], the framework can support continuous learning and professional development for health care professionals worldwide, ultimately improving patient care and outcomes.

Conclusions

This Viewpoint paper has explored the development of a resilient framework for a medical chatbot tailored to radiotherapy education, addressing key aspects such as accuracy, reliability, privacy, ethics, and the potential for innovation. The framework emphasizes the importance of maintaining up-to-date and accurate information, ensuring user trust through robust privacy measures, and fostering an ethical approach to AI in

health care. Through performance evaluation, the chatbot demonstrated its capability to enhance learning outcomes and support health care professionals in their continuous education. Challenges such as bias, user engagement, and integration into existing systems were identified, with strategies proposed to overcome these obstacles.

Looking ahead, the future of LLM-based medical chatbots holds significant promise for radiotherapy education and health care as a whole. These technologies have the potential to revolutionize how complex medical knowledge is disseminated, making education more accessible, personalized, and interactive. By continuing to advance in areas such as NLP, personalized learning, and integration with immersive technologies, LLM-based chatbots can become indispensable tools in both educational and clinical settings. As these chatbots evolve, they will likely play a crucial role in shaping the future of health care, improving patient outcomes, and supporting the ongoing education of health care professionals worldwide. The framework presented in this paper serves as a foundational guide for the responsible and effective implementation of these powerful tools, ensuring that they contribute positively to the field of radiotherapy and beyond.

Acknowledgments

This work was supported by the Planning and Dissemination Grant–Institute Community Support, Canadian Institutes of Health Research, Canada (CHIR PCS–191021). The authors acknowledge the use of ChatGPT, accessed on January 1, 2025, to assist in polishing the English in the paper. After using this tool, the authors reviewed and edited the content as necessary and assume full responsibility for the paper's content.

Authors' Contributions

JCLC and KL were responsible for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, and reviewing and editing the manuscript. JCLC was responsible for visualization and writing the original draft.

Conflicts of Interest

None declared.

References

1. Siddique S, Chow JC. Artificial intelligence in radiotherapy. Rep Pract Oncol Radiother 2020 Jul;25(4):656-666 [FREE Full text] [doi: 10.1016/j.rpor.2020.03.015] [Medline: 32617080]
2. Chow JC. Internet-based computer technology on radiotherapy. Rep Pract Oncol Radiother 2017 Nov;22(6):455-462 [FREE Full text] [doi: 10.1016/j.rpor.2017.08.005] [Medline: 28932174]
3. Siddique S, Chow JC. Machine learning in healthcare communication. Encyclopedia 2021 Feb 14;1(1):220-239. [doi: 10.3390/encyclopedia1010021]
4. Chow JCL, Wong V, Li K. Generative pre-trained transformer-empowered healthcare conversations: current trends, challenges, and future directions in large language model-enabled medical chatbots. BioMedInformatics 2024 Mar 14;4(1):837-852. [doi: 10.3390/biomedinformatics4010047]
5. Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat Med 2023 Aug 17;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
6. Chang Y, Wang X, Wang J, Wu Y, Yang L, Zhu K, et al. A survey on evaluation of large language models. ACM Trans Intell Syst Technol 2024 Mar 29;15(3):1-45. [doi: 10.1145/3641289]
7. Tirumala K, Markosyan A, Zettlemoyer L, Aghajanyan A. Memorization without overfitting: analyzing the training dynamics of large language models. In: Proceedings of the 2022 Conference on Advances in Neural Information Processing Systems. 2022 Presented at: NeurIPS '22; November 28-December 9, 2022; New Orleans, LA p. 35 URL: https://proceedings.neurips.cc/paper_files/paper/2022/hash/fa0509f4dab6807e2cb465715bf2d249-Abstract-Conference.html
8. Dam SK, Hong CS, Qiao Y, Zhang C. A complete survey on LLM-based AI chatbots. arXiv Preprint posted online June 17, 2022 [FREE Full text]
9. Yang R, Tan TF, Lu W, Thirunavukarasu AJ, Ting DS, Liu N. Large language models in health care: development, applications, and challenges. Health Care Sci 2023 Aug 24;2(4):255-263 [FREE Full text] [doi: 10.1002/hcs2.61] [Medline: 38939520]
10. Chakraborty C, Bhattacharya M, Lee SS. Need an AI-enabled, next-generation, advanced ChatGPT or large language models (LLMs) for error-free and accurate medical information. Ann Biomed Eng 2024 Feb 27;52(2):134-135. [doi: 10.1007/s10439-023-03297-9] [Medline: 37368124]
11. Lai T, Shi Y, Du Z, Wu J, Fu K, Dou Y, et al. Supporting the demand on mental health services with AI-based conversational large language models (LLMs). BioMedInformatics 2023 Dec 22;4(1):8-33. [doi: 10.3390/biomedinformatics4010002]
12. Pool J, Indulska M, Sadiq S. Large language models and generative AI in telehealth: a responsible use lens. J Am Med Inform Assoc 2024 Sep 01;31(9):2125-2136. [doi: 10.1093/jamia/ocae035] [Medline: 38441296]

13. Montagna S, Ferretti S, Klopfenstein LC, Florio A, Pengo M. Data decentralisation of LLM-based chatbot systems in chronic disease self-management. In: Proceedings of the 2023 ACM Conference on Information Technology for Social Good. 2023 Presented at: GoodIT '23; September 6-8, 2023; Lisbon, Portugal p. 205-212 URL: https://dl.acm.org/doi/10.1145/3582515.3609536 [doi: 10.1145/3582515.3609536]
14. Bengio Y, Hinton G, Yao A, Song D, Abbeel P, Darrell T, et al. Managing extreme AI risks amid rapid progress. Science 2024 May 24;384(6698):842-845. [doi: 10.1126/science.adn0117] [Medline: 38768279]
15. Chandra RA, Keane FK, Voncken FE, Thomas CR. Contemporary radiotherapy: present and future. Lancet 2021 Jul 10;398(10295):171-184. [doi: 10.1016/S0140-6736(21)00233-6] [Medline: 34166607]
16. Shafiq J, Barton M, Noble D, Lemer C, Donaldson LJ. An international review of patient safety measures in radiotherapy practice. Radiother Oncol 2009 Jul;92(1):15-21. [doi: 10.1016/j.radonc.2009.03.007] [Medline: 19395105]
17. Xu L, Sanders L, Li K, Chow JC. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer 2021 Nov 29;7(4):e27850 [FREE Full text] [doi: 10.2196/27850] [Medline: 34847056]
18. Chow JC, Li K. Ethical considerations in human-centered AI: advancing oncology chatbots through large language models. JMIR Bioinform Biotechnol 2024 Nov 06;5:e64406 [FREE Full text] [doi: 10.2196/64406] [Medline: 39321336]
19. Li J. Security implications of AI chatbots in health care. J Med Internet Res 2023 Nov 28;25:e47551 [FREE Full text] [doi: 10.2196/47551] [Medline: 38015597]
20. Chow JC, Sanders L, Li K. Impact of ChatGPT on medical chatbots as a disruptive technology. Front Artif Intell 2023 Apr 5;6:1166014 [FREE Full text] [doi: 10.3389/frai.2023.1166014] [Medline: 37091303]
21. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ 2023 Jun 01;9:e48291 [FREE Full text] [doi: 10.2196/48291] [Medline: 37261894]
22. Heaven WD. Geoffrey Hinton tells us why he's now scared of the tech he helped build. MIT Technology Review. URL: https://www.technologyreview.com/2023/05/02/1072528/geoffrey-hinton-google-why-scared-ai/ [accessed 2024-04-29]
23. Roumeliotis KI, Tselikas ND. ChatGPT and Open-AI models: a preliminary review. Future Internet 2023 May 26;15(6):192. [doi: 10.3390/fi15060192]
24. Bedi S, Liu Y, Orr-Ewing L, Dash D, Koyejo S, Callahan A, et al. Testing and evaluation of health care applications of large language models: a systematic review. JAMA 2025 Jan 28;333(4):319-328. [doi: 10.1001/jama.2024.21700] [Medline: 39405325]
25. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
26. Giudici M, Padalino L, Paolino G, Paratici I, Pascu AI, Garzotto F. Designing home automation routines using an LLM-based chatbot. Designs 2024 May 13;8(3):43. [doi: 10.3390/designs8030043]
27. Gupta M, Gupta A, Kaur G. Deep learning for medical healthcare: issues, challenges, and opportunities. In: Gupta M, Gupta A, Kaur G, editors. Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications. Hoboken, NJ: Wiley-Blackwell; 2022:28-92.
28. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. J Med Internet Res 2021 Mar 23;23(3):e24850 [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]
29. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
30. Chow JC. Artificial intelligence in radiotherapy and patient care. In: Lidströmer N, Ashrafian H, editors. Artificial Intelligence in Medicine. Cham, Switzerland: Springer; 2021:1-13.
31. Rebelo N, Sanders L, Li K, Chow JC. Learning the treatment process in radiotherapy using an artificial intelligence-assisted chatbot: development study. JMIR Form Res 2022 Dec 02;6(12):e39443 [FREE Full text] [doi: 10.2196/39443] [Medline: 36327383]
32. Chow JC, Wong V, Sanders L, Li K. Developing an AI-assisted educational chatbot for radiotherapy using the IBM Watson assistant platform. Healthcare (Basel) 2023 Aug 29;11(17):2417 [FREE Full text] [doi: 10.3390/healthcare11172417] [Medline: 37685452]
33. Wang A, Qian Z, Briggs L, Cole AP, Reis LO, Trinh QD. The use of chatbots in oncological care: a narrative review. Int J Gen Med 2023 May;Volume 16:1591-1602. [doi: 10.2147/ijgm.s408208]
34. Kovacek D, Chow JC. An AI-assisted chatbot for radiation safety education in radiotherapy. IOPSciNotes 2021 Aug 30;2(3):034002. [doi: 10.1088/2633-1357/ac1f88]
35. Clark M, Bailey S. Chatbots in health care: connecting patients to information. Can J Health Technol 2024 Jan 22;4(1). [doi: 10.51731/cjht.2024.818]
36. Nor A, Nur S, Muhamad K, Rozilawati A, Noorazrul Y, Muhammad F, et al. Radiotherapy communication skills training using chatbot-based prototype (SCIMORT). J Med Imaging Radiat Sci 2022 Dec;53(4):S1. [doi: 10.1016/j.jmir.2022.10.005]

37. Goodman RS, Patrinely JR, Stone CA, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023 Oct 02;6(10):e2336483 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36483] [Medline: 37782499]
38. Luo B, Lau RY, Li C, Si YW. A critical review of state‐of‐the‐art chatbot designs and applications. Wiley Interdiscip Rev Data Min Knowl Discov 2021 Oct 25;12(1):e1434. [doi: 10.1002/widm.1434]
39. Pilato G, Augello A, Gaglio S. A modular system oriented to the design of versatile knowledge bases for chatbots. ISRN Artif Intell 2012 Mar 05;2012:1-10. [doi: 10.5402/2012/363840]
40. Shalaby W, Arantes A, GonzalezDiaz T, Gupta C. Building chatbots from large scale domain-specific knowledge bases: challenges and opportunities. In: Proceedings of the 2020 IEEE International Conference on Prognostics and Health Management. 2020 Presented at: ICPHM '20; June 8-10, 2020; Detroit, MI p. 1-8 URL: https://ieeexplore.ieee.org/document/9187036 [doi: 10.1109/icphm49022.2020.9187036]
41. English A, Ford CA. The HIPAA privacy rule and adolescents: legal questions and clinical challenges. Perspect Sex Reprod 2007 Feb 12;36(2):80-86. [doi: 10.1363/3608004]
42. Hoofnagle CJ, van der Sloot B, Borgesius FZ. The European Union general data protection regulation: what it is and what it means. Inf Commun Technol Law 2019 Feb 10;28(1):65-98. [doi: 10.1080/13600834.2019.1573501]
43. Hasal M, Nowaková J, Ahmed Saghair K, Abdulla H, Snášel V, Ogiela L. Chatbots: security, privacy, data protection, and social aspects. Concurr Comput Pract Exp 2021 Jun 03;33(19):e6426. [doi: 10.1002/cpe.6426]
44. Kim B, Park J, Suh J. Transparency and accountability in AI decision support: explaining and visualizing convolutional neural networks for text information. Decis Support Syst 2020 Jul;134:113302. [doi: 10.1016/j.dss.2020.113302]
45. Ding R, Palomares I, Wang X, Yang GR, Liu B, Dong Y, et al. Large-Scale decision-making: characterization, taxonomy, challenges and future directions from an Artificial Intelligence and applications perspective. Inf Fusion 2020 Jul;59:84-102. [doi: 10.1016/j.inffus.2020.01.006]
46. Guan H, Dong L, Zhao A. Ethical risk factors and mechanisms in artificial intelligence decision making. Behav Sci (Basel) 2022 Sep 16;12(9):343 [FREE Full text] [doi: 10.3390/bs12090343] [Medline: 36135147]
47. Williams R, Cloete R, Cobbe J, Cottrill C, Edwards P, Markovic M, et al. From transparency to accountability of intelligent systems: moving beyond aspirations. Data Policy 2022 Feb 18;4:e7. [doi: 10.1017/dap.2021.37]
48. Xue J, Wang YC, Wei C, Liu X, Woo J, Kuo CC. Bias and fairness in chatbots: an overview. arXiv Preprint posted online September 16, 2023 [FREE Full text] [doi: 10.1561/116.00000064]
49. Srinivasan R, González BS. The role of empathy for artificial intelligence accountability. J Responsib Technol 2022 Apr;9:100021. [doi: 10.1016/j.jrt.2021.100021]
50. Tawfeeq TM, Awqati AJ, Jasim YA. The ethical implications of ChatGPT AI chatbot: a review. J Modern Comput Eng Res 2022;2023:49-57 [FREE Full text]
51. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of artificial intelligence chatbot responses to top searched queries about cancer. JAMA Oncol 2023 Oct 01;9(10):1437-1440. [doi: 10.1001/jamaoncol.2023.2947] [Medline: 37615960]
52. Zhou J, Zhang Y, Luo Q, Parker AG, De Choudhury M. Synthetic lies: understanding ai-generated misinformation and evaluating algorithmic and human solutions. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '23; April 29-May 5, 2022; Virtual Event p. 1-20 URL: https://dl.acm.org/doi/full/10.1145/3544548.3581318 [doi: 10.1145/3544548.3581318]
53. Monteith S, Glenn T, Geddes JR, Whybrow PC, Achtyes E, Bauer M. Artificial intelligence and increasing misinformation. Br J Psychiatry 2024 Feb 26;224(2):33-35. [doi: 10.1192/bjp.2023.136] [Medline: 37881016]
54. Waldrop MM. How to mitigate misinformation. Proc Natl Acad Sci U S A 2023 Sep 05;120(36):e2314143120 [FREE Full text] [doi: 10.1073/pnas.2314143120] [Medline: 37647368]
55. Lee JW, Yoo IS, Kim JH, Kim WT, Jeon HJ, Yoo H, et al. Development of AI-generated medical responses using the ChatGPT for cancer patients. Comput Methods Programs Biomed 2024 Sep;254:108302 [FREE Full text] [doi: 10.1016/j.cmpb.2024.108302] [Medline: 38996805]
56. Lommatzsch A, Katins J. An information retrieval-based approach for building intuitive chatbots for large knowledge bases. In: Proceedings of the 2022 Conference on Lernen, Wissen, Daten, Analysen. 2022 Presented at: LWDA '19; September 30-October 2, 2019; Berlin, Germany p. 1 URL: https://ceur-ws.org/Vol-2454/paper_60.pdf
57. Minkkinen M, Laine J, Mäntymäki M. Continuous auditing of artificial intelligence: a conceptualization and assessment of tools and frameworks. Digit Soc 2022 Oct 04;1(3):1. [doi: 10.1007/S44206-022-00022-2]
58. Davidian K. New space and generative artificial intelligence. New Space 2023 Sep 01;11(3):148-150. [doi: 10.1089/space.2023.29050.editorial]
59. Bateman S, Brooks C, McCalla G. Collaborative tagging approaches for ontological metadata in adaptive e-learning sys-tems. In: Proceedings of the 4th international workshop on applications of semantic web technologies for e-learning. 2006 Presented at: SW-EL '06; June 20, 2006; Virtual Event p. 3-12 URL: https://www.researchgate.net/profile/Gordon-Mccalla/publication/228464205_Collaborative_tagging_approaches_for_ontological_metadata_in_adaptive_e-learning_systems/links/09e415093093c97b7f000000/Collaborative-tagging-approaches-for-ontological-metadata-in-adaptive-e-learning-systems.

pdf?__cf_chl_tk=abcTxToN94_dEPf9TZkF.C1r7ibOKThA6yiPuS5HNGQ-1738762941-1.0.1.1-FB3Ss_5QH_GRML7bRa9ySZbuZRxYEVzqYoOo7vCAP28
60. Woo H, Kim J, Lee W. Analysis of cross-referencing artificial intelligence topics based on sentence modeling. Appl Sci 2020 May 26;10(11):3681. [doi: 10.3390/app10113681]
61. Chen Q, Gong Y, Lu Y, Tang J. Classifying and measuring the service quality of AI chatbot in frontline service. J Bus Res 2022 Jun;145:552-568. [doi: 10.1016/j.jbusres.2022.02.088]
62. Hsu IC, Chang CC. Integrating machine learning and open data into social chatbot for filtering information rumor. J Ambient Intell Humaniz Comput 2021 May 20;12(1):1023-1037 [FREE Full text] [doi: 10.1007/s12652-020-02119-3] [Medline: 32837593]
63. Vijayaraghavan V, Cooper JB. Algorithm inspection for chatbot performance evaluation. Procedia Comput Sci 2020;171:2267-2274. [doi: 10.1016/j.procs.2020.04.245]
64. Shawar BA, Atwell E. Different measurement metrics to evaluate a chatbot system. In: Proceedings of the 2007 Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies. 2007 Presented at: NAACL-HLT-Dialog '07; April 26, 2007; New York, NY p. 89-96 URL: https://dl.acm.org/doi/10.5555/1556328.1556341 [doi: 10.3115/1556328.1556341]
65. Janssen A, Cardona DR, Passlick J, Breitner MH. How to make chatbots productive – a user-oriented implementation framework. Int J Hum Comput Stud 2022 Dec;168:102921. [doi: 10.1016/j.ijhcs.2022.102921]
66. Ko AJ, Lee MJ, Ferrari V, Ip S, Tran C. A case study of post-deployment user feedback triage. In: Proceedings of the 4th International Workshop on Cooperative and Human Aspects of Software Engineeri. 2011 Presented at: CHASE '11; May 21, 2011; Honolulu, HI p. 1-8 URL: https://dl.acm.org/doi/abs/10.1145/1984642.1984644 [doi: 10.1145/1984642.1984644]
67. May R, Denecke K. Security, privacy, and healthcare-related conversational agents: a scoping review. Inform Health Soc Care 2022 Apr 03;47(2):194-210. [doi: 10.1080/17538157.2021.1983578] [Medline: 34617857]
68. Shumanov M, Johnson L. Making conversations with chatbots more personalized. Comput Human Behav 2021 Apr;117:106627. [doi: 10.1016/j.chb.2020.106627]
69. Nicolau A. The impact of AI on internal audit and accounting practices. Int Audit Risk Manag 2021:38-56 [FREE Full text]
70. Følstad A, Araujo T, Law EL, Brandtzaeg PB, Papadopoulos S, Reis L, et al. Future directions for chatbot research: an interdisciplinary research agenda. Computing 2021 Oct 19;103(12):2915-2942. [doi: 10.1007/S00607-021-01016-7]
71. Xin Q, He Y, Pan Y, Wang Y, Du S. The implementation of an AI-driven advertising push system based on a NLP algorithm. Int J Cogn Comput Sci Inf Technol 2023 Dec 30;1(1):30-37. [doi: 10.62051/ijcsit.v1n1.05]
72. Durall Gazulla E, Martins L, Fernández-Ferrer M. Designing learning technology collaboratively: analysis of a chatbot co-design. Educ Inf Technol (Dordr) 2023 Jun 24;28(1):109-134 [FREE Full text] [doi: 10.1007/s10639-022-11162-w] [Medline: 35765268]
73. Kumar R, Ali MM. A review on chatbot design and implementation techniques. Int J Eng Technol 2020;7(11):2791 [FREE Full text]
74. Chow JC, Sanders L, Li K. Design of an educational chatbot using artificial intelligence in radiotherapy. AI 2023 Mar 02;4(1):319-332. [doi: 10.3390/ai4010015]
75. Denecke K, Abd-Alrazaq A, Househ M, Warren J. Evaluation metrics for health chatbots: a Delphi study. Methods Inf Med 2021 Dec 31;60(5-06):171-179. [doi: 10.1055/s-0041-1736664] [Medline: 34719011]
76. Kaur A, Singh S, Chandan JS, Robbins T, Patel V. Qualitative exploration of digital chatbot use in medical education: a pilot study. Digit Health 2021 Aug 31;7:20552076211038151 [FREE Full text] [doi: 10.1177/20552076211038151] [Medline: 34513002]
77. Gabrielli S, Rizzi S, Carbone S, Donisi V. A chatbot-based coaching intervention for adolescents to promote life skills: pilot study. JMIR Hum Factors 2020 Feb 14;7(1):e16762 [FREE Full text] [doi: 10.2196/16762] [Medline: 32130128]
78. Stout NL, Silver JK, Raj VS, Rowland J, Gerber L, Cheville A, et al. Toward a national initiative in cancer rehabilitation: recommendations from a subject matter expert group. Arch Phys Med Rehabil 2016 Nov;97(11):2006-2015. [doi: 10.1016/j.apmr.2016.05.002] [Medline: 27237580]
79. Abd-Alrazaq A, Safi Z, Alajlani M, Warren J, Househ M, Denecke K. Technical metrics used to evaluate health care chatbots: scoping review. J Med Internet Res 2020 Jun 05;22(6):e18301 [FREE Full text] [doi: 10.2196/18301] [Medline: 32442157]
80. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a chatbot (Ada) in the diagnosis of mental disorders: comparative case study with lay and expert users. JMIR Form Res 2019 Oct 29;3(4):e13863 [FREE Full text] [doi: 10.2196/13863] [Medline: 31663858]
81. Shan Y, Ji M, Xie W, Zhang X, Qian X, Li R, et al. Use of health care chatbots among young people in China during the omicron wave of COVID-19: evaluation of the user experience of and satisfaction with the technology. JMIR Hum Factors 2022 Jun 09;9(2):e36831 [FREE Full text] [doi: 10.2196/36831] [Medline: 35576058]
82. Laymouna M, Ma Y, Lessard D, Schuster T, Engler K, Lebouché B. Roles, users, benefits, and limitations of chatbots in health care: rapid review. J Med Internet Res 2024 Jul 23;26:e56930 [FREE Full text] [doi: 10.2196/56930] [Medline: 39042446]

83. Ren R, Zapata M, Castro JW, Dieste O, Acuna ST. Experimentation for chatbot usability evaluation: a secondary study. IEEE Access 2022;10:12430-12464. [doi: 10.1109/access.2022.3145323]
84. Montenegro JL, da Costa CA, Janssen LP. Evaluating the use of chatbot during pregnancy: a usability study. Healthc Anal 2022 Nov;2:100072. [doi: 10.1016/j.health.2022.100072]
85. Ye BJ, Kim JY, Suh C, Choi SP, Choi M, Kim DH, et al. Development of a chatbot program for follow-up management of workers' general health examinations in Korea: a pilot study. Int J Environ Res Public Health 2021 Feb 23;18(4):2170 [FREE Full text] [doi: 10.3390/ijerph18042170] [Medline: 33672158]
86. Singh S, Mahmood A. The NLP cookbook: modern recipes for transformer based deep learning architectures. IEEE Access 2021;9:68675-68702. [doi: 10.1109/access.2021.3077350]
87. Jeong H, Yoo JH, Han O. Next-generation chatbots for adaptive learning: a proposed framework. J Internet Comput Serv 2023;24(4):37-45 [FREE Full text]
88. KC GP, Ranjan S, Ankit T, Kumar V. A personalized medical assistant chatbot: Medibot. Int J Sci Technol Eng 2019;5(7):42-46 [FREE Full text]
89. Zhang H, Zheng J. The application analysis of medical chatbots and virtual assistant. Front Soc Sci Technol 2021;3(2):11-16 [FREE Full text] [doi: 10.25236/fsst.2021.030202]
90. Lin AP, Trappey CV, Luan CC, Trappey AJ, Tu KL. A test platform for managing school stress using a virtual reality group chatbot counseling system. Appl Sci 2021 Sep 29;11(19):9071. [doi: 10.3390/app11199071]
91. Orosoo M, Goswami I, Alphonse FR, Fatma G, Rengarajan M, Bala BK. Enhancing natural language processing in multilingual chatbots for cross-cultural communication. In: Proceedings of the 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks. 2024 Presented at: ICICV '24; March 11-12, 2024; Virtual Event p. 127-133 URL: https://www.computer.org/csdl/proceedings-article/icicv/2024/856400a127/1WKFMzaKGcw [doi: 10.1109/icicv62344.2024.00027]
92. Fournier-Tombs E, McHardy J. A medical ethics framework for conversational artificial intelligence. J Med Internet Res 2023 Jul 26;25:e43068 [FREE Full text] [doi: 10.2196/43068] [Medline: 37224277]
93. Chang CY, Kuo SY, Hwang GH. Chatbot-facilitated nursing education: incorporating a knowledge-based chatbot system into a nursing training program. Educ Technol Soc 2022;25(1):15-27 [FREE Full text]
94. Murad DF, Iskandar AG, Fernando E, Octavia TS, Maured DE. Towards smart LMS to improve learning outcomes students using LenoBot with natural language processing. In: Proceedings of the 6th International Conference on Information Technology, Computer and Electrical Engineering. 2019 Presented at: ICITACEE '19; September 26, 2019 26; Semarang, Indonesia p. 1-6 URL: https://ieeexplore.ieee.org/document/8904311 [doi: 10.1109/icitacee.2019.8904311]
95. Parviainen J, Rantala J. Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. Med Health Care Philos 2022 Mar 04;25(1):61-71 [FREE Full text] [doi: 10.1007/s11019-021-10049-w] [Medline: 34480711]

Abbreviations

AI: artificial intelligence
LLM: large language model
NLP: natural language processing
VR/AR: virtual and augmented reality

Edited by N Cahill; submitted 18.09.24; peer-reviewed by M Chatzimina, M Dursun, H Polat; comments to author 11.12.24; revised version received 15.12.24; accepted 16.01.25; published 18.02.25.

Please cite as:
Chow JCL, Li K
Developing Effective Frameworks for Large Language Model–Based Medical Chatbots: Insights From Radiotherapy Education With ChatGPT
JMIR Cancer 2025;11:e66633
URL: https://cancer.jmir.org/2025/1/e66633
doi:10.2196/66633
PMID:

Original Paper

Ethical Design of Data-Driven Decision Support Tools for Improving Cancer Care: Embedded Ethics Review of the 4D PICTURE Project

Marieke Bak[1,2*], PhD; Laura Hartman[3*], PhD; Charlotte Graafland[3], MA; Ida J Korfage[4], PhD; Alena Buyx[1], PhD; Maartje Schermer[5], PhD; 4D PICTURE Consortium[4*]

[1]Institute of History and Ethics in Medicine, Department of Preclinical Medicine, TUM School of Medicine and Health, Technical University of Munich, München, Germany

[2]Department of Ethics, Law and Humanities, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

[3]Department of Medical Ethics, Philosophy and History of Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

[4]see Acknowledgments

[5]Erasmus School of Health Policy & Management, Rotterdam, The Netherlands

[*]these authors contributed equally

Corresponding Author:
Marieke Bak, PhD
Institute of History and Ethics in Medicine, Department of Preclinical Medicine
TUM School of Medicine and Health
Technical University of Munich
Ismaninger Street 22
München, 81675
Germany
Phone: 49 43480533
Email: marieke.bak@tum.de

Abstract

Oncology patients often face complex choices between treatment regimens with different risk-benefit ratios. The 4D PICTURE (Producing Improved Cancer Outcomes Through User-Centered Research) project aims to support patients, their families, and clinicians with these complex decisions by developing data-driven decision support tools (DSTs) for patients with breast cancer, prostate cancer, and melanoma as part of care path redesign using a methodology called MetroMapping. There are myriad ethical issues to consider as the project will create data-driven prognostic models and develop conversation tools using artificial intelligence while including patient perspectives by setting up boards of experiential experts in 8 different countries. This paper aims to review the key ethical challenges related to the design and development of DSTs in oncology. To explore the ethics of DSTs in cancer care, the project adopted the Embedded Ethics approach—embedding ethicists into research teams to sensitize team members to ethical aspects and assist in reflecting on those aspects throughout the project. We conducted what we call an embedded review of the project drawing from key literature on topics related to the different work packages of the 4D PICTURE project, whereas the analysis was an iterative process involving discussions with researchers in the project. Our review identified 13 key ethical challenges related to the development of DSTs and the redesigning of care paths for more personalized cancer care. Several ethical aspects were related to general potential issues of data bias and privacy but prompted specific research questions, for instance, about the inclusion of certain demographic variables in models. Design methodology in the 4D PICTURE project can provide insights related to design justice, a novel consideration in health care DSTs. Ethical points of attention related to health care policy, such as cost-effectiveness, financial sustainability, and environmental impact, were also identified, along with challenges in the research process itself, emphasizing the importance of epistemic justice, the role of embedded ethicists, and psychological safety. This viewpoint highlights ethical aspects previously neglected in the digital health ethics literature and zooms in on real-world challenges in an ongoing project. It underscores the need for researchers and leaders in data-driven medical research projects to address ethical challenges beyond the scientific core of the project. More generally, our tailored review approach provides a model for embedding ethics into large data-driven oncology research projects from the start, which helps ensure that technological innovations are designed and developed in an appropriate and patient-centered manner.

(JMIR Cancer 2025;11:e65566) doi:10.2196/65566

KEYWORDS

shared decision-making; oncology; IT; ethics; decision support tools; big data; medical decision-making; artificial intelligence

Introduction

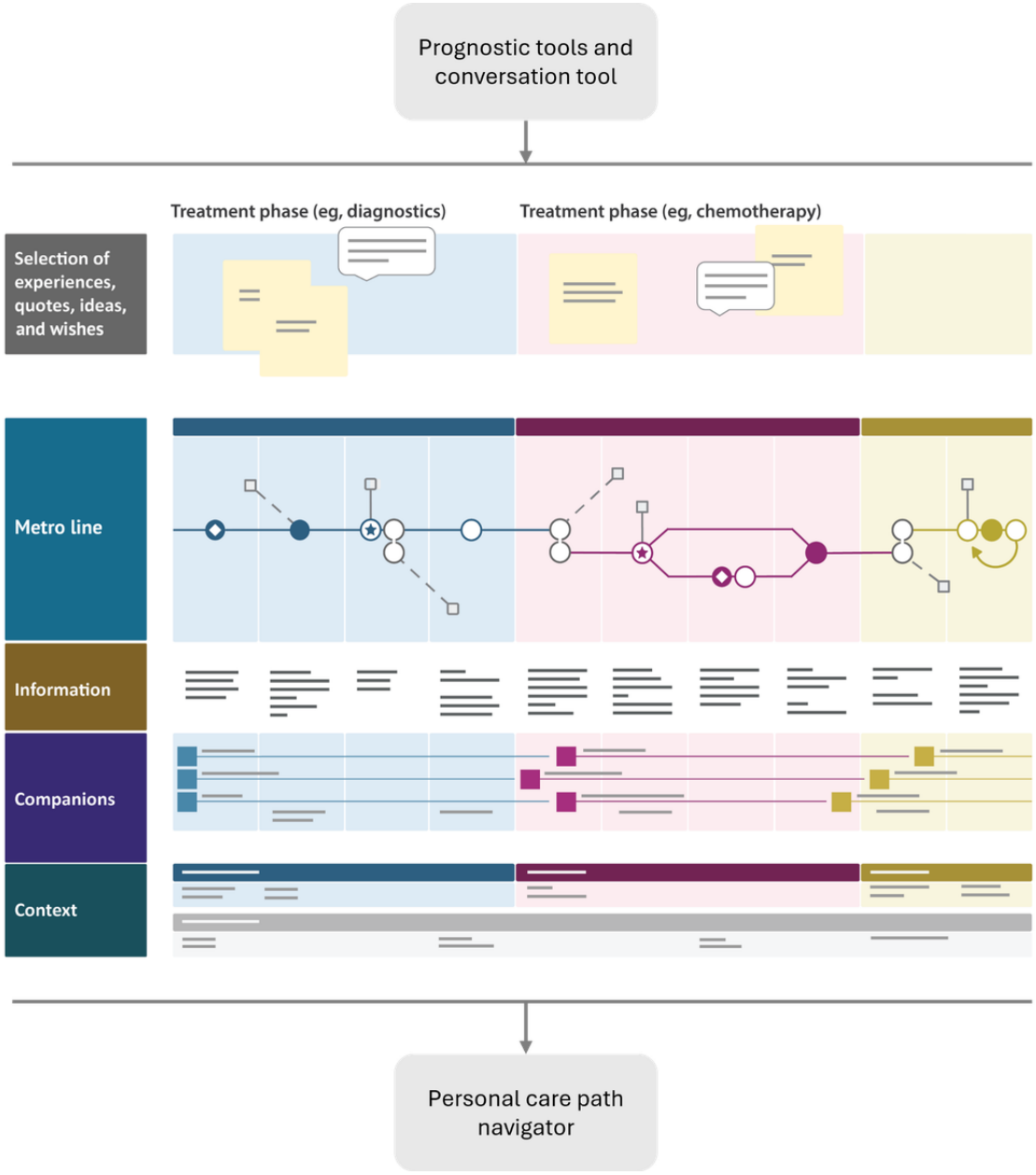
People diagnosed with cancer often face difficult choices regarding their treatment and potential impact on survival and quality of life [1]. Data-driven decision support tools (DSTs) hold significant potential in empowering patients, enhancing personalized care, and fostering health equity [2,3]. Nevertheless, most current DSTs do not account for individual preferences, which hinders their broader integration into clinical practice. To improve cancer treatment decision-making by addressing existing challenges in DSTs, a large European collaboration was started—the 4D PICTURE (Producing Improved Cancer Outcomes Through User-Centered Research) project [4]. Recognizing the complexity that patients face, the consortium seeks to use design methods (particularly the *MetroMapping* methodology, Figure 1) to improve care paths in oncology. This involves the development of innovative prognostic models and conversation tools that consider patient experiences, values, and preferences through models partly based on artificial intelligence (AI). Collaborating with patients and other stakeholders, the project focuses on breast cancer, prostate cancer, and melanoma, aiming to create comprehensive DSTs for these types of cancer. The use of these tools in the *MetroMap* for redesigning cancer care paths will be evaluated on effectiveness and cost-effectiveness, and social and ethical concerns will be addressed throughout the project.

Ethics is highly relevant to the development of data-driven DSTs for personalizing oncology care. For instance, the use of low-quality prognostic models may lead to incorrect and harmful decisions. When using AI-driven DSTs, concerns about quality are particularly warranted as there is still a lack of robust evidence on their effectiveness [5]. As such, normative principles such as data quality, algorithmic fairness, and data privacy are important to consider when developing data-driven DSTs. However, principles alone cannot guarantee that the developed tools are ethical and acceptable to patients and health care providers [6]. What is needed as well is guidance on how researchers can be practically assisted to anticipate, identify, and address ethical issues of data-driven care based on the specific case at hand. This can be done through the *Embedded Ethics* approach, which stimulates close collaborations between ethicists on the one hand and developers, researchers, and clinicians on the other who work together in an iterative and continuous manner [7]. In the 4D PICTURE project, ethicists are embedded into the project in this way to promote guidance and reflection on the ethics of the entire project.

The first task of the ethics *work package* in a large interdisciplinary project is usually to create an overview of potential ethical challenges that can be expected in that project based on the literature. However, we noticed that such literature surveys often come up with the same general issues. A database search on the ethical aspects of data-driven DSTs in medicine is necessarily broad and will provide high-level findings on the aforementioned principles, which still requires translation to the project at hand to derive actionable recommendations. Moreover, the ethical aspects of such interdisciplinary projects are too heterogenous for a systematic review on the "ethics of data-driven DSTs in healthcare"—ethical questions may also arise in parts of the project not directly related to DST development, such as their evaluation or the dissemination of results. Therefore, in the 4D PICTURE project, we took a different approach. As ethicists in the 4D PICTURE project, we discussed ethics in relation to the different work packages with the project researchers and looked for key publications in the ethics literature on the topics that came up in each work package separately. We moved back and forth between literature and practice in an iterative process to be as specific and close to practice as possible. This resulted in an agenda of aspects that may be ethically relevant within the project and serves as a basis for further empirical and theoretical ethics research.

This viewpoint describes this process and has two interrelated aims as it (1) introduces the *embedded review* approach for identifying ethical aspects within an interdisciplinary research consortium and (2) outlines key ethical challenges to be considered when developing data-driven DSTs for more personalized oncology care. In what follows, we describe our methodology before discussing ethical challenges related to data-driven DSTs in the project under study. This paper provides a detailed overview of the ethics of data-driven DSTs because of the link to a particular project, as well as reference to broader ethical aspects (eg, the ethics of interdisciplinary collaboration or psychological safety in research teams) that are often neglected. We find that this work has relevance beyond the 4D PICTURE project as it is the first review-type paper explicitly grounded in the *Embedded Ethics* approach. Our findings show that, even when simply looking for literature on ethical aspects, a lot can be gained when ethics is embedded into a project from the start.

Figure 1. Schematic illustration of the MetroMap that forms the core of the 4D PICTURE (Producing Improved Cancer Outcomes Through User-Centered Research) project. The MetroMap is a comprehensive visualization of the general care trajectory. Feeding into the MetroMap will be the results of two types of models developed in the project: (1) a treatment outcome prediction tool for each cancer type and (2) a conversation tool developed by analyzing patient experiences through text mining. The result of integrating these data-driven tools into the MetroMap will be a personal care path navigator for each patient that serves as a decision aid in shared decision-making.



Methods

Embedded Ethics Approach and Study Design

As ethics is about normative argumentation and conceptual analysis, a systematic literature study usually does not suffice for an overview of the ethical literature on a certain topic. Moreover, the ethical aspects of large interdisciplinary projects such as the 4D PICTURE project are too heterogenous for a traditional systematic or scoping review. Therefore, different methods for reviews have been described in the ethics literature suitable for different purposes, ranging from a *rapid review* to a *critical interpretative review* [8]. For the context of this study, our main aim was to sensitize researchers of the consortium to the ethical issues in their work packages and support them with a shared ethical framework, vocabulary, and argumentation to navigate the various ethical issues of the project deliberately and consciously. Our priority was to ensure that the literature was useful, comprehensible, and relevant for the consortium's needs, thus balancing methodological rigor and depth of analysis with practical applicability. In conducting this review, we aimed to develop directions for specific ethics guidance and further research as well as to "create a shared knowledge base among team members" [9]. Moreover, we wanted to include the ethics *of* the consortium (eg, ensuring psychological safety in research teams) rather than merely looking at the ethics *in* the consortium (eg, avoiding biased outcomes of the prognostic models).

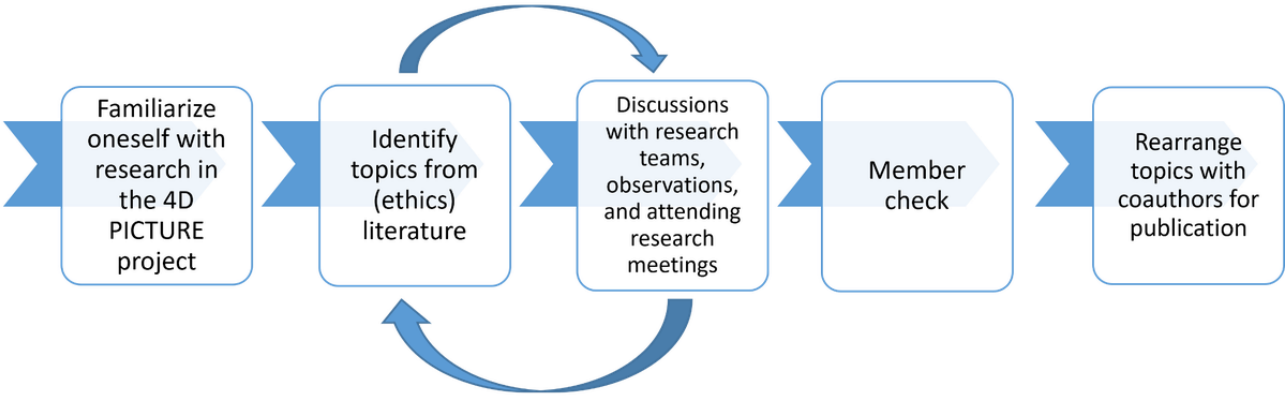
Therefore, we opted for an *embedded review*, which, in the typology of McDougall [8], could be best described as a critical

interpretative review, with the main difference that we established the analytical categories together with the researchers of the consortium. This is in line with the paradigm of *Embedded Ethics*, a relatively new approach for integrating ethics into interdisciplinary health care projects that focuses on delivering guidance for practical ethical dilemmas also when these are unexpected and come up ad hoc. This is done by embedding ethicists into these projects and stimulating close collaboration so that ethical aspects are taken into account in a continuous and iterative manner [7]. Various tools and methods are used to embed ethics into the development of new technologies; usually, a literature review is the starting point. As Willem et al [9] note, a literature review in an embedded ethics project "provides the opportunity to collectively interrogate the project's goals." Thus, the themes described in this paper are based on an iterative approach that we have called an embedded review—going back and forth between research meetings, reading and searching for literature, interactive discussions and meetings with 4D PICTURE researchers, joining trainings, and conducting observations of meetings. Hereafter, we describe how we conducted this embedded review of ethical aspects in the 4D PICTURE project.

Identification of Ethical Challenges and Member Check

First, 2 authors (LH and CG) familiarized themselves with the research objectives and activities of each work package in the 4D PICTURE project. They identified ethical themes that may be relevant for the activities and research of each work package and looked for key publications in the (empirical and theoretical ethics) literature on that particular topic. Some of these publications were known by the authors, whereas others were found through simple searches in PubMed and Google Scholar as well as through further snowball searches. Of note is that the iterative process of *identifying* did not simply consist of extracting issues from the project proposal or the ethics literature but required active interpretation by the authors to link the literature to the intended work in the various parts of the 4D PICTURE project. After having conducted an initial identification of ethical themes, these insights were summarized in such a way that the descriptions were readable and understandable by researchers without a background in ethics. This document was shared within the consortium and also included a *further reading* section for those interested in reading more. The categories of ethical themes were then further refined through conversations with researchers in the 4D PICTURE project. For instance, the points discussed in the *Results* section about the design of the MetroMap were extensively discussed with researchers at an in-person training in the Netherlands by LH and CG. Finally, 2 of the ethicists in the project (LH and MB) presented the findings of this embedded review during a general meeting that was attended by all the researchers of the 4D PICTURE consortium, which took place in 2023 and served as a *member check* to see whether relevant issues were included but also to stimulate researchers in the project to think about ethics in their work package. An overview of these methodological steps is provided in Figure 2. While our aim was explicitly not to conduct a systematic review of the literature, our methodology might still be further improved in future studies deploying this embedded review approach.

Figure 2. Method of conducting an embedded review in the 4D PICTURE (Producing Improved Cancer Outcomes Through User-Centered Research) project—flowchart of the process of identifying and refining ethical challenges based on the literature and discussions with researchers in the project.



In what follows, we discuss 13 specific ethical challenges related to the development of data-driven DSTs for cancer care that arose from our embedded review of the 4D PICTURE project. While we started from general themes (eg, ethical aspects regarding *semantic bias* and *fairness* in relation to text mining patient experiences to develop a conversation tool), we reframed the themes as challenges to provide a slightly more action-oriented overview that can serve as an agenda for embedding ethics into the project (a challenge is then "preserving meaning in the data and including underrepresented groups"). We note that there is no agreed-upon definition of what an ethical challenge is [10], but for the purpose of this paper, we defined it as follows: "an ethical challenge occurs when one does not know how to behave and act in the best way" [11]. To aid readability in our reporting of ethical challenges, in the text of this paper, we combined several 4D PICTURE work packages into more general headings (eg, combining "project management, dissemination and ethics," which are 3 separate work packages in the 4D PICTURE project, into 1 section), but a full overview of the project's work packages and related challenges is presented in Table 1.

Table 1. Ethical challenges related to the work packages (WPs) in the 4D PICTURE (Producing Improved Cancer Outcomes Through User-Centered Research) project. Descriptions of each WP's main objectives were taken from the 4D PICTURE project's website. It should be noted that challenge 13 is relevant across all WPs, although extra responsibility may be put on the project leads in WP 1.

WP	Main objectives	Ethical challenges
WP 2—modeling	The overall objective of WP 2 is to develop algorithms that provide predictions of outcomes for individual patients for each relevant treatment option for 3 major cancers. The algorithms will be translated into decision support tools and included in the MetroMap as developed in WP 4.	• Avoiding biased outcomes due to poor data quality (challenge 1) • Understanding how ethical values are built into models (challenge 2)
WP 3—text mining	The main aim of WP 3 is to use co-design methods to investigate the experiences of patients with cancer and their significant others, drawing upon novel text mining and qualitative analysis methodologies to improve outcomes for patients with cancer and their families and enhance the quality of a conversation tool to be used by citizens, patients, and clinicians across Europe.	• Preserving meaning in the data and including underrepresented groups (challenge 3) • Protecting the privacy of the data of patients with cancer in the public domain (challenge 4)
WP 4—design	The primary task of WP 4 is the redesign of care paths applying the service design method to experience a more individualized and personalized care path with the inclusion of innovative prognostic and conversational tools.	• Incorporating shared decision-making and death into the care path design (challenge 5) • Visualizing care paths responsibly through design justice (challenge 6)
WP 5—practice and WP 6—policy	In WP 5, researchers will evaluate MetroMapping in its entirety using mixed methods designs as well as evaluate the decision support tools to be developed. WP 6 aims to finalize the MetroMapping methodology, assess the generalizability of the methodology and decision support tools of MetroMapping, and provide guidance to policy makers about MetroMapping.	• Reflecting on good criteria for (cost-effectiveness) evaluation (challenge 7) • Balancing the adoption of technology with other values (challenge 8) • Anticipating techno-moral change and developing new ethical frameworks (challenge 9)
WP 1—coordination and WP 8—dissemination	WP 1 entails all aspects of the coordination, management, and progress monitoring of the project. The task of WP 8 is to disseminate project information and results of the research and innovation activities to key stakeholders through various channels and enable access to decision support tools to patients with cancer, their significant others, clinicians, and the public.	• Integrating patients' experiential knowledge to avoid epistemic injustice (challenge 10) • Negotiating shared knowledge in the trading zone between disciplines (challenge 11)
WP 7—ethics	WP 7 is a cross-cutting, integrative WP that establishes an embedded ethics approach within the 4D PICTURE project. It collaborates with all WPs and aims to ensure that ethical and social aspects of the planned tools and implications of their use are considered from the very start of their interdisciplinary development.	• Establishing the position of the embedded ethicist (challenge 12)
All	—[a]	• Ensuring psychological safety in research teams (challenge 13)

[a]Not applicable.

Results

Overview

Through our embedded review of ethics in the 4D PICTURE project, we identified 13 key ethical challenges related to the development of DSTs and the redesign of care paths for more personalized cancer care. These challenges are discussed in sections related to the various parts of this project: the prognostic model development; text mining of patient experiences to develop a conversation tool; the innovative design method of *MetroMapping* to develop a personal care path navigator; the evaluation of the MetroMap and integrated DSTs; and project management, dissemination, and ethics. An overview of how the identified ethical challenges relate to the 8 specific (but interrelated) work packages in the 4D PICTURE project, along with those work packages' main objectives, is provided in Table 1.

Prognostic Models for Individual Treatment Outcomes

Overview

One of the aims of the 4D PICTURE project is to develop models that predict individual patient outcomes for each relevant treatment option for breast cancer, melanoma, and prostate cancer. These prognostic algorithms will be translated into DSTs. Outcome measures will not only be related to survival but also operationalize different aspects of quality of life (eg, side effects or sexual well-being) per treatment option so that patients can make an informed treatment decision based on their personal circumstances and values. However, ethical aspects should be taken into account when developing such prognostic models. Hereafter, we discuss 2 key ethical challenges: the risk of data bias and the need for awareness that societal values are always built into models.

Ethical Challenge 1: Avoiding Biased Outcomes Due to Poor Data Quality

The accuracy of prognostic models depends heavily on the quality and representativeness of the data used. Data bias (ie, results being skewed because of unjust inaccuracies in the data used for modeling) can cause harm to individuals and increase existing inequities in society [12-14]. Biased outcomes can arise due to false assumptions incorporated into data collection, inconsistent definitions, small sample sizes, reproduction of societal trends influencing the data, and the underrepresentation of (minority) groups in datasets [15-17]. In the 4D PICTURE project, bias may be introduced as the input data for the prognostic models come from multiple European countries and from international randomized clinical trials and meta-analyses and, as such, may not be generalizable to, for instance, small patient groups or smaller European countries that lack representation in these datasets (let alone to countries in the Global South, although that is also not the aim of the 4D PICTURE project). For example, data may be predominantly derived from patients with European heritage, and ethnic minority groups may be underrepresented in the data, possibly leading the model to draw misguided conclusions about these patients [17-20]. In addition, models trained on a dataset in one setting often do not perform well in other settings [21]. To mitigate the impact of data bias, the 4D PICTURE researchers will weigh the quality and generalizability of the different data sources.

However, it may turn out that, for specific patient populations, the level of evidence is so low that a prediction tool may cause more confusion than clarity about treatment options for the patient [22,23]. If the evidence is relatively uncertain, does that mean that the clinician can refrain from presenting the patient with the model's outcome to prevent confusion? We find that clinicians should evaluate the usefulness of the model's output for each patient before consultation. However, open questions remain. At what level of certainty is the clinician obliged to share results with the patient to fulfill duties of openness and transparency? In addition, should the use of a prediction model, if available, be incorporated into the standard workflow or rather as an optional step that requires the patient's informed consent before inputting data? Or would providing patients with information and choice lead to an undesirable redistribution of responsibility in which the clinician shifts the burden of making difficult decisions to the patient [24]? These issues call for ethical reflection and careful consideration of how clinical practice changes when prognostic models are introduced in a specific cancer care setting.

Ethical Challenge 2: Understanding How Ethical Values Are Built Into Models

While prognostic models are sometimes thought of as objective calculators, in reality, no algorithm is perfect. Models are not value free and will always have certain undesirable outcomes even when they merely output a prediction without coupling it to advice [25]. Namely, if two developers create a model based on the same database, the resulting models will be different because certain choices about the functioning of the model are made by the developer (eg, whether to accept more false negatives vs more false positives). Clinical risk prediction models are often programmed to prioritize *sensitivity* (fewer false negatives) over *specificity* (fewer false positives) because this reflects the existing tendency of human clinicians to better be safe than sorry (ie, to prefer the risk of overtreatment to the risk of undertreatment) [15]. In the same way, the selection of outcome measures is loaded with values. Overall survival, recurrence-free survival, and progression-free survival are commonly accepted outcome measures. However, not every outcome may be equally important to each patient.

Thus, prognostic models may be perceived within clinical practice as more neutral, objective, certain, and reliable than they actually are. This is akin to what some scholars have called the *automation bias* of humans regarding automated systems—people tend to rely too heavily on automated systems such as AI technologies and forget the cultural and ideological choices that were made during the development of that system [26,27]. It is important for developers and users to understand that these choices often reflect existing societal bias and inequalities. Moreover, developers of prognostic models may be faced with difficult moral dilemmas in which no right solution can be readily modeled [28]. If there is then a lack of ethics guidance, this can lead to arbitrary decision-making based on technical features such as computing power [29]. Even if

developers themselves are very much aware of the limitations, uncertainties, value preferences, and subjective cutoff points of a model, these may not be immediately clear to the clinicians, patients, and policy makers who use the information generated by the model. Prognostic models are value laden and bring about ethical questions, but these are sometimes not recognized as such. Acceptability of certain choices during model development should be based on ethical reasoning and arguments, preferably in consultation with ethicists and a diverse range of stakeholders. The 4D PICTURE project offers a unique possibility to jointly study and discuss the ethical values built into the planned prediction models and their potential influence on actual cancer care practices.

This also includes discussion about *which variables* to include in predictive algorithms. For model developers, the accuracy of predictions has always been most important. However, depending on the intended use, a more accurate model may sometimes reproduce or even increase unfair inequalities in society. Imagine a model that calculates the expected quality of life after the treatment of a stroke. We know that, after a stroke, the average quality of life is much lower in neighborhoods where many people are of a lower socioeconomic status than in neighborhoods where most people are of a high socioeconomic status [30]. Including the postal code in the model might result in more accurate predictions of posttreatment quality of life. The question is then whether it is ethically justified to include postal code as a variable. The answer to this depends not only on the question of whether the variable makes the model more accurate but also on the intended use of the model. The reasoning is different for a model intended to allocate resources in a way that improves care in disadvantaged neighborhoods versus a model intended to decide who to give stroke treatment and for whom stroke treatment is not cost-effective because of a low predicted quality of life after the treatment. The latter use is problematic as it further unfairly disadvantages the already disadvantaged given that the low quality of life likely relates to factors outside individuals' control, such as housing or health literacy. Whether to include variables that might affect treatment options for disadvantaged or protected groups, such as postal code but also gender, ethnicity, disability, BMI, and smoking or diabetes status, is a question that is relevant in the context of cancer care and in need of more interdisciplinary research and ethical reasoning.

Text Mining of Experiences of Patients With Cancer to Develop a Conversation Tool

Overview

Another objective of the 4D PICTURE project is to conduct text mining analyses of *big* data on the experiences of patients with cancer to develop a conversation tool and obtain input for care path redesign. An interdisciplinary approach will be used that combines the strengths of AI tools (ie, text mining and natural language processing techniques), corpus linguistics, and qualitative (narrative) research to efficiently convert the stories of people with cancer and their significant others into usable knowledge. Key ethical challenges revolve around the risk of societal bias and loss of meaning in the data, as well as privacy and ownership questions regarding data scraped from online platforms.

Ethical Challenge 3: Preserving Meaning in the Data and Including Underrepresented Groups

In developing an algorithm for text mining of public forums, there is a risk of reproducing existing biases in society and even exacerbating them. Research has shown that applying machine learning to ordinary human language results in humanlike semantic biases. Namely, *text corpora* or language datasets contain imprints of our societal biases toward gender or race [31]. A well-known example is Google Translate's translation of job descriptions in gender-neutral languages that do not differentiate between *he* and *she* into English—until recently, this produced only biased sentences such as "she takes care of children" and "he is a lawyer." Moreover, subtle differences in meaning might be lost when transforming written experience into classifiable input for a conversation tool. If the eventual 4D conversation tool does not perform as well for each patient due to engrained societal biases or the loss of meaning in language processing, this can have a substantial impact in terms of fairness. For instance, if the tool works suboptimally for a certain minority group, this would not only be unfair but might also serve as a *microaggression*. A simple example of a microaggression is when an automatic soap dispenser cannot identify dark skin, which serves as a small (and unnecessary) reminder for people that their skin is not the *default* skin. Added up, these very small and unexpected daily encounters can truly affect a person's sense of belonging in a society [32]. Thus, it is important that metaphors used for capturing the experience of patients with cancer in the 4D PICTURE project are accurate and reflect not only the metaphors of dominant cultures but also the cultural languages of different minority groups.

Ethical Challenge 4: Protecting the Privacy of the Data of Patients With Cancer in the Public Domain

Data mining of public posts on web forums is not without ethical issues. A recent review [33] mentioned the following aspects: the privacy policies of public forums sometimes lack transparency [34]; users are not always aware of privacy settings or lack the digital skills to manage them according to their preferences [35]; even if users are aware that their posts are public, this does not mean that they agree with their posts being reused for just any purpose [36]; commercial use in particular (which is not part of the 4D PICTURE project) is deemed inappropriate by users [37-39], whereas reuse for the greater good is more accepted; and some users are even willing to put privacy concerns aside for public benefit [40]. There are 2 general issues that stand out. First, the use of social media posts for research brings about privacy risks and, in particular, the potential for reidentification. There is a risk of reidentification both on an individual and group level (eg, identifying a minority group). Patients fear that reidentification may lead to identity theft; could have consequences for employment and pension eligibility; and could lead to increased insurance costs and the use of their data for financial gain, social discomfort, or stigmatization in clinical settings or the community [33,41]. Although there is clear consensus about ensuring anonymity regarding research that scrapes data from web forums, this is

not always possible in practice. This first concern is only indirectly related to 4D PICTURE project, which will not use the data beyond its primary research aims, let alone to reidentify individual patients, but it does highlight the need for good data protection and security measures to protect the sets of data *scraped* from public forums.

Second, it is debated in the literature whether posts on public forums and social media should be considered as being part of the public or private domain given that notions of public and private are changing [40]. Legally, these posts are in the public domain, and no informed consent is needed regarding reuse, so many researchers do not ask for consent for scraping data from online forums [42-44]. However, citizens may have other intuitions about this. The distinction between *private* and *public* is probably too broad to reflect their moral intuitions about data ownership. The discrepancy between regulations and the views of citizens could lead to public outrage and less trust in science in general. Some ethicists have also argued that data may only be collected from social media platforms after explicit consent from the data subject [45]. It may be helpful in this setting to view privacy in terms of *contextual integrity*, a concept by Nissenbaum [46] based on the spheres of justice by Walzer [47]—simply put, contextual integrity says that privacy means something different in an airport security area than in a kindergarten. This is recognized by patients, who regard health data research as part of the (highly regulated) medical sphere and want their data to stay within that sphere [48], but what happens to contextual integrity when patients post their own health information on social media platforms? In the 4D PICTURE project, researchers will further explore these questions together with patient representatives to develop ethical guidance regarding the use of social media data for the project.

MetroMapping to Develop a Personal Care Path Navigator

Overview

In addition to the development of prognostic and conversation tools, the 4D PICTURE project aims to visualize the care trajectory for melanoma, breast cancer, and prostate cancer in 3 countries (the Netherlands, Spain, and Denmark). The new methodology of *MetroMapping*, which involves collaborative meetings with health care workers, is used to create a comprehensive visualization of the care trajectory for each specific cancer, representing various treatment options as metro lines [49]. The resulting MetroMap consists of different layers in addition to the metro line itself, which represents the overall structure of the care trajectory; other layers incorporate patient experiences at various points in this trajectory (eg, highlighting aspects such as magnetic resonance imaging being perceived as frightening) or provide information about the environment (eg, parking options or quiet routes within the hospital to accommodate heightened sensitivity to noise and smell after chemotherapy). The function of the MetroMap is to locate potential for improvements and *redesign* the care paths of people with cancer where needed. In the 4D PICTURE project, the prognostic models and conversation tools will be integrated into the MetroMap to also develop a personal care path navigator for each patient that serves as a decision aid in shared decision-making (SDM). There are 2 main ethical topics relevant here: first, SDM and, second, design justice and the influence of choice architecture.

Ethical Challenge 5: Incorporating SDM and Death Into the Care Path Design

The concept of SDM is increasingly recognized as an important aspect of personalized care, and it lies at the core of the 4D PICTURE project. The goal of SDM is to engage in conversations about relevant treatment options and the patient's values and preferences to arrive at a shared treatment decision aligned with their wishes [50]. In the 4D PICTURE project, the MetroMap design includes designated points labeled as *SDM moments* (indicated by stars). These moments signify important moments (eg, when test results are available) to allow for discussions between health care providers and patients regarding preferred treatment options. However, within actual care practice, SDM is complex and scattered along the care path, extending beyond a specific SDM moment occurring at a specific time point. Patients often struggle to express their values or preferences, necessitating support such as probing questions, multiple conversations, or sources of inspiration to determine what matters most to them in a given situation [51]. Some authors have argued that SDM runs the risk of putting too much responsibility on the patients and leaving patients "abandoned to their autonomy," as O'Neill [52] has said. However, the everyday reality of SDM in actual care practices shows that autonomy has a relational nature—there is an interdependence among patients, their support networks, and their health care providers [53,54]. Thus, when performed well, SDM seems compatible with a notion of autonomy that does not put too much responsibility on patients. However, of note is that patients differ in their preferences for how care is delivered (eg, in the number and duration of consultations in which treatment decisions need to be made [55]), so it is important to keep in mind that one patient may not represent the entire patient population in their preferences for SDM. Moreover, and this is a more general point, social determinants of health—such as economic status, social vulnerability, and access to resources—shape patients' capacity to engage in SDM, and health care systems should address these broader determinants to better support diverse patient needs related to SDM [56].

Another challenging aspect of the personal care path is the death of the patient. Death and mortality raise profound ethical questions, balancing values such as human dignity, patient autonomy, relieving pain, quality of life, and balancing the interests of an individual person within their network. As death carries cultural and social significance and ethical considerations vary across systems and belief systems, it is a notoriously sensitive topic in health care. Both health care workers and the systems of health care are mainly geared toward curing diseases and prolonging life—until relatively recently, clinicians were not taught to discuss end-of-life issues [57]. During the last decades, palliative care and advance care planning have been developed and professionalized [58], but how to incorporate end of life into the cancer care MetroMap is still an open question. Is there an end point? How should death be visualized? These are difficult questions and answers that may also differ by patient, country, and culture [59,60]. Studies show that

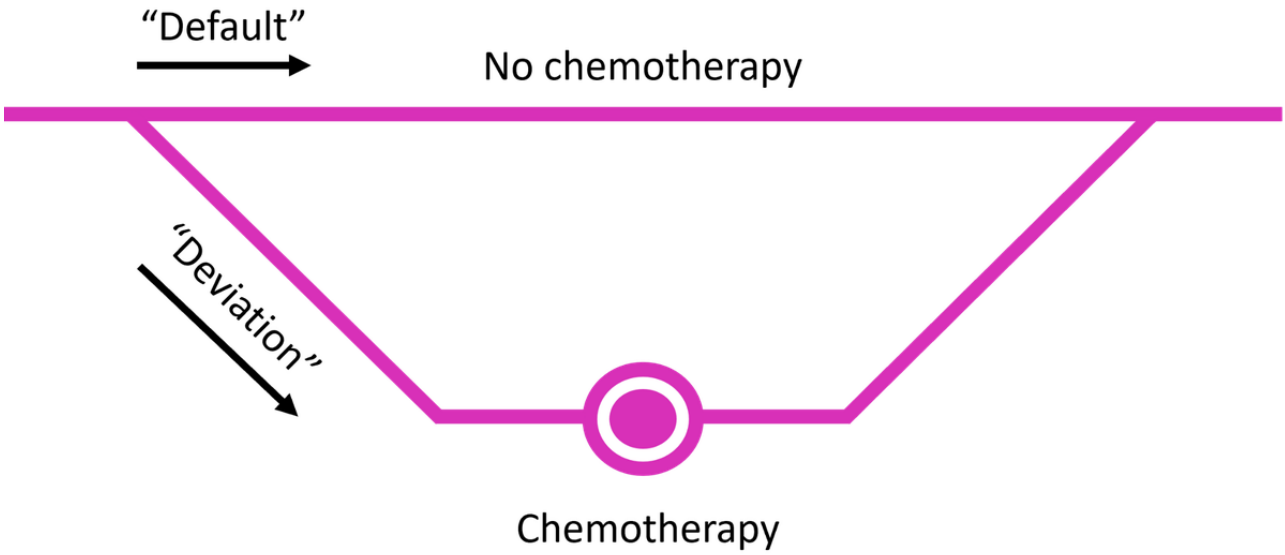
prognostic and conversation tools may be helpful for discussing end-of-life care [61], but their development should be accompanied by reflection on the experience of the patient.

Ethical Challenge 6: Visualizing Care Paths Responsibly Through Design Justice

In the 4D PICTURE project, service design methods will be developed and used by designers to help create the MetroMap. This methodology may give rise to ethical questions, for instance, because the design researchers will work with people with different levels of power (medical team vs patient and family) and need to be aware of sensitivities and vulnerabilities and also because there is a risk of bias on the part of the researcher who leads this process [62]. Moreover, we noted previously that technologies themselves (eg, algorithms) are not neutral but are packed with values [63]. The same is true for the cancer care path design as the way in which it is set up spatially may affect how information is perceived. The lens of *design justice* draws attention to this and shows how design enables or encourages certain actions while excluding or discouraging others. These design aspects are known as *affordances*—properties of an object that suggest possible actions that users can take (eg, a button affords pushing). In the context of the 4D PICTURE MetroMap, questions arise regarding its affordances and disaffordances. For example, if the map relies heavily on color, it may not be accessible to clinicians with color blindness. Identifying such issues within the design process is crucial to ensure that the MetroMap is inclusive. Thus, design justice brings awareness to often unconscious design decisions and seeks to rectify historical and systemic injustices perpetuated by design decisions (eg, a bridge that is too low for a public bus to pass under, allowing only car owners to take a certain road [64]). Design justice challenges power dynamics within design, advocating for the redistribution of resources and opportunities to address social, economic, and environmental disparities [63].

An important part of design justice in the 4D PICTURE project is the impact of what has been dubbed *choice architecture* (ie, the number of options presented simultaneously, or their order, influence which information is best retained by viewers and which choice is finally made [64-68]). To illustrate, it makes a difference whether treatment options are presented as “no treatment vs. treatment” or “no treatment vs. treatment option 1 vs. treatment option 2” because people divide their attention equally in considering all options [66]. The visualization of *default* and *deviated* decision paths (Figure 3) can also have a large effect on which choice is made [69]. There is little evidence from health care on this topic, but cartography research shows that the linearity of map routes matters in route choices made by travelers [70]. The effects of choice architecture are so strong that it works even if decision makers are aware of the mechanism [71]. One paper about default options in oncology concludes that further experimental studies are needed to select which default options successfully change behavior [72], and to this we add that such investigation should include consideration of design justice.

Figure 3. Choice architecture—visualization of the default and deviated decision paths.



Some ethics scholars suggest that choices should be presented in the least directive way, but others have argued for using conscious choice architecture to stimulate unpopular but beneficial choices (ie, nudging) [73]. For example, some types of localized prostate cancer generally present as slow-growing tumors, and active surveillance of progression while withholding or postponing treatment is a beneficial option in this case as treatment has a significant chance of producing complications [65]. However, medicine is biased toward acting rather than waiting, leading some clinicians to present patients more strongly with the different treatment options than with the possibility of active surveillance. In this example, a MetroMap might need to be designed as a choice between active surveillance and treatment first and then as a choice between treatments. MetroMap designers should deliberate regarding which effect they wish the MetroMap to have on decisions and how this is best achieved and underpin the acceptability of this choice with medical-ethical arguments. Although exerting as little influence as possible is preferable in most cases, there are

situations in which some nudging by using ordering effects or default options can be argued to be justified.

Evaluation and Development of Guidance for Policy Makers

Overview

The 4D PICTURE project also contains an evaluation of the MetroMapping design methodology, the MetroMap itself, and the prognostic model and conversation tool, as well as a cost-effectiveness evaluation and an assessment of the generalizability of the methods and developed tools. Moreover, the project will develop guidance for policy makers about MetroMapping using a framework that helps understand the nonadoption; abandonment; and challenges to the scale-up, spread, and sustainability (NASSS) of health technologies by professionals or patients [74]. With regard to the evaluation and policy aspects of the 4D PICTURE project, the following ethical topics are relevant: reflection on the criteria for evaluation, balancing ethics with technology adoption, and techno-moral change.

Ethical Challenge 7: Reflecting on Good Criteria for (Cost-Effectiveness) Evaluation

In the 4D PICTURE project, questionnaires among patients, significant others, and clinicians will be used to compare between the original and redesigned care paths and evaluate the developed tools. As is good practice within quantitative evaluation studies and expected by funding agencies, the criteria based on which the redesigned care paths will be evaluated were defined in advance. Researchers should consider, together with patient representatives, which evaluation method and criteria fit their project best. The same is true for cost-effectiveness analyses. The ultimate goal of such analyses is to help health care decision makers choose between competing alternatives based on some predetermined measure of economic efficiency, such as cost per life saved, cost per year of life saved, or positive net benefits. Just like we saw with prognostic models, an important ethical aspect of cost-effectiveness analyses is that they "expose...hidden assumptions, and require explicit judgements to be made about which ethical position is appropriate in a particular policy context" [75]. Should cost-effectiveness analyses be used at all, or is it inherently unjust to compare costs between persons [76]? Such ethical questions and specific dilemmas regarding cost-effectiveness have been discussed extensively in the medical ethics literature, but further work is still needed to see how existing arguments apply to the novel setting of data-driven and personalized cancer care.

Ethical Challenge 8: Balancing the Adoption of Technology With Other Values

Research shows that many innovations are eventually not adopted [77]. In the 4D PICTURE project, the NASSS framework is used to study how the developed tools can be designed to promote adoption and avoid wasting resources. However, strategies that increase the adoption potential of health innovations can simultaneously pose ethical dilemmas. An example is that of a research laboratory that developed care robots, where the adoption by health care workers was a major issue (eg, some nurses put the robots in a closet because they found them annoying) [78]. The researchers discovered that building gender stereotypes into the robots contributed to their acceptability and adoptability by health care workers. For instance, if they developed a robot that was interpreted as female, the tone of voice was expected to be much more modest and less authoritative than that of robots that were interpreted as male. This poses an ethical dilemma: what is the right thing to do [79]—contributing to acceptability by repeating gender stereotypes into the design or countering gender stereotypes? Ethical reflection with stakeholders is needed on how to balance adoption with other values. This can take the form of structured methods such as moral case deliberation [80] as well as structural collaboration between researchers and groups of patients and the publics, which is discussed under ethical challenge 10.

Ethical Challenge 9: Anticipating Techno-Moral Change and Developing New Ethical Frameworks

Innovations are almost never used in practice as intended, and they often produce unintended, unforeseen, and sometimes even counterproductive consequences—in other words, "things bite back" [81]. A simple example is how, when more highways are built to reduce traffic jams, more people will take their cars, so no reduction in traffic jams is realized. Often, new technologies or other innovations influence our concepts and conceptions of a good life [82,83]. Think of the introduction of the contraceptive pill that contributed to women's emancipation or of blood sugar measurement devices that influence the way in which patients with diabetes relate to their bodies [84]. This process of technology and ethics codeveloping has been referred to as *techno-moral* change [83]. Techno-moral change is notoriously hard to research or predict and may complicate the evaluations and policy development planned in the 4D PICTURE project and similar health technology projects. In addition, the 4D PICTURE project itself may contribute to techno-moral change through the development of prognostic models and the conversation tool and the redesign of care paths—these may influence patients' and clinicians' moral routines in unexpected ways or create new moral dilemmas. Possible techno-moral changes should be anticipated (eg, through qualitative interviews about the expectations of stakeholders or through ethnographic observation studies) during the research phase instead so that they may be acted upon in time.

A timely example of techno-moral change is the incorporation of sustainability into evaluation frameworks and to take the costs, harms, and burdens of health care technology with respect to the environment into account. Health care has always been a system with a significant carbon footprint due to the many single-use products, and now data are also becoming an important factor [85]. The increasing data storage and analysis possibilities bring about new moral questions. For instance, is it proportionate to slightly reduce a certain health risk using a method that puts a considerably larger burden on the environment? However, we currently lack ethical vocabulary, frameworks, and research on this topic. The connection and trade-offs between health care and planetary health need to be further studied, and this important aspect will be included in

the 4D PICTURE project despite not being in the original plan. For instance, to prime researchers about the topic, the authors of this paper organized an interactive session about environmental sustainability during the latest project consortium meeting and asked each work package to present in the next consortium meeting how they will address sustainability in their work.

Project Management, Dissemination, and Ethics

Overview

Similar to most large research projects, the 4D PICTURE project has several *work packages* focused on aspects bordering the science at the project's core—the coordination, management, and progress monitoring of the project, as well as the dissemination of findings and the embedding of ethics throughout the project. Although it is less obvious, even these work packages can give rise to ethical issues. We identified the following issues that are relevant for the 4D PICTURE project and other research projects on data-driven care: epistemic injustice, the *trading zone*, the position of the embedded ethicist, and psychological safety.

Ethical Challenge 10: Integrating Patients' Experiential Knowledge to Avoid Epistemic Injustice

The 4D PICTURE project integrates experiential knowledge from patient and public involvement (PPI) boards in all participating countries aiming to align care paths with patients' needs. However, bridging experiential and academic knowledge presents challenges. How to translate the information shared by the PPI boards into usable knowledge for the research group? What to do with contradictions or differences of opinion [86,87]? A helpful concept is that of epistemic injustice or, in other words, knowledge-related injustice [88]. This encompasses testimonial and hermeneutical injustices [89]. Testimonial injustice arises when dismissing patient experiences as unreliable, emotional, or irrelevant, potentially leading to a loss of confidence that causes patients to stay silent [90]. Hermeneutical injustice arises when patients lack the language or concepts to articulate their experiences [90]. This occurs because these concepts have not been developed yet, because patients do not have access to them, or because the concepts are not recognized by clinicians as the dominant group. In the 4D PICTURE project, the conversation tool is intended to allow patients to express their experiences and bridge this hermeneutical gap between patients and professionals. Awareness of the concept of epistemic injustice may help the researchers notice instances in which it may play a role and search for strategies to minimize epistemic injustice during the research process. Strategies include conveying patient contributions through stories or alternative mediums such as visual art, films, or metaphors [90,91]. Researchers must also undertake "role, emotion, and relationship work," which involves switching between different roles (eg, researcher, facilitator, advocate, relation manager, or coffee maker), handling loaded emotions with care and empathy (ie, rather than distancing oneself from the subject), and fostering relationships [92]. Awareness of these strategies helps minimize epistemic injustice, ensuring that patients' voices are valued and respected in the research process. A final note is that power dynamics leading to epistemic injustice are influenced by existing structural inequalities regarding race or socioeconomic class, for instance. In the 4D PICTURE project, the ethicists and other researchers will work with the PPI boards on appropriate ways (eg, payment of the PPI boards) to alleviate the harmful effects of such structural concerns to help ensure that all voices are heard equally.

Ethical Challenge 11: Negotiating Shared Knowledge in "the Trading Zone" Between Disciplines

Often, professionals from different disciplines use different concepts; ascribe different significance to objects or phenomena; use different conceptual frameworks; and may also have different value systems, accountability rules, and quality indicators. The same is true for clinicians and patients in the cancer care trajectory—clinicians' choices and priorities may conflict with the perspectives of patients (eg, streamlining and speeding up the care path vs preferring more time to deliberate treatment options between multiple consultations [55]). Understanding how to guide collaborations between different scientific disciplines and with patient representatives in the 4D PICTURE project can be based on the metaphor of a *trading zone*, which is often used to study multidisciplinary scientific collaborations [93]. In the *trading zone*, a shared understanding between the different disciplines and different types of (experiential) knowledge should be negotiated. This *trading* can be facilitated by an agent who is familiar with >1 discipline. Sometimes, this could be a nurse; in other cases, it can be an embedded ethicist. To do this effectively and fairly, it is important to have awareness of power differences and implicit value frameworks. Are all relevant voices heard and valued equally [94]? Who has a say in the structure and processes of collaboration? Whose knowledge counts (see the aforementioned concept of *epistemic injustice*)?

To avoid one perspective overshadowing another, collaborators should think about how to handle differences of opinion in meetings and how to ensure that all participants make sufficiently equal contributions. Strategies to divide power equally in the trading zone can consist of making the differences in vocabulary, systems of recognition, and value systems explicit; making use of *boundary objects* that facilitate exchange between multiple worlds [95]; develop meeting structures that stimulate an explicit deliberation; and decide in advance which agents in a project are best placed to facilitate collaboration in the trading zone (this is not always the project lead). Of note is that the trading zone metaphor is relevant not only during the conduct of research but also in the dissemination phase. Dissemination is in itself an ethical imperative, especially when the research was publicly funded, but also comes with ethical sensitivities [96]. Collaborative efforts between different types of partners (eg, clinician researchers, developers, and patient representatives) can help facilitate the *trading* or sharing of knowledge in a way that is valuable for reaching different stakeholders. As such, dissemination can help combat epistemic injustice as well as *demystify* scientific concepts and hypes such as those surrounding AI [97]. Dissemination should always be planned so that the outputs are contextualized and sensitive to the experiences of the group under study (in our case, patients with cancer) [98].

Ethical Challenge 12: Establishing the Position of the Embedded Ethicist

Ethicists are embedded in the 4D PICTURE project to conduct this literature review and several empirical studies about ethical issues, as well as to provide support with ad hoc ethical issues by joining scientific meetings and providing internal trainings. In bioethics, there have been long-standing debates about how empirical data relate to normative reasoning and about the role of the ethicist in health research [99]. Should an ethicist be a critical, distant outsider or be part of the research practice? We find that a more *embedded* role is called for right now. Take the example of AI—while several organizations have developed general principles for AI in health care, the ethical difficulties lie in applying these principles and making trade-offs in actual research practices. Embedded ethics can fill this gap between intentions and actions by engaging in practical and relational work on ethics within a specific research project or development process [7]. However, there are also some risks and disadvantages to embedded ethics. Namely, as the ethicist's normative analysis is so close to research practice, there is a risk that more fundamental ethical questions will not be discussed and researched. An example is the growing influence of *big tech* on our health care systems—as research ethics committees have only focused on topics such as privacy in individual research projects and specific tools to be developed, they lack an ethical framework to meaningfully weigh the broader collaboration with industry. Ethicists can sometimes secure power instead of challenging it because "they locate the source of the problem in individuals or technical systems instead of acknowledging structural power differences and working structurally towards dismantling them" [100]. In certain cases, there is a risk of this leading to *ethics washing* or lip service to industry [7]. Thus, in large interdisciplinary health research projects, one should not forget to consider "the ethics of the ethics work package." Moreover, we find that ethics should be recognized as a shared responsibility—the embedded ethicists can *sensitize* other consortium members to the ethical aspects but cannot be solely responsible for the normative assessments of the research and the tools being developed.

Ethical Challenge 13: Ensuring Psychological Safety in Research Teams

Finally, an underappreciated ethical aspect of research is psychological safety, which is defined as "the belief that one will not be punished or humiliated for speaking up with ideas, questions, concerns, or mistakes, and the team is safe for interpersonal risk taking" [101]. Safe environments are those where it does not feel risky to express one's thoughts, doubts, questions, and errors. As such, psychological safety is a key condition for high-performance research teams as this safe environment promotes takings risks as well as reporting mistakes and learning from them. Psychological safety also influences the degree to which people speak up about (research) misconduct and promotes open discussions about *grey areas* or so-called *questionable research practices* such as cutting corners due to time pressure [102]. In the same vein, it is a precondition for opening up about moral dilemmas in research and development. Strategies for promoting psychological safety include encouraging vulnerability, active listening, appreciating diverse perspectives, promoting a culture of feedback and learning, establishing clear expectations, celebrating experimentation, treating mistakes as learning opportunities, providing tools and training for effective conflict resolution, using metrics to regularly assess psychological safety levels, and leadership actively endorsing psychological safety [101]. Of course, in large projects, the level of psychological safety differs by collaborating partner, but it can still be influenced by the project and work package leads.

Discussion

In this paper, we have described relevant ethical challenges that should be considered when developing data-driven DSTs for more personalized cancer care. We based our review on the European 4D PICTURE research project, and as such, this paper is the first review grounded explicitly in the *Embedded Ethics* approach [7]. Using a collaborative and iterative methodology helped provide a broad overview of the ethics of this heterogeneous and interdisciplinary project. This overview serves as starting point for developing actionable guidance for the project and potentially beyond, as well as for further empirical and theoretical ethics research and future collaborations with developers, clinicians, researchers, and the PPI boards in the 4D PICTURE project. As such, we add to discussions in previous, more systematic literature reviews about data-driven DSTs in health care, which have mostly highlighted general, high-level ethical principles of respect for patient autonomy, prevention of harm, fairness, explicability, and privacy [103], as well as professional autonomy, bias and justice, and explicability [104]. Another study applied an ethical framework on AI-guided clinical decision support and used case examples to illustrate key issues of accountability and transparency, the potential for group harm, efficiency of health care, and conflicts between roles and responsibilities [105]. Such examples are insightful but provide less detail than a review of a complete research project. The iterative process of extracting themes by moving between key papers and discussions with the consortium helped us describe more specific examples of the universal themes in the literature and uncover additional challenges unrelated to the technical and medical core of the project. As such, our embedded review highlights ethical aspects previously neglected in the digital health ethics literature and zooms in on real-world challenges in an ongoing project.

Several identified ethical challenges (challenges 1-4) were related to the data and algorithms needed to develop data-driven DSTs in oncology. Prominent issues were indeed bias and privacy, which have been extensively described in previous literature reviews, yet our analysis gave rise to more specific, real-world questions for further research and advice (eg, "what are the arguments for including or excluding postal code in a prognostic model for cancer treatment outcome?" or "should posts on online patient fora be considered part of the public domain or is consent needed to use them for the development of conversation tools in oncology?"). Other topics (challenges 5 and 6) revolved around the specific design methodology used in the 4D PICTURE project. Questions arose about *design justice* that have not been previously addressed in the context

of DSTs for health care. Furthermore, we described various ethical aspects related to broader policy issues in health care (challenges 7-9). Lysaght et al [105] have previously mentioned the cost-effectiveness of data-driven DSTs and financial sustainability of health care systems as important ethical issues, and we have added that environmental sustainability is also a timely ethical consideration—currently, there is a lack of knowledge on how to apply sustainability in practice and how to balance it with competing values.

Finally, various challenges that we described in this paper do not relate directly to the DST as the outcome of the 4D PICTURE project but rather to the research process itself (eg, to the project management or results dissemination). We highlighted the importance of epistemic justice in the collaboration with patients and between different disciplines, questioned the role of the embedded ethicist who may not be fully independent themselves, and called for attention to psychological safety (challenges 10-13). The latter topic is increasingly discussed in various sectors of work, including health care, but is generally not an agenda point for international consortia even though project leaders seem to have an important role model function. Therefore, a key outcome of this embedded ethics review is that researchers and leaders in medical data-driven research projects should not forget the ethical challenges of the work that surrounds the scientific core of the project.

Our methodology is novel but not without limitations. The literature search was not systematic, so the results are not exhaustive, and the identification of ethical challenges was to a certain extent subjective and colored by the preferences and knowledge of the ethicists that conducted it. We addressed this by conducting a *member check* with 4D PICTURE researchers, but we propose that, in future projects, a check by an external ethicist could serve as additional validation. In addition, the definition of *ethical challenges* could be developed together with participants (in our case, the researchers in the 4D PICTURE project) to avoid variation and achieve clarity in the analysis and interpretation [10]. Moreover, new issues may come up after the publication of this paper as the project is still evolving. While these will be addressed in annual documents shared within the consortium (a sort of *rolling review*), this paper is not as dynamic or adaptable. Further thought is needed on how to better match the publication of ethics research with the developments in medical and data science and the possibility for techno-moral change (eg, should it be possible to update papers after publication in an academic journal to ensure that they remain up-to-date and relevant?). Finally, a general limitation of embedded ethics research is inherently tied to its main strength of providing actionable recommendations within a specific project—namely, ethicists should ensure not to lose sight of the broader structural and systemic issues underlying the particular questions discussed within the scope of a project. For instance, the focus on the 4D PICTURE project limited our discussions on data bias to *the* European context, if that even exists, whereas a huge concern is *data poverty* in low- and middle-income countries that may have fewer means to digitalize health care systems [106]. Another example relates to the more structural socioeconomic inequalities that influence digital inclusion efforts, which cannot be *solved* within applied research projects focused on developing digital tools [107]. We need to consider, in future work, how embedded ethicists can put such structural issues on the agenda when their work is tied to specific ethics work packages in highly delineated projects.

In conclusion, this viewpoint shows that a lot may be gained when ethics is embedded into a project from the start. Analysis of existing literature was deepened, and findings were made more actionable through the iterative collaboration between ethicists and other researchers in a large research consortium. For instance, the work on this embedded ethics review led to further collaboration with the design researchers in the 4D PICTURE project to pre-emptively reflect on ethical challenges that may arise during their fieldwork of *shadowing* patients to analyze oncology service design [62], thus laying the groundwork for responsible visualization and redesign of care paths (challenge 6) while ensuring that patients and their voices are respected in the research process (challenge 10). We suggest that ethicists working in interdisciplinary projects should not automatically opt for systematic or scoping reviews of a single question but rather consider whether a more applied *Embedded Ethics* review strategy might sometimes better fit their practical and theoretical aims. Ethics often formulates abstract, high-level principles, and in this paper, we have shown how these can be operationalized in practice, in particular when designing and developing data-driven support tools for improving cancer care. Taking on board the identified challenges and recommendations will help ensure that data-driven innovations in oncology are developed in an appropriate and patient-centered manner.

Acknowledgments

The authors thank all researchers in the 4D PICTURE (Producing Improved Cancer Outcomes Through User-Centered Research; see Multimedia Appendix 1 for a list of the members' names and affiliations) project for sharing their insights on the ethical challenges discussed in this paper. The research leading to these results was funded by the European Union under the Horizon Europe work program (101057332). However, the views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The UK team of the 4D PICTURE project is funded under the Innovate UK Horizon Europe guarantee program (UK Research and Innovation reference 10041120). The authors thank Panton for providing images of MetroMapping for the 4D PICTURE project.

Authors' Contributions

LH and MS conceptualized this paper. LH and CG collected the literature and conducted the analysis. LH wrote the report that served as the basis for this paper. LH and MB conducted the *member check* among the consortium members. MB wrote the original draft of the paper and led the revision process. MS, IJK, and AB acquired the funding for this study, were involved in the analysis, and reviewed and substantially edited the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
4D PICTURE Collaborator Authorship.
[DOCX File , 15 KB - cancer_v11i1e65566_app1.docx]

References

1. Griffioen IP, Rietjens JA, Melles M, Snelders D, Homs MY, van Eijck CH, et al. The bigger picture of shared decision making: a service design perspective using the care path of locally advanced pancreatic cancer as a case. Cancer Med 2021 Sep 30;10(17):5907-5916 [FREE Full text] [doi: 10.1002/cam4.4145] [Medline: 34328273]
2. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 03;157(1):29-43 [FREE Full text] [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
3. Sperger J, Freeman NL, Jiang X, Bang D, de Marchi D, Kosorok MR. The future of precision health is data‐driven decision support. Stat Anal 2020 Jul 24;13(6):537-543. [doi: 10.1002/sam.11475]
4. Rietjens JA, Griffioen I, Sierra-Pérez J, Sroczynski G, Siebert U, Buyx A, et al. Improving shared decision-making about cancer treatment through design-based data-driven decision-support tools and redesigning care paths: an overview of the 4D PICTURE project. Palliat Care Soc Pract 2024 Feb 12;18:26323524231225249 [FREE Full text] [doi: 10.1177/26323524231225249] [Medline: 38352191]
5. Cresswell K, Callaghan M, Khan S, Sheikh Z, Mozaffar H, Sheikh A. Investigating the use of data-driven artificial intelligence in computerised decision support systems for health and social care: a systematic review. Health Informatics J 2020 Sep 22;26(3):2138-2147 [FREE Full text] [doi: 10.1177/1460458219900452] [Medline: 31964204]
6. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nat Mach Intell 2019 Nov 04;1(11):501-507. [doi: 10.1038/s42256-019-0114-4]
7. McLennan S, Fiske A, Tigard D, Müller R, Haddadin S, Buyx A. Embedded ethics: a proposal for integrating ethics into the development of medical AI. BMC Med Ethics 2022 Jan 26;23(1):6 [FREE Full text] [doi: 10.1186/s12910-022-00746-3] [Medline: 35081955]
8. McDougall R. Reviewing Literature in Bioethics Research: Increasing Rigour in Non-Systematic Reviews. Bioethics 2015 Feb 06;29(7):523-528. [doi: 10.1111/bioe.12149] [Medline: 25655982]
9. Willem T, Fritzsche M, Zimmermann BM, Sierawska A, Breuer S, Braun M, et al. Embedded ethics in practice: a toolbox for integrating the analysis of ethical and social issues into healthcare AI research. Sci Eng Ethics 2024 Dec 24;31(1):3. [doi: 10.1007/s11948-024-00523-y] [Medline: 39718728]
10. Schofield G, Dittborn M, Selman LE, Huxtable R. Defining ethical challenge(s) in healthcare research: a rapid review. BMC Med Ethics 2021 Sep 29;22(1):135 [FREE Full text] [doi: 10.1186/s12910-021-00700-9] [Medline: 34587950]
11. Hem MH, Gjerberg E, Husum TL, Pedersen R. Ethical challenges when using coercion in mental healthcare: a systematic literature review. Nurs Ethics 2018 Feb 01;25(1):92-110. [doi: 10.1177/0969733016629770] [Medline: 26931767]
12. Brusseau J. Using edge cases to disentangle fairness and solidarity in AI ethics. AI Ethics 2021 Sep 20;2(3):441-447. [doi: 10.1007/s43681-021-00090-z]
13. Liao S, Carbonell V. Materialized oppression in medical tools and technologies. Am J Bioeth 2023 Apr 09;23(4):9-23. [doi: 10.1080/15265161.2022.2044543] [Medline: 35262465]
14. Pot M, Kieusseyan N, Prainsack B. Not all biases are bad: equitable and inequitable biases in machine learning and radiology. Insights Imaging 2021 Feb 10;12(1):13 [FREE Full text] [doi: 10.1186/s13244-020-00955-7] [Medline: 33564955]
15. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf 2019 Mar 12;28(3):231-237 [FREE Full text] [doi: 10.1136/bmjqs-2018-008370] [Medline: 30636200]
16. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. JAMA Intern Med 2018 Nov 01;178(11):1544-1547 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3763] [Medline: 30128552]
17. Tsamados A, Aggarwal N, Cowls J, Morley J, Roberts H, Taddeo M, et al. The ethics of algorithms: key problems and solutions. In: Floridi L, editor. Ethics, Governance, and Policies in Artificial Intelligence. Cham, Switzerland: Springer; 2021.

18. Ledford H. Cancer geneticists tackle troubling ethnic bias in studies. Nature 2019 Apr 5;568(7751):154-155. [doi: 10.1038/d41586-019-01080-2] [Medline: 30967668]
19. Tupasela A, Di Nucci E. Concordance as evidence in the Watson for Oncology decision-support system. AI Soc 2020 Feb 01;35(4):811-818. [doi: 10.1007/s00146-020-00945-9]
20. Unger JM, Vaidya R, Hershman DL, Minasian LM, Fleury ME. Systematic review and meta-analysis of the magnitude of structural, clinical, and physician and patient barriers to cancer clinical trial participation. J Natl Cancer Inst 2019 Mar 01;111(3):245-255 [FREE Full text] [doi: 10.1093/jnci/djy221] [Medline: 30856272]
21. Broussard M. More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech. Cambridge, MA: MIT Press; 2023.
22. Garcia-Retamero R, Cokely ET. Designing visual aids that promote risk literacy: a systematic review of health research and evidence-based design heuristics. Hum Factors 2017 Jun 13;59(4):582-627. [doi: 10.1177/0018720817690634] [Medline: 28192674]
23. Kostick-Quenet KM, Gerke S. AI in the hands of imperfect users. NPJ Digit Med 2022 Dec 28;5(1):197 [FREE Full text] [doi: 10.1038/s41746-022-00737-z] [Medline: 36577851]
24. Rosenbaum L. The paternalism preference — choosing unshared decision making. N Engl J Med 2015 Aug 13;373(7):589-592. [doi: 10.1056/nejmp1508418]
25. McDougall RJ. Computer knows best? The need for value-flexibility in medical AI. J Med Ethics 2019 Mar 22;45(3):156-160. [doi: 10.1136/medethics-2018-105118] [Medline: 30467198]
26. Dratsch T, Chen X, Rezazade Mehrizi M, Kloeckner R, Mähringer-Kunz A, Püsken M, et al. Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. Radiology 2023 May 01;307(4):e222176. [doi: 10.1148/radiol.222176] [Medline: 37129490]
27. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012 Jan 01;19(1):121-127 [FREE Full text] [doi: 10.1136/amiajnl-2011-000089] [Medline: 21685142]
28. Paulus JK, Kent DM. Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. NPJ Digit Med 2020 Jul 30;3(1):99 [FREE Full text] [doi: 10.1038/s41746-020-0304-9] [Medline: 32821854]
29. Bak MA. Computing fairness: ethics of modeling and simulation in public health. Simulation 2020 Jun 27;98(2):103-111. [doi: 10.1177/0037549720932656]
30. van der Ende NA, den Hartog SJ, Broderick JP, Khatri P, Visser-Meily JM, van Leeuwen N, et al. Disentangling the association between neurologic deficits, patient-reported impairments, and quality of life after ischemic stroke. Neurology 2023 Mar 28;100(13):e1321-e1328. [doi: 10.1212/wnl.0000000000206747]
31. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. Science 2017 Apr 14;356(6334):183-186 [FREE Full text] [doi: 10.1126/science.aal4230] [Medline: 28408601]
32. Costanza-Chock S. Design Justice: Community-Led Practices to Build the Worlds We Need Information Policy. Cambridge, MA: MIT Press; 2020.
33. Ford E, Shepherd S, Jones K, Hassan L. Toward an ethical framework for the text mining of social media for health research: a systematic review. Front Digit Health 2020 Jan 26;2:592237 [FREE Full text] [doi: 10.3389/fdgth.2020.592237] [Medline: 34713062]
34. Azer SA. Social media channels in health care research and rising ethical issues. AMA J Ethics 2017 Nov 01;19(11):1061-1069 [FREE Full text] [doi: 10.1001/journalofethics.2017.19.11.peer1-1711] [Medline: 29168677]
35. Hunter RF, Gough A, O'Kane N, McKeown G, Fitzpatrick A, Walker T, et al. Ethical issues in social media research for public health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/ajph.2017.304249]
36. Boyd D, Crawford K. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Inf Commun Soc 2012 Jun;15(5):662-679. [doi: 10.1080/1369118X.2012.678878]
37. Bond CS, Ahmed OH, Hind M, Thomas B, Hewitt-Taylor J. The conceptual and practical ethical dilemmas of using health discussion board posts as research data. J Med Internet Res 2013 Jun 07;15(6):e112 [FREE Full text] [doi: 10.2196/jmir.2435] [Medline: 23748147]
38. Monks H, Cardoso P, Papageorgiou A, Carolan C, Costello L, Thomas L. Young people's views regarding participation in mental health and wellbeing research through social media. Int J Emot Educ 2015;7(1):4-19 [FREE Full text]
39. Moreno MA, Grant A, Kacvinsky L, Moreno P, Fleming M. Older adolescents' views regarding participation in Facebook research. J Adolesc Health 2012 Nov;51(5):439-444 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.02.001] [Medline: 23084164]
40. Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. BMC Med Ethics 2016 Apr 14;17:22 [FREE Full text] [doi: 10.1186/s12910-016-0105-5] [Medline: 27080238]
41. Stockdale J, Cassell J, Ford E. "Giving something back": a systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland. Wellcome Open Res 2018 Jan 17;3:6 [FREE Full text] [doi: 10.12688/wellcomeopenres.13531.2] [Medline: 30854470]
42. Azam R. Accessing social media information for pharmacovigilance: what are the ethical implications? Ther Adv Drug Saf 2018 Aug 31;9(8):385-387 [FREE Full text] [doi: 10.1177/2042098618778191] [Medline: 30364758]

43. Gustafson DL, Woodworth CF. Methodological and ethical issues in research using social media: a metamethod of human papillomavirus vaccine studies. BMC Med Res Methodol 2014 Dec 02;14(1):127 [FREE Full text] [doi: 10.1186/1471-2288-14-127] [Medline: 25468265]
44. Taylor J, Pagliari C. Mining social media data: how are research sponsors and researchers addressing the ethical challenges? Res Ethics 2017 Oct 26;14(2):1-39. [doi: 10.1177/1747016117738559]
45. Zimmermann BM, Willem T, Bredthauer CJ, Buyx A. Ethical issues in social media recruitment for clinical studies: ethical analysis and framework. J Med Internet Res 2022 May 03;24(5):e31231 [FREE Full text] [doi: 10.2196/31231] [Medline: 35503247]
46. Nissenbaum H. Privacy as contextual integrity. Wash Law Rev 2004;79(1).
47. Spheres of Justice: A Defense of Pluralism and Equality. New York, NY: Basic Books; 2008.
48. Bak MA, Veeken R, Blom MT, Tan HL, Willems DL. Health data research on sudden cardiac arrest: perspectives of survivors and their next-of-kin. BMC Med Ethics 2021 Jan 28;22(1):7 [FREE Full text] [doi: 10.1186/s12910-021-00576-9] [Medline: 33509184]
49. Stiggelbout A, Griffioen I, Brands J, Melles M, Rietjens J, Kunneman M, et al. Metro mapping: development of an innovative methodology to co-design care paths to support shared decision making in oncology. BMJ Evid Based Med 2023 Oct;28(5):291-294 [FREE Full text] [doi: 10.1136/bmjebm-2022-112168] [Medline: 37236775]
50. Elwyn G. Shared decision making: what is the work? Patient Educ Couns 2021 Jul;104(7):1591-1595. [doi: 10.1016/j.pec.2020.11.032] [Medline: 33353840]
51. Waddell A, Lennox A, Spassova G, Bragge P. Barriers and facilitators to shared decision-making in hospitals from policy to practice: a systematic review. Implement Sci 2021 Jul 31;16(1):74 [FREE Full text] [doi: 10.1186/s13012-021-01142-y] [Medline: 34332601]
52. O'Neill O. Autonomy and Trust in Bioethics. Cambridge, UK: Cambridge University Press; Nov 2009.
53. Gerritse K, Hartman LA, Bremmer MA, Kreukels BP, Molewijk BC. Decision-making approaches in transgender healthcare: conceptual analysis and ethical implications. Med Health Care Philos 2021 Dec;24(4):687-699 [FREE Full text] [doi: 10.1007/s11019-021-10023-6] [Medline: 34008081]
54. Gómez-Vírseda C, de Maeseneer Y, Gastmans C. Relational autonomy: what does it mean and how is it used in end-of-life care? A systematic review of argument-based ethics literature. BMC Med Ethics 2019 Oct 26;20(1):76 [FREE Full text] [doi: 10.1186/s12910-019-0417-3] [Medline: 31655573]
55. Herrmann A, Sanson-Fisher R, Hall A, Wall L, Zdenkowski N, Waller A. A discrete choice experiment to assess cancer patients' preferences for when and how to make treatment decisions. Support Care Cancer 2018 Apr;26(4):1215-1220. [doi: 10.1007/s00520-017-3944-9] [Medline: 29090384]
56. Perron ME, Hudon C, Roux-Levy PH, Poitras ME. Shared decision-making with patients with complex care needs: a scoping review. BMC Prim Care 2024 Nov 05;25(1):390. [doi: 10.1186/s12875-024-02633-9] [Medline: 39501147]
57. Gawande A. Being Mortal: Medicine and What Matters in the End. New York, NY: Metropolitan Books; 2014.
58. Mullick A, Martin J, Sallnow L. An introduction to advance care planning in practice. BMJ 2013 Oct 21;347:f6064. [doi: 10.1136/bmj.f6064] [Medline: 24144870]
59. Chopik WJ. Death across the lifespan: age differences in death-related thoughts and anxiety. Death Stud 2017 Feb;41(2):69-77. [doi: 10.1080/07481187.2016.1206997] [Medline: 27573253]
60. Ergin M. Eternal life as privilege: cultural boundaries and social stratification in death announcements (1950-2010). Death Stud 2023;47(4):461-475. [doi: 10.1080/07481187.2022.2089778] [Medline: 35735249]
61. Xie W, Butcher R. Artificial intelligence decision support tools for end-of-life care planning conversations: CADTH horizon scan. Canadian Agency for Drugs and Technologies in Health. 2023. URL: https://www.cda-amc.ca/artificial-intelligence-decision-support-tools-end-life-care-planning-conversations [accessed 2025-03-20]
62. Renedo Illaguerri E, Sañudo Y, Hartman L, Bak M, Sierra-Pérez J. Embedding ethics in practice: preempting ethical issues in the field by reflecting on the methodology of shadowing within cancer care services. In: Proceedings of the 2024 Design Research Society. 2024 Presented at: DRS 2024; June 23-28, 2024; Boston, MA. [doi: 10.21606/DRS.2024.959]
63. Costanza-Chock S. Design values: hard-coding liberation? Design Justice. 2020 Feb 27. URL: https://designjustice.mitpress.mit.edu/pub/3h2zq86d/release/1 [accessed 2025-03-20]
64. Winner L. Do artifacts have politics? Deadalus 2007;109(1):121-136.
65. Blumenthal-Barby JS, Lee D, Volk RJ. Toward ethically responsible choice architecture in prostate cancer treatment decision-making. CA Cancer J Clin 2015;65(4):257-260 [FREE Full text] [doi: 10.3322/caac.21283] [Medline: 25998159]
66. Münscher R, Vetter M, Scheuerle T. A review and taxonomy of choice architecture techniques. Behav Decis Mak 2015 Aug 14;29(5):511-524. [doi: 10.1002/bdm.1897]
67. Ubel PA, Smith DM, Zikmund-Fisher BJ, Derry HA, McClure J, Stark A, et al. Testing whether decision aids introduce cognitive biases: results of a randomized trial. Patient Educ Couns 2010 Aug;80(2):158-163 [FREE Full text] [doi: 10.1016/j.pec.2009.10.021] [Medline: 20004545]
68. Leonard TC, Richard HT, Cass RS. Nudge: improving decisions about health, wealth, and happiness. Const Polit Econ 2008 Aug 22;19:356-360. [doi: 10.1007/s10602-008-9056-2]

69. Halpern SD, Ubel PA, Asch DA. Harnessing the power of default options to improve health care. N Engl J Med 2007 Sep 27;357(13):1340-1344. [doi: 10.1056/NEJMsb071595] [Medline: 17898105]
70. Bailenson JN, Shum MS, Uttal DH. The initial segment strategy: a heuristic for route selection. Mem Cognit 2000 Mar;28(2):306-318. [doi: 10.3758/bf03213808] [Medline: 10790984]
71. Loewenstein GL, Bryce C, Hagmann D, Rajpal S. Warning: you are about to be nudged. Behav Sci Policy 2015 Apr 01;1(1):35-42. [doi: 10.1353/bsp.2015.0000]
72. Ojerholm E, Halpern SD, Bekelman JE. Default options: opportunities to improve quality and value in oncology. J Clin Oncol 2016 Jun 01;34(16):1844-1847. [doi: 10.1200/JCO.2015.64.8741] [Medline: 26884581]
73. Blumenthal-Barby JS, Cantor SB, Russell HV, Naik AD, Volk RJ. Decision aids: when 'nudging' patients to make a particular choice is more ethical than balanced, nondirective content. Health Aff (Millwood) 2013 Feb;32(2):303-310. [doi: 10.1377/hlthaff.2012.0761] [Medline: 23381523]
74. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
75. Williams A. Cost-effectiveness analysis: is it ethical? J Med Ethics 1992 Mar;18(1):7-11 [FREE Full text] [doi: 10.1136/jme.18.1.7] [Medline: 1573655]
76. Dondorp W. Ethische aspecten van kostenutiliteitsanalyse. Signalering Ethiek en Gezondheid. 2005. URL: https://www.ceg.nl/binaries/ceg/documenten/signalementen/2005/06/29/ethische-aspecten-van-kostenutiliteitsanalyse/Ethische+aspecten+van+kostenutiliteitsanalyse.pdf [accessed 2025-03-20]
77. Greenhalgh T, Abimbola S. The NASSS framework - a synthesis of multiple theories of technology implementation. Stud Health Technol Inform 2019 Jul 30;263:193-204. [doi: 10.3233/SHTI190123] [Medline: 31411163]
78. Radical AP. Can a machine ever be moral? Robot politeness and persuasion with Tom Williams. YouTube. URL: https://youtu.be/BMOY7I-rKTM [accessed 2025-03-20]
79. Weßel M, Ellerich-Groppe N, Koppelin F, Schweda M. Gender and age stereotypes in robotics for eldercare: ethical implications of stakeholder perspectives from technology development, industry, and nursing. Sci Eng Ethics 2022 Aug 09;28(4):34 [FREE Full text] [doi: 10.1007/s11948-022-00394-1] [Medline: 35943612]
80. Weidema FC, Molewijk BA, Kamsteeg F, Widdershoven GA. Aims and harvest of moral case deliberation. Nurs Ethics 2013 Sep;20(6):617-631. [doi: 10.1177/0969733012473773] [Medline: 23426235]
81. Tenner E. Why Things Bite Back: Technology and the Revenge of Unintended Consequences. New York, NY: Knopf Doubleday Publishing Group; 1997.
82. Lucivero F, Swierstra T, Boenink M. Assessing expectations: towards a toolbox for an ethics of emerging technologies. In: The Ethical Challenges of Emerging Medical Technologies. Milton Park, UK: Routledge; 2017.
83. Swierstra T, Stemerding D, Boenink M. Exploring techno-moral change: the case of the ObesityPill. In: Sollie P, Düwell M, editors. Evaluating New Technologies. Dordrecht, The Netherlands: Springer; 2009.
84. Mol A. What diagnostic devices do: the case of blood sugar measurement. Theor Med Bioeth 2000 Jan;21(1):9-22. [doi: 10.1023/a:1009999119586] [Medline: 10927966]
85. Doo FX, Vosshenrich J, Cook TS, Moy L, Almeida EP, Woolen SA, et al. Environmental sustainability and AI in radiology: a double-edged sword. Radiology 2024 Feb 01;310(2):e232030. [doi: 10.1148/radiol.232030] [Medline: 38411520]
86. Abma TA, Pittens CA, Visse M, Elberse JE, Broerse JE. Patient involvement in research programming and implementation: a responsive evaluation of the dialogue model for research agenda setting. Health Expect 2015 Dec;18(6):2449-2464 [FREE Full text] [doi: 10.1111/hex.12213] [Medline: 24889933]
87. Patrick-Lake B, Goldsack JC. Mind the gap: the ethics void created by the rise of citizen science in health and biomedical research. Am J Bioeth 2019 Aug 24;19(8):1-2. [doi: 10.1080/15265161.2019.1639389] [Medline: 31339830]
88. Byskov MF. What makes epistemic injustice an “injustice”? J Soc Philos 2020 May 18;52(1):114-131. [doi: 10.1111/josp.12348]
89. Fricker M. Epistemic Injustice: Power and the Ethics of Knowing. Oxford, UK: Clarendon Press; 2007.
90. Kidd IJ, Carel H. Epistemic injustice and illness. J Appl Philos 2017 Feb 08;34(2):172-190 [FREE Full text] [doi: 10.1111/japp.12172] [Medline: 28303075]
91. Liabo K, Cockcroft EJ, Boddy K, Farmer L, Bortoli S, Britten N. Epistemic justice in public involvement and engagement: creating conditions for impact. Health Expect 2022 Aug;25(4):1967-1978 [FREE Full text] [doi: 10.1111/hex.13553] [Medline: 35774005]
92. Groot B, Abma T. Ethics framework for citizen science and public and patient participation in research. BMC Med Ethics 2022 Mar 13;23(1):23 [FREE Full text] [doi: 10.1186/s12910-022-00761-4] [Medline: 35282819]
93. Galison P. Image and Logic: A Material Culture of Microphysics. Chicago, IL: University of Chicago Press; 1997.
94. Abma T. Ethics work for good participatory action research. Beleidsonderzoek Online 2020 Sep [FREE Full text] [doi: 10.5553/BO/221335502020000006001]
95. Franco-Torres M, Rogers BC, Ugarelli RM. A framework to explain the role of boundary objects in sustainability transitions. Environ Innov Soc Transit 2020 Sep;36:34-48. [doi: 10.1016/j.eist.2020.04.010]
96. Kara H. Ethical dissemination. In: Research Ethics in the Real World. Bristol, UK: Policy Press; 2018.

97. Samuel G, Diedericks H, Derrick G. Population health AI researchers' perceptions of the public portrayal of AI: a pilot study. Public Underst Sci 2021 Feb 21;30(2):196-211 [FREE Full text] [doi: 10.1177/0963662520965490] [Medline: 33084490]
98. Mootz JJ, Taylor L, Wainberg ML, Khoshnood K. Ethical considerations for disseminating research findings on gender-based violence, armed conflict, and mental health: a case study from rural Uganda. Health Hum Rights 2019 Jun;21(1):81-92 [FREE Full text] [Medline: 31239616]
99. Kon AA. The role of empirical research in bioethics. Am J Bioeth 2009 Jun 22;9(6-7):59-65 [FREE Full text] [doi: 10.1080/15265160902874320] [Medline: 19998120]
100. D'Ignazio C, Klein LF. Data Feminism. Cambridge, MA: MIT Press; 2020.
101. Edmondson AC, Lei Z. Psychological safety: the history, renaissance, and future of an interpersonal construct. Annu Rev Organ Psychol Organ Behav 2014 Mar 21;1:23-43. [doi: 10.1146/annurev-orgpsych-031413-091305]
102. Gopalakrishna G, Ter Riet G, Vink G, Stoop I, Wicherts JM, Bouter LM. Prevalence of questionable research practices, research misconduct and their potential explanatory factors: a survey among academic researchers in The Netherlands. PLoS One 2022 Feb 16;17(2):e0263023 [FREE Full text] [doi: 10.1371/journal.pone.0263023] [Medline: 35171921]
103. Karimian G, Petelos E, Evers SM. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. AI Ethics 2022 Mar 28;2(4):539-551. [doi: 10.1007/s43681-021-00131-7]
104. Benzinger L, Ursin F, Balke WT, Kacprowski T, Salloch S. Should artificial intelligence be used to support clinical ethical decision-making? A systematic review of reasons. BMC Med Ethics 2023 Jul 06;24(1):48 [FREE Full text] [doi: 10.1186/s12910-023-00929-6] [Medline: 37415172]
105. Lysaght T, Lim HY, Xafis V, Ngiam KY. AI-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research. Asian Bioeth Rev 2019 Sep 12;11(3):299-314 [FREE Full text] [doi: 10.1007/s41649-019-00096-0] [Medline: 33717318]
106. Holmes W, Miao F. Guidance for Generative AI in Education and Research. Paris, France: UNESCO Publishing; 2023.
107. Coetzer JA, Loukili I, Goedhart NS, Ket JC, Schuitmaker-Warnaar TJ, Zuiderent-Jerak T, et al. The potential and paradoxes of eHealth research for digitally marginalised groups: a qualitative meta-review. Soc Sci Med 2024 Jun;350:116895 [FREE Full text] [doi: 10.1016/j.socscimed.2024.116895] [Medline: 38710135]

Abbreviations

AI: artificial intelligence
DST: decision support tool
NASSS: nonadoption; abandonment; and challenges to the scale-up, spread, and sustainability
PICTURE: Producing Improved Cancer Outcomes Through User-Centered Research
PPI: patient and public involvement
SDM: shared decision-making

Edited by N Cahill; submitted 24.08.24; peer-reviewed by M Dittborn, A Tomicic; comments to author 04.11.24; revised version received 03.01.25; accepted 24.02.25; published 10.04.25.

Please cite as:
Bak M, Hartman L, Graafland C, Korfage IJ, Buyx A, Schermer M, 4D PICTURE Consortium
Ethical Design of Data-Driven Decision Support Tools for Improving Cancer Care: Embedded Ethics Review of the 4D PICTURE Project
JMIR Cancer 2025;11:e65566
URL: https://cancer.jmir.org/2025/1/e65566
doi:10.2196/65566
PMID:

Usability, Acceptability, and Barriers to Implementation of a Collaborative Agenda-Setting Intervention (CASI) to Promote Person-Centered Ovarian Cancer Care: Development Study

Rachel A Pozzar[1,2], APRN, PhD; James A Tulsky[2,3,4], MD; Donna L Berry[5], RN, PhD; Jeidy Batista[2,3], MS; Paige Barwick[6], BA; Charlotta J Lindvall[2,4,6], MD, PhD; Patricia C Dykes[2,3], RN, MA, PhD; Michael Manni[6], BS; Ursula A Matulonis[1,2], MD; Nadine J McCleary[1,2,6], MPH, MD; Alexi A Wright[1,2], MPH, MD

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, BP1143, Boston, MA, United States

[2]Department of Medicine, Harvard Medical School, Boston, MA, United States

[3]Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States

[4]Department of Supportive Oncology, Dana-Farber Cancer Institute, Boston, MA, United States

[5]Department of Biobehavioral Nursing and Health Informatics, University of Washington, Seattle, WA, United States

[6]Department of Informatics and Analytics, Dana-Farber Cancer Institute, Boston, MA, United States

Corresponding Author:
Rachel A Pozzar, APRN, PhD
Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, BP1143, Boston, MA, United States

Abstract

Background: People with advanced ovarian cancer and their caregivers report unmet supportive care needs. We developed a Collaborative Agenda-Setting Intervention (CASI) to elicit patients' and caregivers' needs through the patient portal before a clinic visit and to communicate these needs to clinicians using the electronic health record.

Objective: We aimed to assess the usability and acceptability of the CASI and identify barriers to and facilitators of its implementation.

Methods: We recruited English- and Spanish-speaking patients, caregivers, and clinicians from the gynecologic oncology program at a comprehensive cancer center. Participants used the CASI prototype and then completed individual cognitive interviews and surveys. We assessed usability with the System Usability Scale (scores range 0 - 100, scores ≥70 indicate acceptable usability) and acceptability with the Acceptability of Intervention Measure and Intervention Appropriateness Measure (scores for both measures range from 1 to 5, higher scores indicate greater acceptability). Interviews were audio recorded, transcribed, and analyzed using directed content analysis. Domains and constructs from the Consolidated Framework for Implementation Research comprised the initial codebook. We analyzed survey data using descriptive statistics and compared usability and acceptability scores across patients, caregivers, and clinicians using analyses of variance.

Results: We enrolled 15 participants (5 patients, 5 caregivers, and 5 clinicians). The mean System Usability Scale score was 72 (SD 16). The mean Acceptability of Intervention Measure and Intervention Appropriateness Measure scores were 3.9 (SD 1.0) and 4.1 (SD 0.8), respectively. Participants viewed the CASI content and format positively overall. Several participants appreciated the CASI's integration into the clinical workflow and its potential to increase attention to psychosocial concerns. Suggestions to refine the CASI included removing redundant items, simplifying item language, and adding options to request a conversation or opt out of supportive care referrals. Key barriers to implementing the CASI include its complexity and limited resources available to address patients' and caregivers' needs.

Conclusions: The CASI is usable and acceptable to patients with advanced ovarian cancer, caregivers, and clinicians. We identified several barriers to and facilitators of implementing the CASI. In future research, we will apply these insights to a pilot randomized controlled trial to assess the feasibility of comparing the CASI to usual care in a parallel group-randomized efficacy trial.

(JMIR Cancer 2025;11:e66801) doi:10.2196/66801

KEYWORDS

ovarian neoplasm; ovarian cancer; cancer; oncology; oncologist; metastases; communication; physician-patient relations; electronic health record; EHR; electronic medical record; EMR; implementation science; digital; digital health; digital technology; digital intervention; mobile phone

Introduction

Patients undergoing treatment for advanced ovarian cancer commonly experience burdensome disease- and treatment-related symptoms [1] followed by cancer recurrence after treatment completion [2]. Challenged to adapt to a chronic, life-limiting condition, more than two-thirds of patients with advanced ovarian cancer and their caregivers report at least 1 moderate-to-high unmet supportive care need [3]. Such needs negatively impact patients' and caregivers' health-related quality of life and are associated with a higher likelihood that patients will require emergency department visits and hospitalizations [4]. In 2022, a multidisciplinary expert panel recommended the provision of individualized and timely resources to address the unmet supportive care needs of patients with advanced ovarian cancer and their caregivers [5]. Likewise, health authorities in the United States, Canada, and the United Kingdom have advocated for the delivery of person-centered (alternatively, "patient-centered") cancer care [6-8].

Person-centered care entails eliciting and responding to patients' and caregivers' goals, values, and preferences in a system that supports high quality communication between patients, caregivers, and clinicians [9,10]. Research suggests the provision of person-centered care has the potential to improve health outcomes in patients with cancer [11]. In our prior work, we found that communication that fosters healing patient-clinician relationships is associated with better social and family well-being and lower symptom burdens among people with ovarian cancer [12,13]. Patients in these studies wanted their clinicians to be proactive and attentive to patients' psychosocial concerns and other supportive care needs [13]. However, time constraints, medical complexity, and inadequate resources for follow-up care may challenge clinicians to identify and manage nonmedical needs routinely [14-16].

To overcome these barriers, we used a design thinking approach to develop a Collaborative Agenda-Setting Intervention (CASI) to promote person-centered ovarian cancer care [17]. The CASI is a patient portal- and electronic health record (EHR)–integrated tool that aims to improve patient and caregiver well-being by routinely eliciting patients' and caregivers' values, preferences, and supportive care needs. The CASI supports agenda-setting and person-centered communication between patients, caregivers, and clinicians. The American Society of Clinical Oncology [18] and the European Society for Medical Oncology [19] have published guidelines for person-centered communication which direct clinicians to routinely set an agenda for visits by sharing their goals and eliciting topics that patients and caregivers wish to address. In the primary care setting, agenda-setting has been shown to increase person-centered care without prolonging visit duration [20] and to reduce clinician burden by preventing late-breaking concerns [21]. To date, however, research on agenda-setting interventions in cancer care is limited. The purpose of this study, therefore, was to assess the usability and acceptability of the CASI and identify barriers to and facilitators of implementing the CASI in this setting.

Methods

Study Design

We conducted a cross-sectional study using qualitative and quantitative methods. Our approach to data collection and analysis was guided by principles of design thinking [17] and the Consolidated Framework for Implementation Research (CFIR) [22]. According to the CFIR, the likelihood of successful implementation is affected by 5 domains: the inner setting (ie, the setting in which the innovation is being implemented), the outer setting (ie, the community or system in which the inner setting exists), the characteristics of the innovation itself, the characteristics of the individuals who will interact with the innovation, and the process by which the innovation is implemented [22]. We tested the CASI prototype and elicited stakeholder perspectives on each of the CFIR domains.

Recruitment

We recruited participants from the gynecologic oncology clinic of a National Cancer Institute–designated comprehensive cancer center. Patients were eligible if they were English- or Spanish-speaking adults with stage III, stage IV, or recurrent ovarian cancer. Caregivers were eligible if they were English- or Spanish-speaking adults who self-identified as the caregiver of a person with stage III, stage IV, or recurrent ovarian cancer. Enrollment in the cancer center's patient portal was not required for study participation. Patients and caregivers were purposively sampled to ensure representation of diverse demographic groups. We approached potentially eligible patients and caregivers in person during regularly scheduled clinic visits. Clinicians were eligible if they were an oncologist or advanced practice provider who cared for at least 4 outpatients per month with stage III, stage IV, or recurrent ovarian cancer. We introduced this study at a weekly provider meeting, then approached individual clinicians via email.

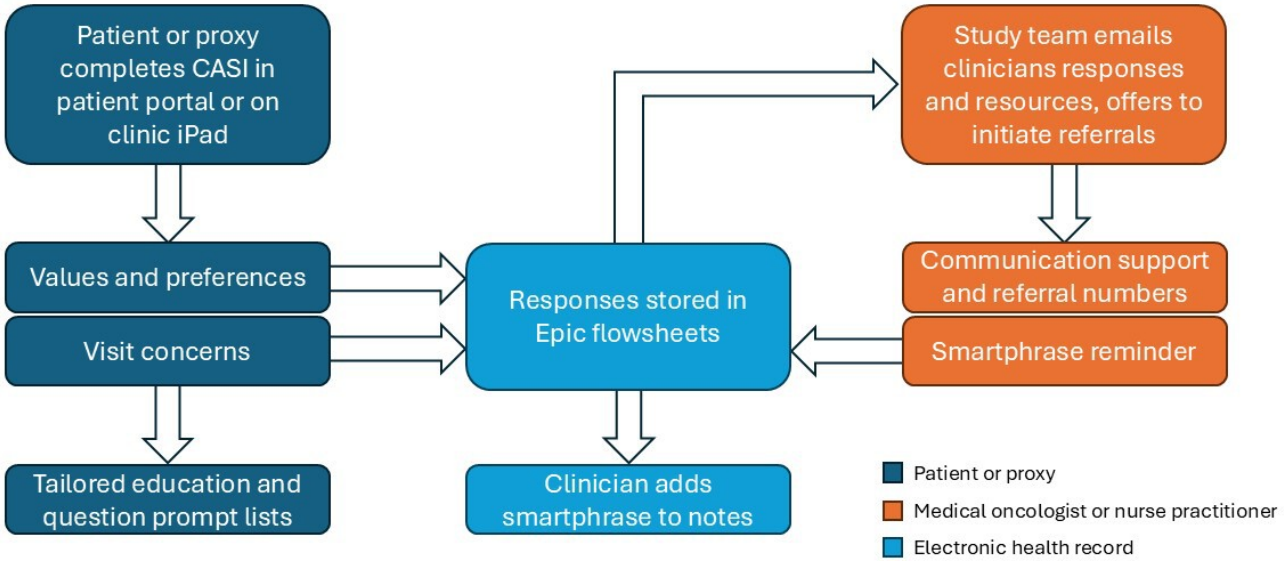
Ethical Considerations

All participants provided written informed consent and received US $25 upon data collection. Study data were deidentified before analysis. The Dana-Farber Cancer Institute Institutional Review Board approved all study procedures (protocol #21 - 322).

Intervention Characteristics

An intervention schema is provided in (Figure 1). The CASI has been integrated into Epic (Epic Systems Corporation), which is the most widely used EHR system in the United States [23]. The CASI is currently available in English and Spanish and has an English-language Flesch-Kincaid reading level of grade 6.9. Patients complete the CASI through the patient portal, which can be accessed on any smartphone, tablet, laptop, or desktop computer. Patients without a personal device may complete the CASI on a tablet provided by the clinic at the time of check-in. Caregivers with proxy access to the patient portal may complete the CASI in the same fashion as patients. Clinicians access the patient's or caregiver's responses to the CASI in Epic hyperspace, which is the clinician-facing EHR platform.

Figure 1. CASI schema. CASI: Collaborative Agenda-Setting Intervention.



Patients and caregivers are prompted through the patient portal to complete the CASI no more than 7 days before an upcoming clinic visit. Completing the CASI involves responding to 2 questionnaires as frequently as once every 3 weeks. The first questionnaire includes items about values, preferences, and communication needs. Selected items were derived from the Control Preferences Scale [24], which is a validated measure of patients' preferred level of involvement in medical decision-making, and the SHARE questionnaire [25], a communication intervention designed to elicit patients' preferences and goals. Patients and caregivers complete the first questionnaire the first time they use the CASI. Thereafter, patients and caregivers have the option to update their responses but are not required to repeat the questionnaire. The second questionnaire includes items derived from the National Comprehensive Cancer Network Distress Thermometer Problem List [26], which asks respondents to select the physical, emotional, social, practical, spiritual, and other concerns experienced during the past 7 days. Patients and caregivers complete the second questionnaire every time they use the CASI. Patients and caregivers receive standardized question prompt lists through the patient portal for each of the concerns they identify. In addition, a member of this study's team assists with follow-up navigation and initiating referrals when warranted.

Responses to both questionnaires are stored in Epic and associated with the appropriate upcoming clinical encounter. Clinicians can view CASI questionnaire responses in the rooming tab of the visit note (which contains patient-reported symptoms, vital signs, and information collected by the clinic assistant before the patient is seen by the physician or advanced practice provider), during chart review, or by using a smart phrase to populate the narrative history of present illness with the most recent CASI questionnaire responses. When a patient or caregiver completes the CASI, a member of this study's team emails the clinician a summary of the CASI responses, offers to initiate referrals, shares links to evidence-based communication guidance, and reminds the clinician how to use the CASI smart phrase to populate their visit note.

Data Collection

Participant Characteristics

All participants self-reported age, gender, ethnicity, and race. Patients and caregivers self-reported marital status, annual household income, and educational attainment. Caregivers self-reported their relationship to the patient and the number of clinic visits they attended with the patient in the last 6 months. Clinicians self-reported their clinical role and specialty, number of patients seen with advanced ovarian cancer per month, and number of years spent caring for patients with advanced ovarian cancer.

Cognitive Interviews

We conducted individual, semistructured cognitive interviews with patients, caregivers, and clinicians in person and over Health Insurance Portability and Accountability Act–compliant video conferencing software (Multimedia Appendix 1). To identify potential usability challenges, participants were instructed to think aloud as they used the CASI in a testing environment that was not linked to the EHR. Trained interviewers (RAP, JB, and PB) observed participants, made note of any sections of the CASI that were difficult for participants to navigate, and invited participants to comment on aspects of the CASI's content and design. Interviews with Spanish-speaking participants were conducted in Spanish by a bilingual member of this study's team (JB). All participants reviewed patient- and caregiver-facing content, but only clinicians reviewed clinician-facing content. To identify potential barriers to implementing the CASI, interviewers asked participants to consider integration of the CASI into their existing routine, identify barriers to using the CASI regularly, and identify strategies to minimize patient, caregiver, and clinician burden. Interviews were audio recorded, professionally transcribed, and professionally translated from Spanish to English when applicable. Following cognitive interviews, participants completed a REDCap (Research Electronic Data Capture; Vanderbilt University) [27,28] survey that included quantitative measures of usability, acceptability, and burden.

We refined the CASI in response to cognitive interview feedback.

Usability

The System Usability Scale (SUS) is a 10-item scale that assesses the usability of electronic systems. The SUS yields a single number representing a composite measure of the overall usability of the system being studied. Scores for individual items are not meaningful on their own. Total scores range from 0 to 100; higher scores indicate better usability [29], and scores of 70 or greater are acceptable [30].

Acceptability

The Acceptability of Intervention Measure (AIM) and Intervention Appropriateness Measure (IAM) are 4-item scales that each assess a single dimension of intervention acceptability. Response options range from 1 (“completely disagree”) to 5 (“completely agree”). Total scores for each measure range from 1 to 5. While cut scores have not yet been established for these measures, higher scores represent better acceptability [31].

Participant Burden

We assessed participant burden by asking participants to rate the extent to which they agreed with the statement “using the CASI placed a considerable burden on me.” Response options ranged from 1 (“strongly disagree”) to 5 (“strongly agree”).

Analysis

Cognitive Interviews

We analyzed transcripts of cognitive interviews using directed content analysis [32]. Directed content analysis involves drawing key constructs from existing theory to develop an initial codebook. When one or more text segments are not represented in the initial codebook, the investigator creates new codes. Our initial codebook was based on the 2009 version of the CFIR domains and constructs. Author RAP, a nurse scientist with expertise in qualitative research, read and coded all transcripts in their entirety. To enhance trustworthiness and foster reflexivity, coding was reviewed by members of this study’s team with expertise in clinical research (JB) and health informatics (PB). Differences in data interpretation were resolved through discussion. We used MAXQDA 2022 (VERBI Software GmbH) to support qualitative data management.

Quantitative Measures

We summarized participant characteristics and SUS, AIM, and IAM scores using descriptive statistics. We performed analyses of variance to identify differences in SUS, AIM, and IAM scores across patients, caregivers, and clinicians and considered values of $P<.05$ significant.

Results

Participant Characteristics

Data collection took place between August 2023 and July 2024. Interviews lasted an average of 40 (SD 10) minutes. A total of 15 participants (5 patients, 5 caregivers, and 5 clinicians) completed this study. Characteristics of each participant group are detailed in Table 1. Briefly, participants were predominantly women (11/15, 73%), White (12/15, 80%), non-Hispanic (11/15, 73%), and working full- or part-time (9/15, 60%). Patients and caregivers were most commonly married or partnered (5/10, 50%) college graduates (5/10, 50%) and reported a range of annual household incomes: less than US $24,000 (2/10, 20%), US $75,000 to US $119,000 (2/10, 20%), or US $120,000 or more (2/10, 20%) per year. Caregivers identified as spouses or partners (2/5, 40%), parents (2/5, 40%), and friends (1/5, 20%) of patients. Caregivers had most often attended between 5 and 9 clinic visits with patients in the last 6 months (2/5, 40%). Clinicians were primarily physicians working in medical oncology (3/5, 60%) who reported caring for more than 20 patients with advanced ovarian cancer each month (5/5, 100%) and had 15 - 19 years of experience caring for patients with advanced ovarian cancer (3/5, 60%).

Table . Participant characteristics.

		Total	Patients	Caregivers	Clinicians
All participants (N=15)					
Age (years), mean (SD)		61 (13)	66 (5)	66 (19)	54 (16)
Gender, n (%)					
	Woman	11 (73)	5 (100)	2 (40)	4 (80)
	Man	4 (27)	0 (0)	3 (60)	1 (20)
Ethnicity, n (%)					
	Hispanic	4 (27)	2 (40)	2 (40)	0 (0)
	Non-Hispanic	11 (73)	3 (60)	3 (60)	5 (100)
Race, n (%)					
	Asian	2 (13)	1 (20)	0 (0)	1 (20)
	Native American	1 (7)	1 (20)	0 (0)	0 (0)
	White	12 (80)	3 (60)	5 (100)	4 (80)
Employment status, n (%)					
	Working	9 (60)	1 (20)	3 (60)	5 (100)
	Retired	4 (27)	2 (40)	2 (40)	0 (0)
	Disabled	2 (13)	2 (40)	0 (0)	0 (0)
Patients and caregivers (n=10)					
Marital status, n (%)					
	Married or partnered	5 (50)	3 (60)	2 (40)	—[a]
	Single or never married	3 (30)	1 (20)	2 (40)	—
	Divorced	2 (20)	1 (20)	1 (20)	—
Annual household income (US $), n (%)					
	Less than $24,000	2 (20)	2 (40)	0 (0)	—
	$45,000-$74,999	1 (10)	1 (20)	0 (0)	—
	$75,000-$119,000	2 (20)	1 (20)	1 (20)	—
	$120,000 or more	2 (20)	1 (20)	1 (20)	—
	Missing	3 (30)	0 (0)	3 (60)	—
Educational attainment, n (%)					
	Did not graduate high school	2 (20)	2 (40)	0 (0)	—
	Graduated high school	1 (10)	0 (0)	1 (20)	—
	Graduated college	5 (50)	2 (40)	3 (60)	—
	Postgraduate degree	2 (20)	1 (20)	1 (20)	—
Caregivers (n=5)					
Relationship to patient, n (%)					
	Spouse or partner	—	—	2 (40)	—
	Parent	—	—	2 (40)	—
	Friend	—	—	1 (20)	—
Clinic visits attended with patient in last 6 months, n (%)					
	5 - 9	—	—	2 (40)	—

		Total	Patients	Caregivers	Clinicians
	10 - 14	—	—	1 (20)	—
	15 - 19	—	—	1 (20)	—
	20 or more	—	—	1 (20)	—
Clinicians (n=5)					
Role, n (%)					
	Physician	—	—	—	3 (60)
	Advanced practice nurse	—	—	—	2 (40)
Patients seen with advanced ovarian cancer per month, n (%)					
	20 or more	—	—	—	5 (100)
Clinical specialty, n (%)					
	Medical oncology	—	—	—	5 (100)
Years caring for patients with advanced ovarian cancer, n (%)					
	10 - 14	—	—	—	1 (20)
	15 - 19	—	—	—	3 (60)
	20 or more	—	—	—	1 (20)

[a]Not applicable.

Usability, Acceptability, and Burden

SUS, AIM, and IAM scores by participant group are reported in Table 2. Briefly, the overall mean SUS score was 72 (SD 16), which is above the threshold of acceptable usability [30]. The overall mean AIM score was 3.9 out of 5 (SD 1), while the overall mean IAM score was 4.1 out of 5 (SD 0.8). There were no statistically significant (P<.05) differences in SUS, AIM, or IAM scores across patients, caregivers, and clinicians. Moreover, 1/15 (7%) participants reported that the CASI was burdensome; this participant was a patient who experienced difficulty adjusting the font size on the CASI questionnaires during usability testing.

Table . Usability and acceptability of the Collaborative Agenda-Setting Intervention.

	Overall (n=15), mean (SD)	Patients (n=5), mean (SD)	Caregivers (n=5), mean (SD)	Clinicians (n=5), mean (SD)	*F* test (*df*)	*P* value
SUS[a]	72 (16)	64 (15)	78 (12)	73 (20)	0.82 (2)	.47
AIM[b]	3.9 (1)	3.8 (0.8)	4 (1)	4 (1.2)	0.05 (2)	.95
IAM[c]	4.1 (0.8)	4 (0.8)	4.2 (0.8)	4.2 (0.8)	0.07 (2)	.93

[a]SUS: System Usability Scale. Possible scores range from 0-100. Higher scores indicate greater usability, and scores ≥70 suggest above-average usability.

[b]AIM: Acceptability of Intervention Measure. Possible scores range from 1 - 5. Higher scores indicate greater acceptability and appropriateness, respectively.

[c]IAM: Intervention Appropriateness Measure. Possible scores range from 1 - 5. Higher scores indicate greater acceptability and appropriateness, respectively.

Cognitive Interview Findings

Suggested Revisions

Participants made numerous suggestions to improve the clarity and helpfulness of CASI questionnaire items. For example, several participants suggested reducing the number of response options on the Control Preferences Scale from 5 to 3. Participants also identified several items they felt were redundant. Caregiver participants suggested explicitly instructing caregivers to respond on behalf of the patient. Clinician participants were concerned about patients and caregivers being "locked in" to a specific preference or decision. For example, these participants suggested the CASI ask patients and caregivers what topics they would like to discuss rather than giving them the opportunity to identify topics they would like to avoid. Furthermore, 1 clinician observed that this modification would allow clinicians to raise topics they believed needed to be addressed even if the patient or caregiver was not previously interested in discussing it. Finally, some participants wanted the option to request a follow-up phone call or to opt out of navigation (ie, a follow-up phone call that would arrange for them to see social work or chaplaincy). A detailed list of suggested and incorporated revisions is provided in Table 3.

Table . CASI[a] revisions derived from cognitive interviews.

Participant recommendations	Number recommending (n)				Actions taken
	PT[b]	CG[c]	CL[d]	All, n (%)	
Rephrase items for clarity	5	4	2	11 (73)	Rephrased items to improve clarity.
Eliminate redundant items	4	0	1	5 (33)	Revised items to eliminate redundancy.
Provide an option for personal interaction	3	2	0	5 (33)	Added item that reads: "Would you like someone to contact you about your options for additional support?"
Avoid locking patients and clinicians into a preference or decision	0	1	2	3 (20)	Rephrased Control Preferences Scales to refer to "most decisions." Removed items allowing patients to indicate they "never" want to talk about a specific topic or undergo a specific procedure.
Consider limiting administration to patients who are not on their first line of treatment	1	0	2	3 (20)	Planned: Will confirm with clinicians whether potential participants are appropriate for pilot RCT.[e]
Revise items to reduce anxiety	0	1	2	3 (20)	Reframed items to allow patients to indicate that they would like to discuss a certain topic with their clinician.
Account for different caregiving roles	1	1	0	2 (13)	Added option for caregiver respondent to report their name and relationship to the patient.
Administer every 3 - 4 weeks	0	1	1	2 (13)	Planned: Will enroll participants receiving treatment every 3 weeks into pilot RCT.
Clarify instructions for caregivers	0	2	0	2 (13)	Added the following instructions: "The following questions should be answered from the patient's perspective. Please respond on behalf of the patient."
Account for preferences for in-person versus digital communication	1	0	0	1 (6.7)	Added item that reads: "Would you like someone to contact you about your options for additional support?"
Allow for preference tracking over time	0	1	0	1 (6.7)	Improved readability of provider-facing Epic flow sheet.

Participant recommendations	Number recommending (n)				Actions taken
	PT[b]	CG[c]	CL[d]	All, n (%)	
Ask an open-ended item about what is meaningful to the patient	0	1	0	1 (6.7)	Added open-ended item that reads "What does a good day look like to you?"
Ask what caregiver is going through	0	1	0	1 (6.7)	Added open-ended item that reads "What, if anything, can we do to help you support the patient?"
Consider removing the option to request exact numbers and statistics	0	0	1	1 (6.7)	Rephrased to "with a lot of detail" and "without a lot of detail."
Differentiate between ongoing and new or acute concerns	0	0	1	1 (6.7)	Planned: In emails to clinicians, study team will highlight changes in concerns.
Ensure mechanism for prompt follow-up	0	0	1	1 (6.7)	Planned: Study team will provide first-line follow-up and will track time spent responding to patient concerns.
Alert clinicians to unmet needs or changes in CASI responses	0	0	1	1 (6.7)	Planned: In emails to clinicians, study team will highlight changes in concerns.
Make follow-up navigation optional	1	0	0	1 (6.7)	Added item that reads: "Would you like someone to contact you about your options for additional support?"
Prompt patients to consider code status discussion	0	0	1	1 (6.7)	Added "the care I would like to receive if my health worsens" as a possible discussion topic.
Reduce number of response options for items about shared decision-making and communication preferences	0	0	1	1 (6.7)	Reduced the number of response options on the Control Preferences Scales from 5 to 3. Revised items related to communication preferences.

[a]CASI: Collaborative Agenda-Setting Intervention.

[b]PT: patient.

[c]CG: caregiver.

[d]CL: clinican.

[e]RCT: randomized controlled trial.

Barriers to and Facilitators of Implementation

Overview

Participants identified barriers to and facilitators of implementing the CASI across 4 CFIR domains: outer setting, inner setting, innovation characteristics, and individual characteristics. Exemplary quotes for the CFIR constructs addressed by participants are provided in Table 4. Definitions of each construct as they pertain to the CASI are incorporated into the Results section below.

Table . Potential barriers to and facilitators of implementing the Collaborative Agenda-Setting Intervention.

CFIR[a] domain and construct		Participants reporting, n (%)	Role in implementation	Exemplary quotes
Outer setting				
	Needs and resources of those served	15 (100)	Facilitator	"I think the tool worked nicely. Provocative, but I think helpful. I think some people-- there is my good friend with cancer who does not want to know anything, absolutely nothing, but what has to happen today-- and I am the one who wants to know the future. So I think it is helpful." [Female patient, aged 62 years]
Inner setting				
	Compatibility	5 (33)	Facilitator	"I do think any way you can streamline information to actually come to the clinician will help a lot in terms of its usability." [Female advanced practice nurse, 10 - 14 years of experience]
	Available resources	5 (33)	Barrier	"I mean there is unfortunately a huge problem now with social work in that there are no social workers. We have one remaining social worker in GI; we have two social workers in GYN, but they are really strapped, and I have found meaningful social work contact almost impossible." [Female medical oncologist, 20+ years of experience]
	Relative priority	3 (20)	Both	"There are very few patients who are like, 'I feel great, and I don't have any of these things,' you know, but the things that they have are often the same and constant from visit to visit." [Female medical oncologist, 15 - 19 years of experience]
	Tension for change	3 (20)	Facilitator	"That would be helpful because you know that question about anxiety and stuff I mean I have anxiety, I have anxiety about the finances I am primarily the one that has been managing the finances and seeing us through all of this." [Male spousal caregiver, age not reported]
	Networks and communications	1 (6.7)	Barrier	"I think it is important for us to know that the patients are approached so we expect what will happen because there are going to be questions which is fine." [Male medical oncologist, 15 - 19 years of experience]

CFIR[a] domain and construct		Participants reporting, n (%)	Role in implementation	Exemplary quotes
Innovation characteristics				
	Complexity	12 (80)	Barrier	"It's not clear to me whether it's me you're asking about or you're asking about her, and I should be answering questions about her." [Female friend, aged 79 years]
	Design quality and packaging	12 (80)	Barrier	"Okay… So now I should press here? Where should I press now?" [Female patient, aged 63 years]
	Relative advantage	5 (33)	Facilitator	"It is nice to know if what kind of person they are so I can start understanding their values." [Female advanced practice nurse, 10 - 14 years of experience]
	Adaptability	4 (27)	Both	"[One concern would be if] you can't go do this other thing that you were supposed to because you made a [different] decision back then." [Male adult child caregiver, age not reported]
	Trialability	4 (27)	Both	"And for a caregiver, how would you present this to them? I mean I'm certainly not on her patient [portal]." [Female friend, aged 79 years]
Individual characteristics				
	Other personal attributes	11 (73)	Barrier	"I like just to talk. I'm not a person that likes filling out answers, to be honest with you." [Female patient, aged 61 years]
	Knowledge and beliefs about the innovation	3 (20)	Barrier	"Even if they are finished with treatment, the last thing they want to think about is a conversation about death and dying when they think they are cured." [Female advanced practice nurse, 15 - 19 years of experience]
	Self-efficacy	3 (20)	Both	"Would I know some of these things? Yes, perhaps, but I would not feel comfortable with [reporting them on the patient's behalf]." [Female friend, aged 79 years]
	Individual stage of change	2 (13)	Barrier	"The only one that would make me change my management style would be the last one, which is, 'I prefer to leave all treatment decisions to my doctor.'" [Female medical oncologist, 15 - 19 years of experience]

CFIR[a] domain and construct		Participants reporting, n (%)	Role in implementation	Exemplary quotes
	Individual identification with organization	1 (6.7)	Barrier	"I have other sources [of health care]." [Female friend, aged 79 years]

[a]CFIR: Consolidated Framework for Implementation Research.

Outer Setting

All participants (15/15, 100%) addressed the needs and resources of those served, referring to the extent to which the needs of end users are accurately known and prioritized by the CASI [22]. Comments related to this construct were overwhelmingly positive. Participants reported that the topics addressed by the CASI are important, and they appreciated that the CASI was brief and easy to use. Further, 1 caregiver was especially enthusiastic about being able to update his or the patient's preferences as they evolve over time. However, 1 caregiver observed that not every caregiver has proxy access to the patient portal, and 1 clinician worried that asking about preferences for prognostic communication may exacerbate patients' anxiety.

Inner Setting

In total 5 (33%) participants addressed the CASI's compatibility, which refers to how well the CASI will fit into existing practice norms, workflows, and systems [22]. Overall, participants approved of the CASI's integration into the clinical workflow. Clinicians especially appreciated having the opportunity to follow up on needs that may not be addressed during a visit and the ease with which they could populate their visit notes with patients' and caregivers' CASI questionnaire responses. Further, 5 (33%) participants discussed the available resources for implementing the CASI [22]. Each of these 4 clinicians and 1 caregiver were concerned there would not be enough staff or supportive care services available to address patient- and caregiver-reported concerns. Furthermore, 3 (20%) participants addressed the CASI's relative priority, which refers to the importance of implementing the CASI relative to other innovations [22]. Moreover, 1 clinician was concerned about the amount of information clinicians already review before a visit. A second clinician noted that, despite the importance of addressing supportive care needs, patients' medical needs would likely take priority. Additionally, 3 (20%) participants addressed tension for change, which refers to the degree to which stakeholders believe current practices need to change [22]. While none of these participants shared a negative opinion of current practice, each acknowledged there is room for improvement in certain aspects of care. For example, 1 patient described an experience of suboptimal communication that was distressing to her, while 1 clinician expressed that it would be helpful to know a patient's decision control preferences. Finally, 1 (6.7%) participant addressed networks and communications, which refers to the nature and quality of formal and informal communication within an organization [22]. This clinician wanted to be sure they would be notified when a patient completed the CASI.

Innovation Characteristics

A total of 12 (80%) participants addressed the CASI's design quality and packaging, which refers to the level of perceived excellence in how the CASI is bundled and presented [22]. Participants appreciated the clean, simple layout of the CASI. However, participants made several suggestions related to font size and screen layout. Further, 12 (80%) participants also addressed the CASI's complexity. As noted above, participants identified items they felt were redundant or overly complex and made suggestions to clarify user instructions. Furthermore, 5 participants addressed the CASI's relative advantage, which refers to the advantage of implementing the CASI versus an alternative solution [22]. Patients appreciated being able to report their concerns in addition to their preferences, and clinicians observed that the CASI addresses topics that are not captured in our institution's existing patient-reported symptom questionnaire. Additionally, 4 (27%) participants addressed the adaptability of the CASI, which refers to the extent to which the CASI can be adapted, tailored, or refined to meet user's needs [22]. Participants provided suggestions to enhance the CASI's adaptability. For example, 2 patients suggested making follow-up navigation optional. Moreover, 4 (27%) participants addressed the CASI's trialability and approved of the iterative process through which it is being developed and tested.

Individual Characteristics

A total of 11 (73%) participants identified personal attributes of end users that may influence adoption of the CASI. Specifically, participants emphasized the need to consider patients' and caregivers' overall literacy, digital literacy, preferences for face-to-face or digital communication, patients' time since diagnosis, and caregivers' caregiving role. Further, 3 (20%) participants expressed knowledge and beliefs about the CASI content that may influence its adoption. For example, as noted above, 1 clinician wondered whether items about prognostic communication would exacerbate patients' anxiety. Furthermore, 3 (20%) participants addressed potential challenges related to patients' and caregivers' self-efficacy to complete the CASI. In addition, 1 clinician suspected that patients who need supportive care services often lack the self-efficacy to seek assistance because they are experiencing symptoms of depression. Besides 1 caregiver did not feel comfortable responding to questions on the patient's behalf. Additionally, 2 (13%) participants addressed their individual stage of change as it pertains to adopting the CASI. These clinicians expressed that they were unsure of the extent to which they would change their practice style in response to patients' or caregivers' CASI responses. Finally, 1 (6.7%) participant addressed her identification with the organization that created the CASI. This caregiver expressed that while she appreciated being asked about her supportive care needs, she would turn to her primary care provider rather than the cancer center for assistance.

Discussion

Principal Results

The findings of our study suggest that the CASI is a usable and acceptable intervention to foster person-centered ovarian cancer care. Participants suggested several revisions that have since been incorporated into the CASI. Participants identified several facilitators of implementing the CASI, including that the CASI meets the needs of patients, caregivers, and clinicians; is well-designed; and is compatible with the existing workflow. Additionally, participants indicated that the CASI offers a relative advantage over usual care and that there is interest among stakeholders in standardizing the way supportive care needs are identified and documented.

Participants also identified potential barriers to implementing the CASI, and the CFIR provided a relevant organizing framework for these barriers. The most frequently identified barriers were the complexity of the CASI and the resources available to address patients' and caregivers' supportive care needs. Our approach to intervention development has preemptively addressed some of these potential barriers. According to the Expert Recommendations for Implementing Change taxonomy [33], intervention complexity can be managed by eliciting local knowledge, conducting small tests of cyclical change, and providing ongoing training. Accordingly, we have prioritized stakeholder engagement throughout the process of intervention development and have taken an iterative approach to refining and testing the CASI [17]. Our next planned study will assess the feasibility of conducting a full-scale efficacy trial of the CASI in a pilot randomized controlled trial (RCT). During this trial, we will aim to manage the perceived complexity of the intervention by training participants to use the CASI and providing written training materials and resources in advance.

To address the potential barrier of limited available resources, the Expert Recommendations for Implementing Change taxonomy recommends securing additional funding and developing resource sharing agreements [33]. In our planned pilot RCT, members of this study's team will be responsible for following up with patients and caregivers who report unmet supportive care needs. We will track the frequency with which these needs arise and will document the resources (eg, time and personnel) needed to address them. Upon completion of the pilot RCT, we hope to have a more robust understanding of the resources that will be required to implement the CASI. These findings will inform our funding application for a subsequent efficacy trial and may eventually inform resource allocations to support the integration of the CASI into routine care.

Strengths and Limitations

There is a growing recognition of the potential for EHR-integrated supportive care interventions to improve health outcomes among patients with advanced cancer. Leaders in communication research have specifically called for the development of interventions that integrate into the clinic workflow and are disseminated through existing mechanisms [34,35]. The CASI meets this need and was developed according to stakeholders' needs and feedback [17]. In the current study, we conducted usability testing in a linguistically and socioeconomically diverse sample of patients with advanced ovarian cancer, caregivers, and clinicians. In addition, our study design was informed by an established implementation science framework [22]. Nevertheless, this was a single-site study with 15 participants, and the generalizability of our findings to other settings may be limited. Additional research is needed to assess the feasibility of conducting a full-scale efficacy trial of the CASI, to compare the effect of the CASI on outcomes to that of usual care, and to assess implementation of the CASI in diverse real-world settings.

Comparison With Prior Work

To our knowledge, the CASI is the first EHR-integrated supportive cancer care intervention to use an agenda-setting approach. Our finding that the CASI is usable by and acceptable to patients, caregivers, and clinicians is consistent with prior studies of EHR-integrated supportive care interventions. In a sample of patients with heterogeneous cancer types and their caregivers, a biopsychosocial care need screening system had above-average usability [36]. Similarly, in a sample of patients with heterogeneous cancer types, a patient-reported symptom and needs monitoring system was usable and relevant [37]. In a sample of more than 100 health professionals, a centralized location for storing patients' values, goals, and preferences was viewed positively [38]. Our study adds to these findings by demonstrating that patients' and caregivers' values, preferences, and supportive care needs can be assessed and communicated as part of the same intervention without sacrificing usability or acceptability.

Widespread implementation of EHR-integrated supportive care interventions has not yet been realized, and research describing barriers to and facilitators of implementing these interventions is limited. However, in an evaluation of patient perspectives related to implementing a patient-reported symptom and needs monitoring system, Lyleroehr et al [37] found that low clinician engagement and suboptimal communication about the intervention were key barriers to implementation. Similarly, Wickline et al [39] found that more than half of patients with advanced ovarian cancer using a remote symptom and quality of life monitoring system felt it was not obvious their clinician used the system's reports. In both studies, patients valued having the option to speak directly to their care team about their concerns [37,39]. In our planned pilot RCT, we will attempt to mitigate these potential barriers by offering patients and caregivers follow-up phone calls, providing clinicians with a summary of patients' and caregivers' responses ahead of a scheduled visit, and offering to initiate referrals to reduce clinician burden.

Conclusions

Agenda-setting is a novel approach to promoting person-centered care for individuals with ovarian cancer. Our findings suggest the CASI is usable and acceptable to patients, caregivers, and clinicians. Guided by an implementation science framework, we identified several barriers to and facilitators of implementing the CASI. In subsequent research, we will assess the feasibility of conducting an efficacy trial comparing the CASI to usual care.

Acknowledgments

RAP received funding from the National Palliative Care Research Center and the National Institute of Nursing Research (K23NR020219-01A1) in support of this work. This study's sponsors did not play a role in this study's design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit this paper for publication.

Data Availability

The datasets generated during this study are available from the corresponding author on reasonable request.

Generative artificial intelligence was not used at any point in the preparation of this paper.

Authors' Contributions

RAP did the conceptualization, methodology, investigation, formal analysis, writing of the original draft, and funding acquisition. JAT, DLB, CJL, and PCD worked on the conceptualization, methodology, and review and editing of the writing. JB and PB handled the investigation, formal analysis, and review and editing of the writing. MM, UAM, and NJM assisted with resources and review and editing of the writing. AAW aided the conceptualization, methodology, formal analysis, and review and editing of the writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Cognitive interview guides.
[DOCX File, 25 KB - cancer_v11i1e66801_app1.docx]

References

1. Webber K, Carolus E, Mileshkin L, et al. OVQUEST - life after the diagnosis and treatment of ovarian cancer - an international survey of symptoms and concerns in ovarian cancer survivors. Gynecol Oncol 2019 Oct;155(1):126-134. [doi: 10.1016/j.ygyno.2019.08.009] [Medline: 31416612]
2. Coleman RL, Spirtos NM, Enserro D, et al. Secondary surgical cytoreduction for recurrent ovarian cancer. N Engl J Med 2019 Nov 14;381(20):1929-1939. [doi: 10.1056/NEJMoa1902626] [Medline: 31722153]
3. DiSipio T, Hartel G, Butow P, Webb PM, Beesley VL. Impact of disease recurrence on the supportive care needs of patients with ovarian cancer and their caregivers. Gynecol Oncol 2024 Jun;185:33-41. [doi: 10.1016/j.ygyno.2024.02.012] [Medline: 38364693]
4. Penedo FJ, Natori A, Fleszar-Pavlovic SE, et al. Factors associated with unmet supportive care needs and emergency department visits and hospitalizations in ambulatory oncology. JAMA Netw Open 2023 Jun 1;6(6):e2319352. [doi: 10.1001/jamanetworkopen.2023.19352] [Medline: 37342038]
5. Temkin SM, Smeltzer MP, Dawkins MD, et al. Improving the quality of care for patients with advanced epithelial ovarian cancer: program components, implementation barriers, and recommendations. Cancer 2022 Feb 15;128(4):654-664. [doi: 10.1002/cncr.34023] [Medline: 34787913]
6. Rathmell WK. Using the national cancer plan to drive innovation in cancer research. Cancer Discov 2024 Apr 4;14(4):555-558. [doi: 10.1158/2159-8290.CD-24-0223] [Medline: 38571429]
7. Moody L, Nicholls B, Shamji H, Bridge E, Dhanju S, Singh S. The person-centred care guideline: from principle to practice. J Patient Exp 2018 Dec;5(4):282-288. [doi: 10.1177/2374373518765792] [Medline: 30574549]
8. Fagan P, Iongh A, Harden B, Wright C. Person-centred approaches. Health Education England. 2020 Nov. URL: https://www.skillsforhealth.org.uk/info-hub/person-centred-approaches-2017/ [accessed 2025-02-19]
9. Person-centered care key concepts. Centers for Medicare & Medicaid Services. 2024. URL: https://www.cms.gov/priorities/innovation/key-concepts/person-centered-care [accessed 2025-02-19]
10. NEJM Catalyst. What is patient-centered care? NEJM Catal 2017 Jan 1 [FREE Full text] [doi: 10.1056/cat.17.0559]
11. Balitsky AK, Rayner D, Britto J, et al. Patient-reported outcome measures in cancer care: an updated systematic review and meta-analysis. JAMA Netw Open 2024 Aug 1;7(8):e2424793. [doi: 10.1001/jamanetworkopen.2024.24793] [Medline: 39136947]
12. Pozzar RA, Xiong N, Hong F, et al. Perceived patient-centered communication, quality of life, and symptom burden in individuals with ovarian cancer. Gynecol Oncol 2021 Nov;163(2):408-418. [doi: 10.1016/j.ygyno.2021.08.007] [Medline: 34454724]

13. Pozzar RA, Xiong N, Hong F, et al. How does patient-centered communication in ovarian cancer care enhance patient well-being? A mixed methods study. Gynecol Oncol 2022 Nov;167(2):261-268. [doi: 10.1016/j.ygyno.2022.08.019] [Medline: 36057465]
14. Schulman-Green D, Lin JJ, Smith CB, Feder S, Bickell NA. Facilitators and barriers to oncologists' conduct of goals of care conversations. J Palliat Care 2018 Jul;33(3):143-148. [doi: 10.1177/0825859718777361] [Medline: 29790421]
15. Prouty CD, Mazor KM, Greene SM, et al. Providers' perceptions of communication breakdowns in cancer care. J Gen Intern Med 2014 Aug;29(8):1122-1130. [doi: 10.1007/s11606-014-2769-1] [Medline: 24599795]
16. Fagerlind H, Kettis Å, Glimelius B, Ring L. Barriers against psychosocial communication: oncologists' perceptions. J Clin Oncol 2013 Oct 20;31(30):3815-3822. [doi: 10.1200/JCO.2012.45.1609] [Medline: 24043746]
17. Pozzar RA, Tulsky JA, Berry DL, et al. Developing a Collaborative Agenda-Setting Intervention (CASI) to promote patient-centered communication in ovarian cancer care: a design thinking approach. Patient Educ Couns 2024 Mar;120:108099. [doi: 10.1016/j.pec.2023.108099] [Medline: 38086227]
18. Gilligan T, Coyle N, Frankel RM, et al. Patient-clinician communication: American society of clinical oncology consensus guideline. JCO 2017 Nov 1;35(31):3618-3632. [doi: 10.1200/JCO.2017.75.2311]
19. Stiefel F, Bourquin C, Salmon P, et al. Communication and support of patients and caregivers in chronic cancer care: ESMO Clinical Practice Guideline. ESMO Open 2024 Jul;9(7):103496. [doi: 10.1016/j.esmoop.2024.103496] [Medline: 39089769]
20. Wolff JL, Roter DL, Boyd CM, et al. Patient-family agenda setting for primary care patients with cognitive impairment: the SAME page trial. J GEN INTERN MED 2018 Sep;33(9):1478-1486. [doi: 10.1007/s11606-018-4563-y] [Medline: 30022409]
21. Hood-Medland EA, White AEC, Kravitz RL, Henry SG. Agenda setting and visit openings in primary care visits involving patients taking opioids for chronic pain. BMC Fam Pract 2021 Jan 4;22(1):4. [doi: 10.1186/s12875-020-01317-4] [Medline: 33397299]
22. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4(1):50. [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
23. Chishtie J, Sapiro N, Wiebe N, et al. Use of Epic electronic health record system for health care research: scoping review. J Med Internet Res 2023 Dec 15;25:e51003. [doi: 10.2196/51003] [Medline: 38100185]
24. Degner LF, Sloan JA, Venkatesh P. The control preferences scale. Can J Nurs Res 1997;29(3):21-43. [Medline: 9505581]
25. Wright AA, Thompson E, Nguyen MV, et al. The SHARE study: pilot study of a communication intervention designed to elicit advanced-stage cancer patients' preferences and goals. J Palliat Med 2022 Nov;25(11):1646-1654. [doi: 10.1089/jpm.2021.0655] [Medline: 35736783]
26. Donovan KA, Handzo G, Corbett C, Vanderlan J, Brewer BW, Ahmed K. NCCN distress thermometer problem list update. J Natl Compr Canc Netw 2022 Jan;20(1):96-98. [doi: 10.6004/jnccn.2022.0004] [Medline: 34991072]
27. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
28. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
29. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry: CRC Press; 1995:189-194.
30. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
31. Weiner BJ, Lewis CC, Stanick C, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017 Aug 29;12(1):108. [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
32. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
33. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21. [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
34. Sisk BA, Schulz GL, Mack JW, Yaeger L, DuBois J. Communication interventions in adult and pediatric oncology: a scoping review and analysis of behavioral targets. PLoS ONE 2019;14(8):e0221536. [doi: 10.1371/journal.pone.0221536] [Medline: 31437262]
35. Tulsky JA, Beach MC, Butow PN, et al. A research agenda for communication between health care professionals and patients living with serious illness. JAMA Intern Med 2017 Sep 1;177(9):1361-1366. [doi: 10.1001/jamainternmed.2017.2005] [Medline: 28672373]
36. Baik SH, Clark K, Sanchez M, et al. Usability and preliminary efficacy of an adaptive supportive care system for patients with cancer: pilot randomized controlled trial. JMIR Cancer 2024 Jul 10;10:e49703. [doi: 10.2196/49703] [Medline: 38986134]

37. Lyleroehr MJ, Webster KA, Perry LM, et al. A mixed methods evaluation of patient perspectives on the implementation of an electronic health record-integrated patient-reported symptom and needs monitoring program in cancer care. J Patient Rep Outcomes 2024 Jul 2;8(1):66. [doi: 10.1186/s41687-024-00742-8] [Medline: 38954112]
38. Desai AV, Michael CL, Kuperman GJ, et al. A novel patient values tab for the electronic health record: a user-centered design approach. J Med Internet Res 2021 Feb 17;23(2):e21615. [doi: 10.2196/21615] [Medline: 33595448]
39. Wickline M, Wolpin S, Cho S, et al. Usability and acceptability of the electronic self-assessment and care (eSAC) program in advanced ovarian cancer: a mixed methods study. Gynecol Oncol 2022 Nov;167(2):239-246. [doi: 10.1016/j.ygyno.2022.09.010] [Medline: 36150917]

Abbreviations

AIM: Acceptability of Intervention Measure
CASI: Collaborative Agenda-Setting Intervention
CFIR: Consolidated Framework for Implementation Research
EHR: electronic health record
IAM: Intervention Appropriateness Measure
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
SUS: System Usability Scale

Edited by N Cahill; submitted 27.09.24; peer-reviewed by B Senst, K Reuter; revised version received 16.01.25; accepted 16.01.25; published 10.03.25.

Please cite as:
Pozzar RA, Tulsky JA, Berry DL, Batista J, Barwick P, Lindvall CJ, Dykes PC, Manni M, Matulonis UA, McCleary NJ, Wright AA
Usability, Acceptability, and Barriers to Implementation of a Collaborative Agenda-Setting Intervention (CASI) to Promote Person-Centered Ovarian Cancer Care: Development Study
JMIR Cancer 2025;11:e66801
URL: https://cancer.jmir.org/2025/1/e66801
doi:10.2196/66801

A Digital Home-Based Health Care Center for Remote Monitoring of Side Effects During Breast Cancer Therapy: Prospective, Single-Arm, Monocentric Feasibility Study

Hanna Huebner[1,2,3], PhD; Lena A Wurmthaler[1,2,3], PhD; Chloë Goossens[1,2,3], PhD; Mathias Ernst[1,2,3], PhD; Alexander Mocker[1,2,3], MD, Dr Med; Annika Krückel[1,2,3], MD, Dr Med; Maximilian Kallert[1,2,3], MS; Jürgen Geck[4], Dipl Ing; Milena Limpert[4], MS; Katharina Seitz[1,2,3], MD, Dr Med; Matthias Ruebner[1,2,3], PhD; Philipp Kreis[1,2,3], MD; Felix Heindl[1,2,3], MD, Dr Med; Manuel Hörner[1,2,3], MD, Dr Med; Bernhard Volz[5], Prof Dr; Eduard Roth[1,2,3], BE; Carolin C Hack[1,2,3], MD, Prof Dr; Matthias W Beckmann[1,2,3], MD, Prof Dr; Sabrina Uhrig[1,2,3*], PhD; Peter A Fasching[1,2,3*], MD, Prof Dr

[1]Department of Gynecology and Obstetrics, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Universitätsstraße 21/23, Erlangen, Germany

[2]Bavarian Cancer Research Center, Erlangen, Germany

[3]Comprehensive Cancer Center Erlangen-EMN, Erlangen, Germany

[4]REFINIO GmbH, Rohr, Germany

[5]Nuremberg Institute of Technology Georg Simon Ohm, Nuremberg, Germany

[*]these authors contributed equally

Corresponding Author:
Peter A Fasching, MD, Prof Dr
Department of Gynecology and Obstetrics, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Universitätsstraße 21/23, Erlangen, Germany

Abstract

Background: The introduction of oral anticancer therapies has, at least partially, shifted treatment from clinician-supervised hospital care to patient-managed home regimens. However, patients with breast cancer receiving oral cyclin-dependent kinase 4/6 inhibitor therapy still require regular hospital visits to monitor side effects. Telemonitoring has the potential to reduce hospital visits while maintaining quality care.

Objective: This study aims to develop a digital home-based health care center (DHHC) for acquiring electrocardiograms (ECGs), white blood cell (WBC) counts, side effect photo documentation, and patient-reported quality of life (QoL) data.

Methods: The DHHC was set up using an Apple Watch Series 6 (ECG measurements), a HemoCue WBC DIFF Analyzer (WBC counts), an iPhone SE (QoL assessments and photo documentation), a TP-Link M7350-4G Wi-Fi router, and a Raspberry Pi 4 Model B. A custom-built app stored and synchronized remotely collected data with the clinic. The feasibility and acceptance of the DHHC among patients with breast cancer undergoing cyclin-dependent kinase 4/6 inhibitor therapy were evaluated in a prospective, single-arm, monocentric study. Patients (n=76) monitored side effects—ECGs, WBC counts, photo documentation, and QoL—at 3 predefined time points: study inclusion (on-site), day 14 (remote), and day 28 (remote). After the study completion, patients completed a comprehensive questionnaire on user perception and feasibility. Adherence to scheduled visits, the success rate of the data transfer, user perception and feasibility, and the clinical relevance of remote measurements were evaluated.

Results: Mean adherence to the planned remote visits was 63% on day 14 and 37% on day 28. ECG measurements were performed most frequently (day 14: 57/76, 75%; day 28: 31/76, 41%). The primary patient-reported reason for nonadherence was device malfunction. The expected versus the received data transfer per patient was as follows: ECGs: 3 versus 3.04 (SD 1.9); WBC counts: 3 versus 2.14 (SD 1.14); QoL questionnaires: 3 versus 2.5 (SD 1.14); and photo documentation: 6 versus 4.4 (SD 3.36). Among patients, 81% (55/68) found ECG measurements easy, 82% (55/67) found photo documentation easy, and 48% (33/69) found WBC measurements easy. Additionally, 61% (40/66) of patients felt comfortable with self-monitoring and 79% (54/68) were willing to integrate remote monitoring into their future cancer care. Therapy-induced decreased neutrophil count was successfully detected (P<.001; mean baseline: 4.3, SD 2.2, $\times 10^9$/L; on-treatment: 1.8, SD 0.8, $\times 10^9$/L). All-grade neutropenia and corrected QT interval prolongations were detected in 80% (55/68) and 2% (1/42) of patients, respectively.

Conclusions: Adherence to scheduled remote visits was moderate, with nonadherence primarily attributed to device-related complications, which may have also affected the success rate of data transfer. Overall, patients considered remote monitoring

useful and feasible. The prevalence of reported adverse events was comparable to existing literature, suggesting clinical potential. This initial feasibility study highlights the potential of the DHHC.

(JMIR Cancer 2025;11:e64083) doi:10.2196/64083

KEYWORDS

breast cancer; digital medicine; telehealth; remote monitoring; cyclin-dependent kinase 4/6 inhibitor; CDK4/6 inhibitor; mobile phone

Introduction

Systemic therapies, such as chemotherapy, targeted therapies, or immunotherapy, are accompanied by several side effects that require continuous and regular monitoring. Monitoring of side effects is particularly important for treatment approaches involving oral medications, as these medications are usually administered at home. Side effect monitoring enables the early detection and prevention of adverse events and is crucial for treatment benefits and adherence to the treatment schedule [1].

In recent years, highly effective oral therapeutic options, such as cyclin-dependent kinase 4 and 6 inhibitors (CDK4/6i), have been introduced for the treatment of patients with hormone receptor-positive breast cancer. Even though these CDK4/6is are administered orally, the associated side effects, such as neutropenia, leukopenia, and corrected QT interval (QTc) prolongation, necessitate regular hospital appointments [2-10]. For patients in rural areas, such appointments present unique challenges. Due to a lack of nearby medical facilities, patients often have to travel long distances to receive adequate medical care, which can be physically and emotionally taxing. If these patients do not receive comprehensive cancer care, including side effect monitoring, they may experience delayed detection and treatment of serious adverse events, potentially affecting their quality of life (QoL), and survival outcomes [11-15]. Remote monitoring and eHealth options may be particularly valuable in addressing these challenges [16].

Remote, home-based monitoring using eHealth options such as apps, wearables, or mobile medical devices can allow health care providers to monitor patients' health status and potential side effects in real time. Multiple studies have shown that remote monitoring of cancer treatment symptoms is linked to improved QoL, fewer treatment disruptions, and increased survival rates [13,17-19]. However, most of these studies included only remote patient-reported outcome assessments. Recently, home- and sensor-based technologies, including various wearable devices, have also been shown to be suitable tools for cancer care. For example, smartwatches and fitness trackers have been used to promote physical activity and monitor heart rate [20-23].

Even though remote monitoring systems, eHealth apps, and wearable devices have the potential to improve cancer care, several challenges still need to be addressed. In particular, while current smartwatch technologies can monitor heart rate and record Food and Drug Administration–approved electrocardiograms (ECGs), and several apps on the market can be used to document patient-reported outcomes, assess QoL, and track side effects, using these individual tools alone is not sufficient to provide comprehensive medical care at home. Therefore, we aimed to establish a digital home-based health care center (DHHC) that includes a smartphone to assess QoL and document visual side effects, a smartwatch to record ECGs, and a white blood cell (WBC) system to analyze a patient's WBCs from capillary blood. The primary focus of this study was to assess the feasibility and acceptance of such a digital remote system for cancer care in order to tailor a patient-centered solution and improve access to quality care.

Methods

Study Design

The SMILER study ("Smart and Interactive Home-Based Health Care Project—A Digital Healthcare Feasibility Pilot Study Including the d.H2C2 Initiative") was a monocentric, single-arm study with the primary objective of assessing the feasibility of remote WBC and ECG measurements, as well as data transmission of remote measurements using a DHHC. The study was conducted at the Department of Gynecology and Obstetrics at the University Hospital Erlangen (Universitätsklinikum Erlangen) in Germany.

Inclusion criteria were an indication for or current treatment with a CDK4/6i (regardless of cycle number) and an age of 22 years or older (in accordance with the minimum age requirements for the use of DHHC devices as specified by their respective manufacturers). Patients could not be included if they had pacemakers or implantable cardioverter defibrillators, severe blood coagulation disorders, abnormalities in the last known ECG, or other comorbidities that might impact at-home measurements. The study was conducted between October 2021 and December 2022.

CDK4/6i therapy was administered according to the Summary of Product Characteristics. All patients who had an indication for CDK4/6i therapy, as determined by the treating physician, or who were already receiving CDK4/6i therapy, were screened for the SMILER study. In general, palbociclib was started at 125 mg/day, and if necessary, reduced to 100 or 75 mg/day. Ribociclib was started at 600 mg/day, and if required, reduced to 400 mg/day and subsequently to 200 mg/day. Abemaciclib was initiated at 300 mg/day, with potential dose reductions to 200 and 100 mg/day.

After study inclusion, participants received the DHHC along with an initial introductory training. Study-relevant measurements were scheduled at study inclusion (on-site), day 14 (d14—remote), and day 28 (d28—remote; Figure 1A). At each of these time points, WBC counts, ECGs, and QoL (Q-5D-3L questionnaire) were monitored. Additionally, photo documentation of the ankle (as an exploratory subproject for the capturing of 3D photo data) was included. The ankle was

chosen as an accessible location for photo acquisition where peripheral edema, a known side effect of CDK4/6i therapy, could be detected.

All participants continued their routine treatment and attended scheduled clinical visits. After the study had been completed, participants filled out a paper questionnaire on the acceptance, success rate, and usability of the DHHC. The SMILER study concluded after the predefined number of patients had been enrolled.

Figure 1. SMILER study ("Smart and Interactive Home-Based Health Care Project—A Digital Healthcare Feasibility Pilot Study Including the d.H2C2 Initiative") design and technology setup. (A) The SMILER study included an initial training session, followed by 2 scheduled at-home tasks on day 14 (d14) and day 28 (d28). (B) Patients received a large case with (C) integrated charging for all devices, (D) specifically designed for at-home use, (E,F) or a smaller and lighter case with foam material to securely hold all devices. (G) The associated SMILER.one app featured a home screen displaying the trial tasks, (H) a data archive for storing all collected data, and (I) functionality for collecting and visualizing specific parameters such as WBC counts. ECG: electrocardiogram; QoL: quality of life; WBC: white blood cell.

Ethical Considerations

The study was conducted in accordance with local guidelines and regulations. Ethical approval was obtained from the Ethics Committee of the Friedrich-Alexander Universität Erlangen-Nürnberg (April 1, 2020: 47_20B). The original protocol was amended on March 22, 2022, to also include patients already receiving CDK4/6i therapy, as previously, patients could only be enrolled in the SMILER study at the start of CDK4/6i therapy. This change was implemented to improve study enrollment. All participants provided written informed consent before participation. Participants did not receive any form of compensation. Data were collected in a pseudonymized manner.

Outcomes

The outcomes of the study were: adherence (primary objective), success rate of the data transfer, usability and feasibility of the DHHC, and clinical relevance. Adherence to scheduled study visits was assessed as the percentage of patients who completed the prescheduled measurements (±2 d around the scheduled visit). The success rate was evaluated as the number of remotely transferred measurements relative to the number of expected measurements. Feasibility was assessed based on the number of enrolled versus screened patients. Furthermore, patient-reported perception and usability were evaluated using a comprehensive paper-based questionnaire at study completion, which included Likert-scale questions on perceived usability. Usability was further assessed using the System Usability Scale (SUS), with the SUS score calculated as the respective outcome measure [24]. Clinical relevance was determined by the number of detected adverse events, specifically neutropenia and QTc prolongations.

DHHC Hardware

The DHHC consisted of the following components: (1) Apple Watch Series 6 (Apple Inc., Cupertino, CA, United States) for ECG measurements, (2) HemoCue WBC DIFF Analyzer (HemoCue AB, Ängelholm, Sweden) for WBC counts, (3) iPhone SE (Apple Inc.) for QoL questionnaire completion and photo documentation, (4) mobile Wi-Fi router TP-Link M7350-4G (TP-Link Corporation Limited, Düsseldorf, Germany), and (5) Raspberry Pi 4 Model B (Raspberry Pi, Cambridge, United Kingdom).

Two cases were designed to enable safe and easy transport and handling of the devices (Figure 1B-F). Case 01 (Figure 1B-D; Fa. Karl Lettenbauer, Erlangen, Germany) featured a plastic base plate to accommodate the HemoCue WBC DIFF Analyzer, Raspberry Pi, power cable, and socket strip (installed under the upper mount with a 14412 - 02 detachable partition protected against tampering), along with all device cables. The iPhone, Apple Watch, and TP-Link were integrated into a raised platform. Case 02 (Figure 1E-F) consisted of a case from MyCaseBuilder.eu (FOAM Studio, the Netherlands) with a custom Pro-Cell interior and Prolife Soft-Cell foam lid. For both cases, the devices could be charged, and WBC measurements could be taken without the devices being removed from the cases.

SMILER.one App and DHHC Software

The custom study app (SMILER.one) was developed by REFINIO GmbH, based on their REFINIO ONE architecture. REFINIO GmbH is a German company specializing in custom software for secure data collection. The software was programmed in TypeScript on NodeJS and had platform abstractions for internet browsers and Linux. All ONE instances of a person formed a federation called the Internet of Me (IoM), where identities, connections, settings, and content could be distributed, ensuring that devices only needed to be registered once.

Data storage in ONE was based on HTML files containing microdata objects, which were stored in individual files within the file system or in the IndexedDB of the browser or WKWebView on mobile devices. The objects were named according to the hash of their content and referenced through their name in parent objects. Data transmission in ONE was facilitated through WebSocket services provided by a commServer, which established connections between devices. Data sharing was based on subtree sharing and conflict-free replicated data types, with encryption occurring at the individual instance level.

The DHHC integrated several REFINIO ONE software components, including the Web Server, SMILER.one Mobile, SMILER.one Pi, SMILER.one Headless, and the SMILER.one representational state transfer application programming interface (REST API). The Web Server was installed and configured with HTTPD software (nginx) to deliver the SMILER.one Progressive Web App and mobile content. SMILER.one Mobile managed patient data in a WebView (IndexedDB) on an iPhone and synchronized it with the clinic's data (SMILER.one Headless) and the patients' IoM instances. It also imported and synchronized ECG data from the Apple HealthKit on patients' iPhones. SMILER.one Pi was installed on the Raspberry Pi, importing WBC data from the HemoCue device. Within the patients' IoM, it acted as a headless replication of the patients' complete dataset. SMILER.one Headless mirrored all settings and storage operations of SMILER.one Mobile and SMILER.one Pi while incorporating the SMILER.one REST API, which provided pseudonymized patient data to the clinic's SQL server (Figure S1 in Multimedia Appendix 1).

The SMILER.one app served as a user interface for study participants. The app featured a registration or login mechanism with password encryption to ensure restricted private access. Within the app, the "My Tasks" screen allowed participants to complete visits (questionnaires, photo documentation, ECG measurements, and WBC counts) and provided an overview of upcoming tasks (Figure 1G). In the "Data Archive," completed data were stored and could be viewed by the participants (Figure 1H). The "Blood Count Chart" displayed a chart of WBC readings from the HemoCue device (Figure 1I). Patients also had the option to enter additional data beyond the scheduled remote visits under "Voluntary Data Entry."

Data Management

Data from the DHHC was stored in a dedicated relational database on an SQL server. The data transfer from the REFINIO

REST API to the SQL database was facilitated via the JSON data format. Parsing of the JSON-formatted data and transformation into a relational tabular format were performed within the database itself. QTc times were calculated from the transmitted ECG curves using the Fridericia formula by a physician (PK). Based on the QTc times and measured neutrophil concentrations, the severity of QTc prolongations and neutropenia was graded according to the Common Terminology Criteria for Adverse Events version 5.0. Clinical data were collected by trained staff and documented in an electronic case report form. Data monitoring was conducted using automated plausibility checks and on-site monitoring. These data included patient and tumor characteristics, as well as details on treatment approaches.

Statistical Analysis

Sample size calculations for this feasibility study were based on the assumption that 70% of ECG measurements and 65% of WBC counts would be successful (two 1-sided exact binomial tests with a significance level of α=2.5%). A total of 212 ECG measurements and 237 blood measurements were required to demonstrate this with a power of 90%, which corresponded to 80 patients, each with three ECG measurements and three WBC count measurements.

The majority of the presented statistics are descriptive. Categorical variables are reported as counts and percentages, while continuous variables are presented as mean (SD). Missing values were omitted from analyses. A 2-sided Wilcoxon rank sum test was used for statistical comparisons, with $P \leq .05$ considered statistically significant. Data are presented as box plots, displaying the median, IQRs, and whiskers representing the 5th and 95th percentiles. Statistical analyses were performed using R (version 4.2.1; The R Foundation) or SPSS Statistics (version 29.0.1.0; IBM Corp). Likert plots were generated using the R library “Likert” (version 1.3.5). The distribution of Likert scale questionnaire scores ranged from 1=I fully agree to 5=I do not agree at all.

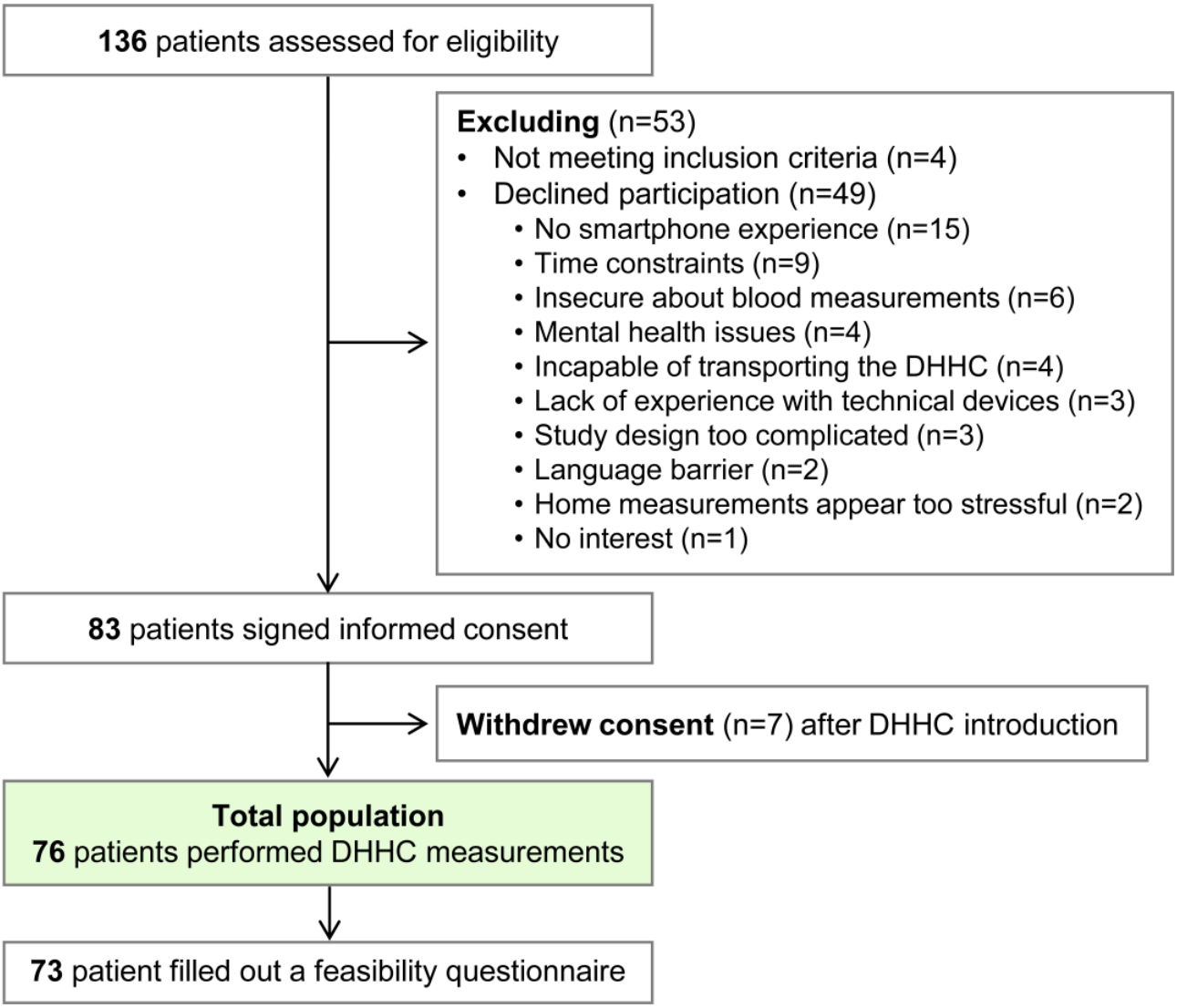
Results

Feasibility

Recruitment

Between October 2021 and December 2022, 136 patients with breast cancer were screened for eligibility. Of the 132 eligible patients, 49 patients declined participation. The most common reasons for nonparticipation were lack of smartphone experience (n=15) and lack of time to complete study procedures (n=9; Figure 2). Of the 83 patients who provided informed written consent, 7 patients withdrew their consent after being introduced to the DHHC (Figure 2), resulting in 76 patients who participated in the SMILER study. Of these 76 patients, 73 (96%) completed the questionnaire on acceptance, success rate, and feasibility of the home-based procedures at study completion.

Figure 2. Patient flowchart. This flowchart illustrates the progression of patients throughout the study, including screening, enrollment, and final study participants. DHHC: digital home health care center.



Patient Characteristics

Participants had a mean age of 58.9 (SD 9.74) years (Table 1). The majority of participants received CDK4/6i for advanced or metastatic disease (54/76, 71%). Abemaciclib was administered to 49% (37/76) of participants, while 39% (30/76) and 12% (9/76) of participants received ribociclib and palbociclib, respectively. Additional baseline patient characteristics are listed in Table 1. Case 01 was assigned to 33% (25/76) of participants, while Case 02 was provided to 67% (51/76) of participants.

Table . Baseline characteristics of study participants.

Characteristics	Study participants (n=76)
Age (years), mean (SD)	58.9 (9.74)
BMI[a] (kg/m^2), mean (SD)	25.7 (4.65)
ECOG[b] index[c], n (%)	
0	68 (91)
1	6 (8)
2	0 (0)
3	1 (1)
Grading[d], n (%)	
G1	7 (10)
G2	40 (54)
G3	27 (36)
ER[e], n (%)	
ER+	76 (100)
ER–	0 (0)
PR[f], n (%)	
PR+	62 (82)
PR–	14 (18)
HER2/neu status, n (%)	
HER2+	6 (8)
HER2–	70 (92)
CDK4/6[h] inhibitor, n (%)	
Abemaciclib	37 (49)
Palbociclib	9 (12)
Ribociclib	30 (39)
Endocrine therapy combination partner, n (%)	
Aromatase inhibitor	54 (71)
Fulvestrant	16 (21)
Other	2 (3)
Metastasis, n (%)	
M0	22 (29)
M1	54 (71)
Line of therapy, n (%)	
1st line	42 (78)
2nd line	7 (13)
3rd line or higher	5 (9)
Highest degree of education[g], n (%)	
No degree	0 (0)
General secondary school	12 (17)
Intermediate secondary school	14 (20)
University of Applied Science entrance certificate	5 (7)

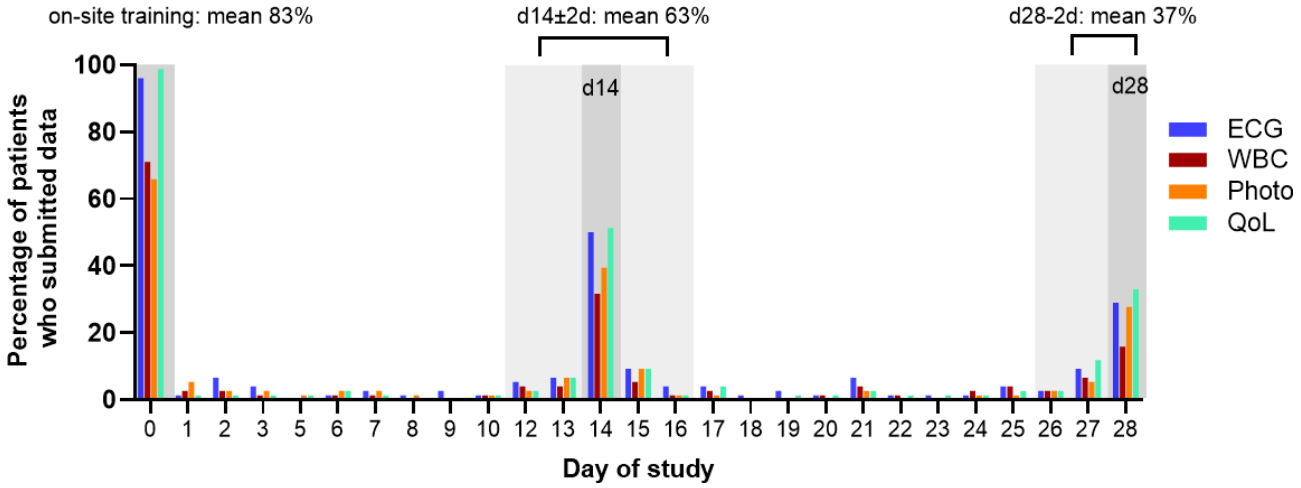
Characteristics	Study participants (n=76)
University entrance certificate	4 (5)
Vocational training	22 (31)
Bachelor's degree	0 (0)
Master's degree or higher	14(20)
SMILER case, n (%)	
Case 01	25 (33)
Case 02	51 (67)

[a]Missing: n=5.
[b]ECOG: Eastern Cooperative Oncology Group.
[c]Missing: n=1.
[d]Missing: n=2.
[e]ER: estrogen receptor.
[f]PR: progesterone receptor.
[g]Missing: n=5.
[h]CDK4/6: cyclin-dependent kinase 4 and 6.

Adherence

Study-relevant measurements were scheduled at study inclusion (introductory on-site training), day 14 (d14—home-based), and day 28 (d28—home-based). The average adherence was 63% at d14±2d and 37% at d28-2d (Figure 3). Among individual remote measurements, ECGs were recorded most frequently (d14±2d: 57/76, 75%; d28-2d: 31/76, 41%), followed by QoL questionnaires (d14±2d: 54/76, 71%; d28-2d: 36/76, 47%).

Figure 3. Patients who submitted data during the study. The percentage of patients who submitted data via the digital home-based health care center is presented for various measurements, including ECG, WBC counts, photo documentation, and QoL questionnaires. The bars indicate the data submission rates over the study duration, with dark gray sections representing planned study visits (day 14, d14; and day 28, d28) and light gray sections representing the days around the study visits (d14±2, d28-2). The mean percentage represents the average data submission rate for each study visit. ECG: electrocardiogram; QoL: quality of life; WBC: white blood cell.



The most common patient-reported reasons for missing scheduled tasks were device malfunction (ECG or Apple Watch: 11/36, 31%; WBC or HemoCue: 16/43, 37%; photo documentation: 5/31, 16%), handling issues (ECG or Apple Watch: 11/36, 31%; WBC or HemoCue: 12/43, 28%; photo documentation: 2/31, 6%), and time constraints (ECG or Apple Watch: 10/36, 28%; WBC or HemoCue: 10/43, 23%; photo documentation: 11/31, 35%; Table 2). Among the patient-reported reasons listed as "other reasons" for failed adherence were "measurements were not performed in time," "unsure about the handling of the devices," "no reception," "mentally too stressed," or "sickness." Among these, "unsure about the handling of the device" was the most common other reason for missed WBC measurements (5/13, 38%).

Table . Patient-reported reasons for not being able to perform scheduled home-based measurements (d14 and d28le)[a].

	ECG[b] (n=36), n (%)	WBC[c] (n=43), n (%)	Photo documentation (n=31), n (%)
Device would not function	11 (31)	16 (37)	5 (16)
Measurement could not be performed properly	11 (31)	12 (28)	2 (6)
Battery was empty	3 (8)	N/A[d]	3 (10)
Safety concerns	N/A	2 (5)	1 (3)
No time	10 (28)	10 (23)	11 (35)
No interest	N/A	N/A	1 (3)
Other reasons	15 (42)	13 (30)	12 (39)

[a]Patients could indicate their reason for nonadherence to the questionnaire upon completion of the study. Participants could choose from predefined reasons, with multiple answers possible. Responses were only requested from study participants who indicated that they were unable to perform all scheduled measurements. When selecting "other reasons," patients could provide additional details in a blank text field.

[b]ECG: electrocardiography.

[c]WBC: white blood cell count.

[d]N/A: not applicable.

Success Rates

Throughout the 28-day study period, data transfer was expected from three ECG measurements, three WBC measurements, three QoL questionnaires, and six photo documentations (each ankle per time point) per patient. The mean number of successfully transferred data per patient during the study period was 3.04 (SD 1.9), ECGs was 2.14 (SD 1.1), WBC measurements was 2.5 (SD 1.1), QoL questionnaires and photo documentations was 4.4 (SD 3.4; Table S1 in Multimedia Appendix 1).

Study participants indicated how many times a measurement had to be repeated before it was successfully completed in the end-of-study questionnaire. The highest number of repetitions was required for WBC measurements, with an average of 1.07 (SD 1.0) additional measurements per patient (Table S1 in Multimedia Appendix 1).

Identified Problems

Several issues may have affected both study adherence and success rates. The incidence of any type of DHHC malfunction (ie, problems with either WBC or ECG measurements) was 24% (18/76) during first use (initial introductory training) and 43% (33/76) during remote measurements (as reported in the end-of-study questionnaire).

For WBC measurements, handling problems with the HemoCue microcuvettes requiring repeat WBC measurements were observed in 8 (out of 76, 10%) patients during on-site training. Of these, 2 (25%) patients also reported handling issues with WBC measurements during remote monitoring. Patients who experienced handling problems during initial training (8/76) and those who reported handling issues as a reason for nonadherence to remote monitoring visits (12/76) appeared to be older than those without handling issues (initial training: handling issues 59.4, SD 9.6 y vs no handling issues 55.0, SD 10.3 y; *P*=.30; remote monitoring: handling issues 63.3, SD 7.6 y vs no handling issues 58.1, SD 9.9 y; *P*=.02).

At initial training, technical problems with the HemoCue WBC DIFF IEC 61010 system (failure to turn on or instant error messaging) prevented measurements in 4 (out of 76, 5%) patients and data transfer from the HemoCue WBC DIFF IEC 61010 system to the SMILER.one app and SQL server in 11 (out of 76, 14%) patients. In two of these cases (2/11, 18%), a lack of an internet connection was reported.

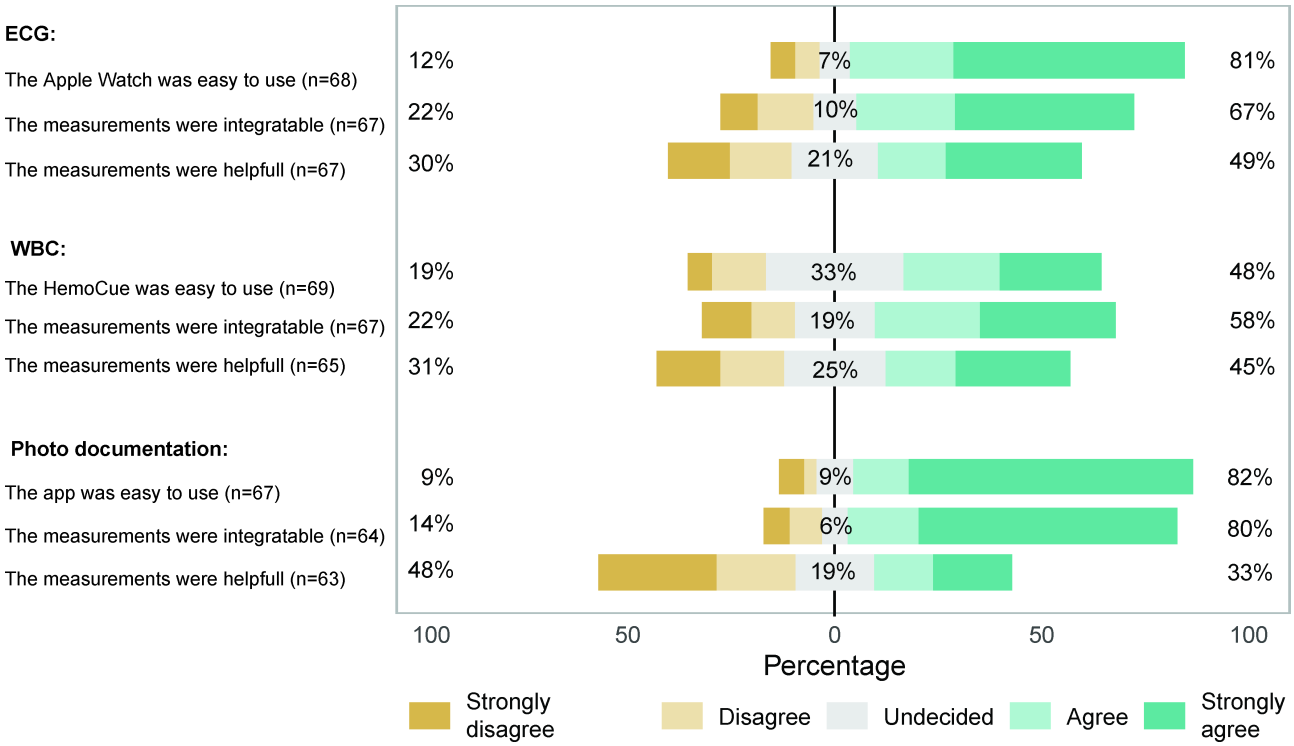
Defective data transfer occurred more often with Case 02 than Case 01 (8/11, 73% vs 3/11, 27%) and in two instances, the same Case 02 DHHC experienced defective data transfer. However, the case type was not statistically associated with defective data transfer (*P*=.80). Among 3 (out of 11, 27%) DHHCs with defective data transfer during on-site training, study participants also reported an inability to perform remote measurements due to device malfunction.

For ECG measurements, defective data transfer was observed in 3 (out of 76, 4%) individual DHHCs during on-site training, and 1 (out of 3, 33%) patient also reported subsequent technical issues with the Apple Watch during remote monitoring. Additionally, 4% (3/76) of patients reported being unable to use the SMILER.one app due to a lack of mobile network coverage.

User Perception of Feasibility

The feasibility of home measurements was assessed at study completion. While most patients found photo documentation with the SMILER.one app (55/67, 82%) and ECG measurements with the Apple Watch (55/68, 81%) easy to use, only 48% (33/69) of patients found the HemoCue System easy to use. Good integration into daily life was most commonly reported for photo documentation (51/64, 80%), followed by ECG (45/67, 67%) and WBC measurements (39/67, 58%). ECG and WBC measurements were seen as helpful by 49% (33/67) and 45% (29/65) of patients, respectively, while only 33% (21/63) of participants found photo documentation helpful (Figure 4).

Figure 4. User perception of the home-based health care system's feasibility. The Likert scale illustrates responses from a user perception questionnaire regarding the Apple Watch ECG, HemoCue WBC measurement, and photo documentation via the SMILER.one app. ECG: electrocardiogram; WBC: white blood cell.

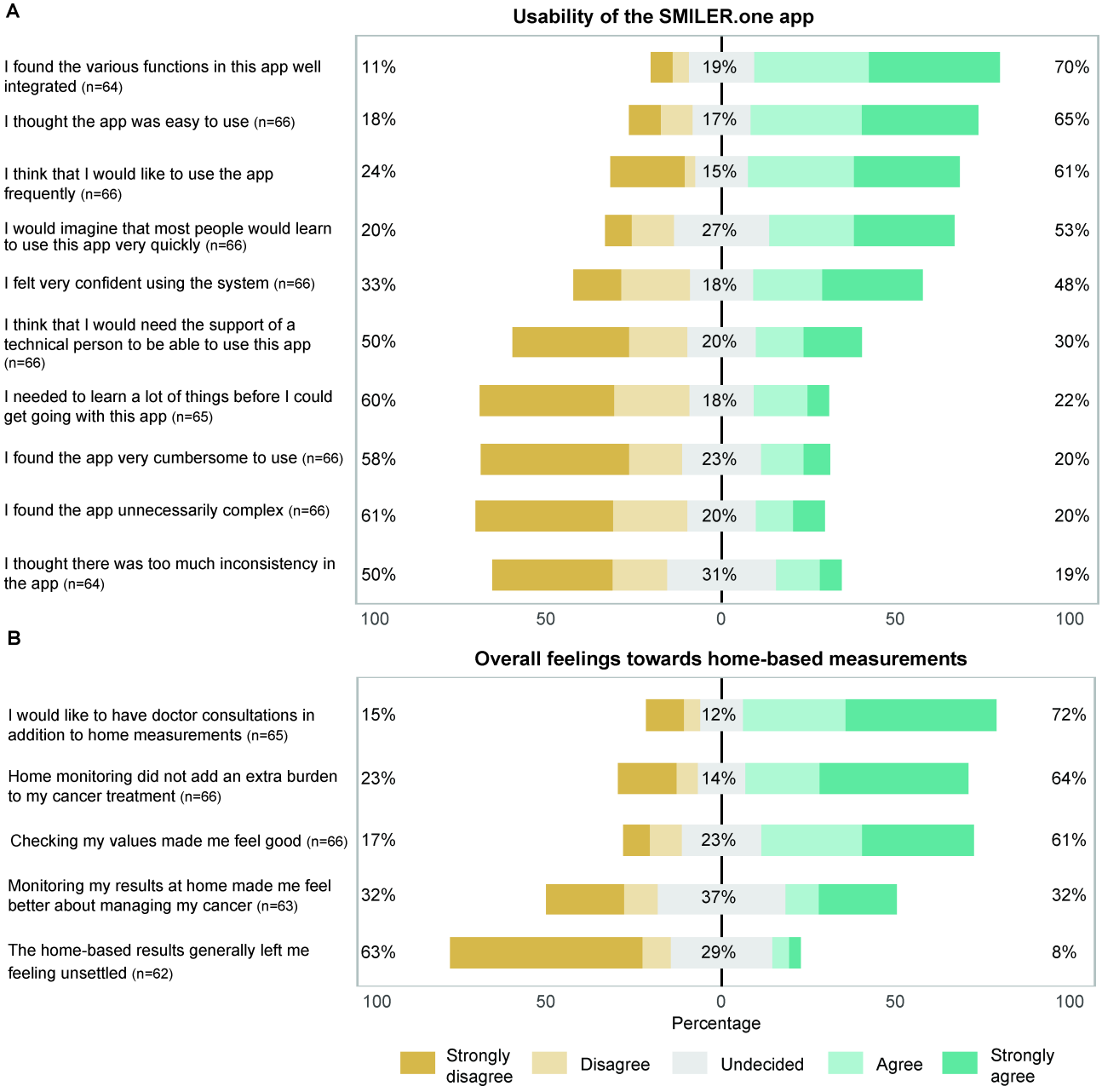


Acceptance

User acceptance of the SMILER.one app and the perception of the DHHC and its future use were assessed through the end-of-study questionnaire. For the SMILER.one app, 70% (45/64) of participants agreed that the technical features of the app were well-integrated, 65% (43/66) of participants found the app easy to use, and 61% (40/66) of participants could imagine using the app regularly. However, 30% (20/66) of participants reported that they would need technical support to use the app (Figure 5A). The corresponding SUS score for the SMILER.one app was 65.2. Regarding the DHHC, 61% (40/66) of participants felt comfortable with self-monitoring, and 64% (42/66) of participants did not consider home-based measurements to be an additional burden. However, 72% (47/65) of participants stated that they would like to have consultations with a doctor in addition to the DHHC measurements (Figure 5B).

When asked about their intention to use, 79% (54/68) of participants expressed a willingness to use home measurements as part of their future cancer care. Additionally, 79% (54/68) of participants were willing to collect data for research purposes at home using the SMILER.one app. Participants who were unwilling or reluctant to integrate the DHHC and remote data collection into future cancer care reported that they would need help interpreting results (9/17, 53%), were concerned about the time required for data collection (11/23, 48%), and would experience difficulties or stress related to the technology (home-based measurements: 3/17, 18%; collection and sharing of data with the SMILER.one app: 7/23, 30%; Table S2 in Multimedia Appendix 1).

Figure 5. User perception and acceptance of the SMILER.one app and the home-based health care system. (A) The Likert scale illustrates responses from a user perception and acceptance questionnaire of the SMILER.one app and (B) the complete home-based health care system.



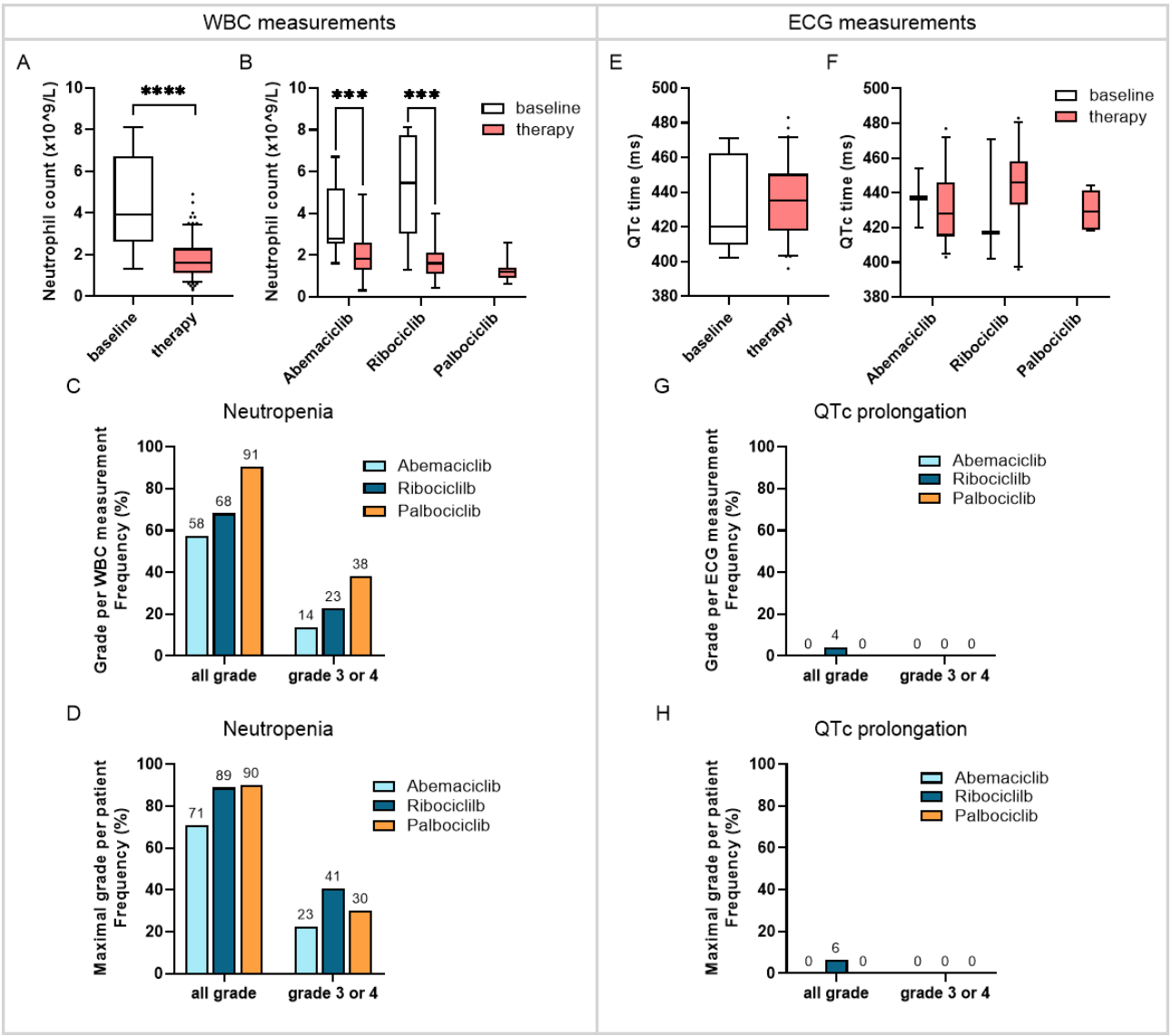
Clinical Relevance

Transferring routine monitoring from the clinical setting to the at-home environment requires the evaluation of WBC counts and ECG values. WBC monitoring is necessary for all CDK4/6i therapies, whereas ECG monitoring, specifically assessing QTc intervals, is required only for ribociclib treatment.

Compared with patients initiating CDK4/6i therapy, those who had been on treatment for more than five days had lower neutrophil counts (mean 4.3, SD 2.2, $\times 10^9$/L vs mean 1.8, SD 0.8, $\times 10^9$/L; P<.001; Figure 6A), regardless of the specific CDK4/6i received (Figure 6B). Mild neutropenia (grade 2) was present in 27% (45/167) of WBC measurements, while severe neutropenia (grade 3 or 4) was detected in 20% (34/167) of WBC measurements (Table S3 in Multimedia Appendix 1). At the patient level, this corresponded to 32% (22/68) of patients experiencing mild neutropenia and 31% (21/68) presenting with severe neutropenia under CDK4/6i therapy, as detected by the DHHC (Table S4 in Multimedia Appendix 1). Regarding individual CDK4/6i therapies, grade 3 or 4 neutropenia was observed in 23% (7/31) of patients receiving abemaciclib, 41% (11/27) of those on ribociclib, and 30% (3/10) of those on palbociclib (Figure 6C-D and Table S5 in Multimedia Appendix 1).

QTc time remained stable under CDK4/6i therapy (Figure 6e,f). Only one measurement indicated QTc prolongation, corresponding to 2% (1/61) of all quantifiable QTc times from ECG measurements and 2% (1/42) of all patients (Tables S5 and S6 in Multimedia Appendix 1). The affected patient was receiving ribociclib (Figure 6g.h and Tables S5 and S6 in Multimedia Appendix 1).

Figure 6. Detection of adverse events with the digital home health care center. Neutrophil counts were measured with the HemoCue system. (A) Neutrophil count before starting CDK4/6i therapy (n=15) and under therapy (>5 days; n=167) of all combined WBC measurements and (B) per CDK4/6i (baseline: abemaciclib, n=9; ribociclib, n=6; under therapy: abemaciclib, n=80; ribociclib, n=66; palbociclib, n=21). Neutropenia was graded according to CTCAE version 5.0. Frequency of neutropenia under CDK4/6i therapy as (C) the grade of each individual WBC measurement (abemaciclib, n=80; ribociclib, n=66; palbociclib, n=21) and (D) the maximal observed grade per patient (abemaciclib, n=31; ribociclib, n=27; palbociclib, n=10). QTc times were calculated from ECGs measurements with the Apple Watch. (E) QTc time before starting CDK4/6i therapy (n=5) and under therapy (>5 days; n=61) of all combined ECG measurements and (F) per CDK4/6i (baseline: abemaciclib, n=2; ribociclib, n=3; under therapy: abemaciclib, n=33; ribociclib, n=23; palbociclib n=5). QTc prolongation was graded according to CTCAE version 5.0. Frequency of QTc prolongation under CDK4/6i therapy as (G) the grade of each individual QTc measurement (abemaciclib, n=33; ribociclib, n=23; palbociclib n=5) and (H) the maximal grade per patient (abemaciclib, n=21; ribociclib, n=16; palbociclib n=5). ***$P \leq .001$, ****$P < .001$; Wilcoxon rank sum test. CDK4/6i: cyclin-dependent kinase 4/6 inhibitor; CTCAE: Common Terminology Criteria for Adverse Events; ECG: electrocardiogram; QTc: corrected QT interval; WBC: white blood cell.



Discussion

Principal Findings

The SMILER study assessed the feasibility and acceptance of a DHHC system for remote monitoring of side effects in patients with breast cancer receiving CDK4/6i therapy. Adherence to remote study visits was moderate and declined over time, with patients who were unable to perform measurements reporting time constraints and handling issues with the devices. Correspondingly, the transfer of remotely collected data was lower than expected, and patients reported needing to perform repeat measurements. Self-monitoring of ECGs and side effect photo documentation was considered easy to use and easily integrable into daily life, whereas WBC measurements were generally found to be more challenging. Nevertheless, the majority of participants felt positive about self-monitoring and expressed a willingness to incorporate home-based measurements into their future cancer care. Additionally, home-based ECG and WBC measurements were effective in detecting QTc time prolongations and neutropenia, demonstrating the clinical relevance of the DHHC.

Comparison to Prior Work

Our findings contribute to the emerging field of digital health and remote monitoring, highlighting the potential of such

interventions in improving patient care and outcomes. Patients receiving oral CDK4/6i therapy can take their anticancer medication at home; however, clinic visits remain necessary to monitor potential side effects [25-27]. Severe neutropenia (grade 3 or 4) is commonly observed under CDK4/6i therapy (ribociclib: 57% - 62% of patients in randomized controlled trials [RCTs] and 15% - 69% of patients in real-world studies [3,28-33]; abemaciclib: ~20% of patients in RCTs and 2% - 24% in real-world studies [9,31,33,34]; palbociclib: 62% - 70% of patients in RCTs and 60% - 63% in real-world studies [5,33,35-37]). Home-based WBC measurements by patients with breast cancer detected severe neutropenia in 41% of patients on ribociclib, 23% of patients on abemaciclib, and 30% of patients on palbociclib, aligning with real-world data for abemaciclib and ribociclib. Severe neutropenia under palbociclib may be underestimated due to the small number of patients receiving palbociclib in this study population. Nevertheless, these findings support the clinical relevance of home-based WBC monitoring with the DHHC.

QTc prolongation is a specific potential side effect of ribociclib therapy and has been reported in 2% - 5% of patients in the different RCTs [2-4]. In this study, prolonged QTc times were detected in one patient, corresponding to 6% (1/76 patients) of those receiving ribociclib therapy, which is in line with RCT findings [2-4]. Patients receiving abemaciclib or palbociclib did not develop QTc prolongations during the SMILER study.

Further research is needed to evaluate the clinical and psychological effects of home-based monitoring. A recent randomized trial involving patients with breast cancer receiving palbociclib therapy found that patients using an eHealth app (CANKADO PRO-React) experienced a longer time to QoL deterioration compared with those who did not use the app [19]. This study further highlights the clinical potential of digital interventions in improving patient outcomes and well-being. Notably, mobile health apps developed by health care professionals appear to have the highest overall quality [38], emphasizing the need to integrate clinical expertise with innovative digital approaches.

User perception and acceptance of the DHHC were generally positive. Both the ECG measurements with the Apple Watch and the photo documentation with the SMILER.one app were considered easy to use. The integration of the measurements into daily life was perceived favorably. However, concerns about result interpretation, time-consuming data collection, and technological difficulties were reported as barriers to engagement. The SUS score also indicated moderate usability. These findings emphasize the importance of providing adequate support and guidance to patients in using and interpreting the collected data. Additionally, the preference for consultations with health care professionals highlights the need to integrate home-based monitoring with clinical care and medical expertise. For future developments, integrating a communication tool into the provided app could be a suitable solution. Implementing a direct, automated digital feedback system based on home-based measurements could offer patients a more immediate and informed understanding of their health status, potentially reducing the need for frequent consultations with doctors. Interestingly, both using medical apps for medical questions and consulting Google appear to result in comparable adverse emotional and behavioral effects associated with cyberchondria [39].

Virtual and remote trials often have high dropout rates or low adherence to visits [40-42]. In the SMILER trial, adherence rates varied across different measurements, with the highest adherence for QoL questionnaires and lower adherence for WBC and ECG measurements. Additionally, adherence to the scheduled remote visits gradually declined over the course of the study, with 63% on d14±2d and 37% on d28-d2. Notably, several patients performed home-based measurements before or after the scheduled visits. Some patients also reported being unable to complete the final measurement due to the automatic deactivation of the SMILER.one app at the end of the study (day 28), suggesting that adherence rates may be underestimated.

The most common reason for missed scheduled measurements was device malfunction, with both technical and handling issues reported. Notably, the incidence of any type of malfunction during the first on-site introductory training was 24% (18/76 participants), and patients reported problems with 43% (33/76) of the DHHCs during remote monitoring. Not all patients with partially malfunctioning DHHC devices at the introductory training reported subsequent issues during remote monitoring, which may indicate that these devices were functional again during remote measurements or that the patient-reported failure rate was incomplete. Technical problems with the HemoCue device and the lack of an internet connection may partly explain DHHC malfunctions. Although the underlying cause of defective data transfer could not be definitively determined, server issues and complications with the Raspberry Pi software may have contributed. It is possible that unplugging the DHHC between remote measurements affected the Raspberry Pi and its software, which facilitates data transfer from the HemoCue device to the SMILER.one app. While there was no clear correlation between case design and the reported issues, data transfer problems occurred at varying rates across the different cases.

Handling problems with the HemoCue microcuvettes, such as an incomplete filling or air bubbles in the sample, was observed in approximately 10% (8/76) of patients during their first on-site WBC measurement. This is comparable to findings from another study that evaluated the feasibility of WBC self-testing with the HemoCue system in patients with cancer [43]. Additionally, handling issues during WBC measurements were associated with older age, highlighting the need for either a more intensive training session or remote assistance from a trained nurse, particularly for older patients. In general, tailoring support to specific patient needs may be essential to ensure successful implementation alongside a robust and reliable technological infrastructure.

Recruitment and patient characteristics also play a crucial role in the success of a study. In the SMILER study, only 76 out of the 136 screened women ultimately participated. Reasons for nonparticipation included lack of smartphone experience, time constraints, and concerns about performing WBC count independently. This underscores the importance of understanding the target population and their readiness to engage with digital health interventions. It is crucial to design digital

interventions in a way that ensures a lack of digital literacy does not contribute to health inequality [44,45].

Strengths

This study demonstrates several strengths in the implementation and evaluation of a DHHC system for patients receiving CDK4/6i therapy. First, it confirms the clinical relevance of remote monitoring by detecting known side effects, such as neutropenia and QTc prolongations, consistent with the established side effect profiles of these therapies. Although not designed to detect differences in the incidence of neutropenia or ECG changes between CDK4/6i, commonly known differences were observed [46]. Patients treated with ribociclib and palbociclib experienced neutropenia more frequently than those treated with abemaciclib. The only patient who had a QTc prolongation was receiving ribociclib.

Second, the study successfully integrated multiple technologies, including WBC measurement systems, smartwatches, and smartphones, into a comprehensive system tailored to the patient's needs. The high level of patient acceptance, with 79% (54/68) of participants expressing a willingness to use such systems in their future care, highlights the feasibility of this patient-centered approach. Moreover, the developed DHHC system and study findings underscore the potential of remote monitoring to reduce disparities in cancer care, particularly for patients in rural areas.

Limitations

This study has several limitations. First, the small sample size, specific patient population, and potential selection bias due to nonparticipation may limit the generalizability of the results. Second, adherence rates were suboptimal, and technical issues were encountered, highlighting challenges in implementing and maintaining home-based health care systems. Further optimization of the remote monitoring system is needed to improve usability. Additionally, a detailed analysis of the reliability of the DHHC could not be performed due to insufficient data. Third, self-reporting bias may have influenced the reported perceptions and experiences of the participants. Fourth, the short study duration limits our understanding of the long-term feasibility and acceptance of remote monitoring interventions. Fifth, the study lacked a control arm with patients receiving traditional nursing care, making direct comparisons with standard care models difficult. Sixth, as this was a pilot study, we did not predefine specific thresholds for the evaluated outcomes, such as adherence, usability, and feasibility, which limits our ability to determine whether the observed rates met predefined success criteria. Finally, this study did not include an analysis of the economic feasibility of the DHHC.

Future Directions

Based on the outcomes of this first feasibility study, several areas for improvement were identified. The technical issues encountered may be related to the DHHC devices, internet connection, server, Raspberry Pi software, and other factors that need to be addressed. Furthermore, longer-term randomized controlled studies are needed to better assess patient compliance, system reliability, and economic impact.

Conclusions

The findings of this study demonstrate that implementing remote monitoring of side effects in the care of patients with breast cancer undergoing CDK4/6i therapy is feasible. Patients with breast cancer generally accepted the idea of remote monitoring, which successfully identified clinically relevant side effects. Future studies with an improved system are required to further evaluate the potential clinical, socioeconomic, and individual benefits of this approach.

Acknowledgments

This research was partially funded by Pfizer (grand 61016789) and supported by the Bavarian State Ministry of Health and Care (grant MGP-2008 - 0003-digiOnko). Pfizer played no role in study design, data collection, analysis, and interpretation of data, or the writing of this manuscript. Adrian Kottas contributed to this publication in partial fulfillment of the requirements for obtaining the doctoral degree "Dr med."

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

HH designed the digital home-based health care center (DHHC) and the feasibility study, analyzed and visualized data, wrote the original manuscript draft, and managed and supervised the project. LAW designed the DHHC and the feasibility study, and managed and supervised the project. CG provided patients with on-site training, analyzed and visualized data, wrote the original manuscript draft, and managed the project. ME analyzed and visualized data, and performed statistical analyses. AM recruited patients for the feasibility study. AK recruited patients for the feasibility study. MK maintained the DHHC software and hardware during the feasibility study. JG and ML developed the DHHC software. KS recruited patients for the feasibility study. MR provided scientific input and designed the DHHC setup. PK analyzed electrocardiogram data. FH recruited patients for the feasibility study. MH recruited patients for the feasibility study. BV helped set up the DHHC hardware and provided scientific input. ER maintained the DHHC software and hardware during the feasibility study. CCH recruited patients for the feasibility study. MWB provided funding and infrastructure. SU supervised the data management and provided scientific input. PAF designed the DHHC and the

feasibility study, supervised the project, wrote the original manuscript draft, and provided funding. All authors reviewed the final manuscript.

Conflicts of Interest

PAF has received institutional funding from BioNTech, Cepheid, and Pfizer and personal honoraria from Novartis, Pfizer, Roche, Daiichi Sankyo, AstraZeneca, Lilly, Eisai, Merck Sharp & Dohme, Pierre Fabre, SeaGen, Agendia, Sanofi Aventis, Gilead, and Mylan for consulting, participation in advisory boards, and steering committees, or lectures. His institution conducts research for Novartis. HH received speaker honoraria from Novartis Pharma GmbH and LEO Pharma GmbH, and grant or research support from Novartis Pharma GmbH. KS received travel support from Gilead and Lilly. CCH received honoraria from AstraZeneca, Daiichi Sankyo, Eisai, Novartis, Pfizer, Roche, Gilead, and MSD, as well as support for attending meetings from Daiichi Sankyo. AK received honoraria from Gilead, as well as support for attending meetings from Lilly. CG received speaker honoraria from Novartis Pharma GmbH and ClinSol GmbH & Co. KG. MH received travel support from AstraZeneca, Lilly Deutschland GmbH, and Novartis. LAW received speaker honoraria from LEO Pharma GmbH.

Multimedia Appendix 1
Combination of supplementary materials.
[DOCX File, 72 KB - cancer_v11i1e64083_app1.docx]

References

1. Conley CC, McIntyre M, Pensak NA, et al. Barriers and facilitators to taking CDK4/6 inhibitors among patients with metastatic breast cancer: a qualitative study. Breast Cancer Res Treat 2022 Apr;192(2):385-399. [doi: 10.1007/s10549-022-06518-2] [Medline: 34997364]
2. Hortobagyi GN, Stemmer SM, Burris HA, et al. Ribociclib as first-line therapy for HR-positive, advanced breast cancer. N Engl J Med 2016 Nov 3;375(18):1738-1748. [doi: 10.1056/NEJMoa1609709] [Medline: 27717303]
3. Slamon DJ, Neven P, Chia S, et al. Overall survival with ribociclib plus fulvestrant in advanced breast cancer. N Engl J Med 2020 Feb 6;382(6):514-524. [doi: 10.1056/NEJMoa1911149] [Medline: 31826360]
4. Im SA, Lu YS, Bardia A, et al. Overall survival with ribociclib plus endocrine therapy in breast cancer. N Engl J Med 2019 Jul 25;381(4):307-316. [doi: 10.1056/NEJMoa1903765] [Medline: 31166679]
5. Finn RS, Martin M, Rugo HS, et al. Palbociclib and letrozole in advanced breast cancer. N Engl J Med 2016 Nov 17;375(20):1925-1936. [doi: 10.1056/NEJMoa1607303] [Medline: 27959613]
6. Slamon D, Lipatov O, Nowecki Z, et al. Ribociclib plus endocrine therapy in early breast cancer. N Engl J Med 2024 Mar 21;390(12):1080-1091. [doi: 10.1056/NEJMoa2305488] [Medline: 38507751]
7. Turner NC, Ro J, André F, et al. Palbociclib in hormone-receptor-positive advanced breast cancer. N Engl J Med 2015 Jul 16;373(3):209-219. [doi: 10.1056/NEJMoa1505270] [Medline: 26030518]
8. Sledge GW Jr, Toi M, Neven P, et al. MONARCH 2: abemaciclib in combination with fulvestrant in women with HR+/HER2- advanced breast cancer who had progressed while receiving endocrine therapy. J Clin Oncol 2017 Sep 1;35(25):2875-2884. [doi: 10.1200/JCO.2017.73.7585] [Medline: 28580882]
9. Goetz MP, Toi M, Campone M, et al. MONARCH 3: abemaciclib as initial therapy for advanced breast cancer. J Clin Oncol 2017 Nov 10;35(32):3638-3646. [doi: 10.1200/JCO.2017.75.6155] [Medline: 28968163]
10. Harbeck N, Rastogi P, Martin M, et al. Adjuvant abemaciclib combined with endocrine therapy for high-risk early breast cancer: updated efficacy and Ki-67 analysis from the monarchE study. Ann Oncol 2021 Dec;32(12):1571-1581. [doi: 10.1016/j.annonc.2021.09.015] [Medline: 34656740]
11. Aapro M, Andre F, Blackwell K, et al. Adverse event management in patients with advanced cancer receiving oral everolimus: focus on breast cancer. Ann Oncol 2014 Apr;25(4):763-773. [doi: 10.1093/annonc/mdu021] [Medline: 24667713]
12. Thill M, Schmidt M. Management of adverse events during cyclin-dependent kinase 4/6 (CDK4/6) inhibitor-based treatment in breast cancer. Ther Adv Med Oncol 2018;10:1758835918793326. [doi: 10.1177/1758835918793326] [Medline: 30202447]
13. Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. JCO 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830]
14. Lustberg MB, Kuderer NM, Desai A, Bergerot C, Lyman GH. Mitigating long-term and delayed adverse events associated with cancer treatment: implications for survivorship. Nat Rev Clin Oncol 2023 Aug;20(8):527-542. [doi: 10.1038/s41571-023-00776-9] [Medline: 37231127]
15. Sammons SL, Topping DL, Blackwell KL. HR+, HER2- advanced breast cancer and CDK4/6 inhibitors: mode of action, clinical activity, and safety profiles. Curr Cancer Drug Targets 2017;17(7):637-649. [doi: 10.2174/1568009617666170330120452] [Medline: 28359238]
16. Banbury A, Roots A, Nancarrow S. Rapid review of applications of e-health and remote monitoring for rural residents. Aust J Rural Health 2014 Oct;22(5):211-222. [doi: 10.1111/ajr.12127] [Medline: 25303412]
17. Kofoed S, Breen S, Gough K, Aranda S. Benefits of remote real-time side-effect monitoring systems for patients receiving cancer treatment. Oncol Rev 2012 Mar 5;6(1):e7. [doi: 10.4081/oncol.2012.e7] [Medline: 25992209]

18. Li Y, Li J, Zhang Y, Ding Y, Hu X. The effectiveness of e-Health interventions on caregiver burden, depression, and quality of life in informal caregivers of patients with cancer: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2022 Mar;127:104179. [doi: 10.1016/j.ijnurstu.2022.104179] [Medline: 35124473]
19. Harbeck N, Fasching PA, Wuerstlein R, et al. Significantly longer time to deterioration of quality of life due to CANKADO PRO-React eHealth support in HR+ HER2- metastatic breast cancer patients receiving palbociclib and endocrine therapy: primary outcome analysis of the multicenter randomized AGO-B WSG PreCycle trial. Ann Oncol 2023 Aug;34(8):660-669. [doi: 10.1016/j.annonc.2023.05.003] [Medline: 37201751]
20. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digital Med 2021 Mar 10;4(1):45. [doi: 10.1038/s41746-021-00418-3] [Medline: 33692479]
21. Coughlin SS, Caplan LS, Stone R. Use of consumer wearable devices to promote physical activity among breast, prostate, and colorectal cancer survivors: a review of health intervention studies. J Cancer Surviv 2020 Jun;14(3):386-392. [doi: 10.1007/s11764-020-00855-1] [Medline: 31933148]
22. Holländer-Mieritz C, Steen-Olsen EB, Kristensen CA, Johansen C, Vogelius IR, Pappot H. Feasibility of using wearables for home monitoring during radiotherapy for head and neck cancer-results from the OncoWatch 1.0 study. Cancers (Basel) 2023 Jan 9;15(2):422. [doi: 10.3390/cancers15020422] [Medline: 36672370]
23. Pope ZC, Zeng N, Zhang R, Lee HY, Gao Z. Effectiveness of combined smartwatch and social media intervention on breast cancer survivor health outcomes: a 10-week pilot randomized trial. J Clin Med 2018 Jun 7;7(6):140. [doi: 10.3390/jcm7060140] [Medline: 29880779]
24. Brooke J. SUS—a quick and dirty usability scale. In: Jordan P, Thomas B, McClelland I, Weerdmeester B, editors. Usability Evaluation In Industry: CRC Press; 1996:189-204.
25. Lux MP, Hartkopf AD, Fehm TN, et al. Update breast cancer 2023 part 2—advanced-stage breast cancer. Geburtshilfe Frauenheilkd 2023 Jun;83(06):664-672. [doi: 10.1055/a-2074-0125]
26. Hartkopf AD, Fehm TN, Welslau M, et al. Update breast cancer 2023 part 1—early stage breast cancer. Geburtshilfe Frauenheilkd 2023 Jun;83(6):653-663. [doi: 10.1055/a-2074-0551] [Medline: 37916183]
27. Jerzak KJ, Bouganim N, Brezden-Masley C, et al. HR+/HER2- advanced breast cancer treatment in the first-line setting: expert review. Curr Oncol 2023 Jun 2;30(6):5425-5447. [doi: 10.3390/curroncol30060411] [Medline: 37366894]
28. Hortobagyi GN, Stemmer SM, Burris HA, et al. Updated results from MONALEESA-2, a phase III trial of first-line ribociclib plus letrozole versus placebo plus letrozole in hormone receptor-positive, HER2-negative advanced breast cancer. Ann Oncol 2018 Jul 1;29(7):1541-1547. [doi: 10.1093/annonc/mdy155] [Medline: 29718092]
29. Tripathy D, Im SA, Colleoni M, et al. Ribociclib plus endocrine therapy for premenopausal women with hormone-receptor-positive, advanced breast cancer (MONALEESA-7): a randomised phase 3 trial. Lancet Oncol 2018 Jul;19(7):904-915. [doi: 10.1016/S1470-2045(18)30292-4] [Medline: 29804902]
30. Lüftner D, Brucker C, Decker T, et al. Real-world efficacy of ribociclib (RIB) plus aromatase inhibitor (AI)/fulvestrant (FUL), or endocrine monotherapy (ET), or chemotherapy (CT) as first-line (1L) treatment (tx) in patients (pts) with hormone receptor–positive (HR+), human epidermal growth factor receptor-2–negative (HER2–) advanced breast cancer (ABC): results of fourth interim analysis (IA) from RIBANNA. J Clin Oncol 2022 Jun 1;40(16_suppl):1065-1065. [doi: 10.1200/JCO.2022.40.16_suppl.1065]
31. Buller W, Pallan L, Chu T, Khoja L. CDK4/6 inhibitors in metastatic breast cancer, a comparison of toxicity and efficacy across agents in a real-world dataset. J Oncol Pharm Pract 2023 Dec;29(8):1825-1835. [doi: 10.1177/10781552231163121] [Medline: 36945886]
32. Fernández-Cuerva C, Chinchilla-Alarcón T, Alcaraz-Sánchez JJ. Real-world effectiveness of ribociclib in metastatic breast cancer patients: does dose affect survival? J Oncol Pharm Pract 2023 Oct;29(7):1619-1627. [doi: 10.1177/10781552221144280] [Medline: 36476018]
33. Queiroz MM, Sacardo KP, Ribeiro MF, et al. Real-world treatment outcomes in HR+ HER2- metastatic breast cancer patients treated with CDK4/6 inhibitors: results from a reference center in Brazil. Cancer Treat Res Commun 2023;35:100683. [doi: 10.1016/j.ctarc.2023.100683] [Medline: 36716534]
34. Johnston SRD, Toi M, O'Shaughnessy J, et al. Abemaciclib plus endocrine therapy for hormone receptor-positive, HER2-negative, node-positive, high-risk early breast cancer (monarchE): results from a preplanned interim analysis of a randomised, open-label, phase 3 trial. Lancet Oncol 2023 Jan;24(1):77-90. [doi: 10.1016/S1470-2045(22)00694-5] [Medline: 36493792]
35. Gnant M, Dueck AC, Frantal S, et al. Adjuvant palbociclib for early breast cancer: the PALLAS trial results (ABCSG-42/AFT-05/BIG-14-03). J Clin Oncol 2022 Jan 20;40(3):282-293. [doi: 10.1200/JCO.21.02554] [Medline: 34874182]
36. Turner NC, Slamon DJ, Ro J, et al. Overall survival with palbociclib and fulvestrant in advanced breast cancer. N Engl J Med 2018 Nov 15;379(20):1926-1936. [doi: 10.1056/NEJMoa1810527] [Medline: 30345905]
37. Hackert MQN, van Uden-Kraan CF, Agterof MJ, et al. Real-world palbociclib effectiveness in patients with metastatic breast cancer: focus on neutropenia-related treatment modification strategies and clinical outcomes. Cancer Treat Res Commun 2023;35:100691. [doi: 10.1016/j.ctarc.2023.100691] [Medline: 36913809]

38. Altmannshofer S, Flaucher M, Beierlein M, et al. A content-based review of mobile health applications for breast cancer prevention and education: characteristics, quality and functionality analysis. Digital Health 2024;10:20552076241234627. [doi: 10.1177/20552076241234627] [Medline: 38528967]
39. Jungmann SM, Brand S, Kolb J, Witthöft M. Do Dr. Google and health apps have (comparable) side effects? An experimental study. Clin Psychol Sci 2020 Mar;8(2):306-317. [doi: 10.1177/2167702619894904]
40. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]
41. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
42. Ali Z, Anderson K, Chiriac A, et al. High adherence and low dropout rate in a virtual clinical study of atopic dermatitis through weekly reward-based personalized genetic lifestyle reports. PLoS One 2020;15(7):e0235500. [doi: 10.1371/journal.pone.0235500] [Medline: 32614886]
43. Mattsson TO, Lindhart CL, Schöley J, Friis-Hansen L, Herrstedt J. Patient self-testing of white blood cell count and differentiation: a study of feasibility and measurement performance in a population of Danish cancer patients. Eur J Cancer Care (Engl) 2020 Jan;29(1):e13189. [doi: 10.1111/ecc.13189] [Medline: 31746087]
44. Kaihlanen AM, Virtanen L, Buchert U, et al. Towards digital health equity—a qualitative study of the challenges experienced by vulnerable groups in using digital health services in the COVID-19 era. BMC Health Serv Res 2022 Feb 12;22(1):188. [doi: 10.1186/s12913-022-07584-4] [Medline: 35151302]
45. Eddison N, Leone E, Healy A, Royse C, Chockalingam N. The potential impact of allied health professional telehealth consultations on health inequities and the burden of treatment. Int J Equity Health 2022 Jun 30;21(1):91. [doi: 10.1186/s12939-022-01689-2] [Medline: 35773695]
46. Nabieva N, Altmann F, Apel K, et al. The endocrine treatment landscape for patients with HR+ HER2- early-stage breast cancer in Germany before the introduction of CDK4/6 inhibitor therapy—a real-world analysis. Geburtshilfe Frauenheilkd 2023 Sep;83(9):1127-1137. [doi: 10.1055/a-2100-0643] [Medline: 37942358]

Abbreviations:

CDK4/6i: cyclin-dependent kinase 4/6 inhibitor
DHHC: digital home-based health care center
ECG: electrocardiogram
IoM: Internet of Me
QoL: quality of life
QTc: corrected QT interval
RCT: randomized controlled trial
REST API: representational state transfer application programming interface
SUS: System Usability Scale
WBC: white blood cell

Edited by N Cahill; submitted 15.07.24; peer-reviewed by Z Ping; revised version received 04.03.25; accepted 06.03.25; published 02.05.25.

Please cite as:
Huebner H, Wurmthaler LA, Goossens C, Ernst M, Mocker A, Krückel A, Kallert M, Geck J, Limpert M, Seitz K, Ruebner M, Kreis P, Heindl F, Hörner M, Volz B, Roth E, Hack CC, Beckmann MW, Uhrig S, Fasching PA
A Digital Home-Based Health Care Center for Remote Monitoring of Side Effects During Breast Cancer Therapy: Prospective, Single-Arm, Monocentric Feasibility Study
JMIR Cancer 2025;11:e64083
URL: https://cancer.jmir.org/2025/1/e64083
doi:10.2196/64083

Co-Designing a User-Centered Digital Health Tool for Supportive Care Needs of Patients With Brain Tumors and Their Caregivers: Interview Analysis

Mahima Kalla[1], PhD; Ashleigh Bradford[2], BBiomed; Verena Schadewaldt[3,4], PhD; Kara Burns[1], PhD; Sarah C E Bray[3,4], PhD; Sarah Cain[3], MBBS; Heidi McAlpine[3], MBBS; Rana S Dhillon[4,5,6], MBBS, MPhil; Wendy Chapman[1], PhD; James R Whittle[7,8,9], PhD; Katharine J Drummond[3,4], MD; Meinir Krishnasamy[2,10,11], PhD

[1]Centre for Digital Transformation of Health, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Grattan Street, Parkville, VIC, Australia

[10]Department of Nursing, School of Health Sciences, Faculty of Medicine, Dentistry and Health, The University of Melbourne, Parkville, VIC, Australia

[11]Victorian Comprehensive Cancer Centre Alliance, Melbourne, VIC, Australia

[2]Department of Health Services Research and Implementation Science, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

[3]Department of Neurosurgery, The Royal Melbourne Hospital, Parkville, VIC, Australia

[4]Department of Surgery, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Parkville, VIC, Australia

[5]Barwon Health, Geelong, VIC, Australia

[6]St Vincent's Hospital Melbourne, Fitzroy, VIC, Australia

[7]Personalised Oncology Division, Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

[8]Department of Medical Biology, The University of Melbourne, Parkville, VIC, Australia

[9]Department of Medical Oncology, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

Corresponding Author:
Mahima Kalla, PhD
Centre for Digital Transformation of Health, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Grattan Street, Parkville, VIC, Australia

Abstract

Background: Brain tumors are characterized by the high burden of disease that profoundly impacts the quality of life in patients and their families. Digital health tools hold tremendous potential to enhance supportive care and quality of life for patients with brain tumors and their caregivers.

Objective: This study aims to generate ideas and concepts, through a co-design paradigm, to inform the development of a digital health tool to address the unmet needs of people affected by brain tumors.

Methods: Patients with brain tumors, caregivers, and health professionals from 2 large public tertiary hospitals in Victoria, Australia, were invited to complete a qualitative interview discussing their unmet needs of care. Overall, 35 qualitative interviews focusing on unmet needs and concepts for a digital health tool were conducted with 13 patients, 11 caregivers, and 11 health professionals. Interviews were audio recorded and transcribed, and a 5-step framework analysis approach was used to analyze data.

Results: Four themes of unmet supportive care needs emerged: (1) emotional and psychological, (2) information, (3) physical and practical, and (4) social connectedness. Participants expressed the desire for early and proactive mental health intervention, noted the importance of providing mental health support to caregivers, and emphasized the need for positive stories and affirmative language. From an information perspective, participants noted a sense of information overload, especially at the beginning. They also underscored the variety of information needed on an ongoing basis, including life after treatment, and comprehensive care assistance to maintain quality of life. Participants also described unmet supportive care needs relating to symptom burden, and practical and administrative support to facilitate the logistics of accessing treatment and accomplishing daily life tasks. Finally, they expressed the desire for greater social connectedness and safe spaces to engage with other people in a similar situation. Our findings are consistent with previous research on this subject and were integrated into the development of a web-based platform.

Conclusions: Participants' perspectives informed the development of content for a web-based digital health platform called "Brain Tumours Online." The platform comprises three pillars—(1) "LEARN": a repository of vetted information about a range of biomedical and psychosocial care topics; (2) "CONNECT": a digital peer support community with a health care professional interface; and (3) "TOOLBOX": an emerging library of validated digital therapeutics for symptom management.

(JMIR Cancer 2025;11:e53690) doi:10.2196/53690

KEYWORDS

brain cancer; unmet needs; supportive care; psychosocial support; digital health; qualitative research; brain tumor; user-centered; patients; caregivers; interview analysis; quality of life; effectiveness; co-design paradigm; ideas; concepts; emotional support; information sharing; social connectedness; health care professionals

Introduction

Despite comprising only 1.2% of all cancers in Australia, the impact of brain tumors on the quality of life is profound [1-3]. Patients with brain tumors experience persistent, distressing, and disabling physical, psychosocial, cognitive, and financial challenges. These challenges are compounded by barriers to connecting and communicating with their treating team, establishing peer support networks, managing their symptoms, and accessing personalized supportive care [4-8]. These issues are exacerbated when patients reside distant from metropolitan, specialist treatment centers [9] without equitable access to comprehensive supportive care services.

Digital health tools hold tremendous potential to revolutionize support that enhances the quality of life for patients with brain tumors and their caregivers within convenient time frames and comfortable environments. Notably, they have the ability to address inadequate cancer services access by overcoming geographical, physical, and psychological barriers, and facilitating treatment access, support, and education [10]. When compared to usual care, health care augmented with digital health interventions has been shown to improve symptom management, reduce distress, decrease unplanned hospitalizations and associated care–related costs, and improve survival and quality of life [11]. While a number of digital supportive care tools for patients with brain tumors have been described [12], there are currently none tailored to the Australian context. In addition, existing tools are limited by a lack of evidence of their effectiveness and impact, development and implementation, and little consumer engagement in their development [7]. Building on our prior research on the quality of life of patients with brain tumors [2,3,13], their patterns of social media use for disease management [8], and needs and expectations from a digital health model of care [7], we engaged with Australian patients with brain tumors, their caregivers, and treating health professionals to co-design a digital health solution to address the unmet supportive care needs of this population. The aim of this study was to generate ideas and concepts to inform the development of a digital health tool to support the needs of Australian patients with brain tumors, their caregivers, and treating health professionals.

Methods

Co-Design Approach

Co-designing with patients and end users is widely recognized as critical to the design and development of digital health interventions and is used extensively across a range of physical and mental health conditions [14]. According to Sanders and Stappers [15], co-design may be conducted at various stages of a project life cycle, for example, (1) predesign (to understand lived experiences), (2) generative (to produce ideas), (3) evaluative (to summatively assess solutions), and (4) postdesign (to assess users' experience of the solution). We adopted a multimodal co-design approach, which included one-on-one interviews, focus groups, workshops, a fortnightly forum called a Design Reference Group, an end-of-life working group to inform design decisions for palliative care resources, and preliminary usability testing of a high-fidelity prototype. This paper reports the findings of one-on-one interviews conducted with patients, informal caregivers, and health care professionals in the project's "generative" phase. The other co-design activities will be addressed in subsequent research papers and are not included within the scope of this paper.

Participants and Recruitment Procedure

Three participant cohorts were recruited for this interview study: (1) adult patients (aged over 18 years) with a primary brain tumor, (2) current and bereaved caregivers of adults with a primary brain tumor, and (3) multidisciplinary health care professionals involved in the care of people affected by brain tumors. We excluded patients with secondary brain tumors from systemic cancer and their caregivers. Neurosurgeons were integral to the predesign phase and research team; thus, the generative phase focused on recruiting a variety of other health care professionals, including nursing and allied health staff.

Purposive sampling was used to identify participants with a wide variety of experiences in relation to aspects such as time since diagnosis, age, sex, postcode, tumor types, and types of treatment received. Researchers AB and VS maintained a spreadsheet of participants with the aforementioned salient information and sought to proactively seek out participants from a wide breadth of lived experience.

Participants were recruited from 2 major metropolitan hospitals in the State of Victoria (Australia). Patients were also asked to nominate a caregiver to take part in the study. Bereaved caregivers were also nominated by brain tumor advocacy organizations and consented by a study researcher. Health professional participants (including medical and radiation oncologists, clinical care coordinators, palliative care physicians, neuro-oncology nurses, and other allied health professionals) were identified through research team networks and invited via email.

Data Collection

The interviews were conducted over a period of 6 months, from October 2021 to March 2022. All consenting participants completed 1 audio-recorded semistructured telephone interview of 20 - 60 minutes with a research assistant. Verbal consent was reconfirmed and recorded prior to the commencement of the interview, along with basic demographic information. Semistructured interview guides were developed for each

participant cohort (Multimedia Appendices 1-3). Patient and caregiver interviews sought to elicit their lived experiences from diagnosis, their unmet supportive care needs, and their desires and preferences for digital solutions to address identified unmet needs. Health care professional interviews explored their perception of patient and caregiver needs, their experience with the provision of digital-based health care, and their perceptions of the need and preferences for digital health solutions to support patients with brain tumors and their caregivers.

Data Analysis

The audio recordings from each interview were transcribed verbatim, and the transcripts were inputted into QSR International's NVivo (version 12.6) for Mac software for coding and qualitative data management. All transcripts were quality-checked and deidentified for analysis by author AB. Data were analyzed using a 5-step framework analysis method [16], as follows:

- Familiarization: Researchers AB and VS read all the interview transcripts to familiarize themselves with the data.
- Identifying a framework: Fitch's 2008 Cancer Supportive Care Framework [17] was used to develop an a priori thematic framework to guide early data analysis. This is an evidence-informed framework that directly relates to the focus of the study to explore and describe unmet supportive care needs as experienced by people affected by brain tumors and their caregivers. The framework articulates 7 domains of supportive care needs: physical, informational, emotional, psychological, social, spiritual, and practical.
- Indexing: The thematic framework was applied to the data. Three researchers (AB, VS, and M Kalla) first indexed an initial set of 10 interview transcripts using Fitch's 7 supportive care categories. After this initial application, overlaps were found across the 7 domains, and the analytical framework was revised into 4 thematic categories for this study: emotional and psychological, information, physical and practical, and social connectedness needs. Subsequently, researchers AB and VS indexed all the interview transcripts independently, meeting regularly to ensure consensus around the final relevant themes and subthemes.
- Charting: The indexed data were organized into a manageable chart format to enable within- and cross-case analyses. NVivo's "Organise" function retrieved charted summaries of the indexed data for each thematic category and individual participant transcripts [18].
- Mapping and interpretation: Finally, mapping and interpretation of the data were conducted through the lens of the 3 pillars of the digital health tool proposed in the original funding proposal, that is, Learn, Connect, and Toolbox. Researchers AB, VS, and M Kalla synthesized the findings into summary presentations, describing how unmet supportive care needs could be addressed through one or more pillars.

Research Rigor

Our research team was comprised of a multidisciplinary team of academics, health service researchers, clinicians, technology developers, and lived experience experts. The multidisciplinary perspectives afforded by our team members strengthened the conceptualization, analysis, synthesis, and dissemination of our research findings. The majority of the interviews were conducted by the second author (AB), who has experience conducting health services research with patients with cancer and their caregivers using qualitative methods, having worked in a nationally reputed cancer treatment center. A small subset of interviews was also conducted by the first author (M Kalla), who is a digital health and qualitative research expert, and the third author (VS), who has experience as a nursing clinician and qualitative health researcher. This study was overseen by the last author (M Krishnasamy), who is a professor of nursing and qualitative research expert. The overall project was led by author KJD, who is the head of neurosurgery at a major metropolitan hospital in Australia.

To ensure credibility in the analysis of the research findings, 2 researchers (AB and VS) independently analyzed the interview transcripts to generate initial codes. Subsequently, a working group involving authors AB, VS, M Kalla, SCEB, and M Krishnasamy met at fortnightly intervals during the entire data collection and analysis process to compare, refine, synthesize, and establish the final themes. Interviews were conducted until data saturation was reached, and no new topics were discovered, resulting in a total of 35 interviews (13 patients, 11 caregivers, and 11 health care professionals), which is within the suggested range of 9 - 17 interviews in qualitative research [19]. Member checking was not conducted with participants to respect the limited available time of patients with brain tumors and caregivers to participate in a research project. Instead, we sense-checked the themes with a lived experience expert (a bereaved caregiver) who was part of our project's steering group. Finally, the findings were presented to the broader research team, including health care professionals and consumer advocates.

Ethical Considerations

The study was approved by the Royal Melbourne Hospital's Human Research Ethics Committee (HREC/77238/MH-2021). The ethics committee is accredited with Australia's National Health and Medical Research Council and is operated in accordance with the National Statement on Ethical Conduct in Human Research. The data presented in this paper are the primary data that were collected as part of this ethical approval. Potential participants were emailed an information and consent form to read and sign to agree to be interviewed. To ensure privacy and confidentiality, all study data were deidentified prior to analysis. Author AB reviewed all interview transcripts in their entirety to ensure that they had been fully deidentified prior to the commencement of analysis. Identifiable participant information was securely stored with password protection and accessible only to researchers authorized under the ethical approval.

Interviews were conducted over the telephone, with costs for the phone calls borne by the research team. Participants were not required to travel, and the interviews were conducted at a time suitable for them. Therefore, participants did not face any financial burden from participating in the study. Thus, reimbursement was not provided for their voluntary participation. The aforementioned arrangements for participant

involvement were approved by the ethics committee, as articulated in the research study protocol submitted at the time of application. As part of the ethics approval, our team was also required to submit annual progress reports and a final report, as well as notify the committee of any adverse events. These reports were provided to the committee in a timely manner. There were no adverse events reported from the conduct of this research.

Results

Overview

A total of 35 participants (13 patients, 11 caregivers, and 11 health care professionals) were recruited (Table 1).

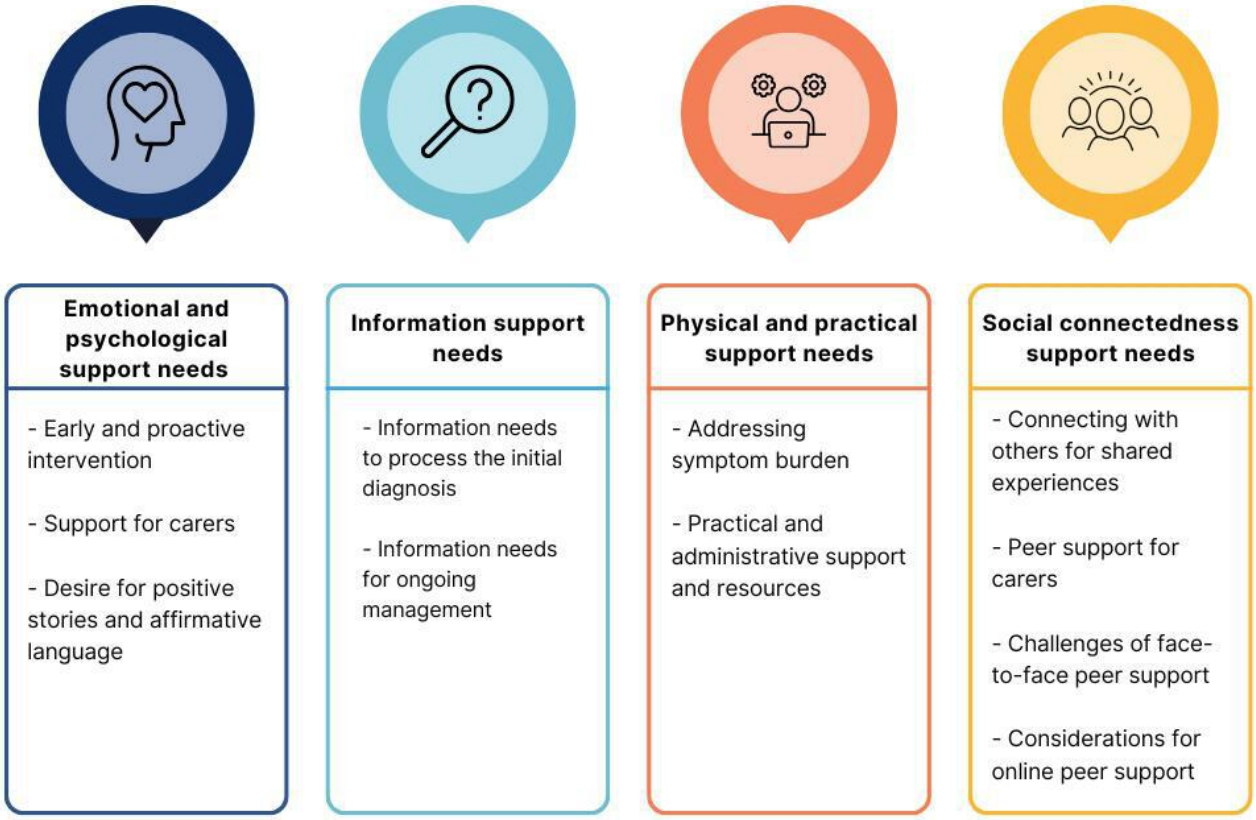
The emergent themes and their subthemes are presented in Figure 1. Detailed and illustrative quotes from each of the 4 support needs domains—emotional and psychological, information, physical and practical, and social connectedness—are presented. All data fit within these themes on completion of the iterative process.

Table . Overview of participant demographics (N=35).

Demographic	Patients (n=13)	Caregivers (n=11)	Health professionals (n=11)
Age (years), mean (range)	42.1 (22 - 67)	55 (32 - 86)	47.1 (40 - 60)
Sex, n (%)			
Male	5 (39)	2 (18)	3 (27)
Female	8 (62)	9 (82)	8 (73)
Location, n (%)			
Metropolitan	9 (69)	4 (36)	8 (73)
Regional or remote	4 (31)	4 (36)	3 (27)
Unanswered	0 (0)	3 (27)	0 (0)
Tumor type, n (%)			
Low-grade glioma	5 (39)	N/A[a]	N/A
High-grade glioma	5 (39)	N/A	N/A
Rare brain cancer	3 (23)	N/A	N/A
Relationship to the patient, n (%)			
Spouse	N/A	6 (55)	N/A
Parent	N/A	4 (36)	N/A
Child	N/A	1 (9)	N/A
Caregiver status, n (%)			
Current	N/A	9 (82)	N/A
Bereaved	N/A	2 (18)	N/A
Role, n (%)			
Neuro-oncology nurse	N/A	N/A	3 (27)
Clinical care coordinator	N/A	N/A	2 (18)
Medical oncologist	N/A	N/A	2 (18)
Neuropsychologist	N/A	N/A	1 (9)
Palliative care physician	N/A	N/A	1 (9)
Exercise physiologist	N/A	N/A	1 (9)
Radiation oncologist	N/A	N/A	1 (9)

[a]N/A: not available.

Figure 1. Overview of emergent themes and subthemes.



Theme 1: Emotional and Psychological Support Needs

Within the theme of patient and caregiver emotional and psychological support needs, 3 subthemes highlighted the need for early and proactive psychological intervention, the importance of emotional support for caregivers, and the desire for positive stories and affirmative language.

Early and Proactive Intervention

Participants reported delayed help-seeking for mental health assistance due to a lack of awareness of relevant services and the benefits of intervention. They described frequent reliance on informal mental health support from family and friends, or those with similar experience. Some participants delayed addressing mental health needs until after treatment. Almost all participants wished they had been encouraged by their treating team to access mental health services or provided with links to resources and tools at diagnosis.

> *I didn't start seeing a psychologist until after treatment which isn't good...I wish someone kind of forced me to see one, not forced...like just did a referral for me to see one...I did like a year of treatment and...you're trusting your family [but] there's only so much they can really do...Before treatment would have been good...Once I was diagnosed, I wish I had saw [sic] someone.* [Patient 3, female, 22 years]

Health care professionals reported varied levels of proactiveness for mental health service referrals. Lack of referral was often driven by insufficient clarity regarding the best service or intervention for their patient's specific needs. Often, health care professionals deferred to brain tumor advocacy organization services and web-based resources.

> *I have...people saying...I'm quite distressed...and then me...going through this list...so we've done the referral to Cancer Council, you're on a wait list for a counselling session...What exactly is the issue, can it be worked through with a counsellor, is it more affecting...do we need psychology support, or do we need some...medication intervention...do I need psychiatry...I think just having that information a bit more available.* [Health care professional 10, neuro-oncology nurse consultant, female, 39 years]

Support for Caregivers

Participants highlighted the importance of mental health support for caregivers, with mental health discussions often focused on the patient and inadequate recognition of caregivers' needs.

> *I think as carers you don't recognise...oh actually your life is going to be significantly impacted in the future and...right here and now you actually need to look after yourself as much as you can, to be the best carer that you can [be].* [Caregiver 2, female, 37 years]

> *I was offered counselling...I reckon it should be available to both parties...I feel that my partner...she was going through more than me.* [Patient 6, male, 31 years]

Desire for Positive Stories and Affirmative Language

Participants expressed the need for information and lived experience stories, framed positively and fostering a sense of hope. Some participants actively sought people with brain tumors on social media who were living fulfilling lives, as it gave them hope for their own future.

> *Just to really emphasise that your life isn't over, and you can still have a normal life after it...I found there's a few people on Instagram that had been diagnosed with it and they're still out there...really doing well.* [Caregiver 4, female, 32 years]

Affirmative explanations of complex medical treatments and procedures were useful in creating a greater understanding of the treatment plan and reducing fear.

> *A lot of people are actually frightened...having brain surgery whilst you're awake, is such a morbid type of experience, where I found it exactly the opposite...It's hard to describe the experience that I went through, but being able to listen to what's going on whilst it's happening...oh we're removing some of the tumour now, having all that conversation...I found that [I was] more in control...more knowledge, more knowing...everything that was going on.* [Patient 7, male, 50 years]

Theme 2: Information Support Needs

Participants described significant initial information needed to process their diagnosis and ongoing information needed to manage the physical, social, and psychological impacts of a brain tumor on their lives and that of their families.

Information Needs to Process the Initial Diagnosis

Information requirements were greatest at diagnosis and reported a sense of overwhelm with "information overload." Patients and caregivers reported difficulty understanding the information presented by their health care professionals and expressed a strong need for tailored, personalized information, appropriate for the type and location of tumor and patient age and sex. Similarly, health care professionals recognized that deconstructing medical jargon while providing the appropriate breadth of information is crucial to support patients and their caregivers in decision-making around treatment options.

> *I think you've got to find a balance between not overloading people with information but giving them access to resources and that's never easy.* [Health care professional 8, medical oncologist, male, 55 years]

At diagnosis, patients and caregivers described turning to internet searches to learn more about the tumor, treatment options (including clinical trials and integrative therapies), life expectancy, and survival statistics. Many participants reported that in these early stages, they struggled to discern credible information and found "Dr Google" overwhelming.

> *I still wanted to know...as much information as I could about...my diagnosis...but I needed accurate information and that's where I say the internet didn't really supply me that accurate information about my particular case.* [Patient 7, male, 50 years]

Information Needs for Ongoing Management

Patients and their caregivers described an ongoing need for information about life after treatment, the impacts of the illness and its treatment, and assistance to maintain quality of life.

> *Help processing what the treatment plan looks like...and knowing where to go for help and answering questions. Not only about diagnosis but also about treatment once they're discharged...The discharge happens pretty quickly from surgery, usually...six to nine days or sometimes sooner.* [Health care professional 5, cancer care coordinator, female, 51 years]

Additionally, patients and caregivers expressed the need for on-demand information in a consolidated space and the ability to learn from experts from the comfort of their homes.

> *Video would be good...hearing from experts and doctors about things would be really helpful...And maybe even a Q&A session where you can ask questions, and they can answer...interactively.* [Caregiver 7, female, 60 years]

Theme 3: Physical and Practical Support Needs

Participants described unmet supportive care needs relating to symptom burden, practical and administrative support to facilitate the logistics of accessing treatment, and accomplishing daily life tasks.

Addressing Symptom Burden

Participants described a variety of symptoms associated with the brain tumor or its treatment, which significantly impacted patient quality of life and caregiver burden. Participants expressed a strong desire for access to resources and interventions to support symptom management, including sleep disturbance, fatigue, cognitive dysfunction, poor concentration, imbalance, and incoordination. Participants indicated that symptoms were clustered, for example, an interaction of sleep disturbance and fatigue.

> *If there was something to help with sleep, I would definitely jump...that would be one of the biggest things...I basically always went to the GP about...I would go to say I can't sleep, I just get...rundown...I can't sleep, I can't sleep, I'm tired...* [Patient 1, male, 53 years]

Seizures were considered particularly burdensome and "frightening" with management options often not well understood by patients or caregivers.

> *Managing seizures and what to do in, in the event of a seizure...I mean not everyone needs to come to hospital if they have a seizure, it depends on...what type...Partial seizures...you can pretty much manage at home. And then...talk to someone the next day or go through the GP...there's certain steps that you could take. I think that would be a really good [pause] thing to put in...if you're uploading information [on*

the platform]. [Health care professional 5, cancer care coordinator, female, 51 years]

We asked about the role of self-monitoring by patients for symptom management, with participants presenting mixed responses. Some health care professionals, particularly in allied health and palliative care contexts, used symptom monitoring to support patients. Nevertheless, participants agreed that symptom monitoring regimes, whether digital or paper-based, would need to be considered contextually within broader clinical workflows to be useful or effective.

> *I guess the key thing would be somebody taking responsibility for that and...following up where concerning symptoms have been presented...I don't know that just collecting symptoms for the sake of collecting symptoms is going to be particularly useful. You've got to have somebody who's going to be reviewing that information and...acting upon it.* [Health care professional 9, radiation oncologist, female, 51 years]

Practical and Administrative Support and Resources

Participants expressed the need for support and resources to manage practical, logistical, or administrative challenges. Information to navigate transport, permission to continue or return to driving, government welfare and advocacy agency support, and accommodation during treatment and management of insurance were commonly difficult to obtain.

> *The thing that really frustrated me is every time I'd try and talk to someone about...things like financial information...life insurance...income protection...just to try and get some information and help with trying to sort through...how I can access [it].* [Patient 1, male, 53 years]

Participants also emphasized the particular need for practical, logistical, and administrative support for patients and caregivers in rural or regional areas.

> *We need to keep in mind the...regional family as well because...there are types of cancer that can't be treated in...regional areas. They need to be treated at a specialist centre. So...demands placed on that around travel...the financial impacts of that...paying for accommodation, parking, travel all those sorts of things. There are subsidies and things that people can get access to but that does have impacts that are ongoing...taking time off work...all those sorts of things are quite huge.* [Health care professional 1, exercise physiologist, male, 42 years]

When asked about telehealth, all patients and caregivers reported having used telehealth and found it time-saving and convenient. However, for important appointments, a face-to-face conversation was preferred.

Theme 4: Social Connectedness Support Needs

Patients and caregivers expressed the need for social connectedness and suggested these may be met by a digital platform. Four subthemes emerged: the importance of connecting with others for shared experiences, the significance of peer support for caregivers, challenges and lessons learned from face-to-face peer support programs, and considerations for creating safe and beneficial digital peer support communities.

Connecting With Others for Shared Experiences

Participants noted that a brain tumor can be alienating, with feelings of isolation further compounded for those in rural or regional areas. Thus, opportunities to connect with others with similar experiences can offer solace.

> *When you find someone who you can relate to and they've got a level head, it just puts your mind at ease.* [Patient 6, male, 32 years]

Participants noted that in addition to helping cope with stressors, peer support could also normalize living with this illness.

> *This is the absence that we have in brain cancers and tumours...you can see the hard, the scary facts, but...it is really to understand what next, who else is out there like me, who else is young and also about to get married, or just about to have a baby, or just got the promotion at work or...I feel like there's a lot of people that sit in this category.* [Caregiver 2, female, 37 years]

Peer Support for Caregivers

Participants also highlighted the need for social connectedness and peer support for caregivers and noted that digital solutions could offer a helpful on-demand avenue for busy caregivers.

> *Online [peer support] is actually like a great entry point...You're so consumed by...needing to do everything for that person...I can't remember how many times I would've been awake at 2 am in the morning or something because my brain was racing, and being able to may be connect with somebody or...something like that, not necessarily to chat right there in that moment.* [Caregiver 2, female, 37 years]

Participants also emphasized the importance of peer support for caregivers in all situations, including those currently undertaking caring duties, as well as bereaved caregivers. Some participants had found established grief support groups, but not for current caregivers, while other participants reported a lack of formalized support structures for bereaved caregivers.

> *Our focus is on the patient and of course we support the partner while they're coming to the centre. But once that patient's passed away our involvement sort of ends, and I think these partners are left...often they're very fatigued because I think [for] the carers of these types of patients...the carer burden is huge.* [Health care professional 4, nurse, female, 46 years]

Challenges of Face-to-Face Peer Support

Participants expressed that despite their desire for social connection with others with similar experiences, they had either not been able to find an in-person support group that was the right fit for them, or they had found the support networks unhelpful or not relatable with members of varying experiences, ages, sexes, or life stages. Additionally, in-person support groups

could be confronting, and digital peer support was considered a helpful alternative.

> *She didn't want to go to a room and sit in a chair around in a circle and tell her story...Some of those support groups can be a bit too full on. So that's why something that isn't quite as full on like an app or online story or...live chat, something...you can still...perhaps talk to someone, so you're...not in a room full of strangers...But something in between they can just have on their smart phone or computer and just send off an enquiry and if that's all you need at the time, then that's fine.* [Caregiver 1, male, 37 years]

Health care professionals with experience in running in-person support groups also highlighted the logistical challenges of facilitating face-to-face groups. They reported high attrition rates, emphasized the need for distress protocols in the event that a participant is emotionally affected, and noted the complexities of running in-person support groups involving people with varying prognoses.

> *If you're meeting in person you've got to have reasonable leaders...Normally the social worker would do it with me...because if someone gets upset, you've got to have enough people to take them out and talk to them separately as well...I think the support groups do help but...you have to be careful, and you can't mix the low grade with the high grade. That's another problem because...the high grade have very immediate needs.* [Health care professional 7, care coordinator, female, 72 years]

Considerations for Digital Peer Support

Challenges to create a safe, relatable, and where possible, customizable web-based peer support platform were recognized as important considerations. Participants commented on the importance of regulation and moderation in digital peer communities. They expressed concerns over harsh language, individuals with domineering personalities and negative responses, or harmful opinions being promoted within an already vulnerable group.

> *You have different...personalities, and I'd be worried if you did an online group that some people might be afraid to speak up because there could be other people that are quite boisterous.* [Caregiver 3, female, 49 years]

Participants emphasized the need for control over the visibility of information and the social media posts of others. They reported that social media news feeds can be jarring and compound emotional distress.

> *My sister joined this Facebook group for...brain tumour and she...added me along. I've sort of been a little bit up and down about that...because it just comes up on my news feed and then you know some...things that I read...is okay, but then...some things that I read that aren't so great, I sort of tune off a little bit you know.* [Patient 9, female, 38 years]

Similar to in-person peer support groups, digital peer communities need to consider the implications of varying prognoses due to different tumor types. When asked about different digital peer support formats, such as a "buddy" program, participants expressed concerns that those who bond with other unwell individuals may open themselves to grief if their "online buddy" dies or becomes too unwell to participate.

> *What happens when something happens to my buddy, you know, like, what does that gonna traumatize? [sic]...I don't know.* [Patient 12, female, 39 years]

Integration of Emergent Findings Into the Platform

The participants' insights were used to inform the development of content for our web-based platform now called "Brain Tumours Online." The Brain Tumours Online platform features three key elements or "pillars": (1) a repository of vetted, evidence-based information about a myriad of health, social care, and administrative supports available to patients and their caregivers ("LEARN"); (2) a digital peer support community to enable connections with other patients, caregivers, and health care professionals ("CONNECT"); and (3) an emerging library of validated digital symptom management therapeutic solutions ("TOOLBOX"). Herein, we describe how participants' insights were synthesized and addressed by the various components or "pillars" of the Brain Tumours Online platform.

Based on the emotional and psychological and social connectedness themes, participants' perspectives on face-to-face and digital peer support informed the design of our web-based peer community under the CONNECT pillar. Our digital peer community mitigates the access barriers (eg, time and geographical constraints, privacy concerns) for people who noted challenges with in-person support groups. Our platform's CONNECT pillar provides a range of avenues for sharing stories, experiences, and knowledge, for example, topic-specific chat forums, users' individual posts, and digital webinars. The CONNECT pillar also provides an avenue for sharing other patients' and caregivers' past hopeful stories, addressing the unmet need for positive stories that foster a sense of hope. The participants' preference for a safe space was also translated into the implementation of a moderator group, which includes moderator-trained patients, caregivers, and health care professionals. The CONNECT pillar also provides health care professional interface, including via digital webinars, and in their capacity as community moderators, to answer any questions on chat forums and vet information.

Within the information theme, participants expressed the unmet need for trusted and bespoke medical and social care information. Our platform's LEARN pillar provides a consolidated repository of brain tumor expert-endorsed information that is relevant to the Australian context, with signposting to specific tumor types, life stages, and personal situations. Examples of information available on the platform range from social care options for childcare, treatment-related travel subsidies for patients in remote and regional areas, eligibility for motor vehicle driving, and what to expect in various biomedical treatment options. The LEARN pillar contains a combination of curated existing external (trustworthy and vetted) resources and new resources freshly developed for

our digital health platform where an existing suitable resource was not already available.

Within the physical and practical theme, participants noted their unmet needs for resources and interventions to manage symptom burden. These needs are addressed by the Brain Tumours Online platform's TOOLBOX pillar. The TOOLBOX is intended to become a growing library of validated digital therapeutic tools for symptom management. The first tool that is currently available in the TOOLBOX is "Somryst," a Food and Drug Administration–approved digital therapeutic tool for chronic insomnia. Additional tools will be incorporated into the platform TOOLBOX in due course, in response to our user community's needs.

Discussion

Principal Findings

In this study, we set out to generate ideas and concepts to inform the development of a digital health tool to support the unmet needs of Australian patients with brain tumors, their caregivers, and treating health professionals. The data revealed 4 themes of unmet support needs that could be addressed by a digital health tool or platform: emotional and psychological, information, physical and practical, and social connectedness. Participants expressed the desire for early and proactive mental health intervention, noted the importance of providing mental health support to caregivers, and emphasized the need for positive stories and affirmative language. From an information perspective, participants noted a sense of information overload, especially at the beginning. They also underscored the variety of information needed on an ongoing basis, including life after treatment, and comprehensive care assistance to maintain quality of life. Participants also described unmet supportive care needs relating to symptom burden, practical and administrative support to facilitate the logistics of accessing treatment, and accomplishing daily life tasks such as work and study. Finally, they expressed the desire for greater social connectedness and safe spaces to engage with other people in a similar situation.

Comparison With Prior Work

The unmet supportive care needs identified in our study echo some of the previous literature published in this area. For example, Janda et al [20] conducted a qualitative exploration of the supportive care needs of patients with brain tumors and their caregivers. They identified the need for greater practical support (eg, support with financial issues and dealing with government agencies), the need for information and coping with uncertainty, as well as support to deal with social isolation. They found that technology could be a helpful avenue for patients in obtaining information and supporting caregivers. Our Brain Tumours Online platform seeks to provide a digital peer support mechanism to help mitigate social isolation and bridge information access gaps.

Indeed, the literature also emphasizes the changing needs of patients and caregivers over time. Previous studies indicate that information and mental health support needs are often greatest at the start and can change over time [4]. Thus, past literature emphasizes the need for bespoke and adaptable information and supportive care resources that can assist patients and caregivers in accordance with their evolving needs over time. Furthermore, past studies also indicate patients and caregivers wish to stay abreast of the latest developments in research and treatment of brain tumors [21]. To this end, our Brain Tumours Online platform seeks to serve as a living resource that can support patients and caregivers at different stages of diagnosis, treatment, and posttreatment living, and present a synthesis of the latest developments in this field.

Based on past research, it appears that a key factor in the mitigation of patients' and their caregivers' unmet needs was the awareness of available psychosocial support services and consequently their service use [22]. Therefore, a key aim of the Brain Tumours Online platform will be to provide a consolidated set of evidence-based resources on a variety of subjects ranging from availing social care services through government agencies to managing physical symptoms through allied health supports.

While our research findings are consistent with the past literature, our team recognized that existing tools for cancer care have often had little direct consumer engagement in their conceptualization, development, and implementation [7]. Thus, our qualitative exploration reported in this study was essential to validate previous literature and ensure that the emergent Brain Tumours Online platform meets the needs of Australian patients with brain tumors and their caregivers.

Limitations

First, it should be noted that some of the unmet needs expressed by our study's participants related to more systemic medical practice challenges, and could not be feasibly addressed through our digital solution, for example, the timeliness of referrals to mental health services. Nevertheless, by providing access to a supportive community, a vetted repository of trusted medical and psychosocial information, and easily accessible digital therapeutics, the Brain Tumours Online platform seeks to bridge comprehensive care and access gaps for a vulnerable patient and caregiver community. Second, we note that this study was aimed at adult participants only. Due to human research ethical constraints, and the scope of our study, pediatric patients were not included. The unmet needs of pediatric patients will need to be explored in a future program of work.

Finally, we note that there were also some limitations associated with our methodological approach. To minimize participant burden, particularly for patients and caregivers, we conducted interviews over the telephone. However, the conduct of interviews over the telephone, as opposed to in-person or via video call, meant that there was limited to no opportunity for identifying nonverbal communication such as body language and facial cues, which can be invaluable in qualitative lived experience research. Similarly, the opportunity for building personal rapport between the researcher and participant is limited in a telephone interview as compared to an in-person interview, which can in turn impact how freely participants share their personal experiences.

Conclusions

To our knowledge, when we began this project, there were no comprehensive digital health supportive care solutions for

Australian patients with brain tumors and their caregivers. Building on our prior research on the quality of life of patients with brain tumors [2,3,13], their patterns of social media use for disease management [8], and needs and expectations from a digital health model of care [7], we engaged with Australian patients with brain tumors, their caregivers, and treating health professionals to co-design a digital health solution to address the unmet supportive care needs of this population. Participants' insights were distilled to develop content for a web-based supportive care platform called "Brain Tumours Online." The platform comprises three pillars—(1) LEARN: a repository of vetted information about a range of biomedical and psychosocial care topics; (2) CONNECT: a digital peer support community with a health care professional interface; and (3) TOOLBOX: an emerging library of validated digital therapeutics for symptom management.

Acknowledgments

The authors would like to thank Meredith Layton, Ann Borda, Olivia Metcalf, Kit Huckvale, and Teresa O'Brien for their input on this paper and its foundational work. The authors thank consumer experts Diana Andrew, Julia Fin, Samantha Wright, and Janet Micaleff for providing their expertise and perspectives to inform this work. Additionally, the authors would like to thank the members of our Design Reference Group and the other lived experience experts who participated in our co-design activities. The study described in this paper was supported by the methodological expertise of the Digital Health Validitron, a collaborative and interdisciplinary research group at the University of Melbourne that assists digital health innovators from health care, academia, and industry to accelerate the creation of evidence that proves the real-world value of their ideas and products. This research received grant funding from the Australian Government through the Medical Research Future Fund 2020 Brain Cancer Survivorship Grant Program (MRFBCII000014).

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the sensitive nature of the data.

Authors' Contributions

M Kalla contributed to conceptualization, methodology, validation, formal analysis, investigation, writing (including original draft, reviews, and editing), and project administration. AB was involved in data curation, formal analysis, investigation, and writing (original draft, review, and editing). VS contributed to conceptualization, methodology, investigation, validation, formal analysis, investigation, writing (original draft, review, and editing), project administration, and funding acquisition. KB was involved in writing, which included the original draft, reviews, and editing. SCEB contributed to conceptualization, methodology, validation, formal analysis, investigation, writing (original draft, review, and editing), and project administration. SC assisted in writing, review, and editing. HM was involved in conceptualization and writing. RSD participated in project conceptualization, writing and editing, and funding acquisition. WC contributed to conceptualization and writing (review and editing). JRW assisted with conceptualization, writing (review and editing), and funding acquisition. KJD was involved in project conceptualization, writing (review and editing), funding acquisition, and overall project supervision. M Krishnasamy helped with project conceptualization, methodology, validation, writing (review and editing), funding acquisition, and overall project supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview questions for health professional cohort.
[DOCX File, 39 KB - cancer_v11i1e53690_app1.docx]

Multimedia Appendix 2
Interview questions for carer cohort.
[DOCX File, 37 KB - cancer_v11i1e53690_app2.docx]

Multimedia Appendix 3
Interview questions for patient cohort.
[DOCX File, 39 KB - cancer_v11i1e53690_app3.docx]

References

1. Brain cancer in Australia statistics. Cancer Australia. URL: https://www.canceraustralia.gov.au/cancer-types/brain-cancer/brain-cancer-australia-statistics [accessed 2025-05-15]

2. Nassiri F, Price B, Shehab A, et al. Life after surgical resection of a meningioma: a prospective cross-sectional study evaluating health-related quality of life. Neuro Oncol 2019 Jan 14;21(Suppl 1):i32-i43. [doi: 10.1093/neuonc/noy152] [Medline: 30649488]
3. Teng KX, Price B, Joshi S, et al. Life after surgical resection of a low-grade glioma: a prospective cross-sectional study evaluating health-related quality of life. J Clin Neurosci 2021 Jun;88:259-267. [doi: 10.1016/j.jocn.2021.03.038] [Medline: 33992194]
4. Halkett GKB, Lobb E, Spilsbury K, Dhillon H, Nowak AK. Brain cancer patients' levels of distress and supportive care needs over time. Psychooncology 2022 Dec;31(12):2074-2085. [doi: 10.1002/pon.6028] [Medline: 36086830]
5. Halkett GKB, Berg MN, Daudu D, et al. Supportive care of patients diagnosed with high grade glioma and their carers in Australia. J Neurooncol 2022 May;157(3):475-485. [doi: 10.1007/s11060-022-03991-z] [Medline: 35397081]
6. Heinsch M, Cootes H, Wells H, et al. Supporting friends and family of adults with a primary brain tumour: a systematic review. Health Social Care Comm 2022 May;30(3):869-887. [doi: 10.1111/hsc.13586]
7. McAlpine H, Joubert L, Martin-Sanchez F, Merolli M, Drummond KJ. A systematic review of types and efficacy of online interventions for cancer patients. Patient Educ Couns 2015 Mar;98(3):283-295. [doi: 10.1016/j.pec.2014.11.002] [Medline: 25535016]
8. McAlpine H, Sejka M, Drummond KJ. Brain tumour patients' use of social media for disease management: current practices and implications for the future. Patient Educ Couns 2021 Feb;104(2):395-402. [doi: 10.1016/j.pec.2020.07.012] [Medline: 32771243]
9. Delavar A, Al Jammal OM, Maguire KR, Wali AR, Pham MH. The impact of rural residence on adult brain cancer survival in the United States. J Neurooncol 2019 Sep;144(3):535-543. [doi: 10.1007/s11060-019-03254-4] [Medline: 31385185]
10. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018 Jan;4. [doi: 10.1177/2055207618770325]
11. Lee K, Kim S, Kim SH, et al. Digital health interventions for adult patients with cancer evaluated in randomized controlled trials: scoping review. J Med Internet Res 2023 Jan 6;25:e38333. [doi: 10.2196/38333] [Medline: 36607712]
12. Cox A, Lucas G, Marcu A, et al. Cancer survivors' experience with telehealth: a systematic review and thematic synthesis. J Med Internet Res 2017 Jan 9;19(1):e11. [doi: 10.2196/jmir.6575] [Medline: 28069561]
13. Nayak MG, George A, Vidyasagar MS, et al. Quality of life among cancer patients. Indian J Palliat Care 2017;23(4):445-450. [doi: 10.4103/IJPC.IJPC_82_17] [Medline: 29123353]
14. Noorbergen TJ, Adam MTP, Roxburgh M, et al. Co-design in mHealth systems development: insights from a systematic literature review. AIS Trans Hum Comput Interact 2021;13(2):175-205. [doi: 10.17705/1thci.00147]
15. Sanders EN, Stappers PJ. Probes, toolkits and prototypes: three approaches to making in codesigning. CoDesign 2014 Jan 2;10(1):5-14. [doi: 10.1080/15710882.2014.888183]
16. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Analyzing Qualitative Data: Routledge; 2002:173-194. [doi: 10.4324/9780203413081_chapter_9]
17. Fitch MI. Supportive care framework. Can Oncol Nurs J 2008;18(1):6-24. [doi: 10.5737/1181912x181614] [Medline: 18512565]
18. Parkinson S, Eatough V, Holmes J, Stapley E, Midgley N. Framework analysis: a worked example of a study exploring young people's experiences of depression. Qual Res Psychol 2016 Apr 2;13(2):109-129. [doi: 10.1080/14780887.2015.1119228]
19. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523. [doi: 10.1016/j.socscimed.2021.114523]
20. Janda M, Eakin EG, Bailey L, Walker D, Troy K. Supportive care needs of people with brain tumours and their carers. Support Care Cancer 2006 Nov;14(11):1094-1103. [doi: 10.1007/s00520-006-0074-1] [Medline: 16710653]
21. Janda M, Steginga S, Dunn J, Langbecker D, Walker D, Eakin E. Unmet supportive care needs and interest in services among patients with a brain tumour and their carers. Patient Educ Couns 2008 May;71(2):251-258. [doi: 10.1016/j.pec.2008.01.020] [Medline: 18329220]
22. Langbecker D, Yates P. Primary brain tumor patients' supportive care needs and multidisciplinary rehabilitation, community and psychosocial support services: awareness, referral and utilization. J Neurooncol 2016 Mar;127(1):91-102. [doi: 10.1007/s11060-015-2013-9] [Medline: 26643806]

Abbreviations

HREC: Human Research Ethics Committee

Edited by N Cahill; submitted 16.10.23; peer-reviewed by A Goodman, E Evbuomwan; revised version received 17.03.25; accepted 20.03.25; published 23.05.25.

Please cite as:
Kalla M, Bradford A, Schadewaldt V, Burns K, Bray SCE, Cain S, McAlpine H, Dhillon RS, Chapman W, Whittle JR, J Drummond K, Krishnasamy M
Co-Designing a User-Centered Digital Health Tool for Supportive Care Needs of Patients With Brain Tumors and Their Caregivers: Interview Analysis
JMIR Cancer 2025;11:e53690
URL: https://cancer.jmir.org/2025/1/e53690
doi:10.2196/53690

Development and Implementation of a Personal Virtual Assistant for Patient Engagement and Communication in Postsurgical Cancer Care: Feasibility Cohort Study

Miguel Bargas-Ochoa[1], MD; Alejandro Zulbaran-Rojas[1], MD; M G Finco[2], PhD; Anthony B Costales[3], MD; Areli Flores-Camargo[1], MD; Rasha O Bara[1], BS; Manuel Pacheco[1], BS; Tina Phan[1], BS; Aleena Khichi[1], BS; Bijan Najafi[1,4], PhD

[1]Digital Health Access Center (DiHAC), Michael E DeBakey Department of Surgery, Baylor College of Medicine, 7200 Cambridge st, Houston, TX, United States

[2]Department of Physical Therapy, University of North Texas Health Science Center, Fort Worth, TX, United States

[3]Division of Gynecologic Oncology, Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, United States

[4]Center for Advanced Surgical & Interventional Technology (CASIT), Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, 700 Westwood Plaza, Suite 2200, Los Angeles, CA, United States

Corresponding Author:
Bijan Najafi, PhD
Digital Health Access Center (DiHAC), Michael E DeBakey Department of Surgery, Baylor College of Medicine, 7200 Cambridge st, Houston, TX, United States

Abstract

Background: Cancer-care complexity heightens communication challenges between health care providers and patients, impacting their treatment adherence. This is especially evident upon hospital discharge in patients undergoing surgical procedures. Digital health tools offer potential solutions to address communication challenges seen in current discharge protocols. We aim to explore the usability and acceptability of an interactive health platform among discharged patients who underwent oncology-related procedures.

Methods: A 4-week exploratory cohort study was conducted. Following hospital discharge, a tablet equipped with an integrated Personal Virtual Assistant (PVA) system was provided to patients who underwent oncology-related procedures. The PVA encompasses automated features that provide personalized care plans, developed through collaboration among clinicians, researchers, and engineers from various disciplines. These plans include guidance on daily specific assignments that were divided into 4 categories: medication intake, exercise**,** symptom surveys, and postprocedural specific tasks. The aim was to explore the acceptability of the PVA by quantification of dropout rate and assessing adherence to each care plan category throughout the study duration. The secondary aim assessed acceptability of the PVA through a technology acceptance model (TAM) questionnaire that examined ease of use, usefulness, attitude toward use, and privacy concerns.

Results: In total, 17 patients were enrolled. However, 1 (5.8%) patient dropped out from the study after 3 days due to health deterioration, leaving 16/17 (94.2%) completing the study (mean age 54.5, SD 12.7, years; n=9, 52% Caucasian; n=14, 82% with a gynecological disease; n=3, 18% with a hepatobiliary disease). At the study end point, adherence to care plan categories were 78% (SD 25%) for medications, 81% (SD 24%) for exercises, 61% (SD 30%) for surveys, and 58% (SD 44%) for specific tasks such as following step-by step wound care instructions, managing drains, administering injectable medications independently, and performing pelvic baths as instructed. There was an 80% patient endorsement (strongly agree or agree) across all TAM categories.

Conclusion: This study suggests the potential acceptability of the PVA among patients discharged after oncology-related procedures, with a dropout rate of less than 6% and fair-to-good adherence to tasks such as medication intake and exercise. However, these findings are preliminary due to the small sample size and highlight the need for further research with larger cohorts to validate and refine the system.

(JMIR Cancer 2025;11:e64145) doi:10.2196/64145

KEYWORDS

digital health; personal virtual assistant; remote patient monitoring; surgical oncology; posthospital discharge; postoperative support; medication adherence postsurgery; patient engagement; mHealth; mobile health

Introduction

Ineffective communication between cancer patients and health care providers can result in heightened distress, compromised quality of life [1], reduced treatment adherence [2,3], and suboptimal quality of care [4]. The complexity of cancer care, coupled with the necessity for a multidisciplinary approach, exacerbates communication challenges in this population [4,5]. For instance, multiple health care professionals including surgeons, medical and radiation oncologists, pathologists, nurses, physical therapists, social workers, and nutritionists, amongst others, can be involved in one single treatment plan [6]. This multidisciplinary approach, while essential to provide comprehensive care, often results in multiple hospital follow-ups, varied medication regimens and heterogeneous instructions, which can lead to misunderstandings, failure of cancer treatment protocols, and physical and psychological harm [4].

These communication issues in cancer patients become especially evident in those undergoing surgical procedures and transitioning from hospital to home settings [7]. Upon hospital discharge, cancer patients are commonly provided with a list of extensive instructions based on written or printed summaries, which are lengthy and tedious [8-11], making it difficult to understand and follow [12,13]. In recent years, there has been a shift toward using computer generated programs to create quick and interactive materials in the form of educational websites, audio, and videos [14]. However, they still fall short in providing interactive and bidirectional communication with health providers.

During the COVID-19 pandemic, digital health emerged as an alternative to address communication challenges through technology platforms and remote health monitoring [15]. Digital tools such as telemonitoring, telemedicine, mobile health apps, and wearable devices [16] became useful tools for improving treatment compliance, symptom management, and patient communication. However, as these tools become more widely adopted, it has become evident that numerous challenges must be addressed to improve their usability, accessibility, and effectiveness, particularly in cancer care. Key challenges in adopting digital technologies for cancer patients include disparities in technology literacy, poor integration into clinical workflows, time-intensive processes, and limited bidirectional communication. Furthermore, content may be biased if created by those without health-related expertise. To address these challenges, a collaboration between academia (Baylor College of Medicine, Houston) and industry (Smartek21, Seattle) developed an interactive digital platform, called Personal Virtual Assistant (PVA). The PVA is designed to virtually coach patients in adhering to their postoperative care plans and enhance communication after hospital discharge. This exploratory study examined the usability and acceptability of the PVA among patients undergoing oncology-related procedures, aiming to identify its successful components, adoption barriers and areas for improvement.

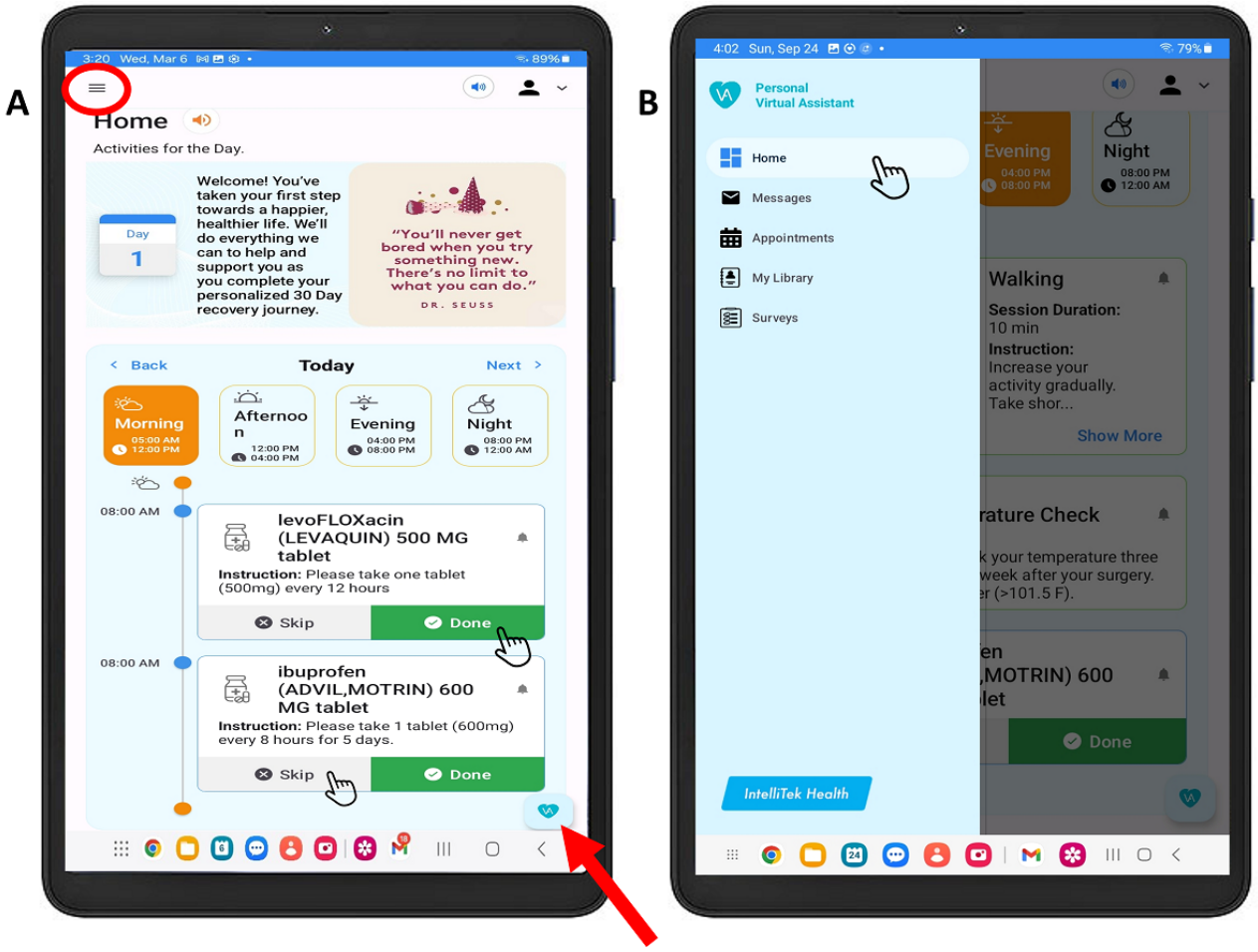
Methods

Personal Virtual Assistant Development

The PVA streamlines posthospital discharge care plans through 3 key elements: an interactive patient platform, a care provider portal, and a secure cloud backend interface.

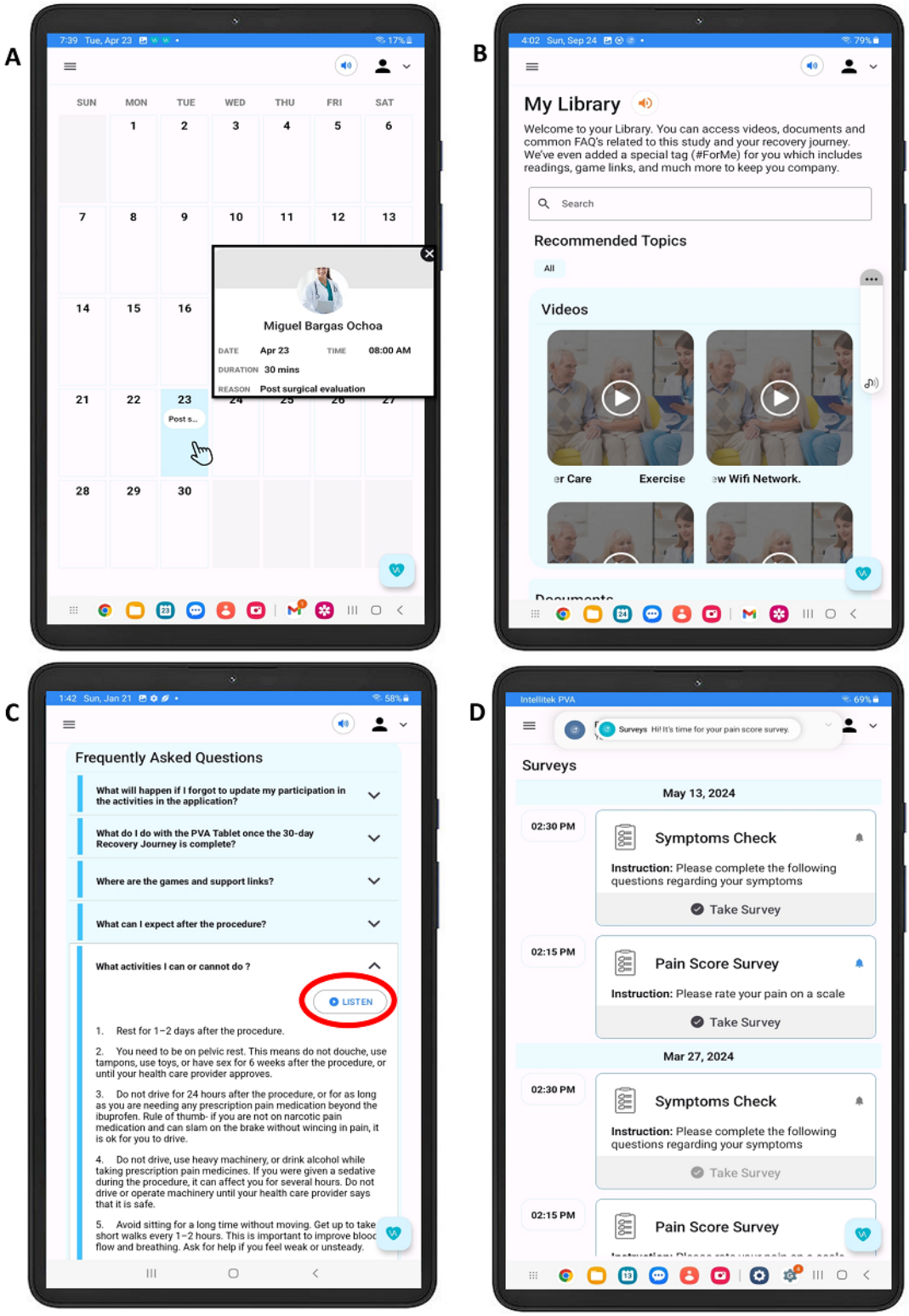
The Interactive Patient Platform uses an app integrated into a tablet that works through internet connection. Upon launching the app, the home screen displays the patient daily care plan (Figure 1A), which can include “prescribed medications,” “exercises,” “symptom surveys,” and “postprocedural specific tasks.” The assignments that appear on the home screen depend on the time and day of recovery. Patients have the option to navigate through these assignments by manual selection or by using natural language voice-commands upon pressing the “voice-command” button and speaking aloud the category name. In addition, the app has an alert notification system to remind patients to complete the assignments by marking each one as “done.”

Figure 1. Personal Virtual Assistant displaying the patients' portal home screen. (**A**) Personal Virtual Assistant displaying the home screen once opened. When scrolling down, all 4 categories (medications, exercises, symptom surveys, and specific tasks) are seen. The red circle represents the button to navigation sidebar and the red arrow depicts the voice-command button. (**B**) Navigation sidebar options displaying additional navigation tools such as messages, appointments, my library, and symptom surveys.



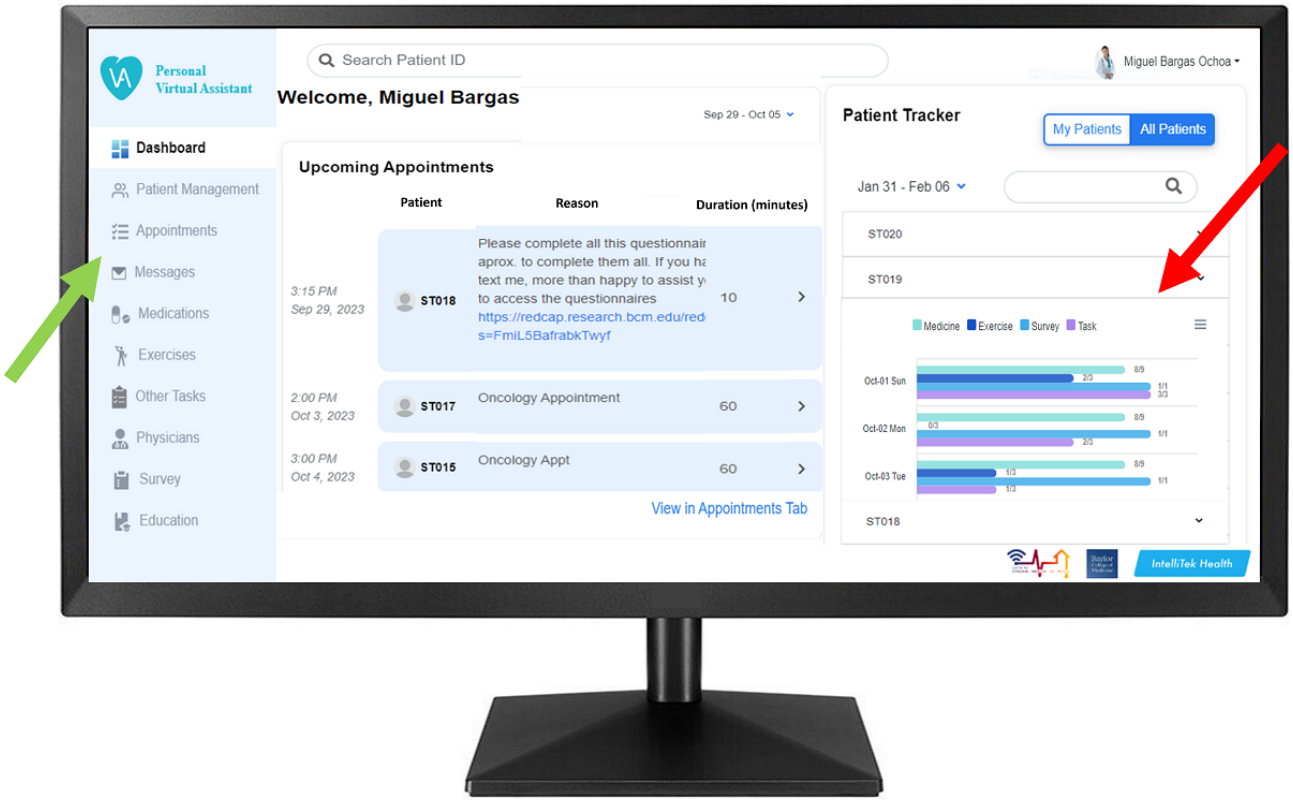
On the top left side of the screen, a menu sidebar button is available to allow patients to navigate into other aspects of their care plan such as "Appointments," "My Library," "Messages," and "Surveys," (Figure 1B). By manually selecting or speaking aloud "Appointments" on the PVA, a calendar appears displaying upcoming clinic visits. Then, more details are available upon selecting the desired appointment (Figure 2A). The "My library" tab offers educational video content covering various topics such as guidance on how to use the tablet, exercise during cancer care, and guidance on postsurgical recovery (Figure 2B). Within this tab, a frequently asked questions section is also available, in which patients have the option to listen to them outload by pressing the "listen" button located above the text (Figure 2C). In addition, the "My Messages" tab opens a chat box where patients can communicate with the user behind the care provider portal, in this case the clinical research team. Finally, the "My Surveys" tab redirects to the home-screen symptom surveys.

Figure 2. Personal Virtual Assistant displaying the patients' portal additional navigation tabs. (**A**) Appointment tab displays a calendar. Details appear upon clicking on the date. (**B**) My Library contains educational content about Personal Virtual Assistant and surgical recovery. (**C**) Frequently Asked Questions about technology troubleshooting and recovery process includes audio option for listening to these explanations. (**D**) Surveys patient must complete that inquire about symptoms and pain.



The Care Provider Portal is accessed through a website on any computer with an internet connection. In this study, the portal was personalized for use by the clinical research team. Once accessed by the user, a dashboard screen listing all patients with their upcoming appointments and weekly logins to the PVA system is displayed (Figure 3). On the left side of the screen, a menu sidebar including 10 different tabs allows the user to individually customize each patient care plan. For instance, in the "Patient Management" tab, the user can customize each patient's posthospital discharge instructions. In the "Medications" and "Exercises" tabs, users can manage drug dose, frequency, duration, and mobility sessions, respectively. All information provided through these tabs are stored as repositories, feeding an electronic library for future care plans.

Figure 3. Personal Virtual Assistant displaying the care provider portal. Main dashboard on any computer with internet connection. Green arrow: points to the navigation sidebar. Red arrow: points at the adherence tracker, which opens a drop-down menu that shows daily adherence for each of the 4 categories in the care plan.



The cloud backend is a secure cloud-based platform that allows for instant updates to care plans, sends alerts, monitors adherence to care plans, arranges appointments, enables communication, and creates a resource library with relevant content for patients and care providers.

Integration of Clinical Content Into the Personal Virtual Assistant

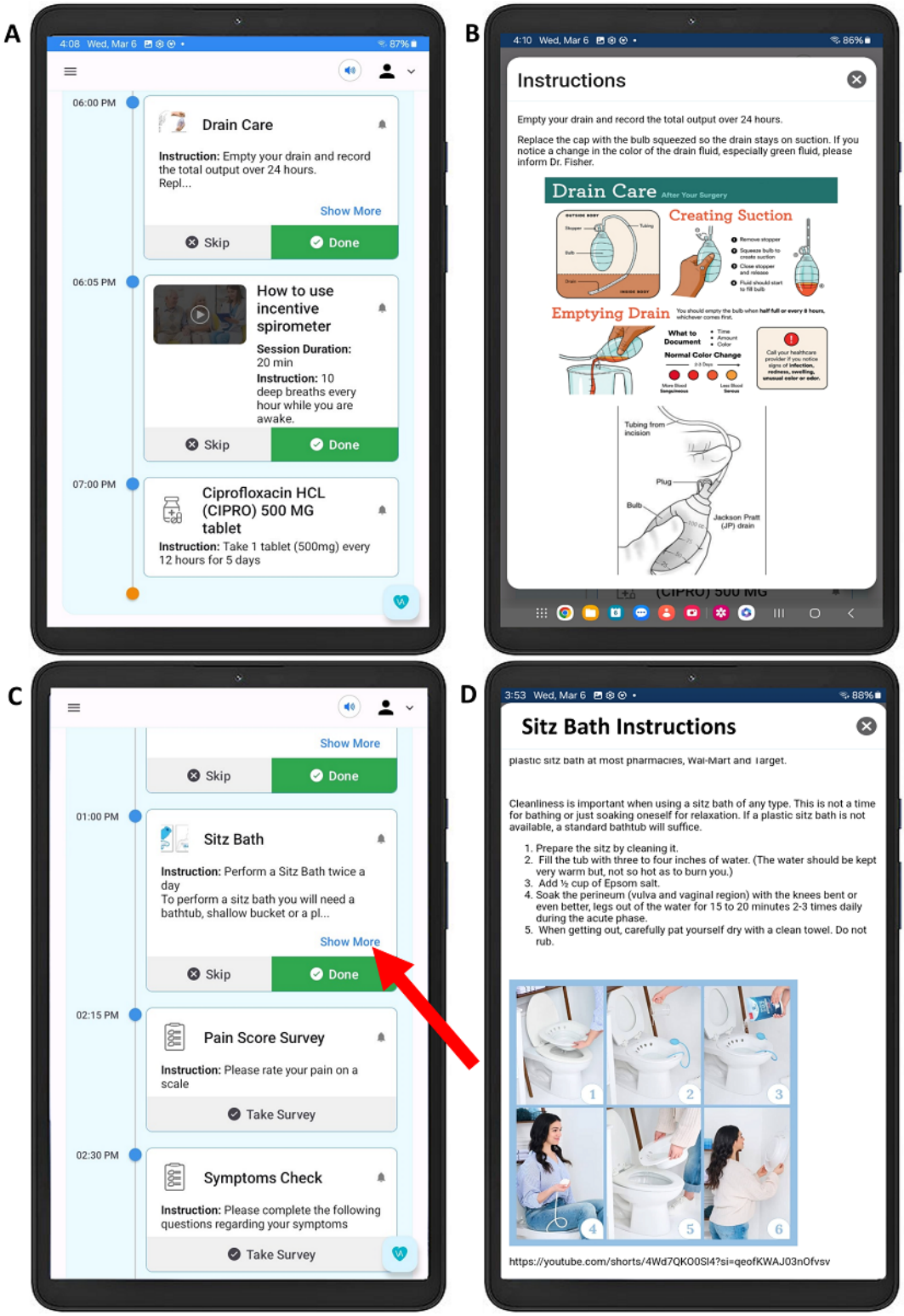
The PVA clinical content was developed by a multidisciplinary team including expert physicians in surgical oncology, medical oncology, pulmonology and critical care, and medical researchers from 1 academic institution (Baylor College of Medicine, Houston). As a result, specific discharge care content was created for 2 medical specialties: hepatobiliary surgery and gynecologic oncology. Engineers from 1 industry (Smartek21, Seattle) assigned the developed content into the different PVA features previously described.

The hepatobiliary surgery content was directed to pancreatic cancer procedures (ie, distal pancreatectomies and Whipple procedure). Content displayed in the home screen included prescribed medications, exercise guidance on respiratory therapy, symptom surveys, and specific tasks detailing wound care for surgical incisions, drain management, and video guides for self-administration of medications (Figure 4A and 4B).

The gynecologic oncology content focused on hysterectomies, radical vulvectomies, and cytoreductive surgeries. Content displayed in the home screen included prescribed medications, exercise guidance on lower extremity mobility (ie, walking, calf pumps, and leg raises), symptom surveys, and specific tasks detailing step-by-step incision care, and revision of instructions for healing of pelvic and genital areas (ie, sitz baths description, Figure 4C and 4D). In addition, information on postsurgery expectations, permissible and prohibited activities, warning signs requiring immediate medical attention, and recommended over-the-counter medications were included in the “My Library” tab.

In total, 2 symptom surveys were included in the content for both specialties, 1 monitored pain through a 10-point Likert scale, and the other monitored deep vein thrombosis (DVT) and assessed bleeding signs from mucosa or surgical incisions, using questions suggested by the clinical team.

Figure 4. Personal Virtual Assistant displaying the patients' portal "specific tasks" tab. (**A**) Home screen displaying the specific task "drain care" for patients undergoing hepatobiliary procedures. (**B**) Explanation of drain care (adapted from Go et al [17] which is published under a Creative Commons Attribution-NonCommercial International License [18]). (**C**) Home screen displaying the specific task "sitz bath" for patients undergoing gynecologic oncology procedures. (D) Upon clicking on the "show more button," an explanation in the form of text and images appears. Image taken from SitzBliss sitz bath (SitzBliss in AMAZON).



Study Design

To test the PVA's feasibility, a 4-week pilot prospective study in individuals undergoing hepatobiliary and gynecologic oncology procedures was performed between November 2022 and October 2023.

Eligible patients were aged ≥18 years old; scheduled to undergo ambulatory surgery due to suspected malignancy; willing to engage with a PVA for 4 weeks following hospital discharge. Patients were excluded if they had: major foot and ankle problems (eg, major amputation and severe neuropathy); unable to provide informed consent; documented (confirmed through

electronical medical records) major cognitive impairment, a psychiatric condition or abnormal laboratory results that, in the judgment of the clinical investigator, would interfere with the ability to participate in the study; active thrombotic condition or were using therapeutic anticoagulants; and were non-English speakers. In addition, individuals with uncorrected severe vision or hearing impairments that prevented them from effectively interacting with the PVA tablet, as determined by the investigators' judgment, were excluded.

Before their surgery, patients received a 10-minute tutorial on PVA navigation. At hospital discharge (baseline), the research team manually customized the patients' PVA content, incorporating the standard care plan detailed in their electronic medical record (EMR) by the attending surgeon. Demographic information was collected, and the patients were given the tablet with instructions to begin using it within 24hrs. After 4 weeks (end point), acceptability, and perceptions of the PVA were assessed by telemedicine, and adherence data was collected through a built-in tracking system within the PVA. Then, patients were asked to ship back the tablet. Weekly phone or video calls by the research team were performed during the study to address questions regarding the PVA functionality, content, or updates on their postoperative care according to their EMR.

Feasibility, Adherence, and Acceptability Outcomes

Feasibility was assessed by quantification of dropout rate set to ≤10% throughout the study period (4 weeks). Adherence was quantified by completion of assignments (marked as "done") in each of the 4 categories included in the care plan (ie, prescribed medications, exercises, symptom surveys, and specific tasks). For each patient, the percentage of each category was calculated by dividing the number of completed assignments by the number of days in which they were assigned during the study period.

Acceptability was assessed using a 12-item Technology Acceptance Model (TAM) questionnaire rated with a 5-point Likert Scale and ranging from strongly agree to strongly disagree, or very easy to very difficult. The TAM assessed 4 key areas: perceived ease of use, perceived usefulness, attitude toward use, and privacy concerns [19].

Furthermore, open-ended questions collected feedback on the user experience and suggestions for enhancing the PVA in Multimedia Appendix 1.

Ethical Considerations

Patients signed an informed consent at the Dan L Duncan Comprehensive Cancer Center at Baylor College of Medicine before study enrollment. The protocol was approved by Baylor College of Medicine's Institutional Review Board (protocol number: H-51654), and research procedures were conducted in accordance with the Declaration of Helsinki.

Results

Patient Characteristics

A total of 212 potential candidates were screened from the Duncan L Cancer Center at Baylor College of Medicine. Out of these, 166 were not eligible for surgery, 12 were non-English speakers, 7 did not respond to calls, and 7 declined to participate. This led to 20 eligible patients who consented and enrolled in the study, however 3 were withdrawn before study initiation due to surgery cancellation. In total, 17 patients initiated the study and received a tablet with the PVA app integrated. All patients were female (mean age 54.5, SD 12.7, years; mean BMI 33, SD 11.5, kg/m2). Out of these, 14 patients had gynecologic oncology procedures, and 3 had hepatobiliary oncology procedures. Ten out of 17 (59%) patients confirmed malignancy on biopsy (Table 1).

Table . Patient demographic information.

Characteristic			Participants (N=17)
Sex (female), n (%)			17 (100)
Age (years), mean (SD)			54.5 (12.7)
Race, n (%)			
	White		11 (6)
	Black		5 (29)
Ethnicity, n (%)			
	African American		5 (29)
	Caucasian		9 (52)
	Hispanic		2 (12)
	Asian		1 (7)
BMI, kg/m^2, mean (SD)			33 (11.5)
Pathology report, n (%)			
	Benign (n=7)		
		Gynecological	7 (100%)
	Malignant (n=10)		
		Gastrointestinal	3 (30%)
		Gynecological	7 (70%)

Feasibility and Adherence

One out of 17 patients (5.8%) dropped out from the study at postoperative day (POD) 3 due to health deterioration (reported feeling very weak and sick) and unwillingness to interact with the tablet. All other patients (16/17, 94.2%) completed the study and were included in the adherence and acceptability analyses. Among the analyzed patients, there was a mean adherence rate of 78% (SD 25%) for prescribed medications, 81 (SD 24%) for exercises, 61 (SD 30%) for symptom surveys, and 58 (SD 44%) for specific tasks. Detailed adherence data for each patient over the 4-week period is available in Multimedia Appendix 1. The responses to the open-ended questionnaire regarding technology acceptance, including perceived ease of use and perceived benefits, are available in Multimedia Appendix 2.

Acceptability

There was an 80% patient endorsement across all TAM categories (Table 2 and Table 3). The highest endorsed items were regarding the simplicity of managing daily tasks (93.8%) and medications (93.8%), in the perceived ease of use category. Perception assessment is depicted in Multimedia Appendix 1.

Table . Responses (n=16) to perceived ease of use in the Technology Acceptance Model (TAM) questionnaire of patients who completed 4 weeks.

Perceived ease of use	Responses, n					Endorsement (very easy + easy), %
	Very easy	Easy	Neutral	Difficult	Very difficult	
Navigating my patient engagement app	11	1	3	1	0	75
Managing my appointments	11	2	1	2	0	81
Managing my medications	12	3	1	0	0	94
Managing daily tasks	11	4	0	1	0	94
Managing my messages	11	3	2	0	0	88
Connecting to my video calls	8	1	6	1	0	56
Accessing my exercise guidance	10	4	4	0	0	88

Table . Responses (n=16) to perceived usefulness, attitude toward use, and privacy concerns in Technology Acceptance Model (TAM) questionnaire of patients who completed 4 weeks.

Category and item		Responses, n					Endorsement (strongly agree + agree), %
		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	
Perceived usefulness							
	I do not need the support of a technical person to use this app	10	1	1	3	1	69
	The patient engagement app is a useful resource in managing my post hospital care	8	4	4	0	0	75
	Reminders via the app effectively reminded me to complete my tasks	8	5	3	0	0	81
Attitude toward use							
	When this patient engagement app becomes available, I will use it	8	4	2	0	0	75
	I would recommend this patient engagement app to other friends or family members who are discharged from the hospital	8	5	3	0	0	81
Privacy concerns							
	I do not have privacy concerns while using this patient engagement app	10	0	4	1	1	62

Reported Events During Study

Events were retrospectively collected via EMR in patients who completed the study. Three events were found; all were noted from patients undergoing gynecologic oncology procedures. One patient reported a falling incident to the research staff by phone call, but not through the PVA. This event did not require hospital readmission. Furthermore, 1 patient visited the emergency room (ER) on POD-17 due to sepsis secondary to acute cystitis or intra-abdominal abscess (per EMR note). Symptoms included pain during urination, lower back pain, and chills, which were not reported in the PVA chat. Symptom surveys recorded pain levels of 2/10 and 3/10 days before the ED visit. Treatment included computed tomography–guided needle aspiration, drain placement, and antibiotics. Another patient visited the ER on POD-28 due to lower abdominal pain and was diagnosed with an intramuscular rectus sheath hematoma (per EMR note). The patient experienced pain the previous week but did not report it in the PVA chat. Pain levels recorded in the PVA surveys were 7/10 just 2 days before the ER visit. The patient reported ineffective pain control through EMR chat (EPIC MyChart) and was advised to visit the ER by the health care team. Treatment consisted of analgesics and hospital observation for 24 hours.

Discussion

Principal Findings

This study explored the usability and acceptability of providing individualized postoperative care plans through a PVA system in patients undergoing hepatobiliary and gynecologic oncology procedures. This system is composed of 3 key elements, a tablet

with an integrated app provided to the patient (patient portal), a website for care providers to manage from any computer (care provider portal), and a secure cloud backend system that links both portals. Through the care provider's portal, the research team entered the patients' postoperative care plans assigned by their attending physician which included guidance on daily personalized assignments divided into 4 categories (ie, medications, exercise, symptom surveys, and specific tasks). Then, a tracking system assessed the adherence to such categories. In total, 16 (93.7%) patients completed the study, showing an 80% acceptability rate evaluated through a technology acceptance model questionnaire. Adherence to the care plan surpassed 70% of assignments completed in the medications and exercise categories.

Digital health applications for postoperative recovery have shown efficacy for enhancing communication between patients and care providers in different areas [20,21]. Strategies used for such enhancement include symptom monitoring surveys [21-29] educational videos for wound care [17] and exercise implementation [30], which support patients in complying with their postoperative management. Other apps may alert care providers to detect warning signs for faster communication of adverse events [21,29,31]. Our multidisciplinary team integrated these features in an interactive, practical, and simple manner through a personalized tablet (Figure 2). This strategy ensures that compliance notifications remain separate from personal devices, thereby minimizing the risk of overlooking care management assignments. In addition, this system's integrated "voice-enabled" navigation commands facilitate usability in those who have challenges on interpreting and inputting text. Emphasis was placed on presenting care plan assignments in an organized format, starting with the home screen, with the objective of creating a daily routine that patients could complete. Importantly, all features were integrated through direct consultation with clinicians, something we believe was crucial for patient engagement. Subhi et al [32] emphasized that without adequate professional input, digital health tools may deliver content that fails to meet patients' needs and deploy interventions that are ineffective. Thus, the success of digital health tools relies on the proper combination of evidence-based systems with realistic content that involves individualized care plans guided by clinical minds.

Today, few authors have explored digital health applications for postsurgical oncologic care, all reporting high feasibility. Graetz et al [31] designed a mobile app for 26 patients undergoing gynecologic oncology procedures, incorporating progressive reminders regarding discharge instructions, medication adherence and completion of symptom surveys. The study was performed in 4 weeks, resulting in 88% of participants completing the study. In a similar population, Temple-Oberle et al [33] conducted a 6-week randomized controlled trial using a mobile app, in which patients uploaded wound pictures, drain volume data, reported symptoms or wound complications, and received unidirectional messages from their physicians. The approach was compared to standard of care, having only 1/36 patients (2%) in the intervention group dropping out of the study. Similarly, the majority of participants in the present 4-week study who used the PVA were those undergoing gynecological oncologic procedures, with 13/14 (93.7%) completing the study. We attribute this rate to the personalized content that was created for each patient. For instance, gynecologic patients undergoing surgery are recommended to perform sitz baths to relieve pain, swelling, and improve wound healing in the pelvic area [34]. However, proper instructions on the right equipment, water level and temperature, duration, and frequency of baths, are often forgotten [35]. The PVA reminds and guides patients to perform this task by notifications and banners on the screen (Figure 4C). This notification system has been shown to increase usability of digital health tools in gynecologic oncology patients [31]. Noteworthy, it is difficult to ascertain if the notification system influenced patient compliance and study completion the most. In addition, the known low-risk of postoperative complications in the group of gynecologic oncology patients could have been another factor contributing to a less pronounced dropout rate [31,33].

Another important aspect for assessing feasibility of digital health tools is the quantification of adherence to such systems. For instance, Mata et al [36] considered adherence as completion of 5 specific daily tasks (ie, early mobilization, gum chewing, consumption of oral liquids, breathing exercises, and consumption of protein drink) by following instructions from a tablet in 40 hospitalized patients who underwent colorectal surgery. Among these, 60% had a cancer diagnosis. Interestingly, a 94% adherence rate was seen on POD-1 but declined to 43% at POD-3. Similarly, Low et al [37] measured in-hospital adherence among patients undergoing abdominal cancer surgery, based on symptom survey responses via a mobile app and quantification of daily usage of a smart band (Fitbit). Adherence rates reported were 22% and 35%, respectively. However, when measuring adherence in the posthospital discharge phase, both rates increased (41 and 65%, respectively). This highlights how the timing and setting of digital health tools usage can impact adherence, particularly in those employed once patients have cleared hospitalization. In this study, posthospital discharge adherence was evaluated upon completing the assignments within the 4 care plan categories. This resulted in a high adherence rate for medications and exercises (78% and 81%, respectively). However, the patients' freedom to mark each assignment as "done" could have biased our results, as it is difficult to verify whether these were truly completed. On the other hand, there was a 61% and 58% adherence rate for symptom surveys and specific tasks, respectively. Perhaps this was reflected on the 46% rate of patients who recommended including fewer surveys and questions in the PVA in Multimedia Appendix 1. Nonetheless, an interventional study evaluating clinical outcomes (ie, faster recovery, adverse events, and hospital readmission) is warranted to confirm associations with adherence.

To evaluate acceptability, the present study used a TAM questionnaire tailored for the PVA system, showing an 88.75% ease of use, 86.62% perceived usefulness, 85% attitude toward use, and 81% privacy concerns rates. No notable differences in technology acceptability were observed between younger (<60 years old) and older (>60 years old) patients, with overall scores appearing similar across age groups. While a small number in both groups indicated needing technical support or gave neutral

responses regarding app usage, these differences were minimal and not statistically analyzed. Other cancer digital health apps using acceptability queries have shown similar high rates. Karlsson et al [38] utilized the System Usability Scale revealing a 77% ease of use score for an app that encourages mobility in patients recovering from abdominal cancer surgery (Pedatim, Phystec), surpassing the threshold (68 points) for high acceptability of this query [39]. Hwang et al [40] used the Patient Satisfaction Survey to evaluate an app that enables remote monitoring of wounds and communication with care providers in patients undergoing breast cancer surgery (Medeo), revealing a 90% ease of use, 90% attitude toward use, 95% perceived usefulness, and 100% of privacy concern rates. These results reflect the potential use of digital health tools in cancer patients recovering from surgical procedures and encourage researchers for future and continuous development of such systems.

Although the present study focused on exploring the PVA's feasibility and acceptability, we sought to retrospectively collect incidents that happened during the study period to understand the challenges that can be addressed for future system improvement of the PVA. For example, one patient was readmitted to the hospital on POD-28 due to rectus sheath hematoma. Interestingly, this patient reported a poor adherence rate in all categories (39% for medications, 61% for exercises, 54% for symptom surveys, and 0% to specific tasks). Despite patients being instructed to directly communicate postoperative incidents or complications to their care providers, our system's alternative monitoring options (ie, bidirectional chat and Likert pain scales) failed to collect such incidents. In fact, neither the patient's care providers nor the PVA system had evidence of the three reported incidents. We attribute this to the lack of warning sign surveys evaluating fever, wound complications, urinary tract infections, hematomas, or ileus (in gynecologic procedures) [41]. We also believe that proper patient education on warning signs should be included in the home screen. In addition, the system could be equipped with advanced monitoring features such as vital signs (ie, temperature, heart rate, and respiratory rate) and symptom reporting surveys integrated with automated alarms to alert users when certain thresholds are exceeded. These technological enhancements could facilitate prompt detection of postoperative complications to avoid hospital readmissions.

The study findings highlight opportunities to enhance the clinical application of this interactive digital technology, particularly in improving adherence to symptom reporting and supporting the successful completion of specific tasks, that may benefit from more detailed, step-by-step education incorporating comprehensive images or videos. Acceptability and usability are essential first steps in deploying interactive digital solutions to support posthospital discharge care plans. While these solutions hold promise for preventing surgical complications and reducing hospital readmissions, such outcomes remain the ultimate goals of these technologies and require further validation. Research suggests that remote patient monitoring during postsurgical recovery phases can significantly lower hospital readmissions and minimize unnecessary clinic visits compared to standard of care [40]. We speculate that the PVA could meaningfully improve clinical outcomes by enhancing patient adherence to postdischarge care plans, including medication adherence, exercise routines, symptom reporting, and task completion. However, this speculation requires validation through future interventional studies. Further research should also explore how the PVA can optimize communication between patients and care providers, fostering a more seamless and effective recovery process.

Limitations

There are several limitations to our study. Major limitations of this study are its exploratory design, small sample size, and the absence of a control group. These factors restrict our ability to evaluate the PVA's effectiveness compared to standard postdischarge care, limiting our understanding of its overall impact. The low dropout rate and high adherence observed could be attributed to the weekly monetary compensation offered to patients, which may have influenced their participation levels. The questions included in the TAM for evaluating participant's adherence and perceptions might have introduced an acquiescence bias, with patients predisposed to agree with the assessment statements. In addition, we did not examine variations in technology acceptance across different age groups, ethnicities, or cancer types. Such an analysis could provide valuable insights and broaden the applicability of our findings. In the next phase of the study, we plan to incorporate demographic analysis to better understand PVA acceptability and tailor its use to diverse patient populations. Another significant limitation was the lack of integration between the PVA system and electronic medical records, requiring research teams to manually enter each participant's discharge plan. This manual process was not only time-consuming but also involved transcribing detailed instructions in each category of the care plan. Even though the system adherence tool identifies which specific assignments within each category of the care plan (ie, medications, exercise, tasks, and surveys) have been completed by the patient, the clinical study coordinator or health care provider must continuously revise this data on a daily basis. In addition, the health care provider portal, although intended for provider use, was managed by the research team. Thus, the absence of direct contact between patients and health care providers could have affected their experience or willingness to engage fully with the PVAs messaging tool. Regarding the surveys used in the PVA (ie, pain scales and DVT symptoms), patients' answers require continuous revision on a daily basis by the user. However, the research team only reviewed these surveys at the end of the study, highlighting a critical limitation. The absence of real-time symptom collection likely affected the ability to assess the PVA's effectiveness in early prevention and detection of common postsurgical complications.

To address this limitation, future iterations will focus on redesigning surveys to facilitate real-time symptom reporting and efficient review by the clinical care team. This redesign aims to enhance usability while avoiding time-consuming processes, ultimately improving the PVA's role in patient care and early intervention. Furthermore, the app requires internet connection to function and receive real-time updates to the care plan. At the time of this study, it was not available on Google Play Store or Apple platforms, which meant that updates had

to be manually downloaded through a link provided by our industry partner (Smartek21) from the web. Finally, while the PVA allows for picture uploads through a secure chat with researchers, this feature was not instructed as part of the study protocol. We believe this feature should be used in future studies, as it is crucial for the prompt identification of clinical warning signs [21,22,24,25,27-29,42].

Conclusion

This exploratory study demonstrated the usability and acceptability of an interactive digital solution designed to provide an organized, step-by-step guide for postsurgical care, and simplify adherence to care plans in patients undergoing gynecological oncology and hepatobiliary oncology procedures. The findings suggest perceived acceptability, ease of use, and intention to use the platform, although privacy concerns remain a limitation for broader scalability. The results showed fair-to-good adherence to certain postdischarge tasks, such as recommended exercises and prescribed medications, while adherence to symptom surveys and specific tasks was notably lower. Integrating additional features, such as notification reminders and voice-enabled systems, appeared promising for improving compliance. Furthermore, information retrospectively gathered from postsurgical complications provided valuable insights for enhancing the PVA system in future iterations. These findings underscore the potential for interactive digital health solutions to improve communication between patients and care providers while coaching patients to adhere to prescribed postdischarge tasks, which may ultimately enhance recovery outcomes. However, these observations are preliminary and need to be confirmed in studies with larger sample sizes. In addition, future research should focus on validating the effectiveness of this solution in improving posthospital discharge outcomes.

Acknowledgments

The authors would like to thank Drs William E Fisher, Ernest Ramsay Camp, Ang Li, Amir Sharafkhaneh, and Feibi Zheng from Baylor College of Medicine for their expert advice in evaluating and building content for the Personal Virtual Assistant (PVA). The authors would also like to thank Martyn Molnar, Paul Costello, and Dr Sammer Sood from SmartTek21 and IntelliTek Health for their in-kind and technical support for PVA used in this study. We also extend our gratitude to Maria Noun and Naima Rodriguez for their support with the institutional review board and operations, and to the Digital Health Access Center team as well as the clinical team at the Baylor College of Medicine (BCM) Dan L Duncan Comprehensive Cancer Center and the BCM St. Luke’s Medical Center for their assistance with patient referral and recruitment. This research was supported by the National Science Foundation's Industry-University Cooperative Research Centers program under award numbers NSF 2052514 and C2SHIP (Center to Stream HealthCare in Place) Y01-BCM-P026. Additional support was provided by the National Institutes of Health, National Cancer Institute under award number 5P30CA125123-17S1, and in-kind support from Smartek21 and IntelliTek Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors.

Authors' Contributions

MBO contributed to writing – original draft, writing – review and editing, patient recruitment, data collection, visualization, PVA connect development, and formal analysis. AZR handled writing – original draft, and writing – review and editing. MGF managed review and editing, and data analysis. ABC handled patient referral, patient care, PVA content development, and review and editing. AFC conducted patient recruitment, data collection, PVA content development, and review and editing. ROB managed patient recruitment, data collection, PVA content development, and review and editing. MP conducted patient recruitment, data collection, and review and editing. TP contributed to patient recruitment, data collection, data analysis, and review and editing. AK managed patient recruitment, data collection, data analysis, and review and editing. BN contributed to conceptualization, writing – review and editing, and supervision.

Conflicts of Interest

BN serves as a consultant for BioSensics LLC on projects unrelated to the scope of this study. The other authors report no conflicts of interest relevant to this study.

Multimedia Appendix 1
Feedback from open-ended questions on the user experience and suggestions for enhancing the Personal Virtual Assistant.
[DOCX File, 16 KB - cancer_v11i1e64145_app1.docx]

Multimedia Appendix 2
Responses to the open-ended questionnaire regarding technology acceptance, including perceived ease of use and perceived benefits.
[DOCX File, 15 KB - cancer_v11i1e64145_app2.docx]

References

1. Thorne SE, Bultz BD, Baile WF, SCRN Communication Team. Is there a cost to poor communication in cancer care?: a critical review of the literature. Psychooncology 2005 Oct;14(10):875-884. [doi: 10.1002/pon.947] [Medline: 16200515]
2. DiMatteo MR. Future directions in research on consumer-provider communication and adherence to cancer prevention and treatment. Patient Educ Couns 2003 May;50(1):23-26. [doi: 10.1016/s0738-3991(03)00075-2] [Medline: 12767580]
3. Tiwary A, Rimal A, Paudyal B, Sigdel KR, Basnyat B. Poor communication by health care professionals may lead to life-threatening complications: examples from two case reports. Wellcome Open Res 2019;4(7):7. [doi: 10.12688/wellcomeopenres.15042.1] [Medline: 31448336]
4. Prouty CD, Mazor KM, Greene SM, et al. Providers' perceptions of communication breakdowns in cancer care. J Gen Intern Med 2014 Aug;29(8):1122-1130. [doi: 10.1007/s11606-014-2769-1] [Medline: 24599795]
5. Clauser SB, Wagner EH, Aiello Bowles EJ, Tuzzio L, Greene SM. Improving modern cancer care through information technology. Am J Prev Med 2011 May;40(5 Suppl 2):S198-S207. [doi: 10.1016/j.amepre.2011.01.014] [Medline: 21521595]
6. Trapani D, Murthy SS, Boniol M, et al. Distribution of the workforce involved in cancer care: a systematic review of the literature. ESMO Open 2021 Dec;6(6):100292. [doi: 10.1016/j.esmoop.2021.100292] [Medline: 34763251]
7. Patel SJ, Landrigan CP. Communication at transitions of care. Pediatr Clin North Am 2019 Aug;66(4):751-773. [doi: 10.1016/j.pcl.2019.03.004] [Medline: 31230621]
8. Berry DL, Cunningham T, Eisenberg S, Wickline M, Hammer M, Berg C. Improving patient knowledge of discharge medications in an oncology setting. Clin J Oncol Nurs 2014 Feb;18(1):35-37. [doi: 10.1188/14.CJON.35-37] [Medline: 24476724]
9. de Leeuw J, Prins JB, Merkx MAW, Marres HAM, van Achterberg T. Discharge advice in cancer patients. Cancer Nurs 2011;34(1):58-66. [doi: 10.1097/NCC.0b013e3181efef18]
10. DeSai C, Janowiak K, Secheli B, et al. Empowering patients: simplifying discharge instructions. BMJ Open Qual 2021 Sep;10(3):e001419. [doi: 10.1136/bmjoq-2021-001419] [Medline: 34521621]
11. Zavala S, Shaffer C. Do patients understand discharge instructions? J Emerg Nurs 2011 Mar;37(2):138-140. [doi: 10.1016/j.jen.2009.11.008] [Medline: 21397126]
12. Chen DW, Davis RW, Balentine CJ, et al. Utility of routine postoperative visit after appendectomy and cholecystectomy with evaluation of mobile technology access in an urban safety net population. J Surg Res 2014 Aug;190(2):478-483. [doi: 10.1016/j.jss.2014.04.028] [Medline: 24880202]
13. Katz MHG, Slack R, Bruno M, et al. Outpatient virtual clinical encounters after complex surgery for cancer: a prospective pilot study of "TeleDischarge". J Surg Res 2016 May 1;202(1):196-203. [doi: 10.1016/j.jss.2015.12.054] [Medline: 27083967]
14. Newnham H, Barker A, Ritchie E, Hitchcock K, Gibbs H, Holton S. Discharge communication practices and healthcare provider and patient preferences, satisfaction and comprehension: a systematic review. Int J Qual Health Care 2017 Oct 1;29(6):752-768. [doi: 10.1093/intqhc/mzx121] [Medline: 29025093]
15. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 6;22(11):e22280. [doi: 10.2196/22280] [Medline: 33079693]
16. Lee K, Kim S, Kim SH, et al. Digital health interventions for adult patients with cancer evaluated in randomized controlled trials: scoping review. J Med Internet Res 2023 Jan 6;25:e38333. [doi: 10.2196/38333] [Medline: 36607712]
17. Go BC, Brewster R, Patel R, Rajasekaran K. Using telemedicine and infographics for physician-guided home drain removal. OTO Open 2020;4(2):2473974X20933566. [doi: 10.1177/2473974X20933566] [Medline: 32548544]
18. Attribution-noncommercial 4.0 international (CC BY-NC 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by-nc/4.0/ [accessed 2025-02-13]
19. Davis FD. A technology acceptance model for empirically testing new end-user information systems: theory and results [Dissertation]. : Massachusetts Institute of Technology Library; 1985 URL: http://hdl.handle.net/1721.1/15192 [accessed 2025-02-03]
20. De La Cruz Monroy MFI, Mosahebi A. The use of smartphone applications (apps) for enhancing communication with surgical patients: a systematic review of the literature. Surg Innov 2019 Apr;26(2):244-259. [doi: 10.1177/1553350618819517] [Medline: 30602332]
21. Felbaum DR, Stewart JJ, Anaizi AN, Sandhu FA, Nair MN, Voyadzis JM. Implementation and evaluation of a smartphone application for the perioperative care of neurosurgery patients at an academic medical center: implications for patient satisfaction, surgery cancelations, and readmissions. Operative Surg 2018;14(3):303-311. [doi: 10.1093/ons/opx112]
22. Armstrong KA, Coyte PC, Brown M, Beber B, Semple JL. Effect of home monitoring via mobile app on the number of in-person visits following ambulatory surgery: a randomized clinical trial. JAMA Surg 2017 Jul 1;152(7):622-627. [doi: 10.1001/jamasurg.2017.0111] [Medline: 28329223]
23. Debono B, Bousquet P, Sabatier P, Plas JY, Lescure JP, Hamel O. Postoperative monitoring with a mobile application after ambulatory lumbar discectomy: an effective tool for spine surgeons. Eur Spine J 2016 Nov;25(11):3536-3542. [doi: 10.1007/s00586-016-4680-4] [Medline: 27349754]
24. Gunter RL, Fernandes-Taylor S, Rahman S, et al. Feasibility of an image-based mobile health protocol for postoperative wound monitoring. J Am Coll Surg 2018 Mar;226(3):277-286. [doi: 10.1016/j.jamcollsurg.2017.12.013] [Medline: 29366555]

25. Higgins J, Semple J, Murnaghan L, Sharpe S, Theodoropoulos J. Mobile web-based follow-up for postoperative ACL reconstruction: a single-center experience. Orthop J Sports Med 2017 Dec;5(12):2325967117745278. [doi: 10.1177/2325967117745278] [Medline: 29318171]
26. Jaensson M, Dahlberg K, Eriksson M, Nilsson U. Evaluation of postoperative recovery in day surgery patients using a mobile phone application: a multicentre randomized trial. Br J Anaesth 2017 Nov 1;119(5):1030-1038. [doi: 10.1093/bja/aex331] [Medline: 29077818]
27. Scott AR, Alore EA, Naik AD, Berger DH, Suliburk JW. Mixed-methods analysis of factors impacting use of a postoperative mHealth app. JMIR Mhealth Uhealth 2017 Feb 8;5(2):e11. [doi: 10.2196/mhealth.6728] [Medline: 28179215]
28. Semple JL, Sharpe S, Murnaghan ML, Theodoropoulos J, Metcalfe KA. Using a mobile app for monitoring post-operative quality of recovery of patients at home: a feasibility study. JMIR Mhealth Uhealth 2015 Feb 12;3(1):e18. [doi: 10.2196/mhealth.3929] [Medline: 25679749]
29. Symer MM, Abelson JS, Milsom J, McClure B, Yeo HL. A mobile health application to track patients after gastrointestinal surgery: results from a pilot study. J Gastrointest Surg 2017 Sep;21(9):1500-1505. [doi: 10.1007/s11605-017-3482-2] [Medline: 28685388]
30. Bäcker HC, Wu CH, Pförringer D, Petersen W, Stöckle U, Braun KF. A review of functional outcomes after the app-based rehabilitation of patients with TKA and THA. J Pers Med 2022 Aug 21;12(8):1342. [doi: 10.3390/jpm12081342] [Medline: 36013291]
31. Graetz I, Anderson JN, McKillop CN, Stepanski EJ, Paladino AJ, Tillmanns TD. Use of a web-based app to improve postoperative outcomes for patients receiving gynecological oncology care: a randomized controlled feasibility trial. Gynecol Oncol 2018 Aug;150(2):311-317. [doi: 10.1016/j.ygyno.2018.06.007] [Medline: 29903391]
32. Subhi Y, Bube SH, Rolskov Bojsen S, Skou Thomsen AS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. JMIR Mhealth Uhealth 2015 Jul 27;3(3):e79. [doi: 10.2196/mhealth.4169] [Medline: 26215371]
33. Temple-Oberle C, Yakaback S, Webb C, Assadzadeh GE, Nelson G. Effect of smartphone app postoperative home monitoring after oncologic surgery on quality of recovery: a randomized clinical trial. JAMA Surg 2023 Jul 1;158(7):693-699. [doi: 10.1001/jamasurg.2023.0616] [Medline: 37043216]
34. Habek D, Cerovac A, Kamerić L, Nevačinović E, Šerak A. Balneogynaecology in the 21st century: increasingly recommended primary and complementary treatment of chronic gynaecological diseases. Med Glas (Zenica) 2020 Oct 9;18(1):1-6. [doi: 10.17392/1263-21]
35. Tejirian T, Abbas MA. Sitz bath: where is the evidence? Scientific basis of a common practice. Dis Colon Rectum 2005 Dec;48(12):2336-2340. [doi: 10.1007/s10350-005-0085-x] [Medline: 15981059]
36. Mata J, Pecorelli N, Kaneva P, et al. A mobile device application (app) to improve adherence to an enhanced recovery program for colorectal surgery: a randomized controlled trial. Surg Endosc 2020 Feb;34(2):742-751. [doi: 10.1007/s00464-019-06823-w] [Medline: 31087175]
37. Low CA, Danko M, Durica KC, et al. A real-time mobile intervention to reduce sedentary behavior before and after cancer surgery: usability and feasibility study. JMIR Perioper Med 2020 Mar 23;3(1):e17292. [doi: 10.2196/17292] [Medline: 33393915]
38. Karlsson P, Nygren-Bonnier M, Henningsohn L, Rydwik E, Hagströmer M. The feasibility of using a digital tool to enhance mobilisation following abdominal cancer surgery-a non-randomised controlled trial. Pilot Feasibility Stud 2023 Aug 23;9(1):147. [doi: 10.1186/s40814-023-01377-0] [Medline: 37612761]
39. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
40. Hwang H. Electronic wound monitoring after ambulatory breast cancer surgery: improving patient care and satisfaction using a smart phone app. BC Med J 2016 [FREE Full text]
41. Clarke-Pearson DL, Geller EJ. Complications of hysterectomy. Obstet Gynecol 2013 Mar;121(3):654-673. [doi: 10.1097/AOG.0b013e3182841594] [Medline: 23635631]
42. Gunter R, Fernandes-Taylor S, Mahnke A, et al. Evaluating patient usability of an image-based mobile health platform for postoperative wound monitoring. JMIR Mhealth Uhealth 2016 Sep 28;4(3):e113. [doi: 10.2196/mhealth.6023] [Medline: 27683059]

Abbreviations

DVT: deep vein thrombosis
EMR: electronic medical record
ER: emergency room
POD: postoperative day
PVA: Personal Virtual Assistant
TAM: Technology Acceptance Model

Edited by J Lau, K Gandhi; submitted 10.07.24; peer-reviewed by A Grilo, S Mashatian; revised version received 27.11.24; accepted 12.12.24; published 18.02.25.

Please cite as:
Bargas-Ochoa M, Zulbaran-Rojas A, Finco MG, Costales AB, Flores-Camargo A, Bara RO, Pacheco M, Phan T, Khichi A, Najafi B
Development and Implementation of a Personal Virtual Assistant for Patient Engagement and Communication in Postsurgical Cancer Care: Feasibility Cohort Study
JMIR Cancer 2025;11:e64145
URL: https://cancer.jmir.org/2025/1/e64145
doi:10.2196/64145

Evaluation of Large Language Models in Tailoring Educational Content for Cancer Survivors and Their Caregivers: Quality Analysis

Darren Liu[1,2], MS; Xiao Hu[1,2], PhD; Canhua Xiao[1,3], RN, PhD; Jinbing Bai[1,3], RN, MSN, PhD; Zahra A Barandouzi[1,3], RN, PhD; Stephanie Lee[1], OCN, RN, MS; Caitlin Webster[1], CPHON, RN, MSN; La-Urshalar Brock[1,3], CNM, FNP-BC, MSN; Lindsay Lee[4], MD; Delgersuren Bold[1,2], MS; Yufen Lin[1,3], RN, PhD

[1]Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Rd NE, Atlanta, GA, United States

[2]Center for Data Science, Emory University, Atlanta, GA, United States

[3]Winship Cancer Institute, Emory University, Atlanta, GA, United States

[4]Department of Medicine, University of Florida, Gainesville, FL, United States

Corresponding Author:
Yufen Lin, RN, PhD
Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Rd NE, Atlanta, GA, United States

Abstract

Background: Cancer survivors and their caregivers, particularly those from disadvantaged backgrounds with limited health literacy or racial and ethnic minorities facing language barriers, are at a disproportionately higher risk of experiencing symptom burdens from cancer and its treatments. Large language models (LLMs) offer a promising avenue for generating concise, linguistically appropriate, and accessible educational materials tailored to these populations. However, there is limited research evaluating how effectively LLMs perform in creating targeted content for individuals with diverse literacy and language needs.

Objective: This study aimed to evaluate the overall performance of LLMs in generating tailored educational content for cancer survivors and their caregivers with limited health literacy or language barriers, compare the performances of 3 Generative Pretrained Transformer (GPT) models (ie, GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo; OpenAI), and examine how different prompting approaches influence the quality of the generated content.

Methods: We selected 30 topics from national guidelines on cancer care and education. GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo were used to generate tailored content of up to 250 words at a 6th-grade reading level, with translations into Spanish and Chinese for each topic. Two distinct prompting approaches (textual and bulleted) were applied and evaluated. Nine oncology experts evaluated 360 generated responses based on predetermined criteria: word limit, reading level, and quality assessment (ie, clarity, accuracy, relevance, completeness, and comprehensibility). ANOVA (analysis of variance) or chi-square analyses were used to compare differences among the various GPT models and prompts.

Results: Overall, LLMs showed excellent performance in tailoring educational content, with 74.2% (267/360) adhering to the specified word limit and achieving an average quality assessment score of 8.933 out of 10. However, LLMs showed moderate performance in reading level, with 41.1% (148/360) of content failing to meet the sixth-grade reading level. LLMs demonstrated strong translation capabilities, achieving an accuracy of 96.7% (87/90) for Spanish and 81.1% (73/90) for Chinese translations. Common errors included imprecise scopes, inaccuracies in definitions, and content that lacked actionable recommendations. The more advanced GPT-4 family models showed better overall performance compared to GPT-3.5 Turbo. Prompting GPTs to produce bulleted-format content was likely to result in better educational content compared with textual-format content.

Conclusions: All 3 LLMs demonstrated high potential for delivering multilingual, concise, and low health literacy educational content for cancer survivors and caregivers who face limited literacy or language barriers. GPT-4 family models were notably more robust. While further refinement is required to ensure simpler reading levels and fully comprehensive information, these findings highlight LLMs as an emerging tool for bridging gaps in cancer education and advancing health equity. Future research should integrate expert feedback, additional prompt engineering strategies, and specialized training data to optimize content accuracy and accessibility.

International Registered Report Identifier (IRRID): RR2-10.2196/48499

(JMIR Cancer 2025;11:e67914) doi:10.2196/67914

KEYWORDS

large language models; GPT-4; cancer survivors; caregivers; education; health equity

Introduction

More than 18.1 million individuals with a history of cancer were alive in the United States in 2022, and that number is projected to reach 26 million by 2040 [1]. Cancer survivors receive a wide range of treatments, often experiencing severe symptoms or side effects, including fatigue, depression, anxiety, sleep disturbance, pain, cognitive impairment, nausea, vomiting, and neuropathy [2-7]. These symptoms negatively impact survivors' functional status, quality of life, and overall survival rates [8-11]. Cancer caregivers, typically family members or significant others offering primary emotional and physical support for cancer survivors, experience an array of similar distressing symptoms [12-14]. These symptoms are linked to high caregiving burden, emotional distress, and communication barriers with cancer survivors and providers [15]. In addition, disparities in health care access further exacerbate the challenges faced by cancer survivors and their caregivers, especially those from disadvantaged communities that have limited health literacy or language barriers [16]. Those with limited health literacy and racial and ethnic minorities facing language barriers are at greater risk for poorer access to care [17-19]. Consequently, they tend to experience a heavier symptom burden and poorer health outcomes during and after cancer treatments [20].

With over 3-quarters of the disadvantaged population owning smartphones or computers [21], technology-based intervention programs can bridge the accessibility gap and promote health equity [22,23]. The advent and growth of artificial intelligence have enabled researchers to design tailored and personalized interventions and educational content to meet individual unmet needs [24]. Large language models (LLMs) are advanced artificial intelligence systems that can understand and generate human-like text by training on vast amounts of data [25]. LLMs perform various language tasks, such as answering questions and translating languages. How questions are asked can significantly affect the performance of LLMs. This process, known as prompt engineering, is crucial for obtaining accurate and relevant responses from LLMs [26,27]. While LLMs have demonstrated remarkable potential in cancer research [28-31], their efficacy in real-world scenarios, such as cancer care and education, which often require advanced levels of comprehension, have yet to be thoroughly assessed.

Recent advancements in LLMs, such as GPT-4 and GPT-4 Turbo (OpenAI) [32,33], have demonstrated their exceptional proficiency in completing various tasks, including coding, design, and content summarization. Previous research [34,35] indicates that LLMs can capture large volumes of text effectively, even without specialized domain knowledge. This ability highlights its sophistication in processing and understanding information across a broad spectrum of topics, and its potential to significantly aid in analyzing unstructured data in clinical environments (eg, clinical notes) [34,35]. However, there are several notable gaps in the current knowledge. First, while LLMs have demonstrated high levels of accuracy in understanding extensive texts [34,36,37], even minor inaccuracies can have detrimental effects on patient outcomes [38], particularly regarding actionable advice. Therefore, the content they generate still necessitates additional expert verification to ensure it is error-free and ready to be presented to patients and their caregivers. Second, although previous research [36,37] has demonstrated promising results in content summarization, these LLMs are often not applied in clinical environments, or they specifically address cancer care and education among disadvantaged groups that has limited health literacy or language barriers [39]. Finally, most educational resources for cancer care are available exclusively in English, which can create comprehension challenges for non-English speakers (eg, Hispanic individuals and immigrants). Also, cancer survivors and their caregivers, already overwhelmed by treatment, often lack the time to read lengthy content. Therefore, it is essential to provide educational content in multiple languages and in concise content to ensure effective communication and education [40].

To address these gaps, our team aimed to evaluate how LLMs perform in tailoring educational content to enhance accessibility and comprehension for cancer survivors and their caregivers. In this study, our primary task was to evaluate and compare the capabilities of multiple GPT-based LLMs in generating concise, low-literacy-level, and multilingual educational content tailored for cancer survivors and their caregivers with limited health literacy or language barriers. Specifically, we aimed to evaluate the overall performance of LLMs in generating tailored educational content that adheres to a strict word limit, a sixth-grade reading level, and high-quality criteria (clarity, accuracy, relevance, completeness, and comprehensibility), compare the performances of 3 GPT models (GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo), and explore how different prompt structures (textual vs bulleted format) influence the quality of the generated content. This approach helps them manage their symptoms more effectively, thereby reducing health disparities and promoting health equity.

Methods

Design

This study involved a multistep methodology that included: (1) specifying the exact task requirements for the LLMs, to produce educational content on 30 selected cancer care topics written at a sixth-grade reading level, limited to 250 words, and translated into Spanish and Chinese; (2) generating tailored educational content using 3 GPT models (GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo) with 2 prompt styles (textual and bulleted); (3) expert evaluation of the generated content's adherence to word count, reading level, and 5 quality criteria; and (4) statistical analyses (ANOVA [analysis of variance] and chi-square test) to compare performance across models and prompt formats.

Prompt Engineering

To promote the accessibility and comprehension of educational content for cancer survivors and their caregivers with limited health literacy and language barriers, we structured prompts to have LLMs produce content at a low reading level, maintain a word limit of 250, and provide Spanish and Chinese translations for each topic, as described below [41].

The Flesch-Kincaid Grade Level (FKG) system [42] was used to assess the readability of content produced by the LLMs. The FKG level is a readability test designed to indicate how difficult a text is to understand. It calculates the grade level required for someone to comprehend the text. The FKG is based on word length and sentence length, providing a numerical score that corresponds to US grade levels [42]. The National Institutes of Health (NIH) and the American Medical Association (AMA) suggest that patient education materials should be written at a reading level no higher than the sixth grade [43]. This recommendation is in place to guarantee that the information is reachable by a broad spectrum of individuals, encompassing those with limited health literacy. Therefore, our research targets an FKG level of 6 to align with this guidance.

We set a 250-word limit for our educational content, recognizing that cancer survivors and their caregivers are frequently preoccupied with treatment schedules and daily responsibilities, leaving them with limited time for reading [44]. This word limit is designed to ensure that participants can complete the reading within 5 minutes, making the task both manageable and feasible within their schedules.

Furthermore, it is shown that prompts exert a considerable impact on the responses generated by LLMs [45]. Therefore, we compared different prompts, including both textual and bulleted formats, to determine which approach yields better results.

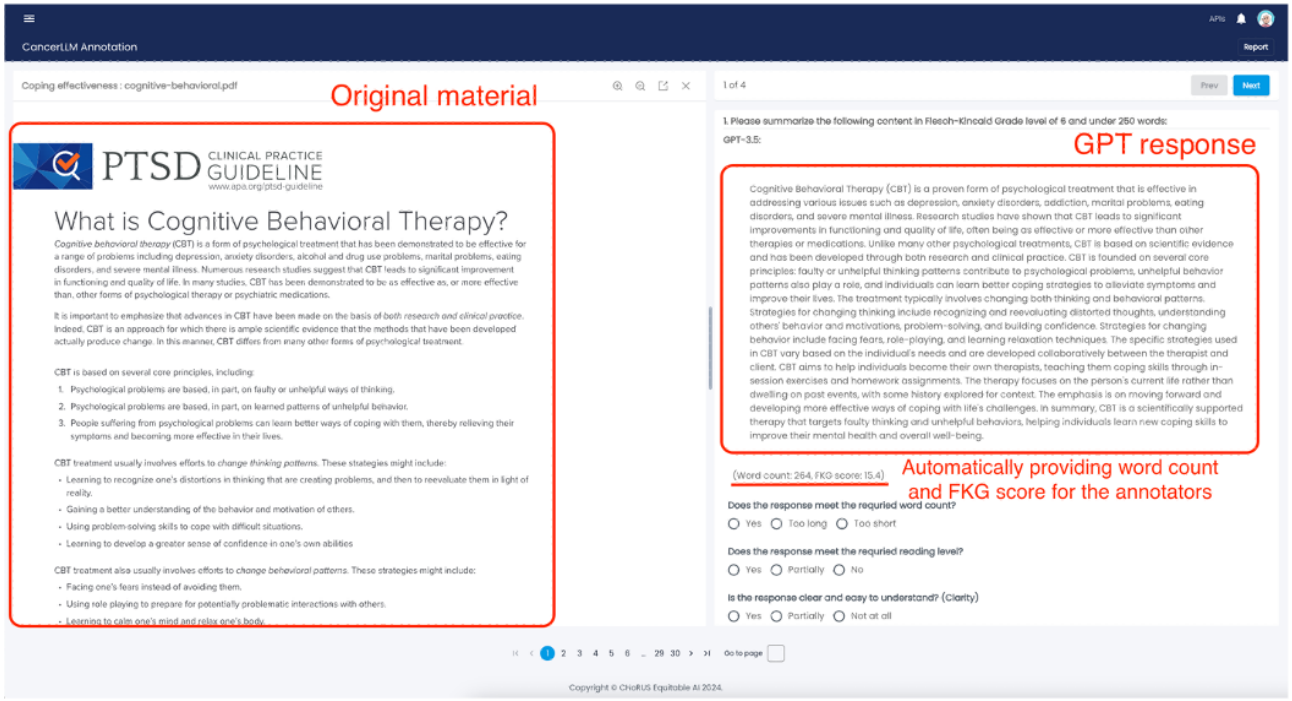
The prompts we used are as follows:

1. Textual format: “Please summarize the following content in Flesch-Kincaid Grade level of 6 and under 250 words: [original text]”
2. Bulleted format: “Please summarize the following content into bullet points in Flesch-Kincaid Grade level of 6 and under 250 words: [original text]”
3. Spanish translation: “Please translate the following content into Spanish: [tailored text]”
4. Chinese translation: “Please translate the following content into Chinese: [tailored text]”

Expert Evaluation

We assembled a panel of 9 oncology experts, comprising 4 oncology professors, 4 doctoral students, and 1 medical resident. Among them, all are fluent in English, with 4 experts proficient in Chinese and 1 proficient in Spanish. Each response generated by the LLMs was evaluated by at least 2 experts to ensure a comprehensive assessment, except for the Spanish translation task, which was evaluated by a single expert. The panel conducted several Zoom meetings: the initial meeting provided training on content evaluation, and 3 additional meetings were held to discuss the results and feedback. Each expert was assigned 10 topics to evaluate and was required to provide feedback on the errors committed by the LLMs. These experts critically reviewed and annotated the LLM-generated content using a web-based Cohort Adjudication and Data Annotation (CADA) application [34] (Figure 1) developed by our team.

Figure 1. A screenshot of Cohort Adjudication and Data Annotation application.



Data Sources

Our primary sources for content generation were cancer survivors and caregiver education materials from the National Cancer Institute and the National Comprehensive Cancer Network guidelines [46,47]. We selected 30 distinct topics covering a range of content such as fatigue, depression, anxiety, pain, cognitive impairment, nutrition, physical activity, healthy lifestyle, family communication, coping skills, and more. The selection of topics was informed by insights from our previous qualitative interviews with cancer survivors and their caregivers [48] and an extensive review of the literature [49-51]. We identified the key areas of interest and specific needs of cancer

survivors and their caregivers with limited health literacy or language barriers, resulting in these 30 topics.

Appraisal Criteria

Based on a previous study of evaluating responses from LLMs [34], we formulated a set of multidimensional criteria to thoroughly assess the performance of LLMs, which include adherence to a word limit of 250 words, achieving a reading level as per the FKG of below 6, and quality assessment: (1) clarity (ie, ease of understanding in the response); (2) accuracy (ie, the response does not contain errors, like medical or language errors, that could negatively impact patients and their caregivers); (3) relevance (ie, the response is fully grounded in the materials we provided); (4) completeness (ie, the response encompasses all critical points from the materials); (5) comprehensibility (ie, the response is understandable that readers can apply it to their daily routine).

In terms of word limit, "yes" refers to a word limit within 250 words, and "no" refers to a word limit of more than 250 words. The reading level was evaluated using "yes" for an FKG level ≤6; "partial" for an FKG level of 6 to ≤8; and "no" for an FKG level >8). The FKG level was calculated by the Python package Textstat (version 0.7.3, Azu). For the quality assessment criteria, we implemented a scoring system in which evaluations were quantified based on their alignment with the expected outcomes. A score of 2 was assigned for "yes" evaluations, indicating full compliance; a score of 1 was given for "partial" evaluations, reflecting partial compliance; and a score of 0 was allocated for "no" evaluations, indicating noncompliance. The quality assessment included 5 criteria (1-5), each contributing a maximum of 2 points, for a total possible score of 10. The overall quality assessment ranged from 0 to 10, with 0 representing the absence or lowest quality and 10 indicating the highest quality. For translation tasks, "yes" indicates a completely accurate translation, "partially" refers to a generally correct and understandable translation with minor errors, and "no" refers to a completely inaccurate translation containing incorrect or misleading information. Accuracy scores are calculated as the proportion of evaluations labeled as "yes."

Data Analysis

Descriptive analyses were conducted to determine the frequencies, percentages (for word limit, reading levels, and translations), mean and SDs (for quality scores) of major variables. Quality scores were determined by calculating the mean scores for each criterion and then obtaining the overall scores through their summation. To compare the differences in each model or prompt, we used ANOVA or chi-square tests, as applicable. Values of $P<.05$ were considered to indicate a significant level. All analyses were conducted using Python statistical packages.

Ethical Considerations

The study protocol (STUDY00004750) was approved with exemptions by the institutional review board at Emory University. Oral consent was obtained from 9 oncology experts, as no protected health information was collected. All participants were informed of the voluntary nature of their participation and their right to withdraw at any time without consequence. No protected health information or personally identifiable information was collected, and all research data were anonymized to maintain confidentiality. Study materials were securely stored and accessible only to authorized research team members. Participants did not receive any monetary or nonmonetary compensation for their involvement. The study was conducted in accordance with the US Common Rule (45 CFR 46) [52].

Results

Overall Performance of Large Language Models

In this study, 360 annotation values were collected from 9 experts. Overall, LLMs have shown excellent performance in tailoring content based on our criteria. For word limit, 267/360 responses (74.2%) were within the word limit (less than 250 words) set for the task. The result indicates the excellent ability of LLMs to produce responses that adhere to specified word limit requirements. Regarding reading levels, LLMs demonstrated moderate performance, with 105/360 responses (29.2%) fully meeting the specified FKG level (FKG level ≤6), 107/360 (29.7%) being partially satisfactory (FKG level of 6 - 8), and 148/360 (41.1%) not aligning with the provided FKG level (FKG level >8).

LLMs demonstrated consistently high average scores across all quality criteria (total score: 8.933 out of 10). The highest average score achieved was 1.91 on relevance, highlighting the LLMs' ability to generate content that was highly pertinent to the given prompts. The lowest average score observed was 1.58 out of 2 in the category of completeness, indicating a moderate adherence to providing responses that capture all key points. In the translation tasks, the LLMs demonstrated high performance, with 87/90 accuracy translations (88%) for Spanish and 73/90 (81%) for Chinese translation.

Three Generative Pretrained Transformer Models Comparisons: GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo

GPT-4 demonstrated a superior capability in adhering to the specified word limit, with 101/120 responses (84.2%) falling within 250 words (Table 1). In contrast, GPT-3.5 Turbo and GPT-4 Turbo exhibited a relatively lower proficiency, with 86/120 (71.7%) and 80/120 (66.7%) responses meeting the word limit, respectively. As shown in Table 2, when comparing the models based on word limit, the chi-square test demonstrated a significant difference among the three models ($P=.006$).

Table . Performance of all models and prompts on the summarization task.

	GPT[a]-3.5 Turbo[bc]			GPT-4[bc]			GPT-4 turbo[bc]		
	Total	Textual format	Bullet points	Total	Textual format	Bullet points	Total	Textual format	Bullet points
Word limit, %	71.7 (86/120)	46.7 (28/60)	96.7 (58/60)	84.2 (101/120)	91.7 (55/60)	76.7 (46/60)	66.7 (80/120)	51.7 (31/60)	81.7 (49/60)
Reading level, %	23.3 (28/120)	18.3 (11/60)	28.3 (17/60)	21.7 (26/120)	21.7 (13/60)	21.7 (13/60)	42.5 (51/120)	53.3 (32/60)	31.7 (19/60)
Accuracy, mean (SD)	1.775 (0.493)	1.767 (0.5)	1.783 (0.49)	1.767 (0.561)	1.8 (0.48)	1.733 (0.634)	1.783 (0.522)	1.8 (0.48)	1.767 (0.563)
Clarity, mean (SD)	1.792 (0.447)	1.833 (0.418)	1.75 (0.474)	1.833 (0.396)	1.867 (0.389)	1.8 (0.403)	1.8 (0.422)	1.883 (0.324)	1.717 (0.49)
Relevance, mean (SD)	1.892 (0.362)	1.883 (0.415)	1.9 (0.303)	1.925 (0.295)	1.883 (0.372)	1.967 (0.181)	1.925 (0.264)	1.9 (0.303)	1.95 (0.22)
Completeness, mean (SD)	1.558 (0.632)	1.533 (0.623)	1.583 (0.645)	1.575 (0.617)	1.483 (0.624)	1.667 (0.601)	1.617 (0.582)	1.583 (0.619)	1.65 (0.547)
Comprehensibility, mean (SD)	1.808 (0.436)	1.817 (0.469)	1.8 (0.403)	1.892 (0.312)	1.883 (0.324)	1.9 (0.303)	1.858 (0.35)	1.9 (0.303)	1.817 (0.39)
Total score, mean (SD)	8.825 (1.643)	8.833 (1.748)	8.817 (1.546)	8.992 (1.247)	8.917 (1.239)	9.067 (1.26)	8.983 (1.195)	9.067 (1.087)	8.9 (1.298)

[a] GPT: Generative Pretrained Transformer.

[b] The performance (%) of GPT-3.5 Turbo was 93.3% (28/30), GPT-4 was 96.7% (29/30), and GPT-4 Turbo was 100% (30/30) for the Spanish translation. The overall performance (%) of the three GPT models in Spanish translation was 96.7% (87/90).

[c] The performance (%) of GPT-3.5 Turbo was 76.7% (23/30), GPT-4 was 86.7% (26/30), and GPT-4 Turbo was 80% (24/30) for the Chinese translation. The overall performance (%) of the three GPT models in Chinese translation was 81.1% (73/90).

Table . Statistical analysis results from analysis of variance and chi-square tests.

Group and criterion	PR(>F)[a]	Chi-square (*df*)
Models		
Accuracy	0.97	—[b] (2)
Clarity	0.721	— (2)
Relevance	0.63	— (2)
Completeness	0.748	— (2)
Comprehensibility	0.215	— (2)
Total score	0.572	— (16)
Word limit	0.006	10.178 (2)
Reading level	<0.001	35.468 (4)
Spanish translation	0.355	2.069 (2)
Chinese translation	0.602	1.015 (2)
Translation	0.481	1.463 (2)
Prompts		
Accuracy	0.213	— (2)
Clarity	0.028	— (2)
Relevance	0.177	— (2)
Completeness	0.154	— (2)
Comprehensibility	0.149	— (2)
Total score	0.939	— (8)

[a] PR(>F): probability that the F-statistic is greater than the observed value under the null hypothesis.

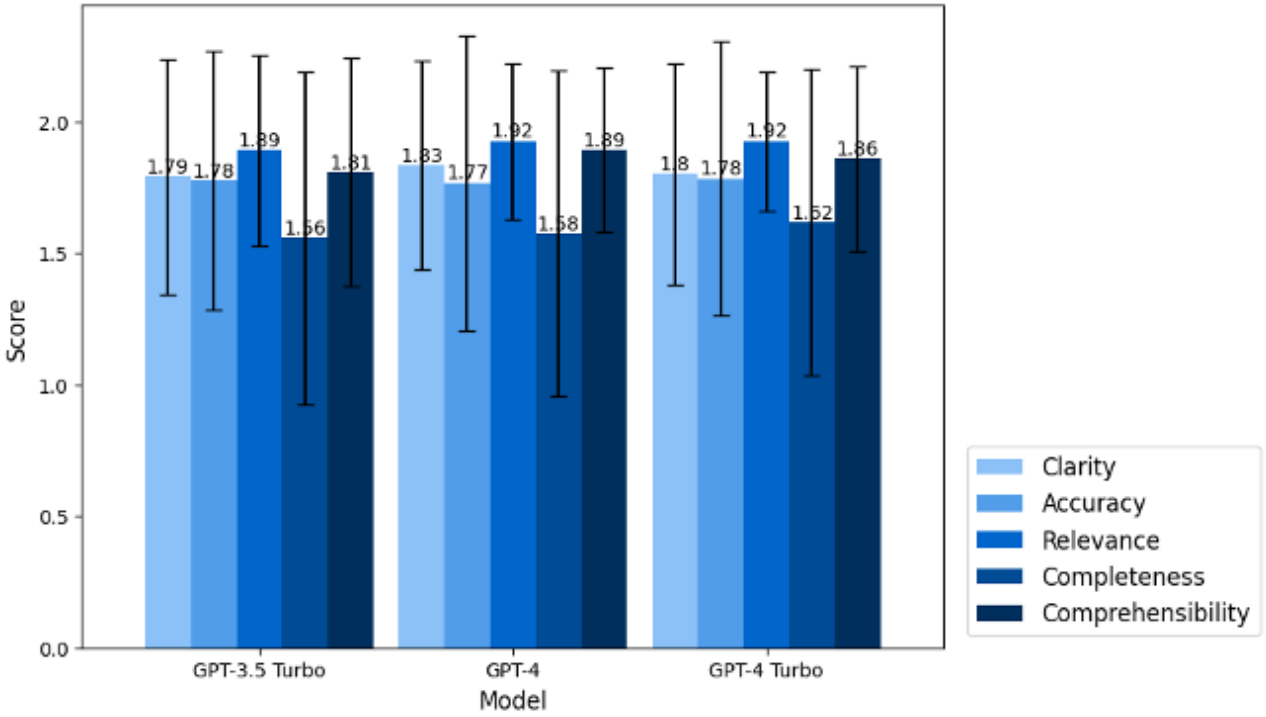
[b] —: not applicable.

Regarding the assessment of reading level, GPT-4 Turbo met the required FKG level of 6 in 51/120 (42.5%) cases, nearly doubling the performance of the other 2 models: 26/120 (21.7%) for GPT-4 and 28/120 (23.3%) for GPT-3.5 Turbo. The result indicated significant discrepancies among the models in adherence to the specified reading level (P<.001), with GPT-4 Turbo performing better compared with the other 2 models.

In terms of quality assessment, each of the LLMs attained a high score exceeding 8.8 out of 10, with GPT-4 and GPT-4 Turbo achieving 8.992 and 8.983, respectively, and GPT-3.5 Turbo trailing slightly at 8.825. Upon evaluation of each criterion, the performance of all models was found to be similar (Figure 2). The application of ANOVA tests to each criterion revealed no significant differences among the 3 models (P=.57).

In the translation tasks, GPT-4 Turbo exhibited perfect accuracy with a 30/30 (100%) success cases in Spanish translation, whereas GPT-4 and GPT-3.5 Turbo exhibited slightly lower, yet commendable success rates of 29/30 (97%) and 28/30 (93%), respectively. For the Chinese translation task, GPT-4 outperformed the other models with an accuracy of 26/30 (87%). In contrast, GPT-3.5 Turbo and GPT-4 Turbo achieved 23/30 (77%) and 24/30 (80%), respectively. The 3 models did not show a significant difference in the translation task (P=.48).

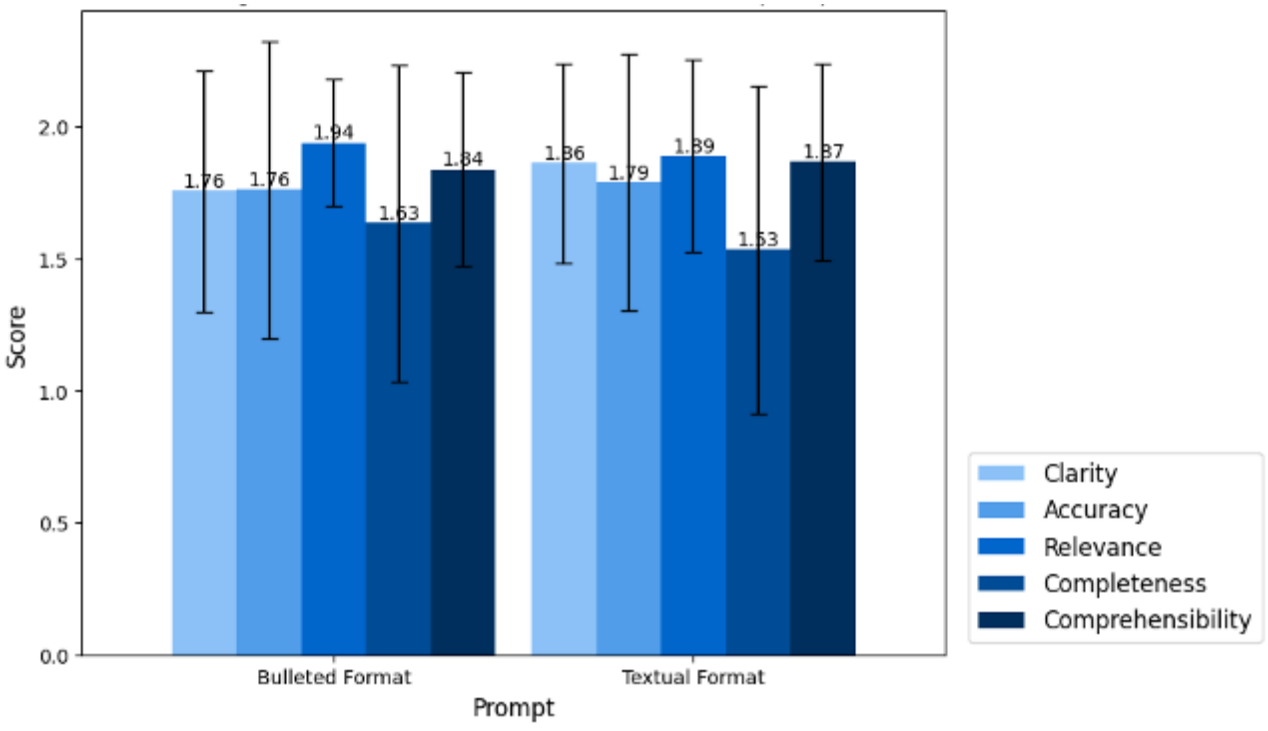
Figure 2. Assessment scores on each criterion between different models. GPT: Generative Pretrained Transformer.



Two Different Prompt Comparisons: Textual and Bulleted Formats

We compared 2 prompting methods in terms of word limits, reading level, and quality assessment. The major difference noted in the comparison of the 2 prompts was that responses generated from prompt 2 (bulleted format) were superior in adhering to the target word limit. Specifically, 153/180 responses (85%) from prompt 2 successfully achieved the word limit, in contrast to 114/180 responses (63.3%) from prompt 1 (textual format) that fully satisfied the word limit. Using prompt 1 resulted in only 56/180 responses (31.1%) meeting our desired reading level, with a slight decrease to 49/180 (27.2%) for prompt 2. For the 5 quality criteria, both prompts achieved high scores (Figure 3). Upon performing an ANOVA test to assess the differences in performance between the 2 prompts (Table 2), it was found that the variations between them were not significant (P=.939). However, the 2 prompt formats demonstrated a significant difference in the clarity criterion (P=.03).

Figure 3. Assessment scores on each criterion between different prompts.



Error Analysis

The errors that LLMs committed were categorized into inaccurate scope, inaccurate definition, inaccurate expression, meaningless points, and inaccurate word. Some examples are shown in Table 3.

Table . Error cases and analysis.

Model	Topic	Output	Error type	Reason
GPT-3.5 Turbo	Nutrition	"It advises limiting animal-based food, processed food, and alcohol consumption."	Inaccurate scope	The chapter only mentions to limit red meat, not all animal-based foods (says it can make up half or less of diet).
GPT-3.5 Turbo	Sexual Health Issues in Men with Cancer	"It is still important to maintain intimacy with a partner."	Inaccurate expression	The tailored content sounds a little judgmental whereas the original document says, "probably still important" and is less assuming.
GPT-3.5 Turbo	Relaxation	"多喝液体"	Inaccurate word	Based on the English sentence: " Drinking plenty of liquids", "liquids" can be better translated into "水."
GPT-4	Mindfulness	"These practices involve focusing the mind on present sensations, such as breathing, a sound, or an image."	Inaccurate definition	It seems to define meditation and mindfulness in one overarching definition, which only defines meditation. The model merged definitions of MBSR[a] and MBCT[b] together and did not include difference between types.
GPT-4	Family Communication	"El apoyo de la comunidad podréda ser beneficioso durante este difícil período."	Inaccurate word	Based on the English sentence: "Support from the community might be beneficial during this difficult period. ", "difícil período" should be "período difícil"
GPT-4 Turbo	Making a Difference	"Learning: Educating yourself about cancer can empower you to assist others. Resources are available online, by phone, and in print."	Meaningless point	The customized content falls short in terms of actionability. The purpose of tailoring content is to educate patients and caregivers, rather than expecting them to educate themselves.

[a]MBSR: Mindfulness-based stress reduction.

[b]MBCT: Mindfulness-based cognitive therapy.

A common error observed with LLMs is their tendency to integrate their own knowledge and interpretation rather than adhering strictly to the provided materials, such as an inaccurate scope. For instance, when the text specified "to limit red meat." in the Nutrition topic, GPT-3.5 Turbo inaccurately generalized this advice to "limiting animal-based food." This interpretation is not entirely correct, as animal-based food encompasses more than just red meat, including white meat such as chicken, which the original material did not intend to restrict.

Other observed errors involve inaccurate expressions. For instance, in the Sexual health issues in men with cancer topic, the original content suggested, "It is probably still important to maintain intimacy with a partner." However, GPT-3.5 Turbo revised this to "it is still important to maintain intimacy with a partner." This alteration results in a tone that may seem judgmental, deviating from the original's more tentative stance.

An example of inaccurate definition was identified within the Mindfulness topic, where GPT-4 defined meditation and mindfulness in one overarching definition for meditation. It also merged definitions of mindfulness-based stress reduction and mindfulness-based cognitive therapy without highlighting differences between the mindfulness interventions.

LLMs may also include information that, while accurate, might not be actionable for patients. For instance, in the Making a difference topic, GPT-4 Turbo correctly sourced from the material that "Learning: Educating yourself about cancer can empower you to assist others. Resources are available online, by phone, and in print." However, this information becomes less useful in the absence of specific links or directions that could guide patients on where to start their education.

Finally, with respect to translation quality, the primary error observed related to inaccurate word choice. In particular, when

an English term offers multiple potential translations, LLMs often encounter difficulty in selecting the most contextually appropriate option. For example, in the Relaxation topic, GPT-3.5 Turbo translated “多喝液体" as “drink more liquids.” Although “液体" does literally translate to “liquids,” the more natural and contextually appropriate term would be “水.”

Discussion

Principal Findings

To our knowledge, this is the first study to evaluate the capability of LLMs in tailoring educational content for cancer survivors and their caregivers with limited health literacy or language barriers. In our study, all 3 LLMs have demonstrated overall excellent performance in most criteria. The more advanced GPT-4 family models showed better overall performance compared with GPT-3.5 Turbo. GPT-4’s high adherence to word limits and GPT-4 Turbo’s better compliance to reading level compliance proved their ability to meet our requirements when tailoring content. Prompting GPTs to produce bulleted-format content is likely to result in better educational content compared with textual-format content. All models exhibit strong capability in generating highly relevant content. However, they fall short in terms of completeness. Overall, it is proven that LLMs are highly effective in tailoring, condensing, and translating educational content for cancer survivors and their caregivers with limited health literacy or language barriers. These findings inform future versions of LLMs to focus more on the reading level and completeness of their output and the development of tailored intervention materials for cancer survivors and their caregivers. These promising results also indicate that LLMs can be a valuable tool in making educational content more accessible and comprehensible to diverse patient populations.

The capabilities of LLMs in text analysis have been well studied. For example, our previous study [34] examined the potential of LLMs to categorize clinical concepts from patient notes. Yet, this study focused solely on the LLMs’ comprehension of patients’ conditions from clinical notes rather than educational content. Study by Veen et al [53] assessed approaches for LLMs to summarize clinical texts. Although it demonstrated overall preferred performance, especially GPT-4, over human experts, the study was limited to the summarization of radiology report findings and confined to 3 attributes: completeness, correctness, and conciseness, whereas our study expanded on this topic by evaluating LLMs against 7 distinct criteria. Furthermore, none of the existing studies focus on education regarding supportive care in cancer, whereas our innovative findings make a significant contribution to the literature in this field.

Despite the excellence of LLMs in adhering to specified word limits and generating high-quality content, several challenges remain. One notable area where LLMs struggle is in adjusting the reading level of the content to accommodate patients from various educational levels. The content tailored by LLMs often does not meet the intended FKG level. This oversight implies that some individuals might find the content overly complex, potentially hindering their understanding of health information and educational content [54,55]. Addressing this challenge is essential for maximizing the applicability of LLMs and ensuring that all cancer survivors receive the support they need to manage their cancer effectively. In future work, in-context learning could be used to offer more detailed guidance to LLMs, focusing on the potential vocabularies frequently appeared in content exceeding the specified FKG level of 6. In addition, retrieval-augmented generation could be implemented to embed vocabularies aligned with an FKG level of 6, thereby enhancing the model’s performance.

It is also observed that the accuracy of Spanish translations is significantly higher than that of Chinese translations. This finding is expected, given the abundance of Spanish content available on the internet compared with Chinese content that can serve as training materials. Previous studies [56,57] have shown that LLMs’ performance in different languages has a clear correlation with the proportion of each language in the pretraining corpus. Without fine-tuning, LLMs have a much higher performance in high-resource languages like German, French, and Spanish, and a significantly lower performance in low-resource languages like Kannada, Occitan, and Western Frisian [56,57]. In future work, integrating high-quality bilingual medical corpora that includes parallel texts of patient education materials, clinical guidelines, and culturally tailored health information could be a promising approach. Fine-tuning LLMs on such specialized corpora may provide them with domain-specific vocabulary and context, thereby increasing their ability to produce accurate, culturally sensitive translations.

The educational content errors could be detrimental to cancer survivors and their caregivers by providing false physical activity, diet, or medication suggestions. Therefore, content produced by LLMs should undergo thorough evaluation and validation before the content is used in a clinical setting [38,58,59]. Our analysis has identified multiple errors in the outputs from LLMs, including inaccuracies in scope, expression, and definition. These types of errors can lead to the dissemination of misinformation, potentially causing harm to patients [60]. Therefore, such inaccuracies must be identified, analyzed, and rectified to prevent any negative impacts on patient care. Our study also detected some meaningless points that were not actionable in LLMs’ outputs, which could increase the reading burden on patients and their caregivers. Recommendations should highlight actionable information for cancer survivors and their caregivers to reduce the burden of reading educational content, emphasizing the need for LLMs to prioritize the use and applicability of the information they present. In addition, education content should be evaluated and validated by content experts before the it is available to cancer survivors and their caregivers.

In addition, both Xiao et al’s and Asthana et al’s studies [36,37] evaluated the performance of fine-tuned LLMs in nonclinical environments. Their results highlighted the significant potential of LLMs in summarizing general text through the adoption of advanced fine-tuning techniques. It is possible that fine-tuning could further improve LLMs’ capacity to analyze educational content specifically tailored for groups such as cancer survivors and their caregivers with limited health literacy or language barriers. With this additional data, more advanced fine-tuning techniques such as instruction tuning [57,61,62] and

parameter-efficient fine-tuning [63] can be implemented, and are likely to further enhance the performance.

Limitations

While the study has shown promising results, it has several limitations. First, the dataset size remains relatively small, which could restrict the generalizability of the findings to broader topics. Second, we lacked participant assessment. Relying solely on oncology experts to evaluate the outputs from LLMs might create obstacles when applying these findings to actual cancer patients and their caregivers. While our oncology experts deeply value caring for disadvantaged populations with limited health literacy or language barriers, it's important to note that they are highly educated and might have unintentional biases. This could make it challenging for them to view educational content from the perspective of individuals with low health education and literacy. Therefore, future studies can be broadened to include a wider range of educational topics and additional annotations from cancer patients and their caregivers. Third, this study was limited to zero-shot learning because of the lack of training data. It could be expanded by collecting tailored content from human experts to serve as training data to incorporate few-shot learning and fine-tuning techniques. In addition, chain-of-thought reasoning and in-context learning also present promising avenues for future exploration, particularly because they do not rely on additional training data. Finally, due to a limited number of annotators from diverse backgrounds, our study was only able to evaluate translations in 2 languages. Our analysis suggests that translation performance can vary between languages, influenced by the availability of content in each language. It is important to note that these findings may not be generalizable to languages spoken by smaller populations, where content availability and linguistic nuances could further affect translation accuracy. In future research, more extensive evaluations of translation tasks involving other languages, especially low-resource languages, should be conducted to expand the applicability.

Conclusions

The study highlights the application of LLMs in cancer care while being cognizant of their potential limitations. All 3 LLMs have demonstrated overall high capability in tailoring educational content for cancer survivors and their caregivers with limited health literacy or language barriers. GPT-4 family models showed better overall performance compared with GPT-3.5 Turbo. Prompting GPTs to produce bulleted-format content can generate better educational content. The findings from this study inform the intervention development and implementation in cancer symptom management and health equity. Additional studies are warranted to expedite the integration of AI-driven solutions into clinical settings.

Acknowledgments

The study was supported by the Oncology Nursing Foundation Research Grant (2022 RE03). The authors also appreciate the support from Emory University School of Nursing and Winship Cancer Institute. During the preparation of this work, the authors used ChatGPT solely to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data Availability

The datasets generated and analyzed during the current study are available upon reasonable request. The corresponding author [YL] will coordinate requests for data and maintain documentation for requests and distributions. Emory University has an established Institutional Data Use Agreement that can easily be adapted and deployed.

Authors' Contributions

DL developed scripts to collect responses from LLMs, build the annotation platform, and wrote the manuscript. XH provided technical support and helped with manuscript writing. CX, JB, ZB, LL, CW, LB, and SW performed annotations and supported manuscript writing. DB supported the development of the annotation platform and manuscript writing. YL supervised the entire project, provided valuable insights, and contributed to manuscript writing and revisions. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Statistics and graphs. National Cancer Institute at the National Institutes of Health. URL: https://cancercontrol.cancer.gov/ocs/statistics [accessed 2025-03-19]
2. Lin Y, Bailey DE, Docherty SL, et al. Distinct morning and evening fatigue profiles in gastrointestinal cancer during chemotherapy. BMJ Support Palliat Care 2023 Nov;13(e2):e373-e381. [doi: 10.1136/bmjspcare-2021-002914]
3. Lin Y, Bailey DE Jr, Docherty SL, et al. Distinct sleep disturbance profiles in patients with gastrointestinal cancers receiving chemotherapy. Cancer Nurs 2022;45(2):E417-E427. [doi: 10.1097/NCC.0000000000000975] [Medline: 34183521]

4. Lin Y, Bailey DE Jr, Docherty SL, et al. Distinct profiles of multiple co-occurring symptoms in patients with gastrointestinal cancers receiving chemotherapy. Support Care Cancer 2021 Aug;29(8):4461-4471. [doi: 10.1007/s00520-020-05946-4] [Medline: 33454824]
5. Lin Y, Bailey DE, Xiao C, et al. Distinct co-occurring morning and evening fatigue profiles in patients with gastrointestinal cancers receiving chemotherapy. Cancer Nurs 2024;47(1):E28-E37. [doi: 10.1097/NCC.0000000000001148] [Medline: 36076314]
6. Lin Y, Bruner DW, Paul S, et al. A network analysis of self-reported psychoneurological symptoms in patients with head and neck cancer undergoing intensity-modulated radiotherapy. Cancer 2022 Oct;128(20):3734-3743. [doi: 10.1002/cncr.34424] [Medline: 35969226]
7. Lin Y, Peng G, Bruner DW, et al. Associations of differentially expressed genes with psychoneurological symptoms in patients with head and neck cancer: A longitudinal study. J Psychosom Res 2023 Dec;175:111518. [doi: 10.1016/j.jpsychores.2023.111518] [Medline: 37832274]
8. Dodd MJ, Miaskowski C, Paul SM. Symptom clusters and their effect on the functional status of patients with cancer. Oncol Nurs Forum 2001 Apr;28(3):465-470. [Medline: 11338755]
9. Tantoy IY, Cooper BA, Dhruva A, et al. Quality of life of patients with gastrointestinal cancers undergoing chemotherapy. Qual Life Res 2018 Jul;27(7):1865-1876. [doi: 10.1007/s11136-018-1860-1] [Medline: 29679368]
10. Lin Y, Docherty SL, Porter LS, Bailey DE Jr. Symptom experience and self-management for multiple co-occurring symptoms in patients with gastric cancer: A qualitative study. Eur J Oncol Nurs 2020 Dec;49:101860. [doi: 10.1016/j.ejon.2020.101860] [Medline: 33126154]
11. Lin Y, Docherty SL, Porter LS, Bailey DE. Common and co-occurring symptoms experienced by patients with gastric cancer. Oncol Nurs Forum 2020 Mar 1;47(2):187-202. [doi: 10.1188/20.ONF.187-202] [Medline: 32078617]
12. Cal A, Avci IA, Cavusoglu F. Experiences of caregivers with spouses receiving chemotherapy for colorectal cancer and their expectations from nursing services. Asia Pac J Oncol Nurs 2017;4(2):173-179. [doi: 10.4103/apjon.apjon_11_17] [Medline: 28503652]
13. Law E, Levesque JV, Lambert S, Girgis A. The “sphere of care”: a qualitative study of colorectal cancer patient and caregiver experiences of support within the cancer treatment setting. PLoS ONE 2018;13(12):e0209436. [doi: 10.1371/journal.pone.0209436] [Medline: 30586391]
14. Lin Y, Porter LS, Chee W, et al. A web-based dyadic intervention to manage psychoneurological symptoms for patients with colorectal cancer and their caregivers: protocol for A mixed methods study. JMIR Res Protoc 2023 Jun 28;12:e48499. [doi: 10.2196/48499] [Medline: 37379055]
15. Howard AF, Lynch K, Beck S, et al. At the heart of It all: emotions of consequence for the conceptualization of caregiver-reported outcomes in the context of colorectal cancer. Curr Oncol 2021 Oct 16;28(5):4184-4202. [doi: 10.3390/curroncol28050355] [Medline: 34677273]
16. Hollis RH, Chu DI. Healthcare disparities and colorectal cancer. Surg Oncol Clin N Am 2022 Apr;31(2):157-169. [doi: 10.1016/j.soc.2021.11.002] [Medline: 35351271]
17. Levy H, Janke A. Health literacy and access to care. J Health Commun 2016;21 Suppl 1(Suppl):43-50. [doi: 10.1080/10810730.2015.1131776] [Medline: 27043757]
18. Murphy-Ende K. Barriers to palliative and supportive care. Nurs Clin North Am 2001 Dec;36(4):843-853. [doi: 10.1016/s0029-6465(22)02675-5] [Medline: 11726357]
19. Sherman ADF, Febres-Cordero S, Johnson KB, et al. Intersectionality in nursing research: a scoping review. Int J Nurs Stud Adv 2023 Dec;5:100155. [doi: 10.1016/j.ijnsa.2023.100155] [Medline: 37982092]
20. Toolkit for patient focused therapy development. National Center for Advancing Translational Sciences. URL: https://toolkit.ncats.nih.gov/glossary/underserved-group/ [accessed 2025-03-24]
21. Vogels EA. Digital divide persists even as americans with lower incomes make gains in tech adoption. Pew Research Center. 2021. URL: https://www.pewresearch.org/short-reads/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption [accessed 2025-03-19]
22. Qan’ir Y, Song L. Systematic review of technology-based interventions to improve anxiety, depression, and health-related quality of life among patients with prostate cancer. Psychooncology 2019 Aug;28(8):1601-1613. [doi: 10.1002/pon.5158] [Medline: 31222956]
23. Song L, Rini C, Deal AM, et al. Improving couples’ quality of life through a Web-based prostate cancer education intervention. Oncol Nurs Forum 2015 Mar;42(2):183-192. [doi: 10.1188/15.ONF.183-192] [Medline: 25806885]
24. Xue VW, Lei P, Cho WC. The potential impact of ChatGPT in clinical and translational medicine. Clin Transl Med 2023 Mar;13(3):e1216. [doi: 10.1002/ctm2.1216] [Medline: 36856370]
25. Strachan JWA, Albergo D, Borghini G, et al. Testing theory of mind in large language models and humans. Nat Hum Behav 2024 Jul;8(7):1285-1295. [doi: 10.1038/s41562-024-01882-z] [Medline: 38769463]
26. Wang J, Liu Z, Zhao L, et al. Review of large vision models and visual prompt engineering. Meta-Radiology 2023 Nov;1(3):100047. [doi: 10.1016/j.metrad.2023.100047]
27. Wang J, Shi E, Yu S, et al. Prompt engineering for healthcare: methodologies and applications. arXiv. Preprint posted online on 2023. [doi: 10.48550/arXiv.2304.14670]

28. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 4;47(1):33. [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
29. Fink MA, Bischoff A, Fink CA, et al. Potential of ChatGPT and GPT-4 for data mining of free-text CT reports on lung cancer. Radiology 2023 Sep;308(3):e231362. [doi: 10.1148/radiol.231362] [Medline: 37724963]
30. Haver HL, Lin CT, Sirajuddin A, Yi PH, Jeudy J. Use of ChatGPT, GPT-4, and bard to improve readability of ChatGPT's answers to common questions about lung cancer and lung cancer screening. AJR Am J Roentgenol 2023 Nov;221(5):701-704. [doi: 10.2214/AJR.23.29622] [Medline: 37341179]
31. Dai W, Mo W, Xu W, Han D, Xu X. Systematic analysis of glutamine metabolism family genes and exploration of the biological role of GPT in colorectal cancer. Aging (Milano) 2023 Nov 15;15(21):11811-11830. [doi: 10.18632/aging.205079]
32. Achiam J, Adler S, Agarwal S, et al. GPT-4 technical report. arXiv. Preprint posted online on 2023. [doi: 10.48550/arXiv.2303.08774]
33. Bhattarai K, Oh IY, Sierra JM, et al. Leveraging GPT-4 for identifying cancer phenotypes in electronic health records: a performance comparison between GPT-4, GPT-3.5-turbo, Flan-T5 and spaCy's rule-based & machine learning-based methods. bioRxiv 2024 Apr 6;2023(2023):27. [doi: 10.1101/2023.09.27.559788] [Medline: 37808763]
34. Liu D, Ding C, Bold D, et al. Evaluation of general large language models in understanding clinical concepts extracted from adult critical care electronic health record notes. arXiv. Preprint posted online on 2024arXiv:240113588.
35. Agrawal M, Hegselmann S, Lang H, Kim Y, Sontag D. Large language models are zero-shot clinical information extractors. arXiv. Preprint posted online on 2022. [doi: 10.48550/arXiv.2205.12689]
36. Xiao L, Chen X, Shan X. Enhancing llm with evolutionary fine tuning for news summary generation. arXiv. Preprint posted online on 2023. [doi: 10.48550/arXiv.2307.02839]
37. Asthana S, Hilleli S, He P, Halfaker A. Summaries, highlights, and action items: design, implementation and evaluation of an LLM-powered meeting recap system. arXiv. Preprint posted online on 2023. [doi: 10.48550/arXiv.2307.15793]
38. Garrouste-Orgeas M, Philippart F, Bruel C, Max A, Lau N, Misset B. Overview of medical errors and adverse events. Ann Intensive Care 2012 Feb 16;2(1):1-9. [doi: 10.1186/2110-5820-2-2] [Medline: 22339769]
39. Gero Z, Singh C, Xie Y, et al. Attribute structuring improves LLM-based evaluation of clinical text summaries. arXiv. Preprint posted online on 2024. [doi: 10.48550/arXiv.2403.01002]
40. Espenshade TJ, Fu H. An analysis of english-language proficiency among U.S. immigrants. Am Sociol Rev 1997 Apr;62(2):288. [doi: 10.2307/2657305]
41. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
42. Solnyshkina M, Zamaletdinov R, Gorodetskaya L, Gabitov A. Evaluating text complexity and Flesch-Kincaid grade level. J Soc Stud Educ Res 2017;8(3):238-248.
43. Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. Respir Care 2008 Oct;53(10):1310-1315. [Medline: 18811992]
44. Abazari A, Chatterjee S, Moniruzzaman M. Understanding cancer caregiving and predicting burden: an analytics and machine learning approach. AMIA Annu Symp Proc 2023;2023:243-252. [Medline: 38222371]
45. Wang L, Chen X, Deng X, et al. Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. NPJ Digit Med 2024 Feb 20;7(1):41. [doi: 10.1038/s41746-024-01029-4] [Medline: 38378899]
46. Patient education publications. National Cancer Institute at the National Institutes of Health. URL: https://www.cancer.gov/publications/patient-education [accessed 2025-03-19]
47. Guidelines for patients. National Comprehensive Career Network. URL: https://www.nccn.org/patientresources/patient-resources/guidelines-for-patients [accessed 2025-03-19]
48. Epari A, Kim K, Xiao C, et al. Perceptions and needs for a technology-based dyadic intervention on symptom management among patients with colorectal cancer and their caregivers: a qualitative study. Cancer Nurs 2024 Nov 29. [doi: 10.1097/NCC.0000000000001429] [Medline: 39625810]
49. Northouse L, Kershaw T, Mood D, Schafenacker A. Effects of a family intervention on the quality of life of women with recurrent breast cancer and their family caregivers. Psychooncology 2005 Jun;14(6):478-491. [doi: 10.1002/pon.871] [Medline: 15599947]
50. Northouse LL, Mood DW, Schafenacker A, et al. Randomized clinical trial of a brief and extensive dyadic intervention for advanced cancer patients and their family caregivers. Psychooncology 2013 Mar;22(3):555-563. [doi: 10.1002/pon.3036] [Medline: 22290823]
51. Northouse LL, Mood DW, Schafenacker A, et al. Randomized clinical trial of a family intervention for prostate cancer patients and their spouses. Cancer 2007 Dec 15;110(12):2809-2818. [doi: 10.1002/cncr.23114] [Medline: 17999405]
52. Part 46 - protection of human subjects. Code of Federal Regulations. URL: https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46 [accessed 2025-03-19]
53. Van Veen D, Van Uden C, Blankemeier L, et al. Clinical text summarization: adapting large language models can outperform human experts. Res Sq 2023 Oct 30:rs.3.rs-3483777. [doi: 10.21203/rs.3.rs-3483777/v1] [Medline: 37961377]

54. Delaney FT, Doinn T, Broderick JM, Stanley E. Readability of patient education materials related to radiation safety: What are the implications for patient-centred radiology care? Insights Imaging 2021 Oct 21;12(1):148. [doi: 10.1186/s13244-021-01094-3] [Medline: 34674063]
55. Davis TC, Mayeaux EJ, Fredrickson D, Bocchini JA, Jackson RH, Murphy PW. Reading ability of parents compared with reading level of pediatric patient education materials. Pediatrics 1994 Mar;93(3):460-468. [doi: 10.1542/peds.93.3.460] [Medline: 8115206]
56. Li Z, Shi Y, Liu Z, Yang F, Liu N, Du M. Quantifying multilingual performance of large language models across languages. arXiv. Preprint posted online on 2024. [doi: 10.48550/arXiv.2404.11553]
57. Zan C, Ding L, Shen L, Zhen Y, Liu W, Tao D. Building accurate translation-tailored llms with language aware instruction tuning. arXiv. Preprint posted online on 2024. [doi: 10.48550/arXiv.2403.14399]
58. Kalra J. Medical errors: impact on clinical laboratories and other critical areas. Clin Biochem 2004 Dec;37(12):1052-1062. [doi: 10.1016/j.clinbiochem.2004.08.009] [Medline: 15589810]
59. Robertson JJ, Long B. Suffering in silence: medical error and its impact on health care providers. J Emerg Med 2018 Apr;54(4):402-409. [doi: 10.1016/j.jemermed.2017.12.001] [Medline: 29366616]
60. Burton-Wood C, Burnell R, Taylor A, Brown D, Robinson B, Garry M. Medical professionals' (mis)remembering of a simulated interaction with a patient: a medical misinformation effect. PsyArXiv. Preprint posted online on 2019. [doi: 10.31234/osf.io/wt8ma]
61. Zhang S, Dong L, Li X, et al. Instruction tuning for large language models: A survey. arXiv. Preprint posted online on 2023. [doi: 10.48550/arXiv.2308.10792]
62. Bai Y, Du X, Liang Y, et al. COIG-CQIA: quality is all you need for chinese instruction fine-tuning. arXiv. Preprint posted online on 2024. [doi: 10.48550/arXiv.2403.18058]
63. Ding N, Qin Y, Yang G, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models. Nat Mach Intell 2023;5(3):220-235. [doi: 10.1038/s42256-023-00626-4]

Abbreviations

AMA: American Medical Association
ANOVA: analysis of variance
CADA: Cohort Adjudication and Data Annotation
FKG: Flesch-Kincaid Grade Level
GPT: Generative Pretrained Transformer
LLM: large language model
NIH: National Institute of Health

Edited by N Cahill; submitted 24.10.24; peer-reviewed by A Deroy, Z Wang; revised version received 27.02.25; accepted 28.02.25; published 07.04.25.

Please cite as:
Liu D, Hu X, Xiao C, Bai J, Barandouzi ZA, Lee S, Webster C, Brock LU, Lee L, Bold D, Lin Y
Evaluation of Large Language Models in Tailoring Educational Content for Cancer Survivors and Their Caregivers: Quality Analysis
JMIR Cancer 2025;11:e67914
URL: https://cancer.jmir.org/2025/1/e67914
doi:10.2196/67914

Identifying Complex Scheduling Patterns Among Patients With Cancer With Transportation and Housing Needs: Feasibility Pilot Study

Allan Fong[1], MS; Christian Boxley[1], BS; Laura Schubel[1], MPH; Christopher Gallagher[2], MD; Katarina AuBuchon[3], PhD; Hannah Arem[1], PhD

[1]MedStar Health Research Institute, 3007 Tilden St, Washington, DC, United States

[2]MedStar Washington Hospital Center, Washington, DC, United States

[3]Lombardi Comprehensive Cancer Center, MedStar Georgetown University, Washington, DC, United States

Corresponding Author:
Allan Fong, MS
MedStar Health Research Institute, 3007 Tilden St, Washington, DC, United States

Abstract

Background: Patients with cancer frequently encounter complex treatment pathways, often characterized by challenges with coordinating and scheduling appointments at various specialty services and locations. Identifying patients who might benefit from scheduling and social support from community health workers or patient navigators is largely determined on a case-by-case basis and is resource intensive.

Objective: This study aims to propose a novel algorithm to use scheduling data to identify complex scheduling patterns among patients with transportation and housing needs.

Methods: We present a novel algorithm to calculate scheduling complexity from patient scheduling data. We define patient scheduling complexity as an aggregation of sequence, resolution, and facility components. Schedule sequence complexity is the degree to which appointments are scheduled and arrived to in a nonchronological order. Resolution complexity is the degree of no shows or canceled appointments. Location complexity reflects the proportion of appointment dates at 2 or more different locations. Schedule complexity captures deviations from chronological order, unresolved appointments, and coordination across multiple locations. We apply the scheduling complexity algorithm to scheduling data from 38 patients with breast cancer enrolled in a 6-month comorbidity management intervention at an urban hospital in the Washington, DC area that serves low-income patients. We compare the scheduling complexity metric with count-based metrics: arrived ratio, rescheduled ratio, canceled ratio, and no-show ratio. We defined an aggregate count-based adjustment metric as the harmonic mean of rescheduled ratio, canceled ratio, and no-show ratio. A low count-based adjustment metric would indicate that a patient has fewer disruptions or changes in their appointment scheduling.

Results: The patients had a median of 88 unique appointments (IQR 60.3), 62 arrived appointments (IQR 47.8), 13 rescheduled appointments (IQR 13.5), 9 canceled appointments (IQR 10), and 1.5 missed appointments (IQR 5). There was no statistically significant difference in count-based adjustments and scheduling complexity bins (χ^2_4=6.296, *P*=.18). In total, 5 patients exhibited high scheduling complexity with low count-based adjustments. A total of 2 patients exhibited high count-based adjustments with low scheduling complexity. Out of the 15 patients that indicated transportation or housing insecurity issues in conversations with community health workers, 86.7% (13/15) patients were identified as medium or high scheduling complexity while 60% (9/15) were identified as medium or high count-based adjustments.

Conclusions: Scheduling complexity identifies patients with complex but nonchronological scheduling behaviors who would be missed by traditional count-based metrics. This study shows a potential link between transportation and housing needs with schedule complexity. Scheduling complexity can complement count-based metrics when identifying patients who might need additional care coordination support especially as it relates to transportation and housing needs.

Trial Registration: ClinicalTrials.gov NCT04836221; https://clinicaltrials.gov/study/NCT04836221

(JMIR Cancer 2025;11:e57715) doi:10.2196/57715

KEYWORDS

patient scheduling; scheduling complexities; temporal data mining; dataset; breast cancer; social determinant of health; oncology; metastasis; cancer patient; social support; community health worker; housing need; care; transportation; algorithm

Introduction

Background

Patients with cancer frequently encounter complex treatment pathways, often characterized by challenges with coordinating and scheduling appointments at various specialty services and locations [1-3]. Previous studies have shown that the burden of scheduling and attending visits across multiple providers and specialties not only burdens patients, but also has ripple effects on families, work, and personal lives [4-6]. In a qualitative study with patients with metastatic breast cancer, patients pointed to the need for someone to coordinate appointments and a need for managing work-related barriers to attending appointments [7]. Furthermore, scheduling complexities do not fall on all patients equally. Patients facing social inequalities, such as unequal access to transportation, housing, and social support, face additional complexities in their cancer care appointments. For instance, patients with cancer without insurance, indicating financial vulnerability, are at high risk of no-show appointments [8,9]. A recent review illustrates that most research on multiappointment scheduling problems in oncology focus on solutions using metaheuristics and multiagent methods to ensure appointment adherence [2]. However, if scheduling complexities reflect underlying socioeconomic barriers, such solutions may not solve the structural issues.

To address structural access challenges around scheduling appointments, some health care institutions employ individuals such as community health workers (CHWs) or patient navigators, who play a pivotal role in guiding patients with cancer through their care journey by offering support for nonmedical needs [10]. CHWs and patient navigators have a wide variety of skills and can provide critical assistance coordinating appointment scheduling and overcoming barriers to attending care [4,11,12]. Patients who might benefit from this additional assistance are largely identified manually by CHWs or by care providers aware of possible challenges and social needs [13], or some clinics may assign all patients to CHWs to screen, a resource intensive process [14-16]. Workflows reliant on staff to identify those who might benefit the most from navigation can be time-consuming and resource-intensive, making it difficult to comprehensively identify patients in need of assistance. While ideally all patients would be offered navigation services, in light of staffing shortages and overall limited patient navigation resources, many institutions may be limited in who they can provide extra supportive services to [17]. A data-driven solution that alleviates burden from support staff (ie, reviewing charts to identify patient needs) or relying on clinician referrals would be ideal to effectively and efficiently allocate limited CHW and patient navigator resources.

Potential of Scheduling Data

A potential way to identify patients with unmet transportation or housing needs is to use scheduling data to examine who is experiencing high scheduling complexities. Scheduling data for most cancer care is electronic, providing detailed data about when appointments are scheduled, cancelled, rescheduled, or no shows. This data is automatically recorded, and thus could be used to identify patients who are struggling to manage the complexity of cancer care. In past research, appointment data has primarily been used to optimize appointment scheduling for patient satisfaction and resource allocation [18-21]. Analyses tend to focus on developing and testing scheduling methods to best balance patient satisfaction (eg, wait times) with clinic resources. For example, using model simulations to optimize the scheduling of oncology visits and chemotherapy treatments [19], or optimizing scheduling rules based on chemotherapy infusion [21]. Other research using scheduling data examines the efficiency of appointment self-scheduling processes [22], optimizing scheduling for cost savings [20], and identifying ways to reduce wait times for patients [18]. A study designed an algorithm that used appointment data to identify patients' primary care physician [23]. However, to our knowledge, researchers have yet to design tools for analyzing scheduling data to identify patients with possible unmet transportation or housing needs during their cancer care.

Contributions

Our study used existing scheduling data to identify patients with complex scheduling patterns which may reflect unmet social needs in transportation and housing. We introduce a novel algorithm to calculate scheduling complexity from scheduling data using a sample of patients with breast cancer with initiating cancer treatment from a larger parent study intervening on comorbidity management [24]. Scheduling complexity is an aggregation of sequence, resolution, and facility components. Each component is motivated by the characteristics of scheduling data, an appointment's anatomy, and possible outcomes. The scheduling complexity algorithm is then applied to the scheduling data of 38 patients with breast cancer as a case example. The resulting scheduling complexities are compared with count-based metrics and call notes between CHWs and patients to identify unmet transportation or housing needs.

Methods

Anatomy of an Appointment

Every appointment has a unique appointment identification (AID) and is scheduled on a specific date and time and scheduled for a specific date, time, and location. An appointment is scheduled for a specific visit reason and is associated with the corresponding visit identification (VID). Typically, 1 date will have 1 appointment scheduled with 1 associated VID, Figure 1 (top). Figure 1 is an illustration of AID and VID possible scenarios. Sometimes there can be multiple appointments with different VIDs scheduled for the same date. This is illustrated in an example patient schedule in Table 1. A magnetic resonance imaging (MRI) and mammogram are both scheduled for January 15, 2023. The MRI and mammogram appointments have different AIDs (AID-5 and AID-6 respectively) and VIDs (VID-4 and VID-5 respectively) because they have different reasons for visit and will be at different locations, 1 on the ground floor and 1 on the second floor of the hospital. There can also be multiple AIDs for different dates associated with the same visit reason and at the same location, VID, Figure 1 (bottom). A common example of this pattern is for daily treatments as illustrated in Table 1. There are 4

appointments for the same treatment at the same location with the same VID (VID-7) but with different AIDs (AID-9, AID-10, AID-11, and AID-12). All the AIDs are for the same treatment and at the same location and would have the same VID.

Figure 1. Illustration of appointment ID (AID) and visit ID (VID) possible scenarios.

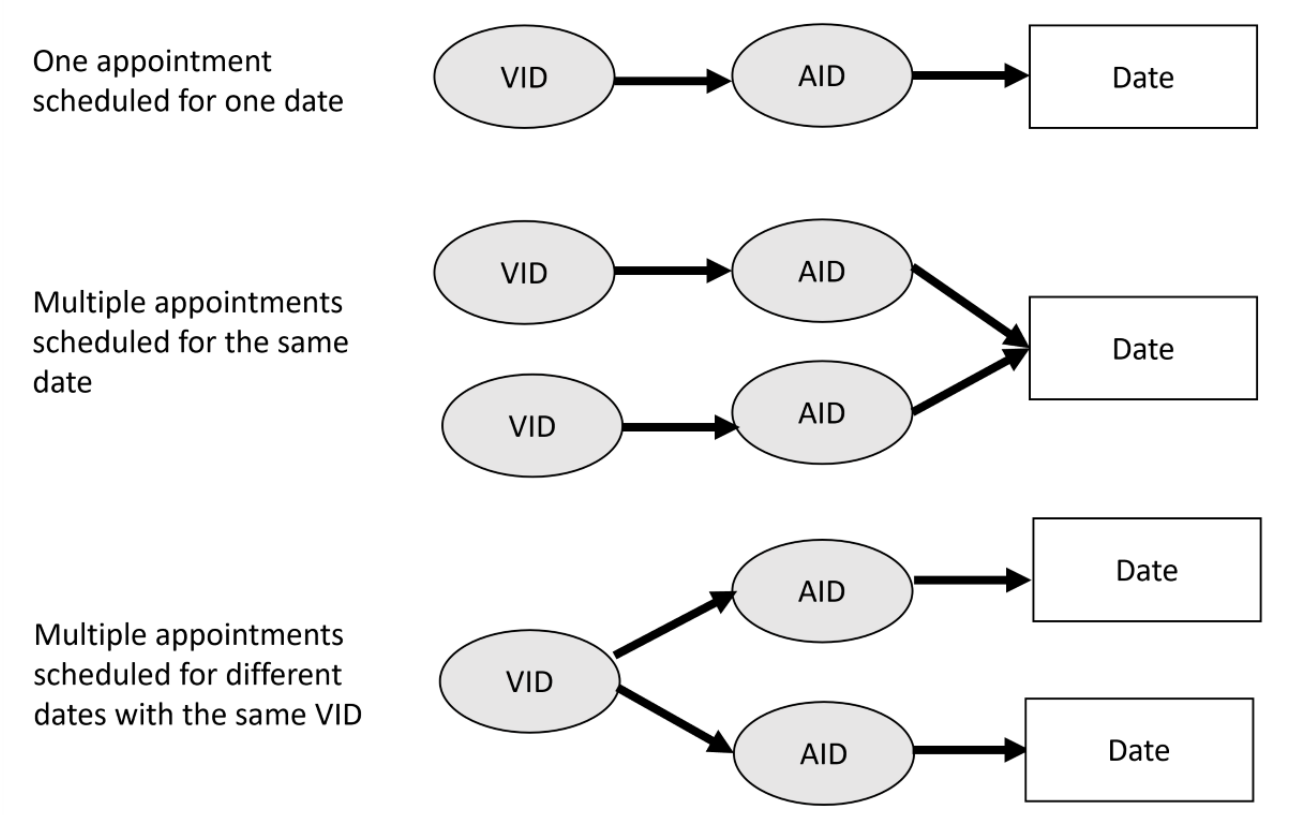


Table . Example individual patient scheduling temporal pattern.

VID[a]	AID[b]	Reason for visit	Location	Scheduled on	Scheduled for	Cancelled on	Rescheduled on	Arrived on
VID-1	AID-1	New consult	Hospital A - 2nd FL[c]	1/1/2023	1/5/2023	—[d]	—	1/5/2023
	AID-2	Colon screening	Hospital B - Ground	1/1/2023	2/1/2023	1/17/2023	—	—
VID-2	AID-3	Skin check	Hospital B - Ground	1/3/2023	1/10/2023	—	—	1/10/2023
VID-3	AID-4	Echocardiogram	Hospital A - Ground	1/5/2023	1/20/2023	—	—	1/20/2023
VID-4	AID-5	C50.912 MRI[e]	Hospital A - Ground	1/10/2023	1/15/2023	—	—	1/15/2023
VID-5	AID-6	LF[f] breast mass - mammo	Hospital A - 2nd FL	1/15/2023	1/15/2023	—	—	1/15/2023
VID-6	AID-7	Follow up	Hospital A - 2nd FL	2/1/2023	2/20/2023	—	2/10/2023	—
VID-6	AID-8	Follow up	Hospital A - 2nd FL	2/10/2023	2/25/2023	—	—	2/25/2023
VID-7	AID-9	Treatment	Infusion center - 2nd FL	4/1/2023	4/5/2023	—	—	4/5/2023
VID-7	AID-10	Treatment	Infusion center - 2nd FL	4/1/2023	4/6/2023	—	—	4/6/2023
VID-7	AID-11	Treatment	Infusion center - 2nd FL	4/1/2023	4/7/2023	—	—	4/7/2023
VID-7	AID-12	Treatment	Infusion center - 2nd FL	4/1/2023	4/8/2023	—	—	4/8/2023

[a]VID: visit identification.

[b]AID:appointment identification.

[c]FL: floor.

[d]Not available.

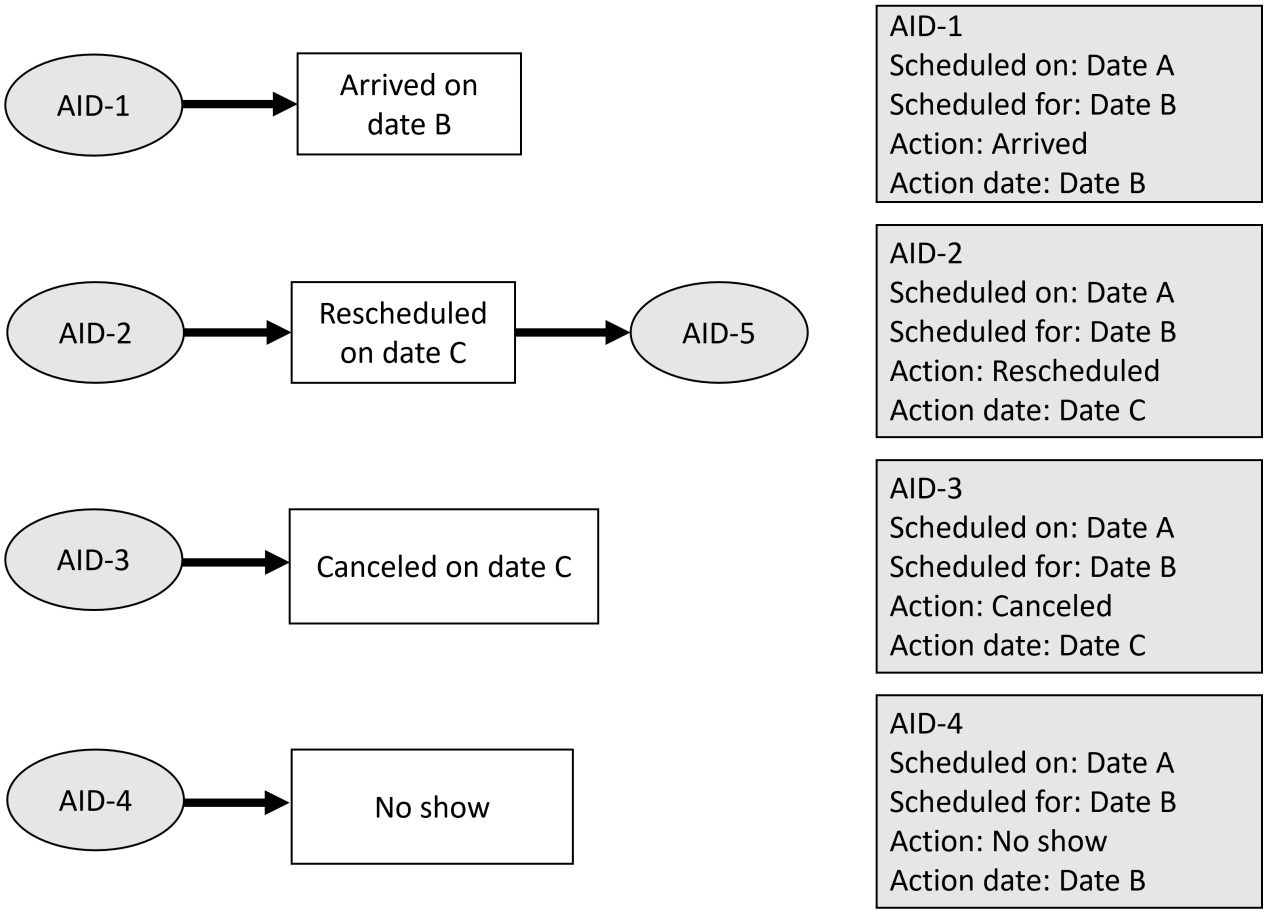
[e]MRI: magnetic resonance imaging.

[f]LF: left.

Appointment Action Outcomes

There are 4 possible action outcomes for an appointment: arrived, rescheduled, canceled, or no show as illustrated in Figure 2. Figure 2 illustrates 4 possible action outcomes for an appointment. Arrived occurs when the patient arrives on the scheduled appointment date. Rescheduled occurs when the appointment needs to be rescheduled for another date. This can be due to multiple reasons, such as patient's preference, medical necessity, financial or transportation issues, circumstances or system related factors, such as being bumped, unresolved insurance authorization, etc. Rescheduling an appointment will result in a new AID. An appointment can also be canceled by the patient or the hospital. For example, a provider could be unavailable due to illness or a scheduling conflict. Similarly, canceled appointments can be caused by a variety of patient or hospital system reasons. Finally, no show occurs when a patient doesn't arrive to an appointment and does not cancel the appointment.

Figure 2. The 4 possible action outcomes for an appointment. AID: appointment ID.



Sources for Scheduling Complexities

Sequence Complexity: Appointment Ordering Sequence

We define schedule sequence complexity as the degree to which appointments are scheduled and arrived to in a nonchronological order. While there are several ways to define temporal complexities, we choose a queuing approach as it most closely aligns with patient scheduling experience [25,26]. As such schedules with low sequence complexity are those where appointments are scheduled and arrived in chronical order. This follows the general queuing rule of first in-first out: appointments scheduled first are arrived to first which minimizes the number of outstanding appointments at any given time. Schedules with more sequence complexity are those where appointments are scheduled and arrived to in a nonchronological order. Using the illustration in Table 1, AID-1 and AID-3 are examples of appointments scheduled and arrived in chronical order. AID-1 is scheduled before AID-3 and AID-1 is arrived to before AID-3. AID-4 and AID-5 are examples of appointments scheduled and arrived in nonchronological order. AID-4 is scheduled before AID-5 but the patient arrived to AID-5 before AID-4. This complexity can be caused by many factors such as appointments scheduled for the far future, canceling and rescheduling of appointments, or emergent appointments. These factors can increase schedule challenges both for the patient and scheduling systems.

Resolution Complexity: Unresolved Appointments

No shows or canceled appointments without rescheduling or reason can increase scheduling complexity in a patient's care. Missing appointments leads to increase patient risk for cancer recurrence and mortality [27,28], and inefficiency for the health care system including lost revenue [29-31]. These unresolved appointments have no resolution, leaving uncertainty about potential delays in treatment and care. However, there are sometimes canceled appointments because of changes in treatment plans or no shows that are resolved through another action. Actions for these resolved appointments often co-occur with action dates for arrived or rescheduled appointments. Hence, we define resolution complexity as the number of no shows or canceled appointments on dates that do not co-occur with other action dates divided by the total number of no shows or canceled appointments.

Location Complexity: Appointments at Multiple Facilities

Having care at multiple facilities or locations can also increase scheduling complexities as this usually means more coordination and travel between facilities. Intuitively, a schedule with lower location complexity will have fewer facilities for care on the same day. A schedule with higher location complexity will more often require the patient to attend different facilities for care on the same day. Location complexity is calculated as the number of arrived dates involving 2 or more different locations divided by the total number of arrived dates.

Calculating Scheduling Complexity

The algorithm for calculating a schedule's scheduling complexity is described below (Textbox 1) . First, schedule data is separated into arrived and not arrived appointments, "ARRIVED" and "NONARRIVED" respectively. ARRIVED appointments are aggregated at the date level. For each AID in ARRIVED, if there exist other AIDs with scheduled on dates preceding the current AID's scheduled on date and these subsequent appointments were attended after the current AID's date, then the count of out-of-order occurrences is increased. Sequence complexity is calculated as the ratio of out-of-order counts to the total count of distinct arrived dates in ARRIVED. Next, for each AID in the NONARRIVED group, an action date is determined, representing the date when an appointment was either canceled, bumped, or scheduled but resulted in a no-show. If this action date does not appear in the dataset of ARRIVED appointments, then the count of unresolved cases is increased by 1. Resolution complexity is then computed as the ratio of unresolved counts to the total count of AIDs within the NONARRIVED group. Location complexity is calculated as the number of arrived dates in ARRIVED involving 2 or most different locations divided by the total number of arrived dates in ARRIVED. Finally, a composite metric scheduling complexity is the harmonic mean of sequence complexity, resolution complexity, and location complexity.

Textbox 1. ALGORITHM: Deriving scheduling complexity

ARRIVED, NONARRIVED ← Separate data into arrived and not arrived appointments

For each AID in **ARRIVED**:

If there are other AID date that was made before current AID date and arrived to after current AID date

Out of order count += 1

sequence complexity = out of order count / total count of unique arrived to dates in **ARRIVED**

For each AID in **NONARRIVED**:

Action date = canceled or bumped date or scheduled date for no-show

If Action date not in Arrived_data:

unresolved += 1

resolution complexity = unresolved count / total count of nonArrived AIDs

For each arrived date with multiple AIDs in **ARRIVED**

If AIDs are at different locations:

Location count += 1

location complexity = location count / total number of arrived dates with multiple AIDs in **ARRIVED**

scheduling complexity = 3 / (1/sequence complexity + 1/resolution complexity + 1/facility complexity)

Case Example and Study Background

To evaluate the use of scheduling complexity, we calculated the scheduling complexity for 38 patients with breast cancer who had hypertension or diabetes, as part of a larger health disparities project to support Black patients with cancer with comorbidities by mobile health and CHW support [24]. The 38 patients with breast cancer were enrolled in a 6-month comorbidity management intervention at an urban hospital in the Washington, DC area that serves low-income patients. This data was collected through the parent study whereby Black patients with breast and prostate cancer were recruited for a 6-month comorbidity management intervention. For tthis analysis, we focused on the association between scheduling patterns and social needs. Given significant differences in course of treatment by cancer site and time since diagnosis we limited the sample to those who had a diagnosis of breast cancer, all of whom had been diagnosed within the previous year. We further limited the sample to the 38 patients with breast cancer whom we had reliable appointment level data. The women in our sample were from an urban hospital in the Washington, DC area that primarily serves low-income patients. Black women with breast cancer in the DC area are a high priority sample, due to the increased mortality rate relative to White women [32].

CHW Call Logs

In addition, as part of the parent study, we conducted a qualitative context analysis of CHWs' call logs to identify social needs that arose throughout the study [33]. We used a deductive approach, first applying discrete categories from the health system screening tool focused on domains of food insecurity, housing instability, transportation, employment, financial strain, and utilities. Additional needs that were documented but did not fit into a predetermined category, such as access to wigs or special bras, were added as new codes using an inductive descriptive coding method to match social needs domain described by the Social Interventions Research and Evaluation Network (SIREN) [34].

Data Analysis

For this pilot evaluation, we use 1 year of scheduling data for each patient starting from their date of diagnosis. Furthermore, 1 year of scheduling data was chosen because the majority of patients completed curative treatment within the first year after diagnosis. Sequence complexity, resolution complexity, location

complexity, and scheduling complexity was calculated for each patient separately. In addition, we calculated count-based metrics: arrived ratio, rescheduled ratio, canceled ratio, and no-show ratio. We define an aggregated count-based adjustment metric as the harmonic mean of rescheduled ratio, canceled ratio, and no-show ratio. Count-based adjustments and scheduling complexities are stratified using quartiles and compared. We stratify patients into high, medium, and low complexities using the upper quartile, middle quartiles, or lower quartiles respectively. We used χ^2 test to compare our scheduling complexity metric to count-based adjustments because they are commonly used in first order analysis of scheduling data. All analysis was done in Python (version 3.0; Python Software Foundation).

Ethical Considerations

This research was approved by the Georgetown University Institutional Review Board (STUDY00003543). This study was registered with ClinicalTrials.gov (NCT04836221). Written informed consent was obtained before conducting all study procedures, which allowed for secondary analysis of participant data without additional consent. A Health Insurance Portability and Accountability Act waiver and access to medical record data was included in the signed informed consent. All data included in this manuscript are deidentified and reported in aggregate. Data obtained through the study adhere to data protection and institutional review board standards as determined by the governing institution. Participants were compensated US $50 at the completion of the study.

Results

Schedule Descriptives

A total of 38 female patients with breast cancer with an average age of 67.1 (SD 8.5), from 3 referring oncology providers had a median of 88 unique AID (IQR 60.3: first quartile [Q1]=59, third quartile [Q3]=119.3), 62 arrived appointments (IQR 47.8: Q1=38.7, Q3=86.5), 13 rescheduled appointments (IQR 13.5: Q1=7.5, Q3=21), 9 canceled appointments (IQR 10.0: Q1=3, Q3=13), and 1.5 missed appointments (IQR 5: Q1=0, Q3=5). The median nonarrived ratio was 0.304 (IQR 0.161: Q1=0.233, Q3=0.394). The median rescheduled ratio was 0.154 (IQR 0.080: Q1=0.127, Q3=0.207). The median canceled ratio was 0.098 (IQR 0.081: Q1=0.046, Q3=0.127) and the median no-show ratio was 0.019 (IQR 0.049: Q1=0, Q3=0.049). Figure 3 is a summary boxplot of nonarrived, rescheduled, canceled, and no-show ratios with nonarrived ratio being the largest. The median sequence complexity was 0.200 (IQR 0.100: Q1=0.140, Q3=0.240). Figure 4 is a summary boxplot of sequence, resolution, location, and scheduling complexities with location complexity being the largest. The median resolution complexity was 0.372 (IQR 0.398: Q1=0.188, Q3=0.586). The median location complexity was 0.464 (IQR 0.371: Q1=0.279, Q3=0.650). Finally, the median scheduling complexity was 0.239 (IQR 0.173: Q1=0.156, Q3=0.329).

Figure 3. Summary boxplots of nonarrived ratio, rescheduled ratio, canceled ratio, and no-show ratio.

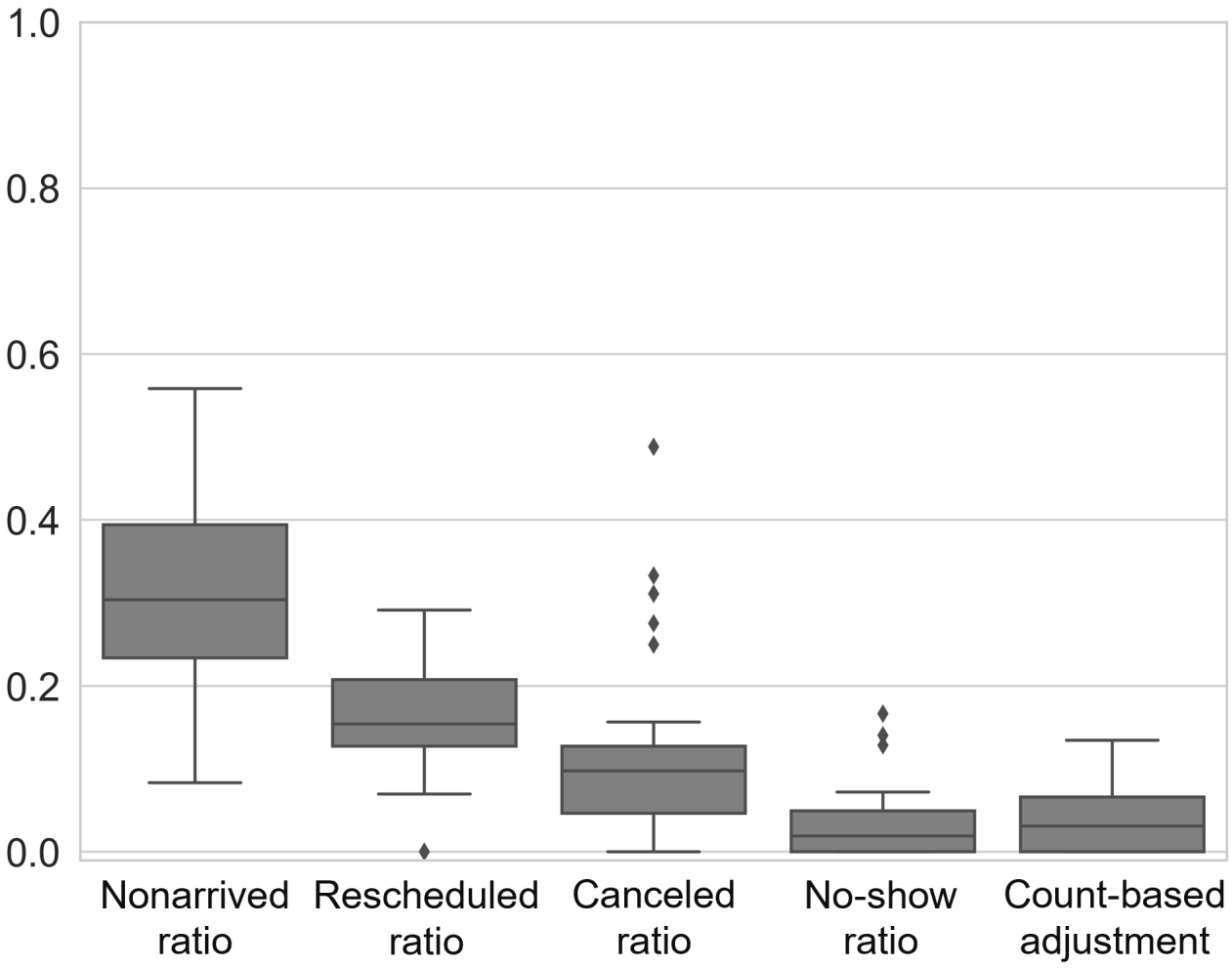
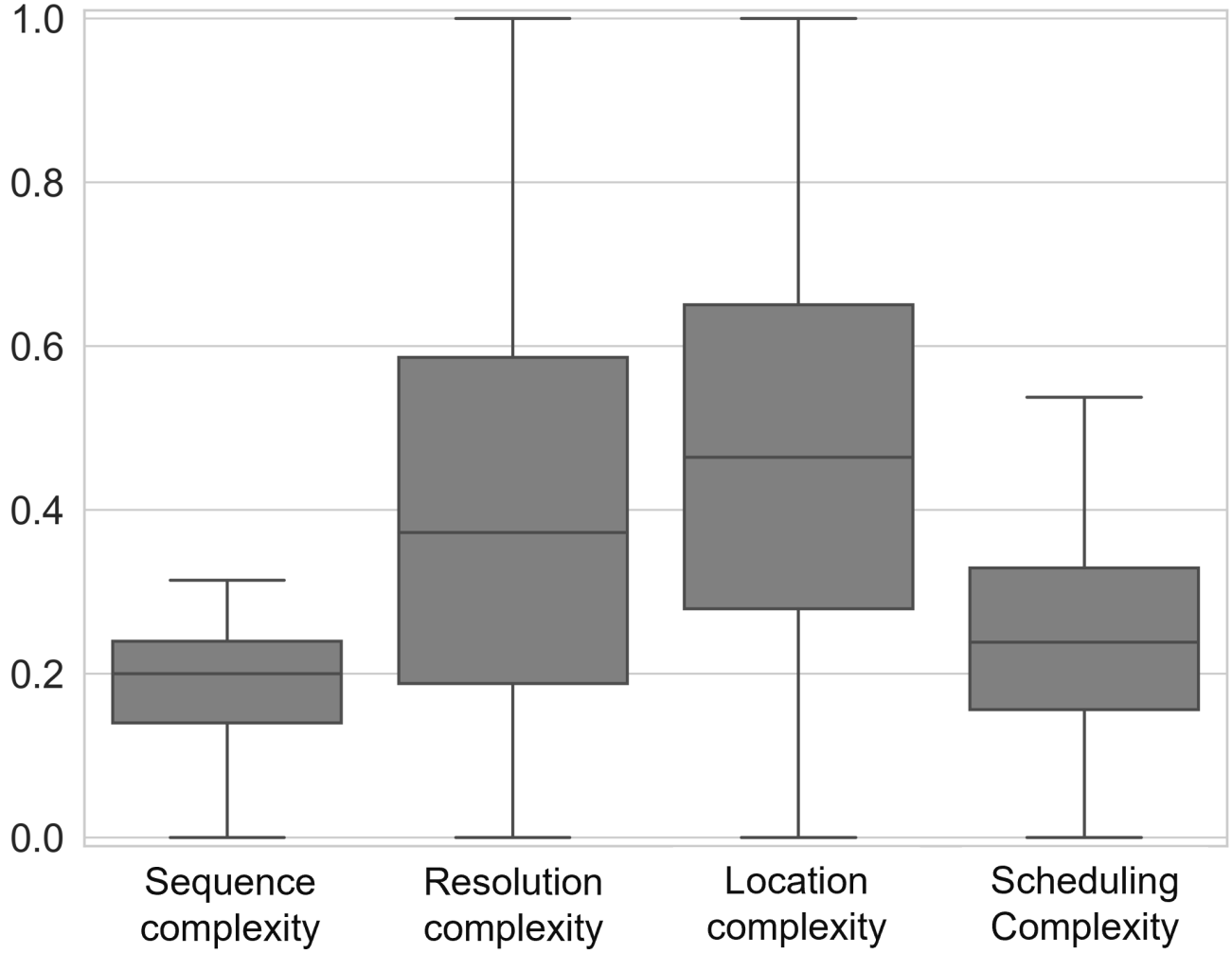


Figure 4. Summary boxplots of sequence complexity, resolution complexity, location complexity, and scheduling complexity.



Comparison of Count-Based Adjustments With Scheduling Complexity

A total of 18 patients exhibited medium count-based complexities, 10 patients exhibited high count-based complexities, and 10 patients exhibited low count-based complexities. Similarity, 18 patients exhibited medium scheduling complexities, 10 patients exhibited high scheduling complexities, and 10 patients exhibited low scheduling complexities. There was no statistically significant difference in count-based adjustments and scheduling complexity bins (χ^2_4=6.296, P=.18). Table 2 shows the count-based and scheduling complexities were the same for 16 patients, 11 of which had both medium scheduling and count-based complexities. Furthermore, 5 patients exhibited high scheduling complexity with low count-based adjustments and 2 patients who exhibited high count-based adjustments with low scheduling complexity.

Table . Correlation of scheduling and count-based complexities binned by low, medium, and high.

Scheduling complexity	Count-based adjustments		
	Low	Medium	High
Low	3	5	2
Medium	2	11	5
High	5	2	3

In addition, Figure 5 gives examples of high scheduling complexity and low count-based adjustments, high count-based adjustments and low scheduling complexity, low scheduling complexity and low count-based adjustments, and high scheduling complexity and high count-based adjustments. Patient A had both low count-based adjustments and low scheduling complexity. Patient A's schedule is a good example of appointments following the first-in first-out pattern, that is appointments scheduled first will be arrived to first. In addition, Patient A only had 3 rescheduled appointments. Patient B had high count-based adjustments but low scheduling complexity; 40% (8/20) of Patient B's AID were rescheduling actions which likely contributing to a high count-based adjustments. However, Patient B had low scheduling complexity because these

rescheduling actions occurred only on 2 separate days and followed a first-in first-out sequence. Patient C had low count-based adjustments and high scheduling complexity. Although Patient C had few rescheduling actions (resulting in a low count-based adjustments), her appointments were largely scheduled and arrived to not in chronical order (resulting in a high scheduling complexity). Patient D had both high count-based adjustments and high scheduling complexity. In addition to nonchronical ordering of action, Patient D had 2 canceled and 1 no-show appointment action outcomes.

Figure 5. Examples of high scheduling complexity and low count-based adjustments, high count-based adjustments and low scheduling complexity, low scheduling complexity and low count-based adjustments, and high scheduling complexity and high count-based adjustments. AID: appointment identification; VID: visit identification.

Patient A
Count-based adjustment: Low
Scheduling complexity: Low

VID	AID	Scheduled on	Scheduled for	Action	Action Date
1	1	8/7	9/9	Arrived	9/9
2	2	8/7	9/11	Arrived	9/11
3	3	9/9	9/17	Arrived	9/17
3	4	9/9	9/17	Arrived	9/17
4	5	9/9	9/30	Arrived	9/30
4	6	9/9	9/30	Arrived	9/30
5	7	9/21	10/13	Arrived	10/13
5	8	9/21	10/13	Arrived	10/13
6	9	10/2	10/23	Arrived	10/23
7	10	9/28	11/11	Arrived	11/11
8	11	11/9	12/18	Arrived	12/18
9	12	11/9	12/29	Arrived	12/29
10	13	1/4	1/8	Rescheduled	1/6
10	14	1/4	1/8	Rescheduled	1/6
10	15	1/4	1/8	Rescheduled	1/6
10	16	1/6	1/19	Arrived	1/19
10	17	1/6	1/19	Arrived	1/19
10	18	1/6	1/19	Arrived	1/19
11	19	1/9	1/20	Arrived	1/20
12	20	1/9	1/21	Arrived	1/21

Patient B
Count-based adjustment : High
Scheduling complexity: Low

VID	AID	Scheduled on	Scheduled for	Action	Action Date
1	1	2/6	3/16	Arrived	3/16
1	2	2/6	4/10	Arrived	4/10
2	3	3/17	4/14	Rescheduled	3/30
2	4	3/17	4/15	Rescheduled	3/30
2	5	3/17	4/16	Rescheduled	3/30
2	6	3/17	4/18	Rescheduled	3/30
2	7	3/17	4/19	Rescheduled	3/30
2	8	3/30	4/21	Arrived	4/21
2	9	3/30	4/22	Arrived	4/22
2	10	3/30	4/23	Arrived	4/23
2	11	3/30	4/24	Arrived	4/24
2	12	3/30	4/25	Arrived	4/25
3	13	6/9	9/10	Arrived	9/10
4	14	7/14	9/22	Rescheduled	9/21
4	15	7/14	9/22	Rescheduled	9/21
5	16	7/14	9/29	Rescheduled	9/21
4	17	9/21	9/24	Arrived	9/24
4	18	9/21	9/24	Arrived	9/24
5	19	9/21	10/2	Arrived	10/2
6	20	9/30	10/3	Arrived	10/3

Patient C
Count-based adjustment : Low
Scheduling complexity: High

VID	AID	Scheduled on	Scheduled for	Action	Action Date
1	1	5/20	5/21	Arrived	5/21
2	2	5/22	5/23	Rescheduled	5/22
2	3	5/22	6/2	Arrived	6/2
3	4	5/24	5/24	Arrived	5/24
4	5	5/21	5/28	Arrived	5/28
5	6	5/21	5/29	Arrived	5/29
6	7	5/21	5/31	Rescheduled	5/23
6	8	5/23	6/5	Arrived	6/5
7	9	6/4	6/4	Arrived	6/4
8	10	5/29	6/4	Rescheduled	6/2
8	11	6/2	6/7	Arrived	6/7
9	12	6/10	7/1	Arrived	7/1
10	13	6/7	7/26	Rescheduled	7/7
10	14	7/7	8/9	Arrived	8/9
11	15	6/21	7/29	Arrived	7/29
12	16	7/31	8/6	Arrived	8/6
13	17	7/1	8/7	Arrived	8/7
14	18	6/21	8/7	Rescheduled	7/20
14	19	7/20	8/10	Arrived	8/10
15	20	8/7	8/9	Arrived	8/9

Patient D
Count-based adjustment : High
Scheduling complexity: High

VID	AID	Scheduled on	Scheduled for	Action	Action Date
1	1	4/7	4/10	Arrived	4/10
2	2	4/2	4/25	Arrived	4/25
3	3	4/25	5/2	Rescheduled	4/30
3	4	4/30	5/1	Arrived	5/1
4	5	5/2	5/5	Arrived	5/5
-	6	4/28	5/5	Canceled	5/2
5	7	5/8	5/8	Arrived	5/8
6	8	4/29	5/11	Arrived	5/11
7	9	4/30	5/12	Rescheduled	5/5
7	10	5/5	5/20	Arrived	5/20
8	11	5/9	5/13	Arrived	5/13
-	12	5/8	5/17	No Show	5/17
-	13	5/8	5/18	Canceled	5/18
9	14	4/28	5/22	Rescheduled	5/12
9	15	4/28	5/22	Rescheduled	5/12
9	16	5/12	5/14	Arrived	5/14
9	17	5/12	5/14	Arrived	5/14
10	18	5/19	5/30	Arrived	5/30
11	19	5/10	6/1	Arrived	6/1
11	20	5/10	6/2	Arrived	6/2

Context From Call Logs: Transportation and Housing Needs

A total of 15 patients specifically indicated transportation or housing insecurity issues. Transportation concerns included "legally blind and worried about metro access," "[patient] feeling unsafe on metro," "transportation challenges to and from appointments," "extensive travel requirements," "making medical transportation rides," and "needing transportation assistance." Housing concerns included "help with finding affordable housing options," concerns related to advocating for tenant rights, home repair needs, "[patient] moving in with relative for a few months to save money," "having to find temporary housing while home is being repaired," "help finding rental assistance programs." Additional examples of identified social needs included a demanding work schedule, complexities with an eye surgery, and the additional responsibility of caring for an ill mother. Scheduling complexity was more sensitive to housing and transportation needs. 86.7% (13/15) of patients specifically indicated transportation or housing insecurity issues were identified as medium or high scheduling complexity compared with 65.2% (15/23) of patients who did not specifically indicate transportation or housing insecurity issues. On the other hand, 60% (9/15) of patients specifically indicated transportation or housing insecurity issues were identified as high with count-based adjustments compared with 82.6% (19/23) of patients who did not specifically indicate transportation or housing insecurity issues Table 3.

Table . Percentage of patients with medium or high complexities by transportation or housing insecurity needs.

Complexity type	Indicated transportation or housing needs (n=15), n (%)	Did not indicate transportation or housing needs (n=23), n (%)
Count-based adjustments	9 (60)	19 (82.6)
Scheduling complexity	13 (86.7)	15 (65.2)

Discussion

Principal Findings

Scheduling complexity stratification provides a novel lens to complement traditional count-based metrics for analyzing scheduling data. The results show that scheduling complexity can identify patients with complex but nonchronical scheduling behaviors missed by traditional count-based metrics. In addition, the study highlights that resolution and location complexity can also serve as an indicator for additional care requirements.

Comparison With Previous Work

This study shows a potential link between transportation and housing needs with schedule complexity. This study reinforces previous research relating social risk factors and schedule complexities [4,9]. Our results complement these findings as it relates to transportation and housing needs and highlights the potential use of the scheduling complexity algorithm to identify patients who might benefit from additional CHW support. Through earlier identification from CHWs, scheduling complexity could help narrow inequities in cancer-care which emerge from social needs. Future studies are needed to better understand temporal sensitivity of this approach, or how quickly in-need patients could be identified.

Support for CHW and Patient Navigators

By examining the temporal patterns of health care use, we gain a more comprehensive view of patients' experiences. Instead of relying solely on infrequent screeners, scheduling complexity can give CHWs and patient navigators a more "real-time" view of patients who might require more support in managing their health care journey, for example patients with changing, complex, and distributed care and changes in living conditions or social needs. In addition, scheduling complexity could also be used to identify care plans that might involve more complexity and preemptively identify patients that might need more support. This data-driven approach can help complement the often manual process for identifying patients who might benefit from additional assistance, potentially affording CHWs and patient navigators more time to directly care and help patients [13]. Additional research is needed to evaluate the utility of this algorithm in near "real-time" applications for CHWs.

Limitations and Future Directions

There are several limitations to consider in this pilot work. First, this research is constrained by its retrospective analysis design, relying on historical data and records. Second, a naïve weighting approach was taken in which sequence, resolution, and location complexity are weighted equally in the algorithm. These components might require different weights depending on circumstances. For example, for cancer care at integrated cancer centers, location complexity would probably be less important than sequence and resolution complexities. Third, the study exclusively focuses on the scheduling system of a single urban cancer institute. As such, the results only reflect the central tendencies given the specific conditions and scheduling pattern characteristics of our patient cohort. The generalizability this approach and interpretation of the metrics to other health care systems and patient conditions will need to be explored. Fourth, while this approach has provided insights, it does not fully capture the entirety of factors that contribute to scheduling intricacies, such as resource allocation, patient preferences, and staff availability. Further validation is needed beyond this case example, particularly with a larger patient sample to better capture variations in the scheduling complexity data. Nevertheless, this method could complement other approaches such as patient and scheduler interviews. This work should also be explored in a larger sample size to further explore our hypotheses generated about scheduling complexity.

Conclusion

Patients facing complex health care journeys often experience significant impacts on various aspects of their lives, including family dynamics and work commitments. While there was no

statistically significant difference in count-based adjustments and scheduling complexity bins, we showed that scheduling complexity can uniquely identify patients with complex and nonchronical schedule behaviors. We highlight the potential use of scheduling complexity in identifying patients who might need care coordination support especially as it relates to transportation and housing needs.

Acknowledgments

The authors acknowledge and appreciate all the work by our community health worker partners. This work was supported by Pfizer and administered by the American Cancer grant (63615141). KA's effort on this paper was supported by the National Cancer Institute (grant T32CA261787).

Disclosure

Generative AI was not used in any portion of the manuscript writing.

Data Availability

The datasets generated during and analyzed during this study are not publicly available due appointment scheduling privacy but are available in a limited format from the corresponding author on reasonable request.

Authors' Contributions

All authors have made substantial contributions to the interpretation of the data, and the drafting, revising, and approval of the work. AF has made substantial contributions to the conception, design of the work, and the acquisition, analysis, and interpretation of the data. CB has made substantial contributions to the acquisition, analysis, and interpretation of the data. LS has made substantial contributions to the conception, design of the work, and the interpretation of the data. CG has made substantial contributions to the conception, design of the work, and the interpretation of the data. KA has made substantial contributions to the interpretation of the data. HA has made substantial contributions to the conception, design of the work, and the acquisition, analysis, and interpretation of the data.

Conflicts of Interest

None declared.

References

1. Hooshangi-Tabrizi P, Contreras I, Bhuiyan N, Batist G. Improving patient-care services at an oncology clinic using a flexible and adaptive scheduling procedure. Expert Syst Appl 2020 Jul;150:113267. [doi: 10.1016/j.eswa.2020.113267]
2. Marynissen J, Demeulemeester E. Literature review on multi-appointment scheduling problems in hospitals. Eur J Oper Res 2019 Jan;272(2):407-419. [doi: 10.1016/j.ejor.2018.03.001]
3. Siekkinen M, Kuokkanen L, Kuusisto H, et al. Work empowerment among cancer care professionals: a cross-sectional study. BMC Health Serv Res 2021 May 25;21(1):502. [doi: 10.1186/s12913-021-06528-8] [Medline: 34034734]
4. Chan RJ, Milch VE, Crawford‐Williams F, et al. Patient navigation across the cancer care continuum: an overview of systematic reviews and emerging literature. CA: A Cancer J Clinicians 2023 Nov;73(6):565-589. [doi: 10.3322/caac.21788]
5. Bairati I, Fillion L, Meyer FA, Héry C, Larochelle M. Women's perceptions of events impeding or facilitating the detection, investigation and treatment of breast cancer. Eur J Cancer Care (Engl) 2006 May;15(2):183-193. [doi: 10.1111/j.1365-2354.2005.00635.x] [Medline: 16643266]
6. Adam R, Nair R, Duncan LF, et al. Treatment burden in individuals living with and beyond cancer: a systematic review of qualitative literature. PLoS One 2023;18(5):e0286308. [doi: 10.1371/journal.pone.0286308] [Medline: 37228101]
7. Alfieri S, Brunelli C, Capri G, Caraceni A, Bianchi GV, Borreani C. A qualitative study on the needs of women with metastatic breast cancer. J Cancer Educ 2022 Oct;37(5):1322-1331. [doi: 10.1007/s13187-020-01954-4] [Medline: 33486712]
8. Luckett R, Pena N, Vitonis A, Bernstein MR, Feldman S. Effect of patient navigator program on no-show rates at an academic referral colposcopy clinic. J Womens Health (Larchmt) 2015 Jul;24(7):608-615. [doi: 10.1089/jwh.2014.5111] [Medline: 26173000]
9. Percac-Lima S, Cronin PR, Ryan DP, Chabner BA, Daly EA, Kimball AB. Patient navigation based on predictive modeling decreases no-show rates in cancer care. Cancer 2015 May 15;121(10):1662-1670. [doi: 10.1002/cncr.29236] [Medline: 25585595]
10. Martinez J, Knickman J. Community Health Workers: A Critical Link for Improving Health Outcomes and Promoting Cost-Effective Care in the Era of Health Reform: New York State Health Foundation; 2010.
11. Budde H, Williams GA, Scarpetti G, Kroezen M, Maier CB. What Are Patient Navigators and How Can They Improve Integration of Care?: European Observatory on Health Systems and Policies; 2022; 2022. URL: https://www.ncbi.nlm.nih.gov/books/NBK577640/ [accessed 2025-01-02]

12. Wang T, Huilgol Y, Black J, et al. Pre-appointment nurse navigation: Patient-centered findings from a survey of patients with breast cancer. CJON 2021 Oct 1;25(5):E57-E62. [doi: 10.1188/21.CJON.E57-E62]
13. Kline RM, Rocque GB, Rohan EA, et al. Patient navigation in cancer: the business case to support clinical needs. J Oncol Pract 2019 Nov;15(11):585-590. [doi: 10.1200/JOP.19.00230] [Medline: 31509483]
14. Vargas RB, Ryan GW, Jackson CA, Rodriguez R, Freeman HP. Characteristics of the original patient navigation programs to reduce disparities in the diagnosis and treatment of breast cancer. Cancer 2008 Jul 15;113(2):426-433. [doi: 10.1002/cncr.23547] [Medline: 18470906]
15. Battaglia TA, Darnell JS, Ko N, et al. The impact of patient navigation on the delivery of diagnostic breast cancer care in the National Patient Navigation Research Program: a prospective meta-analysis. Breast Cancer Res Treat 2016 Aug;158(3):523-534. [doi: 10.1007/s10549-016-3887-8] [Medline: 27432417]
16. Hoffman HJ, LaVerda NL, Young HA, et al. Patient navigation significantly reduces delays in breast cancer diagnosis in the District of Columbia. Cancer Epidemiol Biomarkers Prev 2012 Oct;21(10):1655-1663. [doi: 10.1158/1055-9965.EPI-12-0479] [Medline: 23045540]
17. Payerchin R, Lutton L. Staffing challenges: salary and pay trends, part 1. 2023. URL: https://www.medicaleconomics.com/view/staffing-challenges-salary-and-pay-trends-part-1 [accessed 2025-01-02]
18. Chen PS, Hong IH, Hou Y, Shao YC. Healthcare scheduling policies in a sequence-number based appointment system for outpatients' arrivals: early, on time, or late? Computers & Industrial Engineering 2019 Apr;130:298-308. [doi: 10.1016/j.cie.2019.02.033]
19. Liang B, Turkcan A, Ceyhan ME, Stuart K. Improvement of chemotherapy patient flow and scheduling in an outpatient oncology clinic. Int J Prod Res 2015 Dec 17;53(24):7177-7190. [doi: 10.1080/00207543.2014.988891]
20. Sang P, Begen MA, Cao J. Appointment scheduling with a quantile objective. Computers & Operations Research 2021 Aug;132:105295. [doi: 10.1016/j.cor.2021.105295]
21. Yokouchi M, Aoki S, Sang H, Zhao R, Takakuwa S. Operations analysis and appointment scheduling for an outpatient chemotherapy department. Presented at: The 2012 Winter Simulation Conference (WSC); Dec 9-12, 2012; Berlin, Germany. [doi: 10.1109/WSC.2012.6464990]
22. Woodcock E, Sen A, Weiner J. Automated patient self-scheduling: case study. J Am Med Inform Assoc 2022 Aug 16;29(9):1637-1641. [doi: 10.1093/jamia/ocac087] [Medline: 35652165]
23. Lasko TA, Atlas SJ, Barry MJ, Chueh HC. Automated identification of a physician's primary patients. J Am Med Inform Assoc 2006;13(1):74-79. [doi: 10.1197/jamia.M1876] [Medline: 16221940]
24. Sandling J, Schubel L, Prahbu TM, Hammond T, Gallagher C, Arem H. Identifying health-related social needs among cancer patients with comorbidities. Cancer Survivorship Research & Care 2024 Dec 31;2(1):2362672. [doi: 10.1080/28352610.2024.2362672]
25. Hassin R, Mendel S. Scheduling arrivals to queues: a single-server model with no-shows. Manag Sci 2008 Mar;54(3):565-572. [doi: 10.1287/mnsc.1070.0802]
26. Cormen TH, Leiserson CE, Rivest RL, Stein C. Introduction to Algorithms: MIT Press; 2022.
27. Guay MOD, Tanzi S, Arregui M, Chisholm G, de la Cruz M, Bruera E. Characteristics and outcomes of advanced cancer patients who miss outpatient supportive care consult appointments. Supp Care Cancer 2014 Oct;22(10):2869-2874. [doi: 10.1007/s00520-014-2254-8]
28. Weaver KE, Camacho F, Hwang W, Anderson R, Kimmick G. Adherence to adjuvant hormonal therapy and its relationship to breast cancer recurrence and survival among low-income women. Am J Clin Oncol 2013 Apr;36(2):181-187. [doi: 10.1097/COC.0b013e3182436ec1] [Medline: 22314001]
29. Triemstra JD, Lowery L. Prevalence, predictors, and the financial impact of missed appointments in an academic adolescent alinic. Cureus 2018 Nov 19;10(11):e3613. [doi: 10.7759/cureus.3613] [Medline: 30680269]
30. Husain-Gambles M, Neal RD, Dempsey O, Lawlor DA, Hodgson J. Missed appointments in primary care: questionnaire and focus group study of health professionals. Br J Gen Pract 2004 Feb;54(499):108-113. [Medline: 14965389]
31. Adepoju OE, Angelocci T, Matuk-Villazon O. Increased revenue from averted missed appointments following telemedicine adoption at a large federally qualified health center. Health Serv Insights 2022;15:11786329221125409. [doi: 10.1177/11786329221125409] [Medline: 36186737]
32. Age-adjusted death rate due to breast cancer. DC Health Matters. URL: https://www.dchealthmatters.org/indicators/index/view?indicatorId=94&localeId=130951 [accessed 2025-01-02]
33. Miles MB, Huberman AM. Qualitative Data Analysis: An Expanded Sourcebook: Sage; 1994.
34. De Marchis E, Brown E, Aceves B. State of the Science of Screening in Healthcare Settings: Social Interventions Research and Evaluation Network; 2022.

Abbreviations

AID: appointment identification
CHW: community health worker
MRI: magnetic resonance imaging

Q1: first quartile
Q3: third quartile
VID: visit identification

Edited by N Cahill; submitted 24.02.24; peer-reviewed by A AL-Asadi, L Wreyford, Z Ping; revised version received 05.11.24; accepted 05.11.24; published 17.01.25.

Please cite as:
Fong A, Boxley C, Schubel L, Gallagher C, AuBuchon K, Arem H
Identifying Complex Scheduling Patterns Among Patients With Cancer With Transportation and Housing Needs: Feasibility Pilot Study
JMIR Cancer 2025;11:e57715
URL: https://cancer.jmir.org/2025/1/e57715
doi:10.2196/57715

Original Paper

Patient Voices: Multimethod Study on the Feasibility of Implementing Electronic Patient-Reported Outcome Measures in a Comprehensive Cancer Center

Cinzia Brunelli[1], MSc, PhD; Sara Alfieri[2], PsyD, PhD[‡]; Emanuela Zito[3], BEng; Marco Spelta[4], BEng; Laura Arba[1], RN; Linda Lombi[5], PhD; Luana Caselli[1], PhD; Augusto Caraceni[1,6], MD, PhD; Claudia Borreani[2], PsyD; Anna Roli[7], MDI; Rosalba Miceli[8], MSc, PhD; Gabriele Tine'[8], MSc; Ernesto Zecca[1], MD; Marco Platania[9], MD; Giuseppe Procopio[9], MD; Nicola Nicolai[10], MD; Luigi Battaglia[11], MD; Laura Lozza[12], MD; Morena Shkodra[1], MD, PhD; Giacomo Massa[1], MD; Daniele Loiacono[4], PhD; Giovanni Apolone[13], MD, PhD

[1]Palliative Care, Pain Therapy and Rehabilitation Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

[2]Clinical Psychology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

[3]Information and Communication Technology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

[4]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, Italy

[5]Department of Sociology, Università Cattolica del Sacro Cuore, Milano, Italy

[6]Department of Medical and Community Sciences, Università Degli Studi di Milano, Milano, Italy

[7]Quality, Education and Data Protection Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

[8]Biostatistics for Clinical Research Unit, Epidemiology and Data Science Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

[9]Department of Medical Oncology 1, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

[10]Department of Urology, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

[11]Colorectal Surgery Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

[12]Radiation Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

[13]Scientific Directorate, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milano, Italy

[‡]Patient Voices Study Group

Corresponding Author:
Sara Alfieri, PsyD, PhD
Clinical Psychology Unit
Fondazione IRCCS Istituto Nazionale dei Tumori di Milano
Via Venezian, 1
Milano, 20133
Italy
Phone: 39 0223903179
Email: sara.alfieri@istitutotumori.mi.it

Abstract

Background: "Patient Voices" is a software developed to promote the systematic collection of electronic patient-reported outcome measures (ePROMs) in routine oncology clinical practice.

Objective: This study aimed to assess compliance with and feasibility of the Patient Voices ePROM system and analyze patient-related barriers in an Italian comprehensive cancer center.

Methods: Consecutive patients with cancer attending 3 outpatient clinics and 3 inpatient wards were screened for eligibility (adults, native speakers, and being able to fill in the ePROMs) and enrolled in a quantitative and qualitative multimethod study. Compliance, reasons for not administering the ePROMs, patients' interaction needs, and patient-perceived System Usability Scale (range 0-100) were collected; semistructured interviews were carried out in a subsample of patients.

Results: From June 2020 to September 2021, a total of 435 patients were screened, 421 (96.7%) were eligible, and 309 completed the ePROMs (309/421, 73.4%; 95% CI 69.8%-77.5%; mean age 63.3, SD 13.7 years). Organization problems and patient refusal were the main reasons for not administering the ePROMs (outpatients: 40/234, 17.1% and inpatients: 44/201, 21.9%). Help for

tablet use was needed by 27.8% (47/169) of outpatients and 10.7% (15/140) of inpatients, while the support received for item interpretation was similar in the 2 groups (outpatients: 36/169, 21.3% and inpatients: 26/140, 18.6%). Average System Usability Scale scores indicated high usability in both groups (outpatients: mean 86.8, SD 15.8 and inpatients: mean 83.9, SD 18.8). Overall, repeated measurement compliance was 76.9% (173/225; outpatients only). Interviewed patients showed positive attitudes toward ePROMs. However, there are barriers to implementation related to the time and cognitive effort required to complete the questionnaires. There is also skepticism about the usefulness of ePROMs in interactions with health care professionals.

Conclusions: This study provides useful information for future ePROM implementation strategies, aimed at effectively supporting the routine clinical management and care of patients with cancer. In addition, these findings may be relevant to other organizations willing to systematically collect PROMs or ePROMs in their clinical routines.

Trial Registration: ClinicalTrials.gov NCT03968718; https://clinicaltrials.gov/study/NCT03968718

(JMIR Cancer 2025;11:e56625) doi:10.2196/56625

KEYWORDS

feasibility; oncology; patient-reported outcomes; PROMs; quality of life; mixed methods study; cancer; electronic patient-reported outcomes; patient compliance; barrier; implementation; usability scale; semistructured interview; questionnaire; clinical management; eHealth

Introduction

Patient-reported outcome measures (PROMs) are questionnaires for the self-assessment of patients' symptoms, well-being, and functional status associated with their health condition, without interpretation by a clinician or anyone else [1,2]. Well-validated PROMs are considered the gold standard for the collection of subjective health-related outcomes [3,4].

PROMs were initially developed to be used in research, but interest has been growing in integrating them into cancer clinical practice to facilitate personalized care management [5]. There is now a wealth of evidence indicating that PROMs may improve symptom control, communication, patient satisfaction, quality of life, and overall survival. In addition, consistent use of PROMs may contribute to reduce emergency room access and hospitalization rates [6-11].

Despite the generally positive effects and favorable attitudes reported by health care professionals (HCPs) [12-15], PROMs are not systematically implemented in routine oncology practice [16,17]. Potential facilitators to the routine use of PROMs have been highlighted, such as automatic scoring, immediate availability of above cut-off values, and time-trend visualization, along with automatic triggers and recommendations for clinical action. However, the barriers to PROM implementation act at many levels [18-20]. At the HCP and service level, major barriers include the workload associated with administering questionnaires, the lack of clear guidelines and confidence in routine use, difficulties in scoring and interpreting the results, and integration of PROMs into clinical workflows [17,19].

Electronic PROMs (ePROMs) have been proposed [21] to improve their applicability and acceptability by HCPs. Yet, inadequate IT infrastructures and ePROM systems that are not integrated with the electronic medical record (EMR) [22,23] emerged as critical issues. Relevant among these is the need for HCPs to connect to multiple systems, the increased risk of poor care coordination, inefficiencies in activating clinical pathways and referrals, and missed opportunities for care improvement [18,24].

Patient-related barriers also hinder the use of PROMs and include the negative perception of PROMs as time-consuming and burdensome to complete, difficulties in using electronic devices, lack of adequate explanations and support, and privacy concerns [19,22].

The Patient Voices project started in 2018 at the Fondazione IRCCS Istituto Nazionale dei Tumori (INT) in Milan (Italy), with the purpose of promoting the systematic collection of ePROMs in routine cancer care through a software system integrated into the hospital EMR that has not yet been implemented [25]. The system was designed in compliance with recommendations provided by the European Society for Medical Oncology clinical practice guideline on the role of PROMs in the continuum of cancer care [26]. The aim of the project was 2-fold: on the one hand, to explore the technical viability and the attitude of HCPs toward such an integrated ePROM system; on the other hand, to test its workability in a pilot implementation in routine oncology clinical practice in different settings of the hospital. This study aims at reporting on the assessment of compliance and feasibility of ePROM implementation as well as identifying patient-related implementation barriers.

Methods

Study Design

This study used a multimethod design based on concurrent quantitative and qualitative data collection. The aim was to increase the chances of getting varied and extensive research findings on the feasibility of the systematic use of ePROMs.

Data were collected and analyzed from the following substudies:

- Quantitative longitudinal substudy A: ePROMs were used for symptom screening and monitoring in patients with cancer who attended 3 outpatient clinics at INT, that is, palliative care, genitourinary oncology, and radiotherapy clinics.
- Quantitative cross-sectional substudy B: ePROMs were used to assess psychological distress among inpatients

admitted to urological surgery, medical oncology, and colorectal surgery wards.
- Qualitative feasibility substudy C: Semistructured interviews were carried out with a subgroup of patients involved in the previous quantitative substudies, with the aim to explore in depth the patient-related barriers to successful implementation.

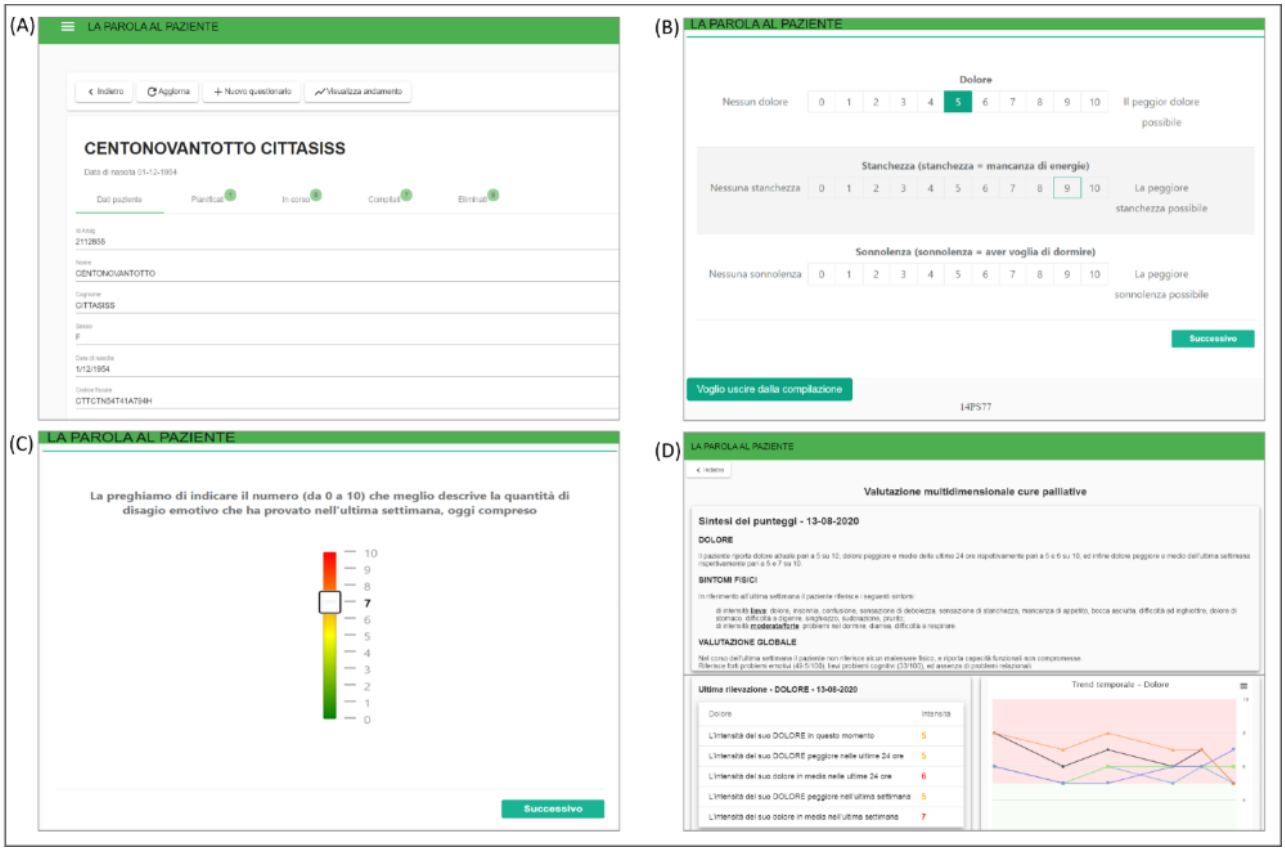
More methodological details are reported elsewhere [25].

Patient Voices ePROM System

The Patient Voices ePROM system was developed in 2020 at INT in collaboration with Politecnico of Milan (Italy). After a predevelopment phase, running an ad hoc analysis of technical and user requirements, the web-based application was integrated with the INT information system, including the EMR, and runs only under the hospital network. The system involves 4 kinds of users [25]: the administrator, who operates through a dedicated dashboard to authorize users to system access and to download PROM data for reports and research; the data collection coordinator (DCC), who registers patients and trains them to use the system; the patient, who completes the assigned questionnaires through a tablet provided by the hospital; and the clinician, who has access to real-time textual and graphical PROM data from the EMR.

Figure 1 shows sample screenshots from the Patient Voices ePROM system for the different users (fictional data).

Figure 1. Screenshots from the Patient Voices electronic patient-reported outcome measure system for the different users. (A) Patients' data registration by the data collection coordinator. (B) Edmonton Symptom Assessment Scale questionnaire compilation by the patient. (C) Distress Thermometer questionnaire compilation by the patient. (D) Visualization of Therapy Impact Questionnaire scores by the clinician.



Questionnaires (ePROMs)

The Patient Voices system was designed to allow for flexibility in the choice of questionnaires to be administered to patients. For the feasibility substudies A and B, the following questionnaires could be electronically administered:

- Edmonton Symptom Assessment Scale (ESAS) [27]: It requires the patient to rate a list of physical and psychological symptoms on a 0-10 numerical scale and was used in genitourinary cancer and radiotherapy clinics in substudy A.
- Therapy Impact Questionnaire (TIQ) [28]: It includes questions on physical symptoms (24 items), overall well-being (1 item), functional and emotional status (3 and 4 items, respectively), as well as cognitive and relational status (2 items each). The reference time is the previous week, and responses are collected on verbal rating scales (no or a little bit or quite a bit or very much). For this study, TIQ was complemented with five 0-10 numerical rating scales assessing pain intensity with different time referral. As this implementation project aimed to meet the needs of clinicians, the TIQ and the pain scales were chosen in place of the ESAS in the palliative care outpatient clinic, as these were the paper-and-pencil PROMs routinely used in that ward.
- Distress Thermometer (DT) [29]: It is a single 0-10 numerical scale on which participants rate their level of distress from any cause in the previous week. The DT is supplemented by a 35-item problem list, which prompts patients to identify their problems in practical,

family-related, emotional, spiritual or religious, and physical domains. Scores of 4 or higher suggest clinically significant distress. The DT was used with inpatients involved in substudy B.

All the above ePROMs are validated in the Italian language [28,30-32].

Quantitative Substudies

Study Participants

Consecutive adult patients with cancer (aged ≥18 years) attending one of the previously listed outpatient clinics and inpatient wards during pre-established days in the enrollment period (1 to 2 months for each ward or clinic) were potentially eligible to be enrolled in substudies A and B. Patients with inability to complete the questionnaires due to physical or cognitive impairment, psychological disturbances, or nonnative language issues were not eligible for the study. All participants attended as volunteers and gave their written informed consent to participate in the research.

Data Collection Procedures

A systematic screening of all patients attending inpatient wards and outpatient clinics was performed by a dedicated research nurse, who acted as the DCC. Reasons for not administering ePROMs were collected and classified as patient related (cognitive impairment, physical conditions, and language issues), institution related (organizational problems and patient already enrolled in another clinical trial), or patient refusal. After the eligibility screening and informed consent collection, the DCC provided basic training on how to use the tablet and explained how to fill in the questionnaire. Patients involved in substudy A filled in the ePROM at all subsequent visits during the data collection period, while patients involved in substudy B filled in the DT on admission only.

The DCC supported the patients while filling out the questionnaire, recording whether and how often the patient needed help in using the device or interpreting the questions. A specific DCC "structured form" was used for collecting such data, which were then analyzed using descriptive statistics. The time needed to complete the questionnaire by each patient was registered by the system.

After ePROM completion, patients were asked to fill in a paper-based questionnaire on their educational level and familiarity with electronic devices and internet use. Patients' demographic and clinical data (sex, date of birth, tumor site, year of cancer diagnosis, and visit or hospitalization reason) were extracted from the institutional data warehouse. Data collection was performed using REDCap (Research Electronic Data Capture; Vanderbilt University) electronic case report form [33,34]. Patients' perception and usability of the Patient Voices system were assessed, only for patients filling in the questionnaire, by the System Usability Scale (SUS), a standardized 10-item questionnaire with a final score ranging from 0 to 100 with higher values indicating higher usability [35].

Study End Points

The main end point of the study was compliance with the Patient Voices system, defined as the percentage of eligible patients completing the ePROMs assigned. Secondary end points were the percentage of screened patients who received the ePROMs; the average perceived system usability measured through SUS; the proportion of patients asking for interaction with the DCC to complete the tasks; the average time to fill in the questionnaire; and the successful administration of the questionnaire during subsequent visits (among patients attending outpatient clinics, at least twice during the follow-up period in substudy A).

Sample Size and Data Analysis

In the hypothesis that the compliance is 50% (hypothesis of maximum variability and then maximum imprecision), a sample size of 200 (in both in- and outpatient settings) allows the estimation of a 95% CI for the percentage of compliance with a precision (half-width) of 6.9% [36]. Basic descriptive statistics were applied to characterize the study sample. Point and interval estimates (95% CI) of proportion and averages described in the study end points were calculated for the whole sample and by inpatient ward or outpatient clinic. Cronbach α is used to measure the internal consistency of the SUS in this sample. Values above 0.70 indicate acceptable reliability [37]. The analyses were performed using the standard software packages Stata (StataCorp) and R (R Foundation for Statistical Computing).

Qualitative Substudy

Study Participants

Patients in the qualitative interview substudy were recruited among those who had completed at least 1 of the ePROMs administered in substudies A or B. Purposive sampling was undertaken to guarantee the inclusion of participants with a broad range of characteristics, such as disease site and stage, age, and sex.

Data Collection Methods and Procedures

A topic guide developed by the research team and based on a literature review [19,24,38,39] was used to structure the interviews with patients. This guide included questions on the feasibility and acceptability of the routine use of ePROMs, the difficulties encountered while completing the questionnaire, the information and help received about ePROMs, and the perception of the impact of these tools on clinical consultation. In addition, there was scope to digress from the guide if participants raised new and relevant topics.

After an initial training session, face-to-face interviews were conducted, audio recorded, and transcribed ad verbatim by volunteers from *Associazione Italiana Malati di Cancro* (AIMAC; Italian Association of Cancer Patients). The interviews were conducted between May and November 2021 and lasted an average of 28 (SD 11.02) minutes. The interviews (and participant involvement) were carried on until thematic saturation had been reached and no new insights could be drawn from additional participants [40].

Data Analysis

Interview transcripts were imported into NVivo (version 12; Lumivero) for data management, coding, and analysis [41]. Patients' names were transformed into textual and numeric strings to ensure data pseudoanonymization. Verbatim transcripts were analyzed using a thematic analysis approach, as described by Braun and Clarke [40]. Initial coding was done by 1 author (L Lombi), then supervised and verified by other 2 authors (C Brunelli and SA). Once the interviews had been initially coded, a finer analysis was conducted to identify themes, sample quotes, and interconnections between themes.

Some results of this qualitative study, particularly regarding participants' perspectives on the implications of using ePROMs for the clinical encounter, were discussed in Lombi et al [42]. In this work, we focused on findings concerning the perceived barriers to ePROM integration into oncological clinical routine.

Ethical Considerations

This study was approved by the institutional ethics committee of Fondazione IRCCS INT (Milan, Italy; INT 167/18) and complied with the Declaration of Helsinki. Participants provided written consent before the visit or during hospitalization. Participants were informed about (1) research purposes, (2) privacy, (3) use of the information obtained, (4) that participation was voluntary and unrewarded, and (5) that they could leave the research at any time without giving explanations. All the data were deidentified before data analysis and storage. There was no compensation for participation in the study.

Results

Quantitative Results

From June 2020 to September 2021, a total of 435 consecutive patients were screened, 421 (96.7%) were eligible, and 309 filled in the ePROMs, which indicates compliance of 73.4% (309/421; 95% CI 69.8%-77.5%). All patients who were administered the ePROM completed it, and no one stopped before completing the questionnaire.

Overall feasibility (percentage of patients, among those screened, who could be administered the ePROMs) was 71% (309/435; 95% CI 66.5%-75.2%). Table 1 shows comparable feasibilities by in- and outpatients; 69.6% (140/201; 95% CI 62.8%-75.2%) and 72.2% (169/234; 95% CI 66%-77.8%), respectively. Organization problems, such as pressure on time and difficult patient flow management, were the main reason for not administering ePROMs among outpatients (40/234, 17.1% vs inpatients: 5/201, 2.5%), whereas patient refusal was the main reason for inpatients (44/201, 21.9% vs outpatients: 10/234, 4.3%).

While the feasibility rates were similar in the 3 inpatient wards (colorectal surgery: 44/60, 73.3%; medical oncology: 48/70, 68.6%; and urological surgery: 48/71, 67.6%), higher heterogeneity was found in outpatient clinics, with the palliative care clinic showing the higher feasibility (59/65, 90.8%), followed by the radiotherapy clinic (42/54, 77.8%), and the genitourinary oncology clinic (68/115, 59.1%; Figure 2).

Table 2 reports demographic, disease characteristics, and the use of technological tools by patients who filled in the ePROMs.

Most of the responders stated that they regularly used smartphones (240/309, 77.7%) and the internet (203/309, 66%). About half of them (159/309, 51.5%) were quite familiar with computer, while tablet use was less common (65/309, 21%). In total, 24 (7.8%) patients reported they were not using any electronic devices or the internet.

Table 3 shows the interaction and help needed by patients during ePROM filling in. More help and interaction with the DCC to use the tablet was needed by outpatients (47/169, 27.8% vs 15/140, 10.7%), and overall, 244 (78.9%) out of 309 patients did not need help to use the tablet. Help received for the interpretation of questions was similar in the 2 settings (outpatients: 36/169, 21.3% and inpatients: 26/140, 18.6%). Of note, 216 (70%) out of 309 patients did not need any kind of interaction to fill in the ePROM (data not shown in table).

In a limited number of patients, although considered eligible for the study in the enrollment phase, the ePROM was filled in by the nurse upon patient interview (10/140, 7.1% vs 13/169, 7.7% in the inpatient and outpatient group, respectively) mainly because of physical issues (2/140, 1.4% of inpatients and 6/169, 3.6% of outpatients) and pain or asthenia (3/140, 2.1% of inpatients and 3/169, 1.8% of outpatients), which prevented them from being completely independent when filling in the questionnaire.

System usability, as measured by the SUS, was above 80 for both inpatient and outpatient groups (mean 86.8, SD 15.8 and mean 83.9, SD 8.9, respectively). Cronbach α was 0.84 (95% CI 0.77-0.88) in the overall sample. On average, the TIQ and DT required more time to be filled in (mean 6.63, SD 5.52 minutes and mean 6.25, SD 4.17 minutes, respectively) compared to the ESAS (mean 3.70, SD 2.79 minutes), as expected due to the different length of the questionnaires.

Figure 3 shows the repeated administration of ePROMs among outpatients who underwent at least 1 follow-up visit (n=225). The maximum number of completed questionnaires was 6 (including baseline ePROMs) of a total of maximum 7 visits recorded for this study, with 76.9% (n=173) of the overall follow-up ePROMs filled out.

Table 1. Feasibility and reasons for not administering ePROMs[a] by setting (N=435).

	Setting	
	Inpatients (n=201), n (%)	Outpatients (n=234), n (%)
ePROM administered		
Yes	140 (69.6)	169 (72.2)
No	61 (30.4)	65 (27.8)
Reasons for not administering ePROMs		
Patient related		
Impaired cognitive status	3 (1.5)	1 (0.4)
Impaired physical condition	0 (0)	1 (0.4)
Nonnative language issues	9 (4.5)	1 (0.4)
Institution related		
Organization problems	5 (2.5)[b]	40 (17.1)[c]
Patient enrolled in a clinical trial	0 (0)	12 (5.1)
Patient refusal	44 (21.9)	10 (4.3)

[a]ePROM: electronic patient-reported outcome measure.

[b]Unavailability of the patients at the bed because of medical procedures.

[c]Unavailability of the patients for delays or anticipation in the timing of the visit or for diagnostic procedures.

Figure 2. Feasibility rates and reasons for not administering the electronic patient-reported outcome measure by ward and outpatient clinic.

Table 2. Baseline sociodemographic and clinical characteristics of patients who filled in the electronic patient-reported outcome measures.

	Setting	
	Inpatients (n=140)	Outpatients (n=169)
Sex, n (%)		
Male	97 (69.3)	100 (59.2)
Female	43 (30.7)	69 (40.8)
Age (years), mean (SD)	60.4 (14.3)	65.8 (12.8)
Educational status, n (%)		
Primary	7 (5)	9 (5.3)
Lower secondary	20 (14.3)	16 (9.5)
Upper secondary	76 (54.3)	42 (24.9)
Postsecondary	34 (24.3)	34 (20.1)
Other	2 (1.4)	3 (1.8)
Missing[a]	1 (0.7)	65 (38.5)
Primary tumor site, n (%)		
Breast	4 (2.9)	34 (20.1)
Lung	13 (9.3)	10 (5.9)
Gastrointestinal	58 (41.4)	11 (6.5)
Urogenital	50 (35.7)	84 (49.7)
Other	14 (10)	25 (14.8)
Missing	1 (0.7)	5 (3)
Reason of the visit or admission, n (%)		
Palliative care	0 (0)	59 (34.9)
Radiotherapy treatment	0 (0)	42 (24.9)
Medical oncologic visit	0 (0)	68 (40.2)
Medical treatment	46 (32.9)	0 (0)
Surgery	86 (61.4)	0 (0)
Other	8 (0.6)	0 (0)
Frequency of use of electronic tools		
Smartphone, n (%)		
Not owing or never using	13 (9.3)	25 (14.8)
Seldom use	18 (12.9)	13 (7.7)
Regular use	109 (77.9)	131 (77.5)
Computer, n (%)		
Not owing or never using	36 (25.7)	55 (32.5)
Seldom use	30 (21.4)	29 (17.2)
Regular use	74 (52.9)	85 (50.3)
Tablet, n (%)		
Not owing or never using	79 (56.4)	112 (66.3)
Seldom use	31 (22.1)	22 (13)
Regular use	30 (21.4)	35 (20.7)
Internet, n (%)		
Never or rarely	29 (20.7)	43 (25.7)
Sometimes	14 (10)	20 (12)

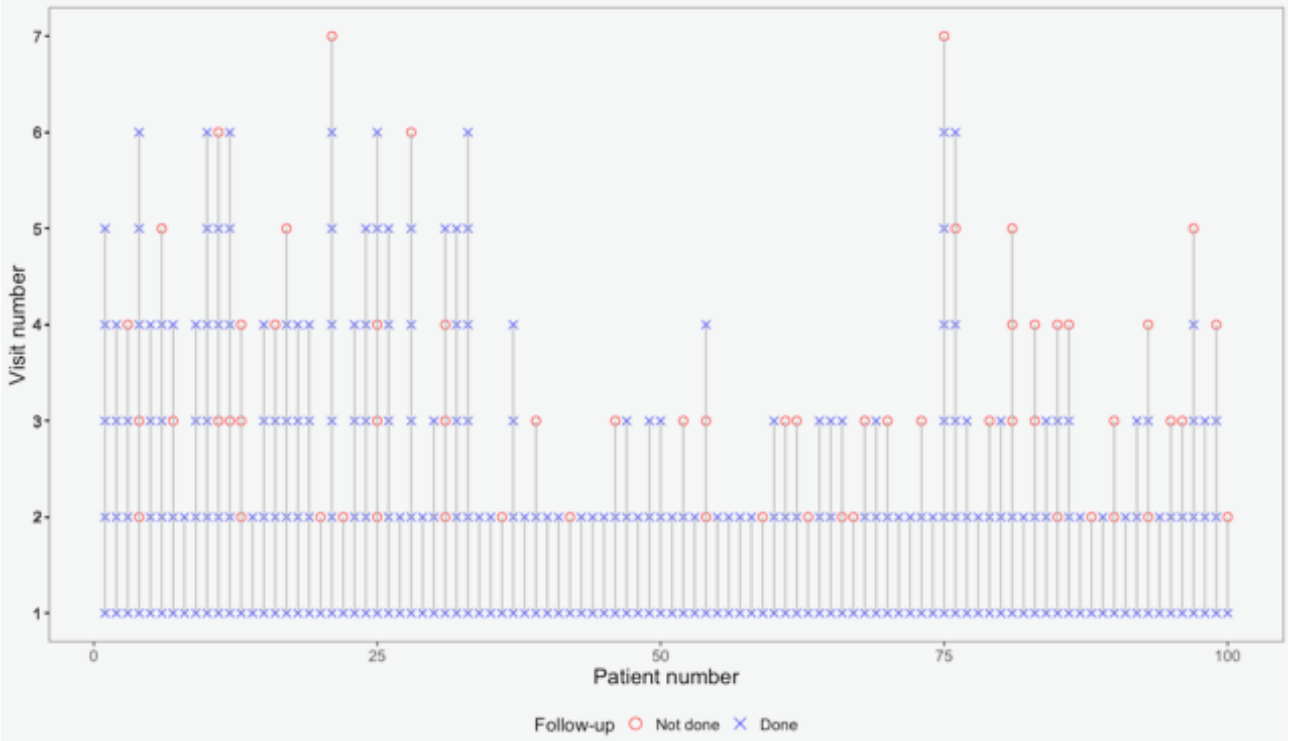
	Setting	
	Inpatients (n=140)	Outpatients (n=169)
Often or every day	97 (69.3)	104 (62.3)
Missing	0 (0)	2 (1.2)

[a]Due to a technical problem, educational status was not collected for patients in 1 outpatient clinic.

Table 3. Interaction and help needed during electronic patient-reported outcome measure filling in by setting.

	Setting	
	Inpatients (n=140)	Outpatients (n=169)
Interaction needed for the use of the tablet, n (%)		
No interaction needed	125 (89.3)	119 (70.4)
Some interaction needed	15 (10.7)	47 (27.8)
Missing	0 (0)	3 (1.8)
If yes, how many times, n (%)		
1-3	5 (3.6)	24 (14.2)
4-6	0 (0)	7 (4.1)
>6	0 (0)	3 (1.8)
Filled in by the nurse	10 (7.1)	13 (7.7)
Interaction needed to interpret questions, n (%)		
No interaction needed	114 (81.4)	129 (76.3)
Some interaction needed	26 (18.6)	36 (21.3)
Missing	0 (0)	4 (2.4)
If yes, how many times, n (%)		
1-3	16 (11.4)	21 (12.4)
4-6	0 (0)	1 (0.6)
>6	0 (0)	1 (0.6)
Filled in by the nurse	10 (7.1)	13 (7.7)
Reason for nurse's compilation, n (%)		
Difficulty with reading	3 (2.1)	1 (0.6)
Difficulty with touchscreen	1 (0.7)	1 (0.6)
Physical issues	2 (1.4)	6 (3.6)
Pain or asthenia	3 (2.1)	3 (1.8)
Missing	1 (0.6)	2 (1.2)
System Usability Scale score (range 0-100)		
Mean (SD)	86.8 (15.8)	83.9 (18.9)
Missing, n (%)	0 (0)	15 (8.9)

Figure 3. Electronic patient-reported outcome measure (ePROM) administration at follow-up for patients undergoing at least 1 follow-up visit. Cross symbols: ePROM completed; circle symbols: ePROM not administered.



Qualitative Results

Sociodemographic Characteristics of Participants

A total of 19 one-to-one interviews were conducted. Participants' age ranged from 35 to 78 (mean 57, SD 12.63) years, 10 (53%) were female, and 11 (58%) had completed secondary education. In total, 9 (47%) participants completed the DT, 8 (42%) completed the ESAS, and 4 (21%) completed the TIQ and pain scales. Some participants completed repeated ePROMs.

Attitude Toward ePROMs and Barriers to Successful Implementation

Participants were generally satisfied with ePROM administration and had positive attitudes toward their use in routine oncology practice. Many of their comments focused on the benefits of ePROMs, including improvement of quality and personalization of care through a holistic approach, increased chances of talking about their symptoms over time, increased awareness of both clinicians and patients about symptoms in real time, aid in rapidly detecting abnormal parameters, and perception to be engaged in scientific research.

Five key dimensions emerged from thematic analysis, relating to potential barriers to successful implementation of ePROMs in routine oncological care. The dimensions are reported in Table 4 alongside significant quotes from patients.

Table 4. Summary of dimensions identified from patient interviews (qualitative substudy C).

Barrier	Illustrative quotes
1. Skepticism toward ePROM[a] use during consultation with health care professional	• "...The way I see it, even the doctor or whoever is visiting you, that is, if he has to study your medical record, read your questionnaire, in one day he has so many patients, that is, it becomes a bit complicated to receive..." [Interview 18, male, 43 years old]. • "It doesn't help anybody. Because often it can happen, so we make a lot of papers, we put a lot of news...On the tablet, on the patient's health record...But if they are not read...They remain a dead letter, as they say, nobody needs them" [Interview 2, female, 74 years old].
2. Skepticism toward ePROMs as tools to examine symptoms	• "(The questionnaire is) very general...We should go into the specifics, everyone has different symptoms according to their history, their life, their way of living and everything...I mean...To their way of living" [Interview 10, female, 72 years old]. • "Definitely a self-limited questionnaire, in the sense, with specific questions, with closed answers, therefore it can be filled in...details should be provided, if necessary. If one feels compelled to write something else, related to each of these issues, the possibility of further compilation could be considered" [Interview 13, male, 38 years old].
3. Cognitive difficulties	• "In fact, the first time...I was wrong, because I went to the right, because...I gave 'No pain = 10', meaning I don't have any. And instead, she told me 'No, look, 10 is the worst'. Well, I hadn't looked closely, I had looked here...I hadn't understood, do you see?" [Interview 11, female, 61 years old]. • "I assumed that at first glance I would have said, probably because of the professional deformation I have...that 0 was the minimum and 10 the maximum. Instead, in some situations it is the opposite...That is, I didn't stop to think that no pain was zero and maximum pain was 10. I said, 'no pain, OK no pain, 10!'. Because I was fine, you know?" [Interview 1, female, 65 years old].
4. Technological issues	• "I'm bad with technology, so I'll tell you, sometimes I mess up so much with the phone that half is enough and so others always have to step in, but whatever." [Interview 4, female, 78 years old]. • "Paper for me is always better...a book for me is a paper book...when I feel paper...I like it! Then at home we have tablets, we have computers, we have everything...but...it's always nice to be able to write...I like it..." [Interview 3, male, 66 years old].
5. ePROMs as time-consuming tasks	• "Yes, it's counterproductive...And then it makes people feel less inclined because the person is there for the visit, not to fill out the form...The form is given to him, he fills it out, but if it takes half an hour to fill out the form when the visit was scheduled at two o'clock and he goes in at half past two he's upset...Because he's wasting time and because it's not correct in short" [Interview 6, male, 59 years old].

[a]ePROM: electronic patient-reported outcome measure.

The five key dimensions are as follows:

1. Skepticism toward ePROM use during consultation with HCPs: During the interviews, several participants complained that their ePROM results were not discussed or even mentioned during the clinical encounter. As a result, many patients doubted that clinicians had consulted their responses likely because of time pressure and work overload.
2. Skepticism toward ePROMs as tools to examine symptoms: Some patients criticized the use of a questionnaire based on close and structured questions to explore their health status and well-being, as these are dimensions that should be investigated through more flexible tools, that is, open questions also including descriptive comments.
3. Cognitive difficulties: Several participants, mostly among those who were asked to complete the TIQ and pain scale, mentioned problems to fill out the questionnaire due to the perceived complexity of the scales.
4. Technological issues: Patients, although older patients, generally stated that they had no difficulties using the tablets to fill out the ePROMs, with the only exception of 1 person who acknowledged limited digital skills and required direct support from the DCC. Two patients said they had no difficulties with the tablet, but they would have preferred to fill out the ePROMs in paper format.
5. ePROMs as time-consuming tasks: Mentioned by only 1 participant, the time required to complete the ePROMs was perceived as potentially lengthening the waiting time and delaying the visit.

Discussion

Principal Findings and Comparison With Prior Work

This study explored the compliance and feasibility of implementing an ePROM system integrated with the EMR in the clinical management of patients with cancer attending hospital wards or outpatient clinics. Potential patient-related barriers to the routine use of ePROMs were analyzed, collecting both reasons for not filling in and patients' views on the difficulties encountered while filling in the questionnaires.

The results showed good compliance and feasibility both by inpatients and outpatients. Most of them (244/309, 78.9%) were able to fill out the questionnaire without any help in using the device, which suggests that the electronic format was not a major barrier. In support of this, many participants stated that they regularly used smartphones, computers, and the internet, indicating a certain degree of familiarity with technology tools. Besides, system usability scores confirmed that the experience with the ePROM system was more than satisfactory by most of the patients in both clinical settings, and the qualitative results

also point in this direction. Indeed, in planning this project, we chose to use a system for ePROM collection that would facilitate the procedure from the patient's standpoint. Patients neither had to download any app or software nor register or log in via password; instead, they were provided with a tablet ready to fill out the questionnaire. Clearly, more proactivity by the patient would be needed in case of remote ePROM assessment on the patient's own personal devices. In any case, as already highlighted by several studies [19,22,43,44], basic IT literacy remains a prerequisite for electronic assessment, and addressing this issue, especially among older patients, is a priority.

The feasibility for outpatients at follow-up visits was also good: three-quarters of the total number of ePROMs administered after the baseline visit were completed, and more than half of the patients filled in the questionnaire at each consecutive visit, indicating an overall positive attitude toward regular use of health technologies. When interviewed, patients reported several benefits from ePROMs, including actively participating in their health care and improving patient-clinician communication.

Approximately 3 in 4 patients successfully filled in the questionnaire regardless of clinical setting. This is a very positive result considering that the administration of ePROMs took place in the context of a research study, which implies that data can be collected only after going through preliminary procedures concerning patient information, privacy, and consent to participation. If, as recommended, ePROMs were part of routine clinical practice, they would be administered to patients by default just like any other medical test or diagnostic noninvasive procedure.

Feasibility was similar in the 2 clinical settings, but the reasons for not filling in the ePROMs were different. In all 3 hospital wards, the main reason was patient refusal, which was 21.9% (44/201). Unfortunately, we could not collect any further information on refusal because patients were not specifically asked and did not spontaneously explain why they preferred not to participate in the study. Another important reason for not filling in was the language barrier for people not fluent in Italian (9/201, 5%), which, however, could be easily overcome by implementing different languages within the same ePROM to offer questionnaires also to nonnative speaking patients. In contrast to inpatient wards, outpatient clinics showed mainly institution-related reasons for not filling in, with organizational problems significantly affecting the feasibility rate. It is noteworthy that the palliative care outpatient clinic showed the highest feasibility (59/65, 90.8%), probably because in this clinical setting, both clinicians and patients are particularly well trained and accustomed to the use of paper-based PROMs that have been part of routine patient care for decades.

Similar studies exploring the compliance and feasibility of an ePROM system integrated into an existing clinical setting have reported good acceptance of both patients and HCPs [43,45-48], suggesting there is potential to use such instruments to improve the quality of information collected from patients by hospital systems. Yet, some challenges need to be addressed, including the patient-related barriers, which are also reported here (ie, time and cognitive burden upon patients to complete the questionnaires and skepticism toward ePROM use during a consultation with HCP) and are highly dependent on contextual factors. For example, a limitation of this study is that the 2 groups of patients were not administered the same ePROMs. This did not allow to distinguish the effect of the clinical setting (inpatient vs outpatient) from the effect of the type of questionnaires offered to inpatients (DT) and outpatients (ESAS and TIQ) on their different refusal rates. It can be assumed that patients awaiting surgery are less willing to take part in a research project than outpatients. On the other hand, filling in a questionnaire on psychological distress may be perceived as more burdensome than reporting about physical symptoms [49]. However, a combined effect of both these aspects on the higher inpatient refusal rate cannot be excluded.

Patients' perception that ePROMs are not valid tools to explore their own symptoms and well-being and share this information with clinicians has been reported by several studies [3,13,19,22,50]. Therefore, efforts should be made to ensure that patients receive adequate information about the questionnaire they are asked to complete and understand the value of PROM collection.

The time needed to fill out the PROMs and difficulties met in completing some items are also relevant barriers already reported in the literature [19,22,23,46]. Patients with cancer may perceive ePROMs as burdensome, as they see them as additional tests that negatively impact on the time already spent in visits, procedures, and treatments. For some patients, the questionnaire can also be difficult to complete without help, as our results showed, and this can be frustrating. Thus, selecting ePROMs that are both sufficiently informative and not overly burdensome for the patient is a critical aspect for their successful implementation.

Limitations and Future Work

This study has some limitations that should be considered. First, we acknowledge that this study was carried out in a single comprehensive cancer center, and this aspect may limit the generalizability of the results. However, existing evidence recommends that individual clinical settings independently examine local barriers in order to adopt ad hoc solutions [19,22,48,50,51].

Second, our results were obtained in the context of a research study. This implies that, on one side, compliance might be underestimated (patient refusal might be about the clinical study and not about filling in the ePROMs); on another side, feasibility might be overestimated, as data collection with a dedicated nurse might be difficult to replicate in real-world implementations due to lack of resources.

For ePROMs to supplement clinician-reported outcomes with useful information and help in patient's care, they should be efficient, effective, and satisfactory for stakeholders [21,52,53]. The European Society for Medical Oncology guideline recommends that PROMs evaluate outcomes that are clinically meaningful and actionable in the reference population and emphasizes that a single software system with PROM functionalities suitable for any stage of cancer disease would be the optimal solution, although technically challenging. However, evidence supporting the implementation of PROM

systems along the entire cancer trajectory is limited and largely based on studies under highly controlled conditions rather than on real-world data from routine clinical settings [26]. Based on these recommendations, this study contributes real-world evidence to support the integration of ePROMs into the hospital information system for their use in routine oncology practice.

Conclusions

At this feasibility stage, the patient-related barriers reported in this study provide useful information for improving future implementation strategies, which will be aimed to effectively support the routine clinical management and care of patients with cancer. In addition, these findings may be relevant to other organizations willing to implement a systematic use of PROMs or ePROMs in their clinical routines. Finally, current evidence suggests that the creation of a cultural infrastructure that values PROMs, that encourages and instructs HCPs to use these tools routinely, and that actively involves patients in their health care process is a key element in fostering the uptake of PROMs in real-world clinical settings [3,48,51].

Acknowledgments

The authors would like to thank Dr Michela Monfredini for her contribution to data collection. The Patient Voices study group includes GA, Marco Bellazzi, Filiberto Belli, C Borreani, C Brunelli, Giuseppi Capri, AC, Paolo Casali, Paolo Corradini, Filippo de Braud, Anna Maria Foschi, Secondo Folli, Marina Garassino, Lisa Licitra, NN, Chiara Pellegrini, MP, GP, AR, Roberto Salvioni, Pierangelo Spada, Riccardo Valdagni, and E Zito and Fondazione IRCCS Istituto Nazionale dei Tumori-Milano. This work was supported by the Italian Ministry of Health (5X1000 funds ID code D171CB).

Data Availability

The datasets generated during this study are available from the corresponding author on reasonable request.

Authors' Contributions

C Brunelli, E Zito, L Lombi, SA, AC, C Borreani, AR, and GA contributed to the study conception and methodological design. C Brunelli, GA, CA, and C Borreani contributed to funding acquisition. C Brunelli, E Zito, and SA contributed to project administration. C Brunelli, E Zito, M Spelta, and DL contributed to software design, development, and validation. C Brunelli, GA, E Zito, SA, MP, GP, NN, LB, LA, L Lozza, M Spelta, GM, and M Shkodra contributed to data acquisition. C Brunelli, SA, RM, GT, and L Lombi contributed to formal data analysis and visualization. All authors contributed to results interpretation. The first draft of the manuscript was written by C Brunelli, SA, and LC, and all authors commented on and contributed to previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

C Brunelli reports grants from the Italian Ministry of Health (5X1000 funds ID code D171CB) and payments from the University of Milan. AC reports consulting fees and honoraria from Mundipharma and Amgen outside the submitted work. All other authors declare that they have no competing interests. RM has received consulting fees from Boeringer outside the submitted work. GP has received grants from Ipsen, Janssen, MSD, and Gilead and consulting fees and honoraria from Amgen, Astellas, AstraZeneca, Bayer, BMS, Eisai, Ipsen, Janssen, Merck, Novartis, Pfizer, and Roche outside the submitted work. All remaining authors declare no conflicts of interest.

References

1. U.S. Department of HealthHuman Services FDA Center for Drug Evaluation and Research, U.S. Department of HealthHuman Services FDA Center for Biologics Evaluation and Research, U.S. Department of HealthHuman Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79 [FREE Full text] [doi: 10.1186/1477-7525-4-79] [Medline: 17034633]
2. Black N, Burke L, Forrest CB, Sieberer UHR, Ahmed S, Valderas JM, et al. Patient-reported outcomes: pathways to better health, better services, and better societies. Qual Life Res 2016;25(5):1103-1112. [doi: 10.1007/s11136-015-1168-3] [Medline: 26563251]
3. Antunes B, Harding R, Higginson IJ, EUROIMPACT. Implementing patient-reported outcome measures in palliative care clinical practice: a systematic review of facilitators and barriers. Palliat Med 2014;28(2):158-175 [FREE Full text] [doi: 10.1177/0269216313491619] [Medline: 23801463]
4. Churruca K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, et al. Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. Health Expect 2021;24(4):1015-1024 [FREE Full text] [doi: 10.1111/hex.13254] [Medline: 33949755]
5. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? Qual Life Res 2009;18(1):115-123. [doi: 10.1007/s11136-008-9430-6] [Medline: 19105048]

6. Barbera L, Sutradhar R, Seow H, Earle CC, Howell D, Mittmann N, et al. Impact of standardized Edmonton Symptom Assessment System use on emergency department visits and hospitalization: results of a population-based retrospective matched cohort analysis. JCO Oncol Pract 2020;16(9):e958-e965. [doi: 10.1200/JOP.19.00660] [Medline: 32463762]
7. Barbera L, Sutradhar R, Seow H, Mittmann N, Howell D, Earle CC, et al. The impact of routine Edmonton Symptom Assessment System (ESAS) use on overall survival in cancer patients: results of a population-based retrospective matched cohort analysis. Cancer Med 2020;9(19):7107-7115 [FREE Full text] [doi: 10.1002/cam4.3374] [Medline: 32794634]
8. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
9. Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of patient-reported outcomes in routine medical care. Am Soc Clin Oncol Educ Book 2018;38:122-134 [FREE Full text] [doi: 10.1200/EDBK_200383] [Medline: 30231381]
10. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. J Natl Cancer Inst 2017;109(9). [doi: 10.1093/jnci/djx029] [Medline: 28423407]
11. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014;32(14):1480-1501 [FREE Full text] [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]
12. Brunelli C, Zito E, Alfieri S, Borreani C, Roli A, Caraceni A, et al. Knowledge, use and attitudes of healthcare professionals towards patient-reported outcome measures (PROMs) at a comprehensive cancer center. BMC Cancer 2022;22(1):161 [FREE Full text] [doi: 10.1186/s12885-022-09269-x] [Medline: 35144569]
13. Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. BMJ 2019;364:k5267. [doi: 10.1136/bmj.k5267] [Medline: 30679170]
14. Amini M, Oemrawsingh A, Verweij LM, Lingsma HF, Hazelzet JA, Eijkenaar F, et al. Facilitators and barriers for implementing patient-reported outcome measures in clinical care: an academic center's initial experience. Health Policy 2021;125(9):1247-1255 [FREE Full text] [doi: 10.1016/j.healthpol.2021.07.001] [Medline: 34311981]
15. Gibbons C, Porter I, Gonçalves-Bradley DC, Stoilov S, Ricci-Cabello I, Tsangaris E, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Cochrane Database Syst Rev 2021;10(10):CD011589 [FREE Full text] [doi: 10.1002/14651858.CD011589.pub2] [Medline: 34637526]
16. Bainbridge D, Seow H, Sussman J, Pond G, Martelli-Reid L, Herbert C, et al. Multidisciplinary health care professionals' perceptions of the use and utility of a symptom assessment system for oncology patients. J Oncol Pract 2011;7(1):19-23 [FREE Full text] [doi: 10.1200/JOP.2010.000015] [Medline: 21532805]
17. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual Saf 2014;23(6):508-518 [FREE Full text] [doi: 10.1136/bmjqs-2013-002524] [Medline: 24505110]
18. Nguyen H, Butow P, Dhillon H, Morris L, Brown A, West K, et al. Using patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs) in routine head and neck cancer care: what do health professionals perceive as barriers and facilitators? J Med Imaging Radiat Oncol 2020;64(5):704-710. [doi: 10.1111/1754-9485.13048] [Medline: 32394587]
19. Nguyen H, Butow P, Dhillon H, Sundaresan P. A review of the barriers to using patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs) in routine cancer care. J Med Radiat Sci 2021;68(2):186-195 [FREE Full text] [doi: 10.1002/jmrs.421] [Medline: 32815314]
20. Schepers SA, Haverman L, Zadeh S, Grootenhuis MA, Wiener L. Healthcare professionals' preferences and perceived barriers for routine assessment of patient-reported outcomes in pediatric oncology practice: moving toward international processes of change. Pediatr Blood Cancer 2016;63(12):2181-2188 [FREE Full text] [doi: 10.1002/pbc.26135] [Medline: 27511830]
21. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
22. Glenwright BG, Simmich J, Cottrell M, O'Leary SP, Sullivan C, Pole JD, et al. Facilitators and barriers to implementing electronic patient-reported outcome and experience measures in a health care setting: a systematic review. J Patient Rep Outcomes 2023;7(1):13 [FREE Full text] [doi: 10.1186/s41687-023-00554-2] [Medline: 36786914]
23. Schepers SA, Sint Nicolaas SM, Haverman L, Wensing M, Schouten van Meeteren AYN, Veening MA, et al. Real-world implementation of electronic patient-reported outcomes in outpatient pediatric cancer care. Psychooncology 2017;26(7):951-959. [doi: 10.1002/pon.4242] [Medline: 27502744]
24. Basch E, Snyder C. Overcoming barriers to integrating patient-reported outcomes in clinical practice and electronic health records. Ann Oncol 2017;28(10):2332-2333 [FREE Full text] [doi: 10.1093/annonc/mdx506] [Medline: 28961852]
25. Brunelli C, Borreani C, Caraceni A, Roli A, Bellazzi M, Lombi L, et al. PATIENT VOICES, a project for the integration of the systematic assessment of patient reported outcomes and experiences within a comprehensive cancer center: a protocol

for a mixed method feasibility study. Health Qual Life Outcomes 2020;18(1):252 [FREE Full text] [doi: 10.1186/s12955-020-01501-1] [Medline: 32723341]
26. Di Maio M, Basch E, Denis F, Fallowfield LJ, Ganz PA, Howell D, et al. The role of patient-reported outcome measures in the continuum of cancer clinical care: ESMO Clinical Practice Guideline. Ann Oncol 2022;33(9):878-892 [FREE Full text] [doi: 10.1016/j.annonc.2022.04.007] [Medline: 35462007]
27. Bruera E, Kuehn N, Miller MJ, Selmser P, Macmillan K. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. J Palliat Care 1991;7(2):6-9. [Medline: 1714502]
28. Tamburini M, Rosso S, Gamba A, Mencaglia E, de Conno F, Ventafridda V. A therapy impact questionnaire for quality-of-life assessment in advanced cancer research. Ann Oncol 1992;3(7):565-570 [FREE Full text] [doi: 10.1093/oxfordjournals.annonc.a058263] [Medline: 1498079]
29. Roth AJ, Kornblith AB, Batel-Copel L, Peabody E, Scher HI, Holland JC. Rapid screening for psychologic distress in men with prostate carcinoma: a pilot study. Cancer 1998;82(10):1904-1908. [doi: 10.1002/(sici)1097-0142(19980515)82:10<1904::aid-cncr13>3.0.co;2-x] [Medline: 9587123]
30. Moro C, Brunelli C, Miccinesi G, Fallai M, Morino P, Piazza M, et al. Edmonton Symptom Assessment Scale: Italian validation in two palliative care settings. Support Care Cancer 2006;14(1):30-37. [doi: 10.1007/s00520-005-0834-3] [Medline: 15937688]
31. Grassi L, Sabato S, Rossi E, Marmai L, Biancosino B. Affective syndromes and their screening in cancer patients with early and stable disease: Italian ICD-10 data and performance of the distress thermometer from the Southern European Psycho-Oncology Study (SEPOS). J Affect Disord 2009;114(1-3):193-199. [doi: 10.1016/j.jad.2008.07.016] [Medline: 18757101]
32. Donovan KA, Grassi L, McGinty HL, Jacobsen PB. Validation of the distress thermometer worldwide: state of the science. Psychooncology 2014;23(3):241-250. [doi: 10.1002/pon.3430] [Medline: 25160838]
33. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
34. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
35. Brooke J. SUS: A 'quick and dirty' usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry. London, United Kingdom: Taylor and Francis; 1996.
36. Machin D, Campbell M, Fayers P, Pinol A. Sample Size Tables for Clinical Studies. New York, NY: John Wiley & Sons; 1997.
37. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol 1993;78(1):98-104. [doi: 10.1037/0021-9010.78.1.98]
38. Baeksted C, Pappot H, Nissen A, Hjollund NH, Mitchell SA, Basch E, et al. Feasibility and acceptability of electronic symptom surveillance with clinician feedback using the patient-reported outcomes version of Common Terminology Criteria for Adverse Events (PRO-CTCAE) in Danish prostate cancer patients. J Patient Rep Outcomes 2017;1(1):1 [FREE Full text] [doi: 10.1186/s41687-017-0005-6] [Medline: 29757324]
39. Girgis A, Durcinoska I, Levesque JV, Gerges M, Sandell T, Arnold A, et al. eHealth system for collecting and utilizing patient reported outcome measures for personalized treatment and care (PROMPT-care) among cancer patients: mixed methods approach to evaluate feasibility and acceptability. J Med Internet Res 2017;19(10):e330 [FREE Full text] [doi: 10.2196/jmir.8360] [Medline: 28970188]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Jackson K, Bazeley P. Qualitative Data Analysis With NVivo. 3rd Edition. Thousand Oaks, CA: Sage Publications; 2019.
42. Lombi L, Alfieri S, Brunelli C. 'Why should I fill out this questionnaire?' A qualitative study of cancer patients' perspectives on the integration of e-PROMs in routine clinical care. Eur J Oncol Nurs 2023;63:102283 [FREE Full text] [doi: 10.1016/j.ejon.2023.102283] [Medline: 36893578]
43. Crockett C, Gomes F, Faivre-Finn C, Howell S, Kasipandian V, Smith E, et al. The routine clinical implementation of electronic patient-reported outcome measures (ePROMs) at the Christie NHS Foundation Trust. Clin Oncol (R Coll Radiol) 2021;33(12):761-764. [doi: 10.1016/j.clon.2021.06.004] [Medline: 34229926]
44. Del Rosario García B, Morales Barrios JA, Viña Romero MM, Ramos Díaz R, Nazco Casariego GJ, Jiménez Sosa A, et al. Patient-reported outcomes and digital literacy of patients treated in an oncology day hospital unit. J Oncol Pharm Pract 2022;28(3):530-534. [doi: 10.1177/1078155221996041] [Medline: 33645328]
45. Neame MT, Reilly D, Puthiyaveetil A, McCann L, Mahmood K, Almeida B, et al. Successful integration of an automated patient-reported outcome measure within a hospital electronic patient record. Rheumatol Adv Pract 2022;6(3):rkac065 [FREE Full text] [doi: 10.1093/rap/rkac065] [Medline: 36071947]

46. Trautmann F, Hentschel L, Hornemann B, Rentsch A, Baumann M, Ehninger G, et al. Electronic real-time assessment of patient-reported outcomes in routine care—first findings and experiences from the implementation in a comprehensive cancer center. Support Care Cancer 2016;24(7):3047-3056. [doi: 10.1007/s00520-016-3127-0] [Medline: 26887586]
47. Rotenstein LS, Agarwal A, O'Neil K, Kelly A, Keaty M, Whitehouse C, et al. Implementing patient-reported outcome surveys as part of routine care: lessons from an academic radiation oncology department. J Am Med Inform Assoc 2017;24(5):964-968 [FREE Full text] [doi: 10.1093/jamia/ocx009] [Medline: 28339771]
48. Briggs MS, Rethman KK, Crookes J, Cheek F, Pottkotter K, McGrath S, et al. Implementing patient-reported outcome measures in outpatient rehabilitation settings: a systematic review of facilitators and barriers using the consolidated framework for implementation research. Arch Phys Med Rehabil 2020;101(10):1796-1812. [doi: 10.1016/j.apmr.2020.04.007] [Medline: 32416149]
49. Granek L, Nakash O, Ariad S, Shapira S, Ben-David M. Oncologists' identification of mental health distress in cancer patients: strategies and barriers. Eur J Cancer Care (Engl) 2018;27(3):e12835. [doi: 10.1111/ecc.12835] [Medline: 29508452]
50. Foster A, Croot L, Brazier J, Harris J, O'Cathain A. The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. J Patient Rep Outcomes 2018;2:46 [FREE Full text] [doi: 10.1186/s41687-018-0072-3] [Medline: 30363333]
51. Anatchkova M, Donelson SM, Skalicky AM, McHorney CA, Jagun D, Whiteley J. Exploring the implementation of patient-reported outcome measures in cancer care: need for more real-world evidence results in the peer reviewed literature. J Patient Rep Outcomes 2018;2(1):64 [FREE Full text] [doi: 10.1186/s41687-018-0091-0] [Medline: 30588562]
52. Velikova G, Wright EP, Smith AB, Cull A, Gould A, Forman D, et al. Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires. J Clin Oncol 1999;17(3):998-1007. [doi: 10.1200/JCO.1999.17.3.998] [Medline: 10071295]
53. ISO 9241-11. International Organization for Standardization. 2018. URL: https://www.iso.org/standard/63500.html [accessed 2023-07-24]

Abbreviations

DCC: data collection coordinator
DT: Distress Thermometer
EMR: electronic medical record
ePROM: electronic patient-reported outcome measure
ESAS: Edmonton Symptom Assessment Scale
HCP: health care professional
INT: Istituto Nazionale dei Tumori
PROM: patient-reported outcome measure
REDCap: Research Electronic Data Capture
SUS: System Usability Scale
TIQ: Therapy Impact Questionnaire

Edited by T de Azevedo Cardoso; submitted 09.02.24; peer-reviewed by J Brooke, E Deja, S Thestrup Hansen, M Eichler; comments to author 10.04.24; revised version received 05.07.24; accepted 23.07.24; published 22.01.25.

Please cite as:
Brunelli C, Alfieri S, Zito E, Spelta M, Arba L, Lombi L, Caselli L, Caraceni A, Borreani C, Roli A, Miceli R, Tine' G, Zecca E, Platania M, Procopio G, Nicolai N, Battaglia L, Lozza L, Shkodra M, Massa G, Loiacono D, Apolone G
Patient Voices: Multimethod Study on the Feasibility of Implementing Electronic Patient-Reported Outcome Measures in a Comprehensive Cancer Center
JMIR Cancer 2025;11:e56625
URL: https://cancer.jmir.org/2025/1/e56625
doi:10.2196/56625
PMID:

Original Paper

Barriers and Facilitators to the Preadoption of a Computer-Aided Diagnosis Tool for Cervical Cancer: Qualitative Study on Health Care Providers' Perspectives in Western Cameroon

Magali Jonnalagedda-Cattin[1,2], MSc; Alida Manoëla Moukam Datchoua[3,4], MSc; Virginie Flore Yakam[5], BSc; Bruno Kenfack[3,5], Prof Dr Med; Patrick Petignat[6], Prof Dr Med; Jean-Philippe Thiran[1], Prof Dr; Klaus Schönenberger[2], PhD; Nicole C Schmidt[6,7], Prof Dr Med, PD, MPH

[1]Signal Processing Laboratory LTS5, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland

[2]EssentialTech Centre, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland

[3]Department of Obstetrics and Gynaecology, Faculty of Medicine and Pharmaceutical Sciences, University of Dschang, Dschang, Cameroon

[4]Institute of Global Health, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[5]Department of Obstetrics and Gynecology, Annex Regional Hospital of Dschang, Dschang, Cameroon

[6]Department of Paediatrics, Gynaecology and Obstetrics, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[7]Faculty of Social Science, Catholic University of Applied Science (KSH), Münich, Germany

Corresponding Author:
Magali Jonnalagedda-Cattin, MSc
Signal Processing Laboratory LTS5
Swiss Federal Institute of Technology Lausanne (EPFL)
EPFL-STI-IEL-LTS5, Station 11
Lausanne, 1015
Switzerland
Phone: 41 21 693 97 77
Email: magali.cattin@epfl.ch

Abstract

Background: Computer-aided detection and diagnosis (CAD) systems can enhance the objectivity of visual inspection with acetic acid (VIA), which is widely used in low- and middle-income countries (LMICs) for cervical cancer detection. VIA's reliance on subjective health care provider (HCP) interpretation introduces variability in diagnostic accuracy. CAD tools can address some limitations; nonetheless, understanding the contextual factors affecting CAD integration is essential for effective adoption and sustained use, particularly in resource-constrained settings.

Objective: This study investigated the barriers and facilitators perceived by HCPs in Western Cameroon regarding sustained CAD tool use for cervical cancer detection using VIA. The aim was to guide smooth technology adoption in similar settings by identifying specific barriers and facilitators and optimizing CAD's potential benefits while minimizing obstacles.

Methods: The perspectives of HCPs on adopting CAD for VIA were explored using a qualitative methodology. The study participants included 8 HCPs (6 midwives and 2 gynecologists) working in the Dschang district, Cameroon. Focus group discussions were conducted with midwives, while individual interviews were conducted with gynecologists to comprehend unique perspectives. Each interview was audio-recorded, transcribed, and independently coded by 2 researchers using the ATLAS.ti (Lumivero, LLC) software. The technology acceptance lifecycle framework guided the content analysis, focusing on the preadoption phases to examine the perceived acceptability and initial acceptance of the CAD tool in clinical workflows. The study findings were reported adhering to the COREQ (Consolidated Criteria for Reporting Qualitative Research) and SRQR (Standards for Reporting Qualitative Research) checklists.

Results: Key elements influencing the sustained use of CAD tools for VIA by HCPs were identified, primarily within the technology acceptance lifecycle's preadoption framework. Barriers included the system's ease of use, particularly challenges associated with image acquisition, concerns over confidentiality and data security, limited infrastructure and resources such as the internet and device quality, and potential workflow changes. Facilitators encompassed the perceived improved patient care, the potential for enhanced diagnostic accuracy, and the integration of CAD tools into routine clinical practices, provided that

infrastructure and training were adequate. The HCPs emphasized the importance of clinical validation, usability testing, and iterative feedback mechanisms to build trust in the CAD tool's accuracy and utility.

Conclusions: This study provides practical insights from HCPs in Western Cameroon regarding the adoption of CAD tools for VIA in clinical settings. CAD technology can aid diagnostic objectivity; however, data management, workflow adaptation, and infrastructure limitations must be addressed to avoid "pilotitis"—the failure of digital health tools to progress beyond the pilot phase. Effective implementation requires comprehensive technology management, including regulatory compliance, infrastructure support, and user-focused training. Involving end users can ensure that CAD tools are fully integrated and embraced in LMICs to aid cervical cancer screening.

(JMIR Cancer 2025;11:e50124) doi:10.2196/50124

KEYWORDS

qualitative research; technology acceptance; cervical cancer; diagnosis; computer-assisted; decision support systems; artificial intelligence; health personnel attitudes; Cameroon; mobile phone

Introduction

Background

In low- and middle-income countries (LMIC), visual inspection with acetic acid (VIA) is a common low-cost method for screening and triage. This method involves applying diluted acetic acid to the cervix during gynecological examination. This induces tissue whitening, which a trained observer assesses to guide the diagnosis. Despite VIA's cost-effectiveness and accessibility, a major limitation is its high subjectivity owing to variability in the training and experience of the observer [1].

With advancements in artificial intelligence (AI), its potential to assist in diagnosis has been extensively explored and investigated [2]. Tools that aid health care providers (HCPs) in detecting diseases and identifying abnormalities are clinical decision support (CDS) systems and, more specifically, computer-aided detection and diagnosis (CAD) systems [3]. In cervical cancer, CAD tools can mitigate the subjectivity inherent in VIA by providing standardized evidence for clinical decision-making [4].

Despite the promising potential of CAD systems, their implementation and sustained use in LMICs are limited, irrespective of the target disease. These barriers include the risk of workflow disruption, dependency on computer literacy, poor data quality, lack of transportability and interoperability (ie, system compatibility), and financial challenges [5]. In addition, frameworks exist to guide the implementation, evaluation, and regulation of these digital health tools [6-10]. However, the lack of harmonization across different entities compounds their development complexity. The technology may not progress beyond the pilot stage because of the previously mentioned reasons—a common phenomenon referred to as "pilotitis" [11].

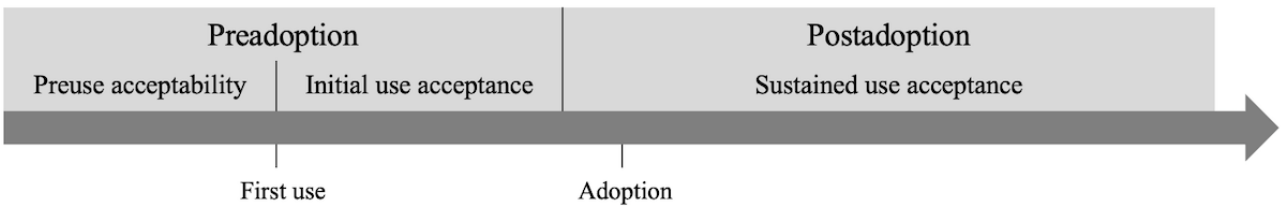
This study examines the barriers and facilitators to deploying a VIA CAD tool to enhance and standardize cervical cancer diagnosis using AI. User perspectives were collected to understand the challenges and enablers of implementing this technology. This study provides insights into deploying AI-enhanced diagnostic tools in LMICs to bridge the gap between technological development and real-world implementation.

Theoretical Framework

In the context of VIA CAD tools for LMICs, the following definitions are used throughout the paper: (1) acceptability: "the quality of being satisfactory and able to be agreed to or approved of," according to the Cambridge Dictionary [12]; (2) acceptance: "general agreement that something is satisfactory or right," as stated by the Cambridge Dictionary [13]; and (3) adoption: "a multiphase process starting with deciding to adopt [a technology] (selecting, purchasing, or committing to use it) and then achieving persistent use," provided by Carroll et al [14].

These 3 concepts are combined in successive steps in the technology acceptance lifecycle (TAL) proposed by Nadal et al [15]. The model (Figure 1) reveals that technology preadoption is a 2-stage process. First, acceptability before use is preuse acceptability, followed by initial use acceptance once the technology has been used for the first time. Postadoption is considered sustained use acceptance, implying that the device is fully adopted and sustainably used.

Figure 1. Technology acceptance lifecycle adapted from Nadal et al [15].



The TAL model was chosen because it captures key factors in the preadoption phase, which is crucial previous to integrating CAD tools in resource-constrained health care settings. Furthermore, the postadoption phase addresses specific barriers, such as infrastructure and data security, while simultaneously supporting sustained adoption and mitigating pilotitis risk.

Objective

This study focused on the challenges perceived by HCPs that would prevent the sustained use of CAD tools for VIA in clinical settings. The primary objective of this study was to identify the common barriers to adopting CAD tools for VIA, and facilitators were the secondary objective. Barriers and facilitators were studied within the framework of the TAL, especially in the preadoption phase. The study was conducted in Western Cameroon; however, the results are critically interpreted for generalizability to other geographical settings within sub-Saharan Africa or other LMICs.

Methods

Overview

This study is reported per the COREQ (Consolidated Criteria for Reporting Qualitative Research; Multimedia Appendix 1) and SRQR (Standards for Reporting Qualitative Research; Multimedia Appendix 2) checklists.

Study Settings and Recruitment

This study was conducted as part of a cervical cancer screening program in the Dschang district, Western Cameroon. A larger study, the 3T program (for test-triage-treatment), was initiated in 2018 at Dschang District Hospital with the support of the Ministry of Health and in collaboration with Geneva University Hospital [16]. This program includes patient recruitment through awareness campaigns in rural and urban areas, tests for HPV, followed by VIA triage (if the HPV test is positive), and treatment, if necessary. The cervical cancer screening was performed by midwives who welcomed the patients, explained the study, conducted the gynecological examination, and, if required, administered treatment. In the case of severe lesions, the patient was treated by a midwife and gynecologist. Within the scope of this program, images were collected during the colposcopy for peer review of the diagnoses and to develop a CAD tool for VIA. This technology relies on image processing and machine learning [17].

Participants were recruited from trained teams at the Dschang District Hospital and the Regional Hospital of Bafoussam, a second 3T program site approximately 60 km east of Dschang. Every HCP in the 3T program is Cameroonian, better to understand the local context and its specific culture. Participants were compensated for their transportation expenses and provided with refreshments during discussions.

Study Design and Procedure

A qualitative methodology was chosen to capture nuances in participants' views, experiences, and behaviors regarding technology and their environment [18]. Compared with a quantitative approach, a qualitative approach provides better insight into participants' perspectives and provides the flexibility required for participants to highlight information they may not have anticipated [19,20].

A mixed approach composed of individual interviews and mini–focus group discussions (mFGDs) was adopted. Individual interviews contributed to in-depth data collection [21], and mFGDs facilitated the exchange of participants' perspectives and the compilation of collective perceptions [20]. The term "mini" denotes the small group size of 3 participants, which was chosen because of the limited availability of individuals with relevant expertise and the sensitivity of the topic, which benefits from an intimate setting [22]. Due to the HCPs' various educational levels, training, and daily duties, the participants were grouped into homogeneous professional mFGDs. The aim was to create a climate of trust, enabling all participants to express themselves freely and reduce authority bias [23].

Therefore, 2 mFGDs, each with 3 midwives and 3 individual interviews with gynecologists, were conducted between March and May 2022. All discussions were moderated in French—the official language—by a female Cameroonian anthropologist who had worked on the 3T program for several years and had intimate knowledge and familiarity with the professional community. During focus group discussions, in which the perspectives of the HCPs were collected at the Dschang District Hospital, a second female Cameroonian anthropologist took notes to capture nonverbal cues and additional context. An individual interview was conducted via videoconferencing because the medical doctor was traveling. The remaining discussions occurred in clinical settings, either in a confidential conference room at the hospital or a private medical office.

Each interview began with an introduction to the CAD tools for VIA, and the procedure was described in detail. The slides supported the verbal explanation and clarified the following steps. During the gynecological examination, the HCP first applied diluted acetic acid to the cervix, corresponding to the routine VIA. Second, cervical images were recorded using a dedicated mobile app. Subsequently, the algorithm integrated into the mobile app processed the sequence of images and provided an analysis to the user. Finally, the HCP interpreted the results to determine whether treatment was necessary. The entire procedure can be conducted offline; nonetheless, internet-based synchronization with a server is also feasible, ensuring data backup and compliance with data privacy and confidentiality requirements.

A mobile app was specifically developed for the demonstration to concretize the concept for the participants. The mobile app allowed users to capture a series of pictures and generate simulated predictions of cervical cancer. The participants could manipulate the mobile app and take pictures of their surroundings to familiarize themselves with CAD tools for cervical cancer. A semistructured questionnaire was then used to (1) investigate HCPs' perceptions of smartphone use for medical applications, cervical cancer, and AI and (2) identify the challenges and facilitators for integrating CAD tools for VIA in clinical settings.

All interview guide questions were fully addressed, and the participants were given opportunities to ask questions. Finally, participants could share their thoughts and comments. All discussions were audio-recorded, transcribed, and anonymized. Triangulation was applied within and across sessions to enhance data reliability.

Data Processing and Analysis

A total of 4 audio recordings were transcribed verbatim in French, categorized, and coded using content analysis [24]. Therefore, coding was initially deductive based on a codebook, guided by the topics outlined in the interview guide, followed by the generation of inductive codes directly from the transcriptions. Around 2 female coresearchers (MJC and AMDM) independently coded transcripts using the computer-assisted qualitative data analysis software (ATLAS.ti, version 22.2.3) to support scientific rigor and reflexivity. The data were double-coded by a researcher from a different cultural context to address the potential reporting bias that could arise from the shared experience between the anthropologist and the HCPs. The coding was then compared, discussed when diverging, combined, and finally assessed for consistency following the open discussion method [25].

The coded data were analyzed to identify relationships between the categories and hierarchically ordered, with the TAL framework applied as a theoretical lens for interpreting the coded data. Data were analyzed with team members from diverse professional and cultural backgrounds. This combination of expertise and cultural insights strengthened the study design and analysis. Data saturation was confirmed through consistent responses across the mFGDs and interviews. Each mFGD was included, and all sessions were audio-recorded while maintaining confidentiality. Suppose quotations were selected for publication; in that case, they were translated from French into English.

Ethical Considerations

This study was approved by the Ethical Cantonal Board of Geneva, Switzerland (CCER, 2017-01110 and CER-amendment no. 4) and the Cameroonian National Ethics Committee for Human Health Research (2022/12/1518/CE/CNERSH/SP). The participants verbally consented after being informed of the study's purpose, topics, duration, benefits, and risks. Participants were compensated for their transportation expenses. All transcripts were anonymized and analyzed in a fully deidentified manner to ensure privacy and confidentiality. The achievement of data saturation, the presence of 2 researchers during the mFGDs, audio recordings, and triangulation of information attest to the rigor and trustworthiness of the data collection and analysis.

Results

Participants

The study involved all HCPs from both 3T program sites, encompassing 6 midwives and 2 gynecologists, of whom 5 were females and 3 were male participants aged 30-55 years. All participants specialized in cervical cancer, with experience between 6 months and over 10 years, providing insights for evaluating the technology across different stages of professional development. They all owned a smartphone, used it daily, and were familiar with smartphone apps in clinical settings through the 3T program. None of the participants dropped out of the study, and the discussion duration was 42-92 minutes.

Main Results

HCPs' perceptions of adopting a CAD tool for VIA were mainly positive. The HCP highlighted 8 facilitators and 5 barriers that should be addressed in the future.

The barriers were:

- Restriction of the HCP's movement due to the smartphone's position.
- Constraining requirements for the image quality.
- Confidentiality concerns for sensitive data, especially in case of data breach.
- Workflow changes may encourage HCPs to heavily depend on the technology.
- Limited access to internet connection and smartphones.

The facilitators were:

- Improved patient care through rapid and reliable diagnosis.
- Improved diagnosis.
- Reduced workload because of improved efficiency.
- Reinforcement of visual inspection with acetic acid (VIA) protocol.
- Clinical evaluation of the technology in conditions reflecting real-life conditions.
- Automated assessment of image quality.
- Training of the users.
- Provision of smartphones and tripods.

Barriers to CAD Tools for VIA

System Usability Challenges

The smartphone was positioned on a tripod between the gynecological chair and HCP to capture the images. This positioning restricted the movement of professionals during the examination.

> *When pipetting [acetic acid] and the whole cervix is not captured on the smartphone but the recording already started, then we need to twist and turn to finish. For those difficulties, we could pull the trip away, pipet to remove all the acetic acid well, and then position the device.* [P6]

In addition, image quality requirements entail acquisition conditions that are difficult to achieve. These quality assessment criteria can be affected by video movements, light changes, reflections, and the presence of blood or mucus. Furthermore, some HCPs raised concerns that the high-quality images required for an accurate diagnosis would be difficult to achieve, possibly leading to misdiagnosis.

> *[…] So that artificial intelligence provides an accurate result, one will need to fulfill all acquisition conditions. It means that if you make a small mistake in the process, the risk of a false positive or false negative will be high […]* [P2]

Confidentiality

Most HCPs expressed apprehension regarding the handling and storage of sensitive data. Their concerns revolved around potential breaches of medical data confidentiality and the unauthorized sharing of VIA images. Furthermore, they feared

the loss of images due to the misfunctioning or mishandling of smartphones.

> *[…] If I send a picture to someone, they can easily forward it to someone else without asking my permission. There were so many scandals in the field of medicine because someone took a picture with their smartphone, which was shared multiple times and leaked […]* [P8]

An HCP emphasized that framing the picture around the cervix alone can aid confidentiality concerns because it is anonymizing.

> *When using a smartphone to take photos of the cervix, it is beneficial as long as it preserves the woman's privacy by capturing solely the cervix. That way, the woman cannot feel frustrated about her face potentially being seen elsewhere.* [P2]

In addition, using a private mobile phone to take pictures during the gynecological examination was not allowed. The device should be dedicated to only clinical settings and not leave the screening area.

> *[…] the personal mobile phone is never fully private… when you drop it at home, kids may search into it and access the image […] Personally, I believe the mobile phone used needs to be professional only and to stay at the screening site, never reaching someone's home […]* [P8]

Workflow Changes

Several HCPs were concerned that the adaptation of the device might change the workflow, and users might heavily depend on CAD tools, neglecting their expertise. An HCP explained:

> *For me, and my colleagues, if this device is given to medical doctors, they might become lazy. They will not have to do their work fully. As soon as there is a case, they would take the smartphone and would not think about the diagnosis themselves.* [P1]

Therefore, their gynecological knowledge might be affected, possibly limiting patient care in complicated cases.

Limited Infrastructure and Resources

Another challenge regarding adopting the technology is the limited access to resources. Some HCPs were concerned about access to reliable internet connections and smartphones, along with their quality. The wear of the smartphone battery was also a concern.

> *[…] Everyone cannot access a smartphone, everyone cannot access the internet […]* [P8]

Facilitators of CAD Tools for VIA

Patient Care Improvement

Most HCPs agreed that the technology would improve patient care by providing a rapid and reliable diagnosis. According to them, accelerating the diagnosis speed benefited the patients and HCPs. First, patients would not return home with unanswered questions and directives to wait, which could induce stress and unnecessary worry.

> *They [patients] will accept the technology since when coming out of here [the hospital], they will know their diagnosis; they will think, 'They told me that I come back home with a clear head, not like before when I came and I had to wait for a phone call with the results […]* [P3]

Second, the technology could help HCPs provide a diagnosis or mitigate the lack of resources for conducting biopsies. Some professionals also believed that the workload could be reduced using technology that would enable the screening of more women for cervical cancer.

Furthermore, some HCPs mentioned that adopting a CAD tool for cervical cancer could enhance diagnosis by identifying lesions that are challenging to detect with the human eye. This can reduce the potential for human error.

> *If it is efficient, then it would be a very important tool because visual inspection is now very subjective. But if we manage to generate a diagnosis from a simple image, that would be a very good progress.* [P7]

This technology may also reinforce the protocol for performing appropriate VIA. For example, the predetermined duration of the recording constrains the user to wait until the recording is completed. Diagnosis improvements may also reduce the rate of overtreatment, which is currently considered crucial, as illustrated by one of the HCPs.

> *[…] in the approach to screen and treat, we are afraid, on the one hand, not to treat someone who should have, until the point it develops into a cancer. On the other hand, we may treat someone who did not need it, which is called overtreatment, but if we have this technology helping, then we will win in both cases […]* [P7]

Clinical Evaluation

Before this technology can be used in clinical settings, approval and validation by health care professionals is required. This implies comparing the technology's diagnosis in real conditions and that of HCPs, histopathology results, and a usability study regarding the tool's features and interface. The HCPs confirmed the need to validate the technology because they were concerned about its performance and their ability to distinguish accurate diagnoses from misdiagnoses.

Automated Assessment of Image Quality

Some HCPs apprehend that automating the technology would induce misdiagnosis and false-positive and false-negative predictions because of cervical abnormalities such as blood, mucus, and benign lesions. Therefore, an automated pipeline for assessing image quality was indicated as an essential technological feature preventing misdiagnosis by not allowing the application of the algorithm to images that are of extremely low quality or ineligible.

> *For instance, we could have a feature that gives us an ok when all acquisition criteria are fulfilled and that the analysis of the images can be conducted. Or, if criteria are not fulfilled, then there would be a*

message telling you not to launch the analysis and to, maybe, retake images. [P4]

Training and Resources Provision

The participants highlighted the importance of comprehensive training on properly using the CAD tool for its clinical adoption. In addition, resources such as smartphones and tripods are required for effective implementation.

[…] One needs to ensure that the user of the technology is well trained to create conditions allowing the capture of a good image because… bad acquisition conditions result in bad image quality. [P7]

Discussion

Principal Findings

This study identified 4 barriers and 4 facilitators each to adopting CAD tools for VIA during the preadaptation phase of the TAL model (Figure 1), either related to preuse acceptability or to initial use acceptance (Table 1). These included structural components such as limited infrastructure, confidentiality, data management, and personal factors such as workflow changes and user interaction with the CAD tool. HCPs emphasized the overall benefits, including improved patient care through better clinical evaluation, enhanced training, and technical advantages, such as ease of use of the system. The following section discusses the barriers and facilitators identified in both phases, contextualized within this literature, and highlights the barriers observed in previous studies but not encountered in this one.

Table 1. Synthesis of principal findings under the preadoption phases of the technology acceptance life cycle.

	Preuse acceptability	Initial use acceptance
Barriers	• Confidentiality and data management • Limited infrastructure and resources	• User interaction • Change in workflow
Facilitators	• Patient care improvement • Training	• Clinical evaluation • Ease of use

Barriers to Preuse Acceptability

Confidentiality and Data Management

Confidentiality concerns were highlighted as a significant factor affecting pre-use acceptability. HCPs expressed reservations about data loss and breaches and the acquisition of cervical images with smartphones. Patients also expressed concerns about the image frame. However, these concerns are not specific to the use of AI but are raised by the use of digital, sensitive images. Lodhia et al [26] observed that assuring patients of confidentiality was crucial to their mHealth intervention study for eye care in Kenya and recommended a robust data protection system. Prioritizing confidentiality, incorporating a secure data management system, and transparent communication about security measures are essential for alleviating privacy concerns and building trust among HCPs and patients. Practically, some HCPs have suggested showing the acquired data to patients to reassure them about the content of the images.

Limited Infrastructure and Resources

A lack of infrastructure and resources can affect the preuse acceptability of CAD tools for VIA in cervical cancer. Equipping HCPs with the necessary devices (such as smartphones and tripods) and ensuring regular maintenance of the devices is essential for limiting technical challenges and providing uninterrupted and effective health care services to patients.

Limited access to reliable internet connections, smartphone quality, and smartphone battery wear concerns during prolonged use have been highlighted as barriers by HCPs. These challenges align with findings from the existing literature in which infrastructure-related challenges have been identified in other health care settings. A study by Elahi et al [27] on a CAD tool for traumatic brain injury in Uganda emphasized the importance of internet connection and the associated costs. Spence et al [28] and Knoble and Bhusal [29] reported reliable electricity as a key challenge in the studies of childhood pneumonia diagnostic tools and electronic diagnostic algorithms.

A viable solution to overcome internet-related constraints is developing a CAD tool that can function offline. Internet access might be occasionally needed for maintenance, updates, and data sharing or backup; however, the device would provide CDS, irrespective of internet coverage and connectivity. In addition, ensuring access to electricity is essential to keep the smartphones charged. Therefore, adopting good practices, such as switching off the device at night or when it is not in use for several days, can help preserve battery life. Furthermore, optimizing the CDS algorithm to minimize smartphone power consumption can contribute to its ease of use without frequent recharging. By implementing these strategies, HCPs can confidently use CAD tools, even in remote areas with limited internet access and electricity.

Facilitators for Preuse Acceptability

Patient Care Improvement

CAD tools for VIA have diverse benefits that may contribute to improvements in cervical cancer screening and diagnosis [30]. A key advantage is rapid and reliable diagnosis, which reduces waiting time for patients who can receive test results and treatment in a single appointment. This point-of-care

approach can enhance patient satisfaction and positively impact mental health by providing a timely, near–real-time diagnosis [31]. In addition, this technology serves as a valuable supplementary test, assisting HCPs and making the diagnostic process less subjective. This technology limits cases of overtreatment and mitigates the risk of unnecessary treatment-related complications, such as pregnancy-related morbidity, by enhancing diagnosis accuracy [32].

Training

Comprehensive training is essential for users to understand and use the technology accurately. HCPs should not rely solely on the suggested diagnosis but use it as an assistive device that provides supplementary evidence to be considered in addition to their clinical knowledge and experience. Training must be tailored to the specific contexts of use and encompass geographical, cultural, and educational considerations [33]. The objective is to equip users with the ability to comprehend, interpret, and integrate recommendations from CAD tools into their decision-making processes.

Barriers to Initial Use Acceptance

User Interaction

The initial acceptance of the system can be hindered by difficulty in use. Demanding acquisition conditions are challenging for HCPs to comply with, potentially hindering technology adoption. The CDS algorithm requires adherence to specific acquisition conditions to address issues, such as movement, light, and mucus or blood in the cervix. The HCPs suggest that this could be facilitated by an automated quality assessment pipeline that assesses image quality in real time and provides immediate feedback to users, indicating whether the image quality is sufficient. Such a process could assess the visibility of the cervix, monitor movement, detect blurriness, or detect external objects (eg, pipettes).

Beyond the acquisition process, interaction with the device must be satisfactory to the HCPs. Knoble [29] conducted a comprehensive study on electronic diagnostic algorithms in Nepal and uncovered that HCPs found the device's size too small and the touchscreen sensitivity too low. This highlights the importance of considering all usability aspects while developing the technology and iteratively testing it with users.

Workflow Changes

Potential changes in the HCPs workflow also need to be assessed because they could hinder the initial acceptance of CAD tools for VIA [34]. Furthermore, a few HCPs expressed concerns about the risk of neglecting their expertise because of heavy reliance on the new technology. This observation is consistent with current literature. Despite the potential educational purposes of AI-driven CDS tools, the tendency to rely heavily on automation, described as automation bias, has been highlighted by Khera et al [35]. Furthermore, Jabbour et al [36] observed a decline in diagnostic performance when clinicians used AI support, even when visual explanations of AI-driven diagnoses were provided.

Introducing a novel technology can also increase the workload during the initial learning phase. Melas et al [37] observed that reduced time consumption was a vital factor influencing clinicians' intentions to adopt it. In contrast, Ellington et al [38] reported that HCPs need to practice using technology before using it with patients to improve efficiency. This preparatory practice would allow HCPs to become familiarized with the technology's functionalities and workflows, ensuring a smoother and more efficient integration of the technology into their clinical practice.

Facilitators for Initial Use Acceptance

Clinical Evaluation

The technology should undergo rigorous evaluation to assess its clinical and cost benefits to ensure initial use acceptance and foster sustained use acceptance [9]. The tool's performance can be evaluated using metrics such as sensitivity, specificity, accuracy, and false-positive rate. Evaluating the efficiency of the device and its integration into patient care contributes to assessing its clinical benefits. This involves examining the clinical utility of the technology, its compatibility with specific clinical settings, and its impact on patient outcomes. In addition, a user-centered design highlights the importance of user feedback for validating the technology, particularly its interface, features, and ease of use [39]. Cost benefits should also be assessed by estimating cost savings and implementation costs, including initial setup, training, and maintenance costs.

Laka et al [40] highlighted that evaluation frameworks should consider the dynamic nature of clinical settings and CDS (eg, through software updates). However, such devices need to be continuously monitored to ensure their safe and effective integration in clinical settings [41]. Papadopoulos et al [42] illustrated this challenge in a systematic literature review, revealing that only a few studies included "any form of evaluation," with often insufficient methodologies.

Ease of Use

In addition to facilitating the acquisition and quality assessment processes, simplicity in using CAD tools is crucial for their successful adoption in clinical settings [28]. Panicker et al [43] indicated that easy-to-use CAD tools save time because of their simple operation, fast access, effective recording, and information retrieval. These features contribute to the perceived usefulness and potential for extensive system adoption in clinical practice.

Comparison to Previous Studies

The barriers identified mostly align with those in the existing literature. However, these studies also reported barriers not mentioned by the participants of this study.

A common barrier highlighted in the literature is the limited resources to finance the technology, cope with increased diagnosis, and provide follow-up and treatment when necessary [26-28,44]. In the 3T program, external funding covers all costs, including salaries, clinical materials (gloves, speculum, acetic acid, etc), equipment (smartphone, tripod, etc), and patient travel cost compensation, eliminating immediate financial constraints.

Another barrier emerging from the literature, but not specifically addressed in this study, is the risk of mobile device theft [26].

This could be a relevant consideration in less controlled settings. Community and political support are also lacking in the literature [38,43]. The 3T program receives support from the Health Ministry, albeit financed through foreign grants.

The literature indicates a lack of user experience or skill (ie, phone literacy) as a potential barrier, even though the participants were comfortable handling smartphones [26,29,38,43]. This may result in misuse or an increased workload. In the 3T program, all HCPs follow comprehensive training, ensuring their proficiency in using the technology.

In summary, the absence of these barriers identified in the international scientific literature but not encountered in the 3T program might be attributed to the specific context and controlled environment provided by the program, as well as the active involvement of the study participants in it.

Further Recommendations

The World Health Organization (WHO) proposes 3 main solutions for barriers to the general use of medical devices [45]. First, the technology should be designed to fit the context of use. Second, the device system should be managed comprehensively, including its regulatory aspects, installation, maintenance, and monitoring. Finally, professionals need to be trained for proper use, maintenance, and presentation to patients. In alignment with the WHO recommendations to avoid pilotitis, this study identified barriers and facilitators to integrating CAD tools for VIA in cervical cancer diagnosis.

Addressing specific barriers to deploying CAD systems involves the fundamental element of trust. Patients and HCPs must trust the technology to ensure its successful implementation and sustained use. Trust might motivate HCPs to integrate CAD systems into their workflows and enhance patients' acceptance and comfort through a diagnostic process facilitated by CAD tools.

Trust can be built by involving users from the outset of technological development and integrating their feedback. Early engagement allows the understanding of their specific needs and ensures that the technology addresses relevant challenges. Involving various user profiles (eg, profession, degree, and working experience) from different clinical settings introduces a range of perspectives that can be leveraged to tailor the technology to diverse contexts.

In addition, providing evidence of the technology's trustworthiness is key to building trust [46]. Explainability, which makes the decision-making process of the technology transparent and understandable to users, builds confidence in its reliability. Traditionally, CDS are knowledge-based and rely on medical literature and conditional logic [47]. This approach tends to be more transparent than the current AI-leveraged CDS systems. Comprehensive training further reinforces trust for knowledge- and AI-based CDS by ensuring the appropriate use of the technology through stepwise instructions and presentation of its limitations.

The TAL framework provides valuable perspectives on the current stages of technological adoption. However, it should be considered as an iterative tool, acknowledging that perceptions and acceptance may evolve as the technology progresses. Continuous monitoring of these changes is essential to ensure the successful adoption and use of CAD tools for VIA in clinical practice.

Strengths and Limitations

To our knowledge, this is one of the first studies in sub-Saharan Africa that investigates the adoption of CAD tools for cervical cancer from the perspective of HCPs. The various professional backgrounds of the participants, years of experience in cervical cancer screening, and sex allowed us to gain in-depth insights into the barriers and facilitators present during the preadaptation phase.

The qualitative method allowed for collecting different perspectives; however, several limitations were observed. First, the study involved a limited number of nonrandom samples of participants familiar with the use of smartphones for cervical cancer diagnosis. In addition to potential selection bias and limited generalizability, mini–focus groups have been acknowledged in the scientific literature, especially among participants with intense experience with a topic, which was the case in our study, allowing for a more in-depth discussion [48]. The mFGDs were complemented by individual interviews to capture a wider range of perspectives and mitigate this bias. Second, the study was conducted in a controlled program in Western Cameroon, which may not reflect other settings with varying resources and infrastructure. Therefore, we recommend replicating the study in other contexts for broader validation. Third, this study focused on HCPs; however, patient perspectives, crucial for understanding technology adoption, were not included but addressed in a separate study [49]. Finally, barriers can be attributed to the use of CAD tools or smartphones. This study was conducted during the preadaptation phase; however, the technology and user attitudes have evolved. Iterative research is needed to determine whether factors can be attributed to specific digital tools and capture changing perceptions after implementing the CAD tool.

Some of the encountered barriers (such as workflow changes) might be addressed through adaptation and facilitator reinforcement (such as training); therefore, the findings of this study provide a crucial foundation for future work. In summary, further qualitative and quantitative studies should be conducted among HCPs with less smartphone experience and in various clinical settings to enhance the generalizability of findings and guide tailored strategies for implementing CAD tools in cervical cancer diagnosis.

Conclusions

Interviewing HCPs from various backgrounds with diverse working experience—midwives and gynecologists—was essential to better understand their concerns, perceptions, and expectations regarding a CAD tool for cervical cancer and to avoid "pilotitis."

Even if a CAD tool is designed to fit its context of use perfectly, understanding the barriers and facilitators is key to its successful implementation and use. In this study, the CAD tool was appreciated by the HCP for its ease of use and crucial role as an assistive device to improve patient care and support clinical

decisions. The barriers encountered (confidentiality and data management, limited infrastructure and resources, workflow changes, and user interaction) are specific to the study setting and might vary with the context. However, these challenges can be addressed in line with the WHO recommendations [40] by ensuring proper management of the technology (eg, regulation and maintenance) and involving the end user at every step of developing the solution. HCPs will help define the appropriate way of introducing new technology to patients and their peers, as well as how to train them for proper use. These measures aim to derisk the deployment of the technology and contribute to overcoming pilotitis.

Acknowledgments

The authors would like to thank all health care providers (HCPs) who dedicated some of their time to discussing with anthropologists from the project. This work was supported by the Swiss Federal Institute of Technology Lausanne through its Programme Tech4Dev.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to confidentiality and privacy concerns but from the corresponding author on reasonable request.

Authors' Contributions

All authors contributed to the conceptualization and design of the study. The methodology was developed collaboratively by AMDM, MJC, and NCS. For the investigation, AMDM conducted participant recruitment and data collection, while VFY attended the mini–focus group discussions and took notes. In terms of data curation, AMDM and VFY transcribed the audio recordings. The formal analysis and interpretation of the data were carried out by MJC, AMDM, and NCS. MJC wrote the original draft, while all authors contributed to reviewing and editing multiple versions of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consolidated criteria for reporting qualitative research (COREQ) checklist.
[DOCX File , 24 KB - cancer_v11i1e50124_app1.docx]

Multimedia Appendix 2
Standards for Reporting Qualitative Research (SRQR) checklist.
[DOCX File , 22 KB - cancer_v11i1e50124_app2.docx]

References

1. Hilal Z, Tempfer CB, Burgard L, Rehman S, Rezniczek GA. How long is too long? Application of acetic acid during colposcopy: a prospective study. Am J Obstet Gynecol 2020;223(1):101.e1-101.e8. [doi: 10.1016/j.ajog.2020.01.038] [Medline: 31981505]
2. Mirbabaie M, Stieglitz S, Frick NRJ. Artificial intelligence in disease diagnostics: a critical review and classification on the current state of research guiding future direction. Health Technol 2021;11(4):693-731 [FREE Full text] [doi: 10.1007/s12553-021-00555-5]
3. Petrick N, Sahiner B, Armato SG, Bert A, Correale L, Delsanto S, et al. Evaluation of computer-aided detection and diagnosis systems. Med Phys 2013;40(8):087001 [FREE Full text] [doi: 10.1118/1.4816310] [Medline: 23927365]
4. Viñals R, Jonnalagedda M, Petignat P, Thiran J, Vassilakos P. Artificial intelligence-based cervical cancer screening on images taken during visual inspection with acetic acid: a systematic review. Diagnostics (Basel) 2023;13(5):836 [FREE Full text] [doi: 10.3390/diagnostics13050836] [Medline: 36899979]
5. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
6. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016. URL: https://iris.who.int/handle/10665/252183 [accessed 2025-01-14]
7. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://apps.who.int/iris/handle/10665/344249 [accessed 2025-01-14]

8. Machine learning-enabled medical devices: key terms and definitions. International Medical Device Regulators Forum. 2022. URL: https://www.imdrf.org/documents/machine-learning-enabled-medical-devices-key-terms-and-definitions [accessed 2025-01-14]
9. Generating evidence for artificial intelligence based medical devices: a framework for training validation and evaluation. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240038462 [accessed 2025-01-14]
10. The AI act explorer. EU Artificial Intelligence Act. 2024. URL: https://artificialintelligenceact.eu/ai-act-explorer/ [accessed 2025-01-14]
11. Greve M, Brendel AB, van Osten N, Kolbe L. Overcoming the barriers of mobile health that hamper sustainability in low-resource environments. J Public Health 2022;30(1):49-62 [FREE Full text] [doi: 10.1007/s10389-021-01536-8]
12. Acceptability. Cambridge Dictionary. URL: https://dictionary.cambridge.org/dictionary/english/acceptability [accessed 2025-01-16]
13. Acceptance. Cambridge Dictionary. URL: https://dictionary.cambridge.org/dictionary/english/acceptance [accessed 2025-01-14]
14. Carroll J, Howard S, Peck J, Murphy J. From adoption to use: the process of appropriating a mobile phone. Australas J Inf Syst 2003;10(2) [FREE Full text]
15. Nadal C, Sas C, Doherty G. Technology acceptance in mobile health: scoping review of definitions, models, and measurement. J Med Internet Res 2020;22(7):e17256 [FREE Full text] [doi: 10.2196/17256] [Medline: 32628122]
16. Levy J, de Preux M, Kenfack B, Sormani J, Catarino R, Tincho EF, et al. Implementing the 3T-approach for cervical cancer screening in Cameroon: preliminary results on program performance. Cancer Med 2020;9(19):7293-7300 [FREE Full text] [doi: 10.1002/cam4.3355] [Medline: 32757469]
17. Viñals R, Vassilakos P, Rad MS, Undurraga M, Petignat P, Thiran JP. Using dynamic features for automatic cervical precancer detection. Diagnostics (Basel) 2021;11(4):716 [FREE Full text] [doi: 10.3390/diagnostics11040716] [Medline: 33920732]
18. Tenny S, Brannan JM, Brannan GD, Bayot ML. Qualitative Study. Treasure Island, FL: StatPearls Publishing; 2022.
19. Verhoef M, Casebeer AL. Broadening horizons: integrating quantitative and qualitative research. Can J Infect Dis 1997;8(2):65-66 [FREE Full text] [doi: 10.1155/1997/349145] [Medline: 22514478]
20. Gill P, Stewart K, Treasure E, Chadwick B. Methods of data collection in qualitative research: interviews and focus groups. Br Dent J 2008;204(6):291-295. [doi: 10.1038/bdj.2008.192] [Medline: 18356873]
21. Ryan F, Coughlan M, Cronin P. Interviewing in qualitative research: The one-to-one interview. International Journal of Therapy and Rehabilitation 2009 Jun;16(6):309-314 [FREE Full text] [doi: 10.12968/ijtr.2009.16.6.42433]
22. Nyumba TO, Wilson K, Derrick CJ, Mukherjee N. The use of focus group discussion methodology: insights from two decades of application in conservation. Methods Ecol Evol 2018;9(1):20-32 [FREE Full text] [doi: 10.1111/2041-210x.12860]
23. Morgan D. Planning Focus Groups. Washington DC: SAGE Publications, Inc; 1998:59-63.
24. Krippendorff K. Content Analysis: An Introduction to Its Methodology. Washington DC: SAGE; 2019.
25. Chinh B, Zade H, Ganji A, Aragon C. Ways of qalitative coding: a case study of four strategies for resolving disagreements. 2019 Presented at: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; May 02, 2019; Glasgow, Scotland URL: https://doi.org/10.1145/3290607.3312879
26. Lodhia V, Karanja S, Lees S, Bastawrous A. Acceptability, usability, and views on deployment of peek, a mobile phone mHealth intervention for eye care in Kenya: qualitative study. JMIR Mhealth Uhealth 2016;4(2):e30 [FREE Full text] [doi: 10.2196/mhealth.4746] [Medline: 27160779]
27. Elahi C, Spears CA, Williams S, Dunn TW, Najjuma JN, Staton CA, et al. An attitude survey and assessment of the feasibility, acceptability, and usability of a traumatic brain injury decision support tool in Uganda. World Neurosurg 2020;139:495-504. [doi: 10.1016/j.wneu.2020.04.193] [Medline: 32376375]
28. Spence H, Baker K, Wharton-Smith A, Mucunguzi A, Matata L, Habte T, et al. Childhood pneumonia diagnostics: community health workers' and national stakeholders' differing perspectives of new and existing aids. Glob Health Action 2017;10(1):1290340 [FREE Full text] [doi: 10.1080/16549716.2017.1290340] [Medline: 28485694]
29. Knoble S, Bhusal MR. Electronic diagnostic algorithms to assist mid-level health care workers in Nepal: a mixed-method exploratory study. Int J Med Inform 2015;84(5):334-340. [doi: 10.1016/j.ijmedinf.2015.01.011] [Medline: 25670230]
30. Hou X, Shen G, Zhou L, Li Y, Wang T, Ma X. Artificial intelligence in cervical cancer screening and diagnosis. Front Oncol 2022;12:851367 [FREE Full text] [doi: 10.3389/fonc.2022.851367] [Medline: 35359358]
31. Singh M, Maheu C, Brady T, Farah R. The psychological impact of the rapid diagnostic centres in cancer screening: a systematic review. Can Oncol Nurs J 2017;27(4):348-355 [FREE Full text] [doi: 10.5737/23688076274348355] [Medline: 31148761]
32. Nakisige C, de Fouw M, Kabukye J, Sultanov M, Nazrui N, Rahman A, et al. Artificial intelligence and visual inspection in cervical cancer screening. Int J Gynecol Cancer 2023;33(10):1515-1521 [FREE Full text] [doi: 10.1136/ijgc-2023-004397] [Medline: 37666527]
33. Chen Z, Liang N, Zhang H, Li H, Yang Y, Zong X, et al. Harnessing the power of clinical decision support systems: challenges and opportunities. Open Heart 2023;10(2):e002432 [FREE Full text] [doi: 10.1136/openhrt-2023-002432] [Medline: 38016787]

34. Renukappa S, Mudiyi P, Suresh S, Abdalla W, Subbarao C. Evaluation of challenges for adoption of smart healthcare strategies. Smart Health 2022;26:100330 [FREE Full text] [doi: 10.1016/j.smhl.2022.100330]
35. Khera R, Simon MA, Ross JS. Automation bias and assistive AI: risk of harm from AI-driven clinical decision support. JAMA 2023;330(23):2255-2257. [doi: 10.1001/jama.2023.22557] [Medline: 38112824]
36. Jabbour S, Fouhey D, Shepard S, Valley TS, Kazerooni EA, Banovic N, et al. Measuring the impact of AI in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. JAMA 2023;330(23):2275-2284 [FREE Full text] [doi: 10.1001/jama.2023.22295] [Medline: 38112814]
37. Melas C, Zampetakis LA, Dimopoulou A, Moustakis V. Modeling the acceptance of clinical information systems among hospital medical staff: an extended TAM model. J Biomed Inform 2011;44(4):553-564 [FREE Full text] [doi: 10.1016/j.jbi.2011.01.009] [Medline: 21292029]
38. Ellington LE, Najjingo I, Rosenfeld M, Stout JW, Farquhar SA, Vashistha A, et al. Health workers' perspectives of a mobile health tool to improve diagnosis and management of paediatric acute respiratory illnesses in Uganda: a qualitative study. BMJ Open 2021;11(7):e049708 [FREE Full text] [doi: 10.1136/bmjopen-2021-049708] [Medline: 34281930]
39. ISO 9241-210:2019, Ergonomics of human-system interaction, Part 210: Human-centred design for interactive systems. International Organization for Standardization (ISO). URL: https://www.iso.org/standard/77520.html [accessed 2025-01-14]
40. Laka M, Carter D, Merlin T. Evaluating clinical decision support software (CDSS): challenges for robust evidence generation. Int J Technol Assess Health Care 2024;40(1):e16. [doi: 10.1017/S0266462324000059] [Medline: 38328905]
41. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearb Med Inform 2019;28(1):128-134 [FREE Full text] [doi: 10.1055/s-0039-1677903] [Medline: 31022752]
42. Papadopoulos P, Soflano M, Chaudy Y, Adejo W, Connolly T. A systematic review of technologies and standards used in the development of rule-based clinical decision support systems. Health Technol 2022;12(4):713-727 [FREE Full text] [doi: 10.1007/s12553-022-00672-9]
43. Panicker RO, Soman B, Gangadharan KV, Sobhana NV. An adoption model describing clinician's acceptance of automated diagnostic system for tuberculosis. Health Technol 2016;6(4):247-257 [FREE Full text] [doi: 10.1007/s12553-016-0136-4]
44. Maixenchs M, Anselmo R, Zielinski-Gutiérrez E, Odhiambo FO, Akello C, Ondire M, et al. Willingness to know the cause of death and hypothetical acceptability of the minimally invasive autopsy in six diverse African and Asian settings: a mixed methods socio-behavioural study. PLoS Med 2016;13(11):e1002172 [FREE Full text] [doi: 10.1371/journal.pmed.1002172] [Medline: 27875532]
45. Medical devices: managing the mismatch: an outcome of the priority medical devices project. World Health Organization. 2010. URL: https://apps.who.int/iris/handle/10665/44407 [accessed 2025-01-14]
46. O'Neill O. Linking Trust to Trustworthiness. International Journal of Philosophical Studies 2018 Apr 25;26(2):293-300 [FREE Full text] [doi: 10.1080/09672559.2018.1454637]
47. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA 2018;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
48. Krueger RA, Casey MA. Participants in a Focus Group. In: Focus Groups: A Practical Guide for Applied Research. Washington DC: SAGE Publications; 2009.
49. Sachdeva M, Datchoua AM, Yakam VF, Kenfack B, Jonnalagedda-Cattin M, Thiran J, et al. Acceptability of artificial intelligence for cervical cancer screening in Dschang, Cameroon: a qualitative study on patient perspectives. Reprod Health 2024;21(1):92 [FREE Full text] [doi: 10.1186/s12978-024-01828-8] [Medline: 38937771]

Abbreviations

AI: artificial intelligence
CAD: computer-aided diagnosis CDS: clinical decision support
COREQ: Consolidated Criteria for Reporting Qualitative Research HCP: health care provider LMIC: low- and middle-income country mFGD: mini–focus group discussion
SRQR: Standards for Reporting Qualitative Research TAL: technology acceptance lifecycle VIA: visual inspection with acetic acid WHO: World Health Organization

Edited by N Cahill; submitted 13.10.23; peer-reviewed by T Salzmann, M Zahmatkeshan, L Guo; comments to author 06.06.24; revised version received 31.07.24; accepted 23.12.24; published 05.02.25.

Please cite as:
Jonnalagedda-Cattin M, Moukam Datchoua AM, Yakam VF, Kenfack B, Petignat P, Thiran JP, Schönenberger K, Schmidt NC
Barriers and Facilitators to the Preadoption of a Computer-Aided Diagnosis Tool for Cervical Cancer: Qualitative Study on Health Care Providers' Perspectives in Western Cameroon
JMIR Cancer 2025;11:e50124
URL: https://cancer.jmir.org/2025/1/e50124
doi:10.2196/50124
PMID:

Association Between Risk Factors and Major Cancers: Explainable Machine Learning Approach

Xiayuan Huang[1], PhD; Shushun Ren[2], MS; Xinyue Mao[3], BS; Sirui Chen[3]; Elle Chen[2]; Yuqi He[4], PhD; Yun Jiang[2], MS, PhD

[1]Department of Biostatistics, Yale University, New Haven, CT, United States

[2]School of Nursing, University of Michigan–Ann Arbor, 400 North Ingalls Street, Ann Arbor, MI, United States

[3]College of Literature Science and the Arts, University of Michigan–Ann Arbor, Ann Arbor, MI, United States

[4]University Library, San Jose State University, San Jose, CA, United States

Corresponding Author:
Yun Jiang, MS, PhD
School of Nursing, University of Michigan–Ann Arbor, 400 North Ingalls Street, Ann Arbor, MI, United States

Abstract

Background: Cancer is a life-threatening disease and a leading cause of death worldwide, with an estimated 611,000 deaths and over 2 million new cases in the United States in 2024. The rising incidence of major cancers, including among younger individuals, highlights the need for early screening and monitoring of risk factors to manage and decrease cancer risk.

Objective: This study aimed to leverage explainable machine learning models to identify and analyze the key risk factors associated with breast, colorectal, lung, and prostate cancers. By uncovering significant associations between risk factors and these major cancer types, we sought to enhance the understanding of cancer diagnosis risk profiles. Our goal was to facilitate more precise screening, early detection, and personalized prevention strategies, ultimately contributing to better patient outcomes and promoting health equity.

Methods: Deidentified electronic health record data from Medical Information Mart for Intensive Care (MIMIC)–III was used to identify patients with 4 types of cancer who had longitudinal hospital visits prior to their diagnosis presence. Their records were matched and combined with those of patients without cancer diagnoses using propensity scores based on demographic factors. Three advanced models, penalized logistic regression, random forest, and multilayer perceptron (MLP), were conducted to identify the rank of risk factors for each cancer type, with feature importance analysis for random forest and MLP models. The rank biased overlap was adopted to compare the similarity of ranked risk factors across cancer types.

Results: Our framework evaluated the prediction performance of explainable machine learning models, with the MLP model demonstrating the best performance. It achieved an area under the receiver operating characteristic curve of 0.78 for breast cancer (n=58), 0.76 for colorectal cancer (n=140), 0.84 for lung cancer (n=398), and 0.78 for prostate cancer (n=104), outperforming other baseline models (*P*<.001). In addition to demographic risk factors, the most prominent nontraditional risk factors overlapped across models and cancer types, including hyperlipidemia (odds ratio [OR] 1.14, 95% CI 1.11 - 1.17; *P*<.01), diabetes (OR 1.34, 95% CI 1.29 - 1.39; *P*<.01), depressive disorders (OR 1.11, 95% CI 1.06 - 1.16; *P*<.01), heart diseases (OR 1.42, 95% CI 1.32 - 1.52; *P*<.01), and anemia (OR 1.22, 95% CI 1.14 - 1.30; *P*<.01). The similarity analysis indicated the unique risk factor pattern for lung cancer from other cancer types.

Conclusions: The study's findings demonstrated the effectiveness of explainable ML models in assessing nontraditional risk factors for major cancers and highlighted the importance of considering unique risk profiles for different cancer types. Moreover, this research served as a hypothesis-generating foundation, providing preliminary results for future investigation into cancer diagnosis risk analysis and management. Furthermore, expanding collaboration with clinical experts for external validation would be essential to refine model outputs, integrate findings into practice, and enhance their impact on patient care and cancer prevention efforts.

(JMIR Cancer 2025;11:e62833) doi:10.2196/62833

KEYWORDS

electronic health record; EHR; cancer risk modeling; risk factor analysis; explainable machine learning; machine learning; ML; risk factor; major cancers; monitoring; cancer risk; breast cancer; colorectal cancer; lung cancer; prostate cancer; cancer patients; clinical decision-making

Introduction

Cancer is a life-threatening disease and leading cause of death worldwide. In 2024, 611,000 people were estimated to have died from cancer in the United States, and the estimated new cancer cases will reach more than 2 million for the first time [1]. This surge includes rising incidence rates for major cancers, including breast, prostate, lung, and colorectal cancers, which display the trend of increasingly affecting younger individuals who have many more years of life expectancy [1]. The US Preventive Services Task Force modified the recommended age for colorectal cancer screening from 50 to 45 years for people at average risk in 2021 and adjusted the recommendation for breast cancer screening for all women to start at the age of 40 years in 2024 [2,3]. Similar upward trends in the incidence of early-onset cancers are observed in other high-income countries, suggesting shared risk factors and exposures across these regions. However, besides those uncontrollable risk factors, such as previous cancer diagnosis, family history of cancer, and genetics or inherited cancer syndrome, many cancer risk factors, including lifestyle factors, are modifiable and can be managed to decrease people's risk for cancer [4].

Extensive evidence highlights the potential benefits of early identification of individuals at high risk for cancer, which can contribute to improved prevention, more effective treatment, reduced cancer burden, and better long-term outcomes. However, demonstrating a clear survival advantage [5] from screening remains challenging, with notable exceptions such as cervical cancer [6]. It is essential to address biases like lead-time and length bias in screening, as they can overestimate its benefits, ensuring accurate evaluations [7]. In the context of breast cancer, it was estimated that early access to treatment services following breast cancer screening could have reduced breast cancer mortality by 25% - 40% [8]. Given the tremendous benefits of early identification of high-risk patients, an increasing number of cancer risk prediction models have been developed [9]. However, Traditional models used for cancer risk prediction, such as logistic regression (LR) and Cox regression, often demonstrate moderate discrimination accuracy, with an area under the receiver operating characteristic curve (AUC) ranging from 0.53 to 0.64 [10-13]. These models frequently emphasize family history and may have limited generalizability, potentially introducing biases when applied to specific subpopulations [14,15]. Furthermore, nontraditional risk factors, such as chronic diseases, are often overlooked, despite evidence suggesting that chronic conditions can elevate cancer risk similarly to lifestyle factors [16]. This highlights the need for more advanced methods to enhance cancer diagnosis risk prediction and support effective cancer prevention strategies.

Machine learning has shown promising potential in cancer prediction by leveraging electronic health record (EHR) data to identify risk factors [17]. Current applications range from developing predictive models for early cancer detection to personalized treatment recommendations and outcome predictions, based on various patient characteristics and biomarkers. Despite these advancements, several challenges remain in cancer prediction using machine learning [18]. A key issue is the need for a deeper understanding of risk factors within and across different cancer types [19]. As research progresses, explainable machine learning offers a meaningful step forward in improving the efficacy and transparency of predictive models [20-22]. These models not only enhance predictive accuracy but also provide interpretable insights into how predictions are made, fostering trust and facilitating clinical decision-making [23]. By systematically identifying and excluding irrelevant features, explainable approaches can reduce noise and streamline the prediction process. However, it is important to recognize that feature selection algorithms can be sensitive to dataset characteristics, where small changes in the data may lead to differing results [24]. This underscores the importance of carefully selecting features that are most relevant, contributing to a deeper understanding of cancer diagnosis risk factors and improving predictive performance.

Hence, this study presented comprehensive research aimed at uncovering the association between pivotal factors and the risks of 4 major cancer diagnoses (breast, prostate, lung, and colorectal) through the use of explainable machine learning techniques on penalized LR, random forest (RF), and multilayer perceptron (MLP). Our primary objective was to pinpoint the significant features that exert an influence on the risks associated with the diagnosis of these major cancers and to delineate the patterns of risk factors corresponding to each cancer type. Such insights can contribute to enhanced risk monitoring and patient stratification and provide valuable support for clinicians in their decision-making processes, ultimately improving the quality patient care. By elucidating these critical factors and their associated risk factor patterns, we provided clinicians valuable insights through rigorous analysis for enhancing risk monitoring and patient care across various cancer types.

Methods

Experimental Dataset

Our study was conducted using data from Medical Information Mart for Intensive Care (MIMIC)–III, a comprehensive, structured, longitudinal EHR dataset that is publicly available [25]. This dataset contains deidentified, detailed clinical data from intensive care unit (ICU) admissions between 2001 and 2012 at Beth Israel Deaconess Medical Center in Boston, Massachusetts, and is accessible to the global research community under a data use agreement. We used the most recent version (v2.0 released in January 2023) for this work which contains a broad spectrum of data, including information on individual patients' health and health care from various inpatient and outpatient visits, such as diagnoses, prescriptions, lab tests, and procedures. These visits include emergency room admissions and subsequent hospital transfers, where a patient's transfer to a ward or subsequent re-admission to the ICU within the same hospitalization period was considered a single visit. In total, this dataset contains 58,976 admissions of 46,520 patients.

Additionally, we investigated the health status and prevalence of a few common chronic diseases for the MIMIC-III dataset, compared with the prevalence of these chronic diseases in the US population. The MIMIC-III dataset shows that hypertension

affects 47.97% of ICU patients, while in the US population, prevalence ranges from 46.9% to 49.4% [26,27]. Diabetes mellitus is present in 21.20% of MIMIC-III patients, whereas it affects 11.6% of the US population and 14.7% of adults [28]. Hypercholesterolemia appears in 14.94% of ICU cases, with US estimates between 10% and 11.4% [29,30]. Congestive heart failure is recorded in 27.38% of MIMIC-III patients, while the lifetime risk in the US is 24% [31]. Esophageal reflux affects 15.33% of ICU patients and 20% of people in the US [32]. Pneumonia is diagnosed in 12.46% of ICU patients, while 24.9% of US adults have reported cases [33]. Anemia affects 14.02% of ICU patients, while 5.6% of the US population has the condition [34]. Acquired hypothyroidism is observed in 10.71% of MIMIC-III patients and 4.6% of US adults [35]. Tobacco use is recorded in 7.76% of ICU cases, while 19.8% of US adults report smoking [36]. Depressive disorders affect 8.17% of ICU patients, while 9.5% of American adults have been diagnosed [37]. Chronic airway obstruction is reported in 10.24% of MIMIC-III cases, while national estimates range from 6.0% to 6.1% [38].

Data Preprocessing

We included patients with 4 types of cancers (breast, colorectal, lung, and prostate) identified using *International Classification of Diseases, Ninth Revision* (*ICD-9*) codes associated with the diagnosis of each type of cancer (Table S1 in Multimedia Appendix 1).

We took a few steps to preprocess the experimental dataset, starting with the consolidation of 3 main tables from the MIMIC-III database. These included: (1) foundational patient information, capturing demographics and initial hospital admission data; (2) a reference table for *ICD-9* codes, detailing both codes and corresponding diagnostic labels; and (3) logs of patient visit sequences with associated *ICD-9* codes. This consolidation linked the records via patient IDs to construct a detailed longitudinal dataset. Figure 1 illustrates the data processing workflow of this study. Patients' ages were determined by deducting their date of birth from their initial hospital admission date, with the result rounded to the nearest year. Any patient records missing demographic details (such as ethnicity, marital status, or religion) were omitted, narrowing the dataset to a total of 21,372 unique individuals. Our study focused on patients who had multiple hospital visits prior to their cancer diagnosis presence in the record to identify potential risk factors. After a cancer diagnosis code was recognized, further visits were disregarded. These records were combined with those of patients without a cancer diagnosis. A label was created as 1 if a visit included an *ICD-9* code for a cancer diagnosis and 0 if not. To ensure a balanced dataset in terms of cancer diagnosis, the study matched patients diagnosed with cancer with those without cancer using propensity score matching based on demographic factors. Table 1 contains a detailed description of patient characteristics for 4 cancer types.

Figure 1. Medical Information Mart for Intensive Care (MIMIC)–III data processing pipeline. EHR: electronic health record; *ICD-9*: *International Classification of Diseases, Ninth Revision*.

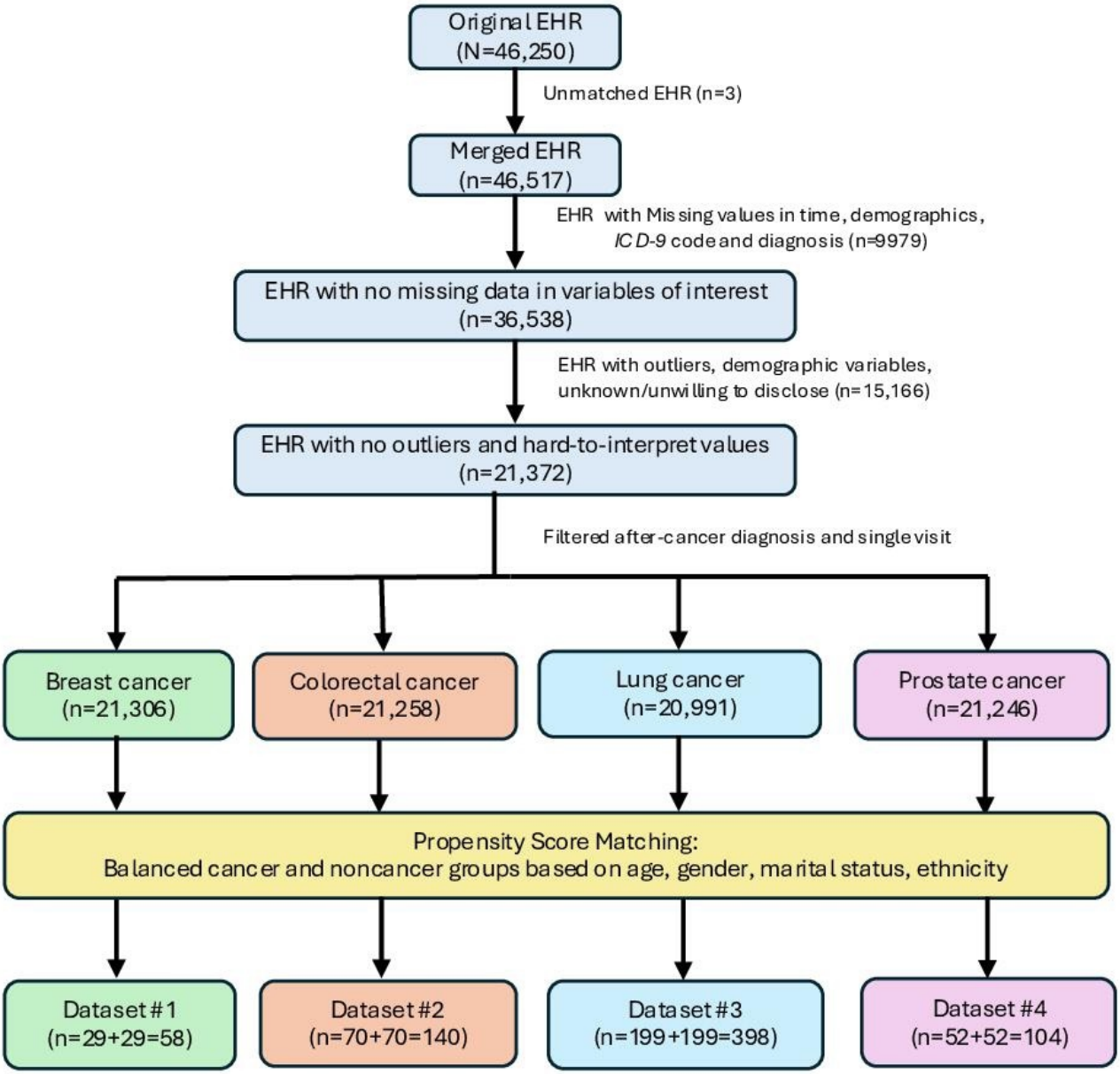


Table . Characteristics of patients for 4 types of cancer.

	Breast cancer (n=58)		Colorectal cancer (n=140)		Lung cancer (n=398)		Prostate cancer (n=104)	
	With cancer (n=29)	Without cancer (n=29)	With cancer (n=70)	Without cancer (n=70)	With cancer (n=199)	Without cancer (n=199)	With cancer (n=52)	Without cancer (n=52)
Age (years), median (range)	60 (40 - 86)	60 (39 - 86)	76 (21 - 87)	75.5 (29-87)	69 (39 - 88)	69 (39 - 87)	74.5 (52-88)	73.5 (52-88)
Sex, n (%)								
Female	27 (93.1)	27 (93.1)	35 (50.0)	33 (47.1)	93 (46.7)	83 (41.7)	0 (0)	0 (0)
Male	2 (6.9)	2 (6.9)	35 (50.0)	37 (52.9)	106 (53.3)	116 (58.3)	52 (100)	52 (100)
Race, n (%)								
White	20 (69.0)	21 (72.4)	51 (72.9)	52 (74.3)	162 (81.4)	157 (78.9)	41 (78.8)	39 (75.0)
Non-White	9 (31.0)	8 (27.6)	19 (27.1)	18 (25.7)	37 (18.6)	42 (21.1)	11 (21.2)	13 (25.0)
Marital status, n (%)[a]								
Married	16 (55.2)	15 (51.7)	37 (52.9)	41 (58.6)	109 (54.8)	113 (56.8)	31 (59.6)	32 (61.5)
Not married	13 (44.8)	14 (48.3)	33 (47.1)	29 (41.4)	90 (45.2)	86 (43.2)	21 (40.4)	20 (38.5)
Religion, n (%)								
Catholic	15 (51.7)	13 (44.8)	35 (50.0)	31 (44.3)	111 (55.8)	107 (53.8)	19 (36.5)	19 (36.5)
Jewish	7 (24.1)	7 (24.1)	17 (24.3)	18 (25.7)	33 (16.6)	31 (15.6)	11 (21.2)	10 (19.2)
Protestant Quaker	7 (24.1)	7 (24.1)	14 (20.0)	11 (15.7)	42 (21.1)	38 (19.1)	16 (30.8)	17 (32.7)
Other	0 (0)	3 (10.3)	4 (5.7)	10 (14.3)	13 (6.5)	23 (11.6)	6 (11.5)	6 (11.5)
ICU[b] visits, n								
Mean	2.5	1.5	2.6	1.5	2.6	1.6	2.5	1.5
Maximum	5	5	6	6	10	12	7	5
Minimum	2	1	2	1	2	1	2	1
ICD-9[c] codes for each patient, n								
Mean	25	14	27	16	26	15	30	15
Maximum	51	68	81	71	82	96	63	66
Minimum	3	3	9	3	6	2	5	4

[a]Categories of marital status include “single”, “divorces”, “widowed”, and “separated”.

[b]ICU: intensive care unit.

[c]*ICD-9*: *International Classification of Diseases, Ninth Revision*.

Feature Selection

Our experiment’s initial dataset comprised thousands of diagnosis codes intended for predicting cancer diagnosis risk. Aware of some features’ potential redundancy and less informative nature, we did a feature selection process. This involved assessing the relevance and importance of each feature in relation to 4 specific types of cancer. We performed a correlation-based feature selection process to identify a subset of features that were highly correlated with the target cancer outcomes. This was followed by a thorough review of relevant literature and consultation with experts to validate and refine the selected features.

Framework

In this work, we applied 3 advanced models, penalized LR, RF, and MLP, based on their demonstrated accuracy and robustness in handling high-dimensional datasets. RF and MLP excel at identifying complex, nonlinear interactions among variables without requiring predefined interaction terms. This capability is crucial for analyzing interactions between risk factors and cancer outcomes. Our choice of RF and MLP was determined by a desire to balance complexity with interpretability, as well as to ensure computational efficiency. Both methods are straightforward and offer high interpretability, which makes

them excellent foundational models for exploring how different features influence cancer diagnosis risk.

Since the task aimed at forecasting cancer diagnosis risk by considering important and relevant risk factors, we evaluated the efficacy of our methodologies by employing several critical performance metrics: AUC, accuracy, specificity, sensitivity, and the F_1-score for each model. We partitioned the dataset into 3 sections for model development: 70% for training, 10% for validation, and 20% for testing. The model that exhibited the best results on the validation set was further subjected to an in-depth analysis of the test set, using a 3-fold cross-validation technique to calculate its AUC precisely. To enhance our understanding of how our machine learning models contribute to cancer prevention, we also quantified the impact of each feature on the prediction of 4 cancer types. We then ranked these features according to their significance. All statistical analyses and model implementations were coded using Python, with the scikit-learn library serving as the foundation for our predictive framework [39]. To assess the generalizability of the model, we validated its performance using an independent ICU dataset from MIMIC-IV-ED ((Medical Information Mart for Intensive Care), which represents an extended patient population. For each cancer type, we randomly sampled 200 cases and 200 matched controls from MIMIC-IV-ED, ensuring no patient overlap with the MIMIC-III experimental dataset.

To investigate the similarity of features ranking by different cancer types, we applied rank biased overlap (RBO) [40], a similarity measure of 2 ranked lists. The RBO score ranges between 0 and 1, where a higher score indicates greater similarity between the lists. A score of 1 implies perfect overlap, meaning the 2 lists are identical in both order and content. On the other hand, a score of 0 suggests no overlap between the lists.

Mathematically, let xi be the high-dimensional feature input. Let yi∈ {0,1} be the corresponding label. yi=0 means not affected, and yi=1 means affected. Our goal is to learn a predictive function f that best classifies the data. We built 3 state-of-the-art models for 4 cancer types respectively in this study:

- Penalized LR: given M training instances, we considered L1 regularized LR by minimizing the following function: ∑i=1M−loglogp(x(i);θ)+β||θ||1.
- RF [41]: a robust ensemble learning method that constructs multiple decision trees during training to improve prediction accuracy and prevent overfitting, where f is the decision tree as base learners. The RF model was trained by iteratively selecting features from root to leaf nodes and aggregating multiple trees with the weights from a subset of the training instances. The nodes and the weights in the model reflect their importance to the final prediction.
- MLP [42]: a type of artificial neural network that consists of at least 3 layers of nodes: an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, in one layer, connects with a certain weight to every node in the following layer, and nodes do not connect within the same layer. The nonlinear activation functions, such as the sigmoid, or Rectified Linear Unit, are applied to the weighted sum of inputs to a neuron, determining its output signal.

To rank the impact on predictive models of the features, relative to all 3 models, we used a permutation importance score to rank all features in the training models for MLP [43]. The scores were defined by the mean decrease in accuracy of the trained model when each feature was permuted.

Ethical Considerations

MIMIC-III data are the result of a collaboration between Beth Israel Deaconess Medical Center (BIDMC) and Massachusetts Institute of Technology. Data collected at BIDMC as part of routine clinical care are deidentified, transformed, and made available to researchers who have completed training in human research and signed a data use agreement. The Institutional Review Board (HUM00230096) at the BIDMC granted a waiver of informed consent and approved the sharing of the research resource. This study was determined to be exempt from further ethical review. The contributing author, XH, obtained the necessary authorization to access the anonymized dataset and oversaw the meticulous data extraction process.

Results

Feature Selection

We conducted a feature selection process to refine thousands of diagnosis codes for predicting cancer diagnosis risk, using correlation-based selection to identify the most relevant features for 4 cancer types. Through this rigorous analysis, we aimed to distill the dataset down to a more manageable and meaningful subset of features. Eventually, we identified 33 features (recategorized into 20 factors for further analysis, Table 2) that emerged as particularly crucial for accurately predicting cancer diagnosis risk. These features were meticulously curated, ensuring that only the most informative and pertinent variables were retained for our predictive models.

Table . Features selected for predicting cancer diagnosis risks.

Features	Factors
Acidosis	Acidosis
Acute kidney failure, unspecified	Acute kidney failure
Age	Age
Anemia, unspecified	Anemia
Acute posthemorrhagic anemia	Anemia
Depressive disorder, not elsewhere classified	Depressive disorder
Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled	Diabetes
Esophageal reflux	Esophageal reflux
Ethnicity	Ethnicity
Gender	Gender
Cardiac complications, not elsewhere classified	Heart disease
Aortocoronary bypass status	Heart disease
Coronary atherosclerosis of native coronary artery	Heart disease
Old myocardial infarction	Heart disease
Congestive heart failure, unspecified	Heart disease
Atrial fibrillation	Heart disease
Subendocardial infarction, initial episode of care	Heart disease
Pure hypercholesterolemia	Hyperlipidemia
Other and unspecified hyperlipidemia	Hyperlipidemia
Unspecified essential hypertension	Hypertension
Other iatrogenic hypotension	Hypotension
Unspecified acquired hypothyroidism	Hypothyroidism
Marital status	Marital status
Religion	Religion
Acute respiratory failure	Respiratory or pulmonary diseases
Unspecified pleural effusion	Respiratory or pulmonary diseases
Pneumonia, organism unspecified	Respiratory or pulmonary diseases
Pneumonitis due to inhalation of food or vomitus	Respiratory or pulmonary diseases
Pulmonary collapse	Respiratory or pulmonary diseases
Chronic airway obstruction, not elsewhere classified	Respiratory or pulmonary diseases
Unspecified septicemia	Sepsis
Personal history of tobacco use	Tobacco use
Urinary tract infection, site not specified	Urinary tract infection (UTI)

Model Performance

For each predicted cancer outcome, we carried out the experiment by predicting cancer using the entire diagnosis history of the patient by building LR, RF, and MLP models. Table 3 illustrates the accuracy, specificity, sensitivity, and F_1-score of these 3 models for breast, colorectal, lung, and prostate cancers. Figure 2 shows the receiver operating characteristic plots of 3 models for 4 types of cancer, respectively. Both Table 3 and Figure 2 show that within the 3 models, MLP performs the best, RF falls in the middle, and LR ranks last. It is worth noting that MLP achieved an AUC of 0.78 for breast cancer, 0.76 for colorectal cancer, 0.84 for lung cancer, and 0.78 for prostate cancer, demonstrating a higher AUC over traditional risk factor-based models and a statistically significant superiority over random chance. The underwhelming results from the LR model led us to investigate the complexity of risk factors for prediction. Compared with LR, MLP reveals the intricate, nonlinear associations between risk factors and the likelihood of cancer, offering meaningful insights into the

collective influence of these risk factors on cancer diagnosis risk.

Table . Comparison of model performance across 4 types of cancer on Medical Information Mart for Intensive Care (MIMIC)–III.

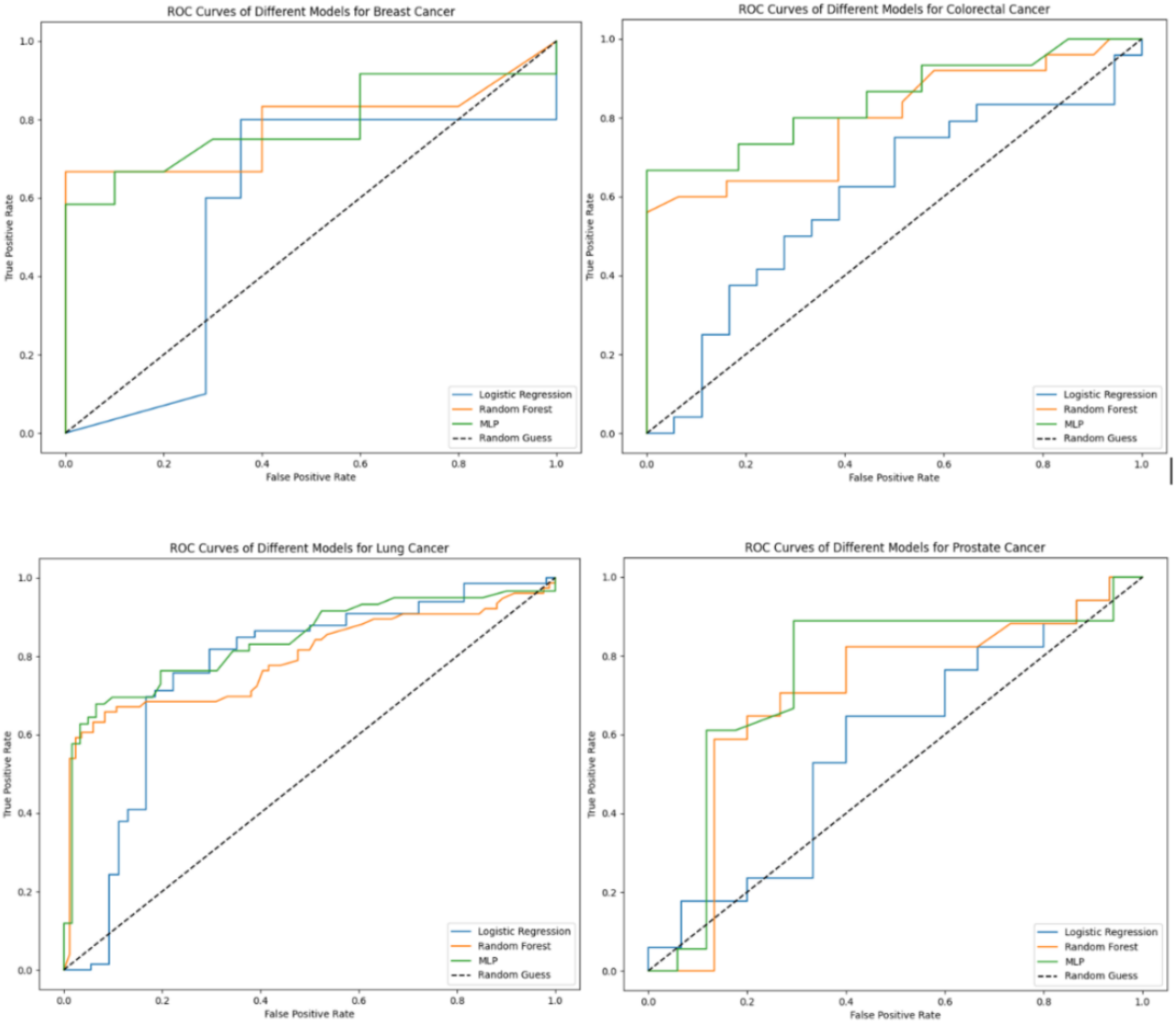
	Breast cancer			Colorectal cancer			Lung cancer			Prostate cancer		
	LR[a]	RF[b]	MLP[c]	LR	RF	MLP	LR	RF	MLP	LR	RF	MLP
Accuracy	0.56	0.73	0.78	0.60	0.70	0.76	0.74	0.80	0.83	0.59	0.72	0.78
Specificity	0.45	0.70	0.80	0.67	0.61	0.81	0.61	0.92	0.87	0.53	0.80	0.84
Sensitivity	0.71	0.75	0.75	0.54	0.80	0.73	0.85	0.68	0.80	0.65	0.65	0.72
F_1-score	0.56	0.75	0.75	0.60	0.70	0.79	0.78	0.78	0.84	0.63	0.71	0.76

[a]LR: logistic regression.

[b]RF: random forest.

[c]MLP: multilayer perceptron.

Figure 2. Area under the receiver operating characteristic curve (AUC) performance of the 3 binary classification models (logistic regression [LR], random forest [RF], and multilayer perceptron [MLP]). The figure shows AUC curves of breast cancer, colorectal cancer, lung cancer, and prostate cancer for LR, RF, and MLP, respectively.



Additionally, Table S2 in Multimedia Appendix 1 presents the AUC, accuracy, specificity, sensitivity, and F_1-score for the 3 models across breast, colorectal, lung, and prostate cancers. Among the models evaluated, MLP demonstrated the highest performance, achieving an AUC of 0.88 for breast cancer, 0.83

for colorectal cancer, 0.90 for lung cancer, and 0.85 for prostate cancer.

Feature Importance Analysis

We analyzed the feature importance for each cancer type further to investigate the potential impact of risk factors on cancer. Tables 4 and 5 present the feature importance analysis of RF and MLP, showcasing the top-ranked risk factors for each type of cancer. The ranks of these factors were relatively different by model and cancer type, although some consistency can be observed across cancer types. Age emerged as the top risk factor across all 4 types of cancer; race/ethnicity ranked among the top 10 factors for all cancers from all models except for the RF-based lung cancer and prostate cancer models; gender was ranked among the top 10 in MLP-based models but not in any RF-based models; marital status and religion were presented for some types of cancer in some of the models; and tobacco use as an important factor for patients with lung and prostate cancer exclusively. However, all these demographic risk factors were included in the top 20 factors for all cancer types (Table S3 in Multimedia Appendix 1). Similarly, RF-based models identified hypertension, heart diseases, respiratory/pulmonary diseases, and acute kidney failure as the common top risk factors for all types of cancers, while MLP-based models highlighted hyperlipidemia, diabetes, depressive disorder, and heart diseases. We calculated the odds ratio (OR) for each highlighted feature to assess its association with overall cancer diagnosis risk across 4 cancer types. The results indicated that hyperlipidemia had an OR of 1.14 (95% CI 1.11 - 1.17; P<.001), while diabetes showed a stronger association with an OR of 1.34 (95% CI 1.29 - 1.39; P<.01). Similarly, depressive disorders were linked to an OR of 1.11 (95% CI 1.06 - 1.16 P<.01), and heart diseases exhibited the highest association with an OR of 1.42 (95% CI 1.32 - 1.52; P<.01). Last, anemia was also significantly associated with cancer diagnosis risk, with an OR of 1.22 (95% CI 1.14 - 1.30; P<.01). These findings suggest a statistically significant relationship between these conditions and an increased risk of developing these 4 types of cancer. In MLP-based models, respiratory/pulmonary diseases and acute kidney failure were only presented as the top 10 for lung cancer. Both RF and MLP-based models pinpointed anemia as the top risk for breast cancer. Figure 3 shows the RBO similarity scores of risk factors for 4 types of cancer according to MLP-based models. Low similarity scores are presented between lung cancer and any other 3 cancer types, all around 0.58, suggesting distinct patterns of risk factors associated with lung cancer. Risk factors for breast and prostate cancers show the most similar ranking with an RBO similarity score of 0.76. A moderate similarity score between colorectal and breast cancers is about the same as the score between colorectal and prostate cancer, both around 0.70.

Table . Top-10 ranked features generated across 4 different cancer types in random forest.

Ranking	Breast cancer	Colorectal cancer	Lung cancer	Prostate cancer
1	Age	Age	Age	Age
2	Hypertension	Respiratory or pulmonary diseases[a]	Hypertension	Hypertension
3	Religion	Hypertension	Religion	Religion
4	Marital status	Acute kidney failure	Hyperlipidemia	Heart diseases[c]
5	Respiratory or pulmonary diseases	Diabetes	Heart diseases	Marital status
6	Heart diseases	Heart diseases	Acute kidney failure	UTI[c]
7	Race or ethnicity	Hyperlipidemia	UTI	Respiratory or pulmonary diseases
8	Depressive disorders	Race or ethnicity	Respiratory or pulmonary diseases	Anemia
9	Acute kidney failure	Religion	Marital status	Hyperthyroidism
10	Anemia	Acidosis	Anemia	Diabetes

[a]Respiratory or pulmonary diseases include pneumonia, acute respiratory failure, chronic airway obstruction, and other respiratory or pulmonary complications.

[b]Heart diseases include atrial fibrillation, myocardial infarction, congestive heart failure, coronary atherosclerosis, and other cardiac complications.

[c]UTI: urinary tract infection.

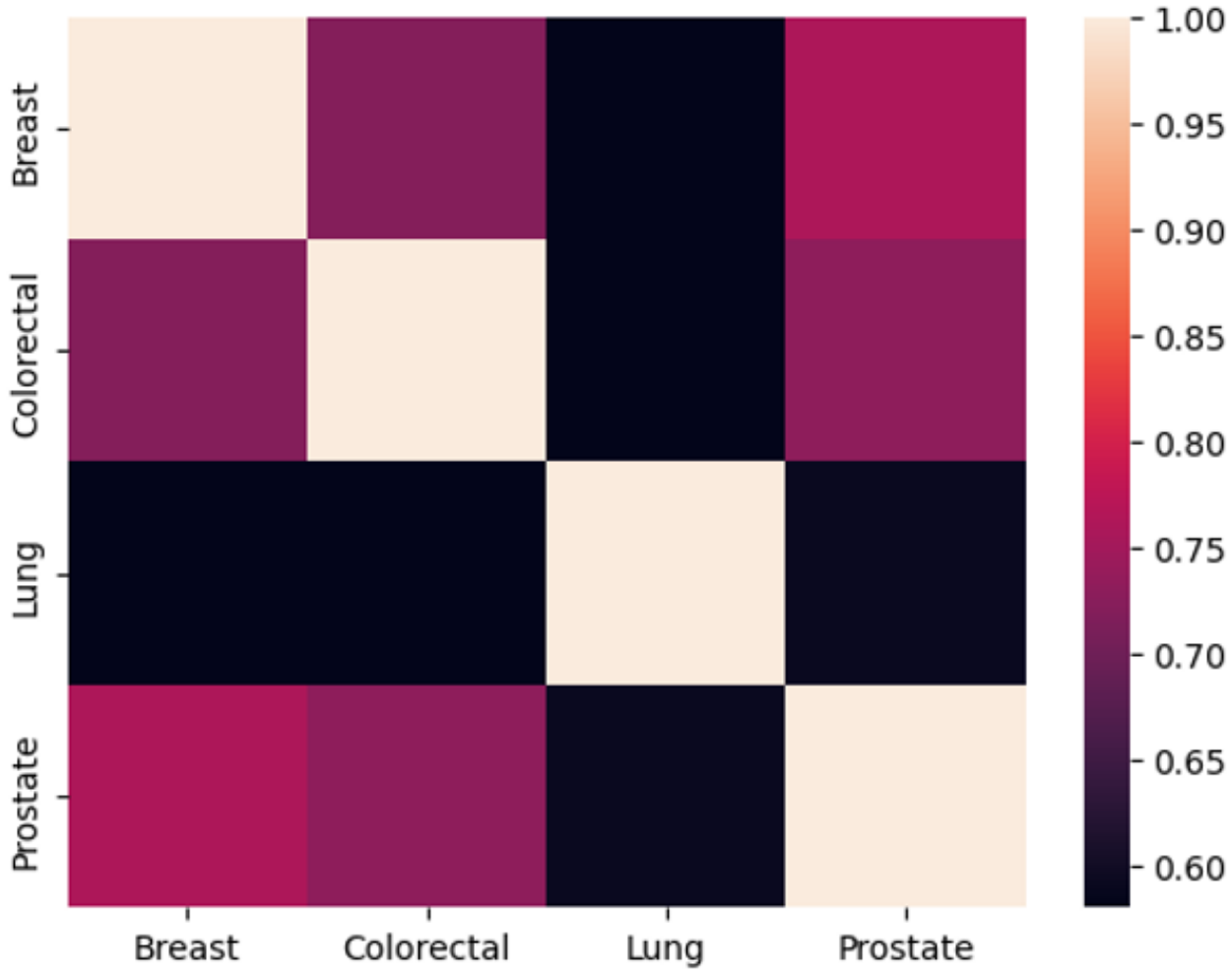
Table . Top-10 ranked features generated across 4 different cancer types in multilayer perceptron.

Ranking	Breast cancer	Colorectal cancer	Lung cancer	Prostate cancer
1	Age	Age	Tobacco use	Age
2	Gender	Diabetes	Age	Gender
3	Hyperlipidemia	Anemia	Respiratory or pulmonary diseases[a]	Race or ethnicity
4	Heart diseases[b]	Acidosis	Gender	Tobacco use
5	Race or ethnicity	Hyperlipidemia	Race or ethnicity	Diabetes
6	Marital status	Sepsis	Diabetes	Hyperlipidemia
7	Depressive disorder	Gender	Hyperlipidemia	Heart diseases
8	Religion	Race or ethnicity	Hypertension	Marital status
9	Anemia	Marital status	Heart diseases	Religion
10	Hypothyroidism	Depressive disorder	Acute kidney failure	Depressive disorder

[a]Respiratory or pulmonary diseases include pneumonia, acute respiratory failure, chronic airway obstruction, and other respiratory or pulmonary complications.

[b]Heart diseases include atrial fibrillation, myocardial infarction, congestive heart failure, coronary atherosclerosis, and other cardiac complications.

Figure 3. Rank biased overlap similarity score of risk factors for 4 cancer types. A high value represents high similarity, and a low value represents low similarity of risk factor ranks between 2 cancer types.



Discussion

Principal Findings

This study used comprehensive patient diagnosis histories to evaluate the association between key risk factors and cancer outcomes and identify risk factor patterns across different cancer types using penalized LR, RF, and MLP models. The analysis identified the top-ranking risk factors, including nontraditional risk factors such as the diagnosis of hyperlipidemia, diabetes, depressive disorders, heart diseases, and anemia, in addition to demographic factors such as age, sex, race/ethnicity, for the most prevalent 4 types of cancer, including breast, colorectal, lung, and prostate cancers. The model performance evaluation revealed the valuable potential of neural network-based models, especially MLPs, in oncology for predicting cancer diagnosis risks across cancer types. MLPs exhibit a strong capability to model complex, nonlinear interactions among diverse risk factors, making them potentially valuable tools for identifying patterns in cancer diagnosis risk and informing early detection strategies. However, their application in clinical interventions should be guided by a solid scientific rationale and supported by pathological models that explain the role of these risk factors in disease development. Additionally, validation across different cohorts and, ideally, prospective studies are necessary to ensure their reliability and clinical utility. This advantage is particularly important given the model's capacity to integrate and interpret the intricate relationships between clinical factors present in EHRs. In contrast to simpler models like LR, which struggle with the multidimensional nature of risk factors on cancer diagnosis in many cases, MLPs offer a more detailed and comprehensive analysis, enhancing our understanding of how these factors collectively impact cancer diagnosis risk and improving the precision of preventive strategies in clinical settings. Last, this study does not aim to establish causal inference but rather to examine significant overlapping risk factors that may contribute to cancer diagnosis risk, particularly those observed in patients with other medical conditions. While these diagnoses are not independent causal determinants of cancer, their presence may be associated with an increased risk. Careful consideration of these associations is essential for a comprehensive understanding of cancer risk factors and their potential interactions.

Comparison to Prior Work

Prior cancer risk prediction models usually focus on lifestyle factors like smoking, diet, alcohol consumption, physical activity, and sun exposure as key variables [44-46]. Some models have also incorporated genetic risk factors [47,48]. However, many of these models reported less optimal performance, such as a high specificity but low sensitivity [46] or a low AUC of around 0.65 [48]. Chronic diseases are often overlooked as risk factors for cancer, and they are not often targeted in cancer prevention strategies. The association between some of these diseases and cancers may partly be due to shared risk factors, such as aging, obesity, diet, and physical inactivity. However, they can also be independent risk factors for cancer. For example, diabetes mellitus has been identified as an independent risk factor for colon and rectal cancer in a meta-analysis of studies that either controlled for smoking and obesity, or smoking, obesity, and physical exercise [49]. As nontraditional risk factors, the influence of certain chronic conditions on cancer has been brought to researchers' attention in the past decade. A prospective cohort study with 405,878 participants followed for an average of 8.7 years demonstrated that 8 common chronic diseases accounted for more than 20% of cancer risk, which are comparable to 5 major lifestyle factors, such as smoking and lack of physical activity [16]. These 8 chronic diseases or markers included blood pressure, total cholesterol, heart rate, diabetes, proteinuria, glomerular filtration rate, pulmonary disease, and gouty arthritis marker [16]. However, as these diseases or markers were pre-selected by the researchers based on their disease burden worldwide, some other essential influential conditions might be missed. Our models confirmed most of these 8 diseases as the top-ranking risk factors. Additionally, some new conditions were revealed in our models among the top 10 factors for 4 types of cancer, such as depressive disorder, anemia, hypothyroidism, sepsis, urinary tract infection, and acidosis, which encourages further exploration. Certainly, some of these diagnoses may be directly related to the cancer itself. For example, anemia is a common symptom of metastatic breast cancer and a side effect of chemotherapy [50]. In addition, sepsis and colorectal cancer have demonstrated a complex relationship and may have shared pathophysiological traits and potential bacterial associations reported by the literature [51]. Notably, tobacco use and respiratory/pulmonary diseases emerged as pivotal risk factors, specifically for lung cancer, which is not surprising based on our knowledge in the field. Diabetes and anemia were highlighted as significant risk factors for colorectal cancer, which is congruent with the literature [52,53]. Iron deficiency has been recognized long-term as an independent predictor of colorectal cancer, which may be due to chronic blood loss from the gastrointestinal tract and the inflammation associated with malignancy [54,55]. These conditions could have shared risk factors with cancer. However, emerging evidence implies that they may have more complicated relationships, including shared pathophysiological mechanisms that need further exploration [56]. Moreover, cancer prevention strategies should consider the impact of comorbid conditions on the incidence of cancer and particularly their joint impact on cancer risks [53].

Understanding the relationships between various risk factors and cancer diagnosis risk is pivotal for the early detection and prevention of cancer. In this context, our feature importance analysis using RF and MLP models pinpointed critical risk factors for different cancer types and explored patterns of these risk factors across various cancers. Although the ranks of risk factors for cancers were slightly different by the RF and MLP-based models, similar patterns were presented among the top 10 factors (Tables 4 and 5), which are interpretable and supported by the literature. Both models highlighted age as the predominant risk factor across all 4 types of cancer, which is evident that as age increases, the incidence rates for cancer overall climb steadily, and alongside age, demographic variables such as gender, race/ethnicity, marital status, and religion emerged within the top 10 features [57]. Racial/ethnic disparities in cancer incidence and outcomes are well-known. Employing culturally tailored community awareness and education programs may increase cancer screening to improve early-stage

diagnoses and modify risk behaviors for cancer prevention [58]. Although there may not be existing evidence to confirm that marital status is an independent risk factor for cancer, observational studies demonstrate that married status is associated with reduced risk of cancer-specific and all-cause mortality [59,60]. Religion and spirituality are important in patient cancer care, and specifically, a systematic review suggests a positive association between religious attendance and cancer screening use [61]. Our models not only confirmed the significance of these risk factors for each cancer type but also our RF-based model facilitated an interpretable analysis, allowing us to clearly rank the significance of each risk factor, while the MLP-based model provided deeper insights into complex, nonlinear interactions among the risk factors. This approach enriches our understanding of how specific risk factors influence cancer diagnosis, enhancing the potential for developing tailored intervention strategies that address the unique risk profiles associated with different cancer types and potentially shared risk patterns across prevalent cancer types.

The analysis of the similarity among risk factors for the diagnosis of 4 types of cancer also revealed interesting findings. As breast and prostate cancer are both hormone-dependent cancers, it is understandable that their importance-ranked risk factors share a high level of similarity. However, lung cancer had more unique ranked risk factors than other types of cancer, which may be because lung cancer is more sensitive to environmental risk factor exposure. The findings from our analysis underscore the shared risk factors and heterogeneous nature of cancer and highlight the importance of considering unique risk profiles for different cancer types. This also urges us to address the fundamental mechanism of risk factors leading to cancers. Such insights are crucial for developing tailored prevention strategies, optimizing screening protocols, and informing personalized treatment approaches to mitigate the burden of lung cancer and improve patient outcomes.

Limitations

First, the use of the MIMIC-III dataset in this study on explainable machine learning for cancer risk prediction presents certain limitations that may impact the generalizability of the findings, Since the data are derived from ICU patient records, it primarily represents individuals with severe conditions, and the available *ICD* codes may not fully capture disease complexity, potentially leading to incomplete representations of patient conditions. Additionally, the limited sample size for patients with cancer may impact predictive performance and increase the risk of overfitting. Both limitations may affect the generalizability of the findings. To enhance the robustness of future research, integrating more recent and varied data sources and validating findings across different cohorts are essential steps. Second, one limitation comes from the application of explainable machine learning models for cancer risk prediction. Employing advanced techniques like penalized LR, RF, and MLP, this research seeks to optimize predictive accuracy. However, each model inherently embodies trade-offs: while more complex models, such as multi-layer perceptron, may enhance performance, they often compromise on interpretability. This presents significant challenges in clinical settings, where understanding the reasoning behind model predictions is crucial for acceptance and trust by medical practitioners. Third, another limitation of this study arises from the inherent nature of machine learning models, which are primarily designed to detect correlations in data and associations between features and the outcome rather than establish causal relationships. These models rely on the quality and comprehensiveness of the input data, and while they can reveal significant associative patterns, they do not focus on differentiating whether the associations observed are causal. Meanwhile, given the limited availability of patient lifestyle and socioeconomic information, additional factors related to social determinants of health, such as socioeconomic status, employment, and family size, can be considered as potential confounders within the model for future improvement. To address all the above, future work should integrate causal inference frameworks to validate the relationships suggested by the machine learning predictions and provide insights into underlying mechanisms.

Conclusions

In conclusion, our study established a predictive framework using EHR data to assess the association between risk factors and cancer outcomes using explainable ML models across major cancer types. We reported critical nontraditional chronic condition risk factors in addition to common demographic risk factors and outlined distinct patterns for each of the 4 cancer types studied. Additionally, we explored the similarities and differences in risk factor patterns across these cancers. These insights contribute to a better understanding of cancer risk profiles and benefit in improving cancer diagnosis and risk monitoring, offering supportive guidance for clinical decision-making.

Acknowledgments

The authors would like to thank the Beth Israel Deaconess Medical Center and the MIT Laboratory for Computational Physiology for making the Medical Information Mart for Intensive Care (MIMIC)–III database available to the research community. The authors would also acknowledge the support from the University of Michigan School of Nursing and the Yale School of Public Health throughout this study.

Data Availability

The datasets analyzed during this study are available on PhysioNet [62].

Authors' Contributions

XH and YJ conceived the study. XH and SR implemented the algorithm, conducted the experiments, and performed all the analyses. SR generated results visualization. XH and YJ supervised the study. XH, SR, EC, YH, and YJ wrote the manuscript. All authors provided feedback and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables on the *ICD-9* codes description, model performance, and top 20 ranked features for the 4 cancer types
[DOCX File, 32 KB - cancer_v11i1e62833_app1.docx]

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74(1):12-49. [doi: 10.3322/caac.21820] [Medline: 38230766]
2. Colorectal cancer: screening. The US Preventive Services Task Force. 2021. URL: https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/colorectal-cancer-screening [accessed 2025-04-30]
3. Breast cancer: screening. The US Preventive Services Task Force. 2024. URL: https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/breast-cancer-screening#:~:text=What%20is%20the%20Task%20Force,This%20is%20a%20B%20grade [accessed 2025-04-30]
4. Mansour R, Al-Ani A, Al-Hussaini M, Abdel-Razeq H, Al-Ibraheem A, Mansour AH. Modifiable risk factors for cancer in the middle East and North Africa: a scoping review. BMC Public Health 2024 Jan 18;24(1):223. [doi: 10.1186/s12889-024-17787-5] [Medline: 38238708]
5. Fitzgerald RC, Antoniou AC, Fruk L, Rosenfeld N. The future of early cancer detection. Nat Med 2022 Apr;28(4):666-677. [doi: 10.1038/s41591-022-01746-x] [Medline: 35440720]
6. Byskov Petersen G, Sadolin Damhus C, Ryborg Jønsson AB, Brodersen J. The perception gap: how the benefits and harms of cervical cancer screening are understood in information material focusing on informed choice. Health Risk Soc 2020 Feb 17;22(2):177-196. [doi: 10.1080/13698575.2020.1778645]
7. Lee CI, Elmore JG. Beyond survival: a closer look at lead-time bias and disease-free intervals in mammography screening. J Natl Cancer Inst 2024 Mar 7;116(3):343-344. [doi: 10.1093/jnci/djad254] [Medline: 38145456]
8. Duffy SW, Tabár L, Yen AMF, et al. Mammography screening reduces rates of advanced and fatal breast cancers: results in 549,091 women. Cancer 2020 Jul 1;126(13):2971-2979. [doi: 10.1002/cncr.32859] [Medline: 32390151]
9. Cintolo-Gonzalez JA, Braun D, Blackford AL, et al. Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications. Breast Cancer Res Treat 2017 Jul;164(2):263-284. [doi: 10.1007/s10549-017-4247-z] [Medline: 28444533]
10. Gao Y, Li S, Jin Y, et al. An assessment of the predictive performance of current machine learning-based breast cancer risk prediction models: systematic review. JMIR Public Health Surveill 2022 Dec 29;8(12):e35750. [doi: 10.2196/35750] [Medline: 36426919]
11. Nayan M, Salari K, Bozzo A, et al. A machine learning approach to predict progression on active surveillance for prostate cancer. Urol Oncol 2022 Apr;40(4):161. [doi: 10.1016/j.urolonc.2021.08.007] [Medline: 34465541]
12. Huang T, Le D, Yuan L, Xu S, Peng X. Machine learning for prediction of in-hospital mortality in lung cancer patients admitted to intensive care unit. PLoS One 2023;18(1):e0280606. [doi: 10.1371/journal.pone.0280606] [Medline: 36701342]
13. Leonard G, South C, Balentine C, et al. Machine learning improves prediction over logistic regression on resected colon cancer patients. J Surg Res 2022 Jul;275:181-193. [doi: 10.1016/j.jss.2022.01.012] [Medline: 35287027]
14. Guan Z, Huang T, McCarthy AM, et al. Combining breast cancer risk prediction models. Cancers (Basel) 2023 Feb 8;15(4):1090. [doi: 10.3390/cancers15041090] [Medline: 36831433]
15. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. J Thorac Dis 2019 Mar;11(Suppl 4):S574-S584. [doi: 10.21037/jtd.2019.01.25] [Medline: 31032076]
16. Tu H, Wen CP, Tsai SP, et al. Cancer risk associated with chronic diseases and disease markers: prospective cohort study. BMJ 2018 Jan 31;360:k134. [doi: 10.1136/bmj.k134] [Medline: 29386192]
17. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform 2018 Sep;22(5):1589-1604. [doi: 10.1109/JBHI.2017.2767063]
18. Wang R, Li Z, Liu S, Zhang D. Global, regional and national burden of inflammatory bowel disease in 204 countries and territories from 1990 to 2019: a systematic analysis based on the Global Burden of Disease Study 2019. BMJ Open 2023 Mar 28;13(3):e065186. [doi: 10.1136/bmjopen-2022-065186] [Medline: 36977543]

19. Steinberg J, Yap S, Goldsbury D, et al. Large-scale systematic analysis of exposure to multiple cancer risk factors and the associations between exposure patterns and cancer incidence. Sci Rep 2021 Jan 27;11(1):2343. [doi: 10.1038/s41598-021-81463-6] [Medline: 33504831]
20. Belle V, Papantonis I. Principles and practice of explainable machine learning. Front Big Data 2021;4:688969. [doi: 10.3389/fdata.2021.688969] [Medline: 34278297]
21. Gurmessa DK, Jimma W. Explainable machine learning for breast cancer diagnosis from mammography and ultrasound images: a systematic review. BMJ Health Care Inform 2024 Feb 2;31(1):e100954. [doi: 10.1136/bmjhci-2023-100954] [Medline: 38307616]
22. Shulha M, Hovdebo J, D'Souza V, Thibault F, Harmouche R. Integrating explainable machine learning in clinical decision support systems: study involving a modified design thinking approach. JMIR Form Res 2024 Apr 16;8:e50475. [doi: 10.2196/50475] [Medline: 38625728]
23. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Sci Rep 2021 Mar 26;11(1):6968. [doi: 10.1038/s41598-021-86327-7] [Medline: 33772109]
24. Huang P, Kong Z, Wang L, Han X, Yang X. Efficient and stable unsupervised feature selection based on novel structured graph and data discrepancy learning. IEEE Trans Neural Netw Learning Syst 2024;36(4):6229-6243. [doi: 10.1109/TNNLS.2024.3385838]
25. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035. [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
26. Hypertension - health, United States. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/nchs/hus/topics/hypertension.htm [accessed 2025-02-04]
27. FastStats - hypertension. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/nchs/fastats/hypertension.htm [accessed 2024-02-04]
28. National Diabetes Statistics Report. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/diabetes/php/data-research/index.html [accessed 2025-02-04]
29. High cholesterol facts. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/cholesterol/data-research/facts-stats/index.html [accessed 2025-02-04]
30. QuickStats: prevalence of high total cholesterol* among adults aged ≥20 years,† by age group and sex - national health and nutrition examination survey, 2015-2018. MMWR Morb Mortal Wkly Rep 2020 Jun 5;69(22):690. [doi: 10.15585/mmwr.mm6922a5] [Medline: 32497032]
31. Bozkurt B, Ahmad T, Alexander KM, et al. Heart failure epidemiology and outcomes statistics: a report of the heart failure society of America. J Card Fail 2023 Oct;29(10):1412-1451. [doi: 10.1016/j.cardfail.2023.07.006] [Medline: 37797885]
32. Antunes C, Aleem A, Curtis SA. Gastroesophageal reflux disease. In: StatPearls: StatPearls Publishing; 2023.
33. FastStats - pneumonia. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/nchs/fastats/pneumonia.htm [accessed 2025-02-04]
34. Moawad H. Anemia facts and statistics: what you need to know. Verywell Health. 2022. URL: https://www.verywellhealth.com/facts-about-anemia-6503636#:~:text=According%20to%20one%20study%2C%205.6%25%20of%20the%20U.S.,and%20people%20of%20any%20age%20who%20are%20undernourished [accessed 2025-02-04]
35. Wyne KL, Nair L, Schneiderman CP, et al. Hypothyroidism prevalence in the United States: a retrospective study combining national health and nutrition examination survey and claims data, 2009-2019. J Endocr Soc 2022 Nov 17;7(1):bvac172. [doi: 10.1210/jendso/bvac172] [Medline: 36466005]
36. Current cigarette smoking among adults in the United States. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/tobacco/php/data-statistics/adult-data-cigarettes/index.html [accessed 2025-02-04]
37. Mental health disorder statistics. The Johns Hopkins University. URL: https://www.hopkinsmedicine.org/health/wellness-and-prevention/mental-health-disorder-statistics [accessed 2025-02-04]
38. Liu Y, Carlson SA, Watson KB, Xu F, Greenlund KJ. Trends in the prevalence of chronic obstructive pulmonary disease among adults aged ≥18 years - United States, 2011-2021. MMWR Morb Mortal Wkly Rep 2023 Nov 17;72(46):1250-1256. [doi: 10.15585/mmwr.mm7246a1] [Medline: 37971940]
39. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825-2830.
40. Sarica A, Quattrone A. Introducing the rank-biased overlap as similarity measure for feature importance in explainable machine learning: a case study on parkinson's disease. Presented at: 15th International Conference on Brain Informatics; Jul 15-17, 2022; Padua, Italy. [doi: 10.1007/978-3-031-15037-1_11]
41. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
42. ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 1958 Nov;65(6):386-408. [doi: 10.1037/h0042519] [Medline: 13602029]
43. Nirmalraj S, Antony ASM, Srideviponmalar P, et al. Permutation feature importance-based fusion techniques for diabetes prediction. Soft Comput 2023. [doi: 10.1007/s00500-023-08041-y]

44. Mertens E, Barrenechea-Pulache A, Sagastume D, Vasquez MS, Vandevijvere S, Peñalvo JL. Understanding the contribution of lifestyle in breast cancer risk prediction: a systematic review of models applicable to Europe. BMC Cancer 2023 Jul 21;23(1):687. [doi: 10.1186/s12885-023-11174-w] [Medline: 37480028]
45. Usher-Smith JA, Sharp SJ, Luben R, Griffin SJ. Development and validation of lifestyle-based models to predict incidence of the most common potentially preventable cancers. Cancer Epidemiol Biomarkers Prev 2019 Jan 1;28(1):67-75. [doi: 10.1158/1055-9965.EPI-18-0400]
46. Harrison H, Thompson RE, Lin Z, et al. Risk prediction models for kidney cancer: a systematic review. Eur Urol Focus 2021 Nov;7(6):1380-1390. [doi: 10.1016/j.euf.2020.06.024] [Medline: 32680829]
47. Carver T, Hartley S, Lee A, et al. CanRisk tool-a web interface for the prediction of breast and ovarian cancer risk and the likelihood of carrying genetic pathogenic variants. Cancer Epidemiol Biomarkers Prev 2021 Mar;30(3):469-473. [doi: 10.1158/1055-9965.EPI-20-1319] [Medline: 33335023]
48. Thomas M, Sakoda LC, Hoffmeister M, et al. Genome-wide modeling of polygenic risk score in colorectal cancer risk. Am J Hum Genet 2020 Sep 3;107(3):432-444. [doi: 10.1016/j.ajhg.2020.07.006] [Medline: 32758450]
49. Yuhara H, Steinmaus C, Cohen SE, Corley DA, Tei Y, Buffler PA. Is diabetes mellitus an independent risk factor for colon cancer and rectal cancer? Am J Gastroenterol 2011 Nov;106(11):1911-1921. [doi: 10.1038/ajg.2011.301] [Medline: 21912438]
50. Muthanna FMS, Karuppannan M, Abdulrahman E, Uitrakul S, Rasool BAH, Mohammed AH. Prevalence and associated factors of anemia among breast cancer patients undergoing chemotherapy: a prospective study. Adv Pharmacol Pharm Sci 2022;2022:7611733. [doi: 10.1155/2022/7611733] [Medline: 35464620]
51. Masood L, Müller A, Ali NZ, et al. A narrative literature review on sepsis: a primary manifestation of colorectal neoplasm. Cureus 2023 Sep;15(9):e44803. [doi: 10.7759/cureus.44803] [Medline: 37809261]
52. Stan MC, Georgescu D, Mire tean CC, Bădulescu F. Cancer and diabetes: predictive factors in patients with metabolic syndrome. Diagnostics (Basel) 2023 Aug 11;13(16):2647. [doi: 10.3390/diagnostics13162647] [Medline: 37627906]
53. Soltani G, Poursheikhani A, Yassi M, Hayatbakhsh A, Kerachian M, Kerachian MA. Obesity, diabetes and the risk of colorectal adenoma and cancer. BMC Endocr Disord 2019 Oct 29;19(1):113. [doi: 10.1186/s12902-019-0444-6] [Medline: 31664994]
54. Chardalias L, Papaconstantinou I, Gklavas A, Politou M, Theodosopoulos T. Iron deficiency anemia in colorectal cancer patients: is preoperative intravenous iron infusion indicated? a narrative review of the literature. Cancer Diagn Progn 2023;3(2):163-168. [doi: 10.21873/cdp.10196] [Medline: 36875314]
55. Hamilton W, Lancashire R, Sharp D, Peters TJ, Cheng KK, Marshall T. The importance of anaemia in diagnosing colorectal cancer: a case-control study using electronic primary care records. Br J Cancer 2008 Jan 29;98(2):323-327. [doi: 10.1038/sj.bjc.6604165] [Medline: 18219289]
56. de Boer RA, Meijers WC, van der Meer P, van Veldhuisen DJ. Cancer and heart disease: associations and relations. Eur J Heart Fail 2019 Dec;21(12):1515-1525. [doi: 10.1002/ejhf.1539] [Medline: 31321851]
57. Age and cancer risk. National Cancer Institute. 2021. URL: https://www.cancer.gov/about-cancer/causes-prevention/risk/age [accessed 2025-04-30]
58. Zavala VA, Bracci PM, Carethers JM, et al. Cancer health disparities in racial/ethnic minorities in the United States. Br J Cancer 2021 Jan;124(2):315-332. [doi: 10.1038/s41416-020-01038-6] [Medline: 32901135]
59. Zhu S, Lei C. Association between marital status and all-cause mortality of patients with metastatic breast cancer: a population-based study. Sci Rep 2023 Jun 5;13(1):9067. [doi: 10.1038/s41598-023-36139-8] [Medline: 37277464]
60. Chen ZH, Yang KB, Zhang YZ, et al. Assessment of modifiable factors for the association of marital status with cancer-specific survival. JAMA Netw Open 2021 May 3;4(5):e2111813. [doi: 10.1001/jamanetworkopen.2021.11813] [Medline: 34047792]
61. Kretzler B, König HH, Brandt L, Weiss HR, Hajek A. Religious denomination, religiosity, religious attendance, and cancer prevention. a systematic review. Risk Manag Healthc Policy 2022;15:45-58. [doi: 10.2147/RMHP.S341085] [Medline: 35079226]
62. PhysioNet. URL: https://physionet.org/ [accessed 2025-04-28]

Abbreviations

AUC: area under the receiver operating characteristic curve
BIDMC: Beth Israel Deaconess Medical Center
EHR: electronic health record
***ICD-9*:** *International Classification of Diseases, Ninth Revision*
ICU: intensive care unit
LR: logistic regression
MIMIC: Medical Information Mart for Intensive Care
MLP: multilayer perceptron
OR: odds ratio

RBO: rank biased overlap
RF: random forest

Edited by N Cahill; submitted 02.06.24; peer-reviewed by D Bracken-Clarke, R Benson; revised version received 08.03.25; accepted 20.03.25; published 02.05.25.

Please cite as:
Huang X, Ren S, Mao X, Chen S, Chen E, He Y, Jiang Y
Association Between Risk Factors and Major Cancers: Explainable Machine Learning Approach
JMIR Cancer 2025;11:e62833
URL: https://cancer.jmir.org/2025/1/e62833
doi:10.2196/62833

Original Paper

AI-Based Identification Method for Cervical Transformation Zone Within Digital Colposcopy: Development and Multicenter Validation Study

Tong Wu[1], MS; Yuting Wang[1], BM; Xiaoli Cui[2], MM; Peng Xue[1], MPH, PhD; Youlin Qiao[1], MD, PhD

[1]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[2]Liaoning Cancer Hospital and Institute, Department of Gynecologic Oncology, Cancer Hospital of China Medical University, Shenyang, China

Corresponding Author:
Youlin Qiao, MD, PhD
School of Population Medicine and Public Health
Chinese Academy of Medical Sciences and Peking Union Medical College
31 Yard, Beijige Santiao
Beijing, 100730
China
Phone: 86 10 8778 8489
Email: qiaoy@cicams.ac.cn

Abstract

Background: In low- and middle-income countries, cervical cancer remains a leading cause of death and morbidity for women. Early detection and treatment of precancerous lesions are critical in cervical cancer prevention, and colposcopy is a primary diagnostic tool for identifying cervical lesions and guiding biopsies. The transformation zone (TZ) is where a stratified squamous epithelium develops from the metaplasia of simple columnar epithelium and is the most common site of precancerous lesions. However, inexperienced colposcopists may find it challenging to accurately identify the type and location of the TZ during a colposcopy examination.

Objective: This study aims to present an artificial intelligence (AI) method for identifying the TZ to enhance colposcopy examination and evaluate its potential clinical application.

Methods: The study retrospectively collected data from 3616 women who underwent colposcopy at 6 tertiary hospitals in China between 2019 and 2021. A dataset from 4 hospitals was collected for model conduction. An independent dataset was collected from the other 2 geographic hospitals to validate model performance. There is no overlap between the training and validation datasets. Anonymized digital records, including each colposcopy image, baseline clinical characteristics, colposcopic findings, and pathological outcomes, were collected. The classification model was proposed as a lightweight neural network with multiscale feature enhancement capabilities and designed to classify the 3 types of TZ. The pretrained FastSAM model was first implemented to identify the location of the new squamocolumnar junction for segmenting the TZ. Overall accuracy, average precision, and recall were evaluated for the classification and segmentation models. The classification performance on the external validation was assessed by sensitivity and specificity.

Results: The optimal TZ classification model performed with 83.97% classification accuracy on the test set, which achieved average precision of 91.84%, 89.06%, and 95.62% for types 1, 2, and 3, respectively. The recall and mean average precision of the TZ segmentation model were 0.78 and 0.75, respectively. The proposed model demonstrated outstanding performance in predicting 3 types of the TZ, achieving the sensitivity with 95% CIs for TZ1, TZ2, and TZ3 of 0.78 (0.74-0.81), 0.81 (0.78-0.82), and 0.8 (0.74-0.87), respectively, with specificity with 95% CIs of 0.94 (0.92-0.96), 0.83 (0.81-0.86), and 0.91 (0.89-0.92), based on a comprehensive external dataset of 1335 cases from 2 of the 6 hospitals.

Conclusions: Our proposed AI-based identification system classified the type of cervical TZs and delineated their location on multicenter, colposcopic, high-resolution images. The findings of this study have shown its potential to predict TZ types and specific regions accurately. It was developed as a valuable assistant to encourage precise colposcopic examination in clinical practice.

(JMIR Cancer 2025;11:e69672) doi:10.2196/69672

KEYWORDS

artificial intelligence; AI; cervical cancer screening; transformation zone; diagnosis and early treatment; lightweight neural network

Introduction

Cervical cancer remains the fourth most prevalent cancer among women worldwide [1], and it continues to be a leading cause of morbidity and mortality threatening women's health. Since cervical precancerous lesions often progress to invasive cancer over an extended period, early detection is critical for cervical cancer prevention. Colposcopy serves as a crucial component of cervical cancer screening, providing a preliminary diagnosis for patients based on screening results, which then guides subsequent biopsy and treatment. Although it is a fundamental technique that health care providers can easily teach and implement, the strong subjective nature of colposcopy diagnosis makes it difficult for colposcopists with different qualifications to perform standardized diagnoses and make effective clinical decisions [2,3]. Artificial intelligence (AI) diagnostic technology could resolve the disparities in expertise among clinicians and enhance screening efficiency [4].

The transformation zone (TZ) is where a stratified squamous epithelium develops from the metaplasia of simple columnar epithelium and is the most common site of precancerous lesions. More than 90% of cervical cancers develop within the TZ [5], making it a critical region for cervical intraepithelial neoplasia (CIN) diagnosis and early treatment. According to the visibility of the squamocolumnar junction (SCJ), the TZ can be classified into three types: TZ1 (SCJ fully visible), TZ2 (SCJ fully visible under endocervical speculum), and TZ3 (SCJ partially visible or not visible) [6]. Accurately identifying the TZ is crucial for diagnosing and treating cervical precancerous lesions. As the TZ moves into the cervical canal with increasing age, endocervical curettage (ECC) is necessary for biopsy-guided pathology [7]. If TZ types are not classified, the importance of ECC for the canal may be neglected, leading to missed diagnosis of lesions during colposcopic examination. In addition, excision of the entire TZ is a standard treatment for cervical precancerous lesions. For excisional treatment, TZ types determine the length and depth of the cervix to be excised. In destructive treatments, a prerequisite is that the TZ must be either type 1 or type 2. Therefore, the type and location of the TZ are the determinants of treatment choices, and accurately assessing the TZ is essential for guiding more effective biopsies and precise treatment.

However, in underserved population, the skills of colposcopists are generally suboptimal, with colposcopic finding accuracy being significantly lower than desired [8,9]. AI-assisted technology could effectively enhance the competencies of colposcopists in these underserved areas. Current evaluation studies [10,11] have demonstrated that junior or less-experienced colposcopists can detect abnormal cervical lesions more effectively with AI assistance, which indicates its potential to help reduce missed diagnoses. However, the functions of AI cannot be limited. All colposcopic features were the indicators for assessing colposcopist performance. Among them, the accurate identification of TZ types is essential for implementing effective colposcopy diagnostics and treatment procedures, potentially reducing the number of missed diagnoses and unnecessary biopsy procedures. However, current AI-assisted colposcopy systems or developed AI diagnosis models do not include TZ or SCJ detection during model conduction. Therefore, it is essential to integrate important clinical features into AI model training to improve colposcopy diagnosis efficiency.

AI colposcopy diagnostic systems remain challenging to distinguish among benign, CIN1, CIN2, and CIN3+ cases during colposcopy examination [12-14], although they achieved over 80% accuracy in detecting high-grade squamous intraepithelial lesions. This difficulty is attributable to the lack of specificity in the CIN-related acetowhite staining characteristics. Typically, normal cervical features, such as immature squamous metaplasia, congenital TZ, inflammation, and epithelium regeneration, may exhibit mild acetowhite reactions similar to those associated with CIN. This similarity implies that relying solely on acetowhite area features can easily lead to misclassification either for AI or in less experienced colposcopist. In AI model training, standardized annotated images of different acetowhite morphologies or fine-grained lesion descriptions may help with a more precise assessment of acetowhite characteristics [15]. However, current AI-assisted colposcopy research is constrained by the lack of standardized annotated colposcopy images. From a clinical perspective, the colposcopy guidelines issued by the International Federation for Cervical Pathology and Colposcopy (IFCPC) emphasize that CIN-related acetowhite changes are most commonly found in the TZ, and near the new SCJ, with clear demarcation from the surrounding epithelium [6]. Therefore, the TZ region can be used as an indicator to identify lesion areas and can be developed as a learned feature for diagnostic model development, thereby resolving the problem of the lack of annotated colposcopy images. As a result, the multiclassification accuracy of AI-guided colposcopic diagnostic systems may be significantly improved. Therefore, in AI model development, accurately identifying TZ types and the SCJ is a crucial step toward improving AI diagnostic accuracy and guiding biopsies.

In this study, an AI method is developed and validated for the classification and delineation of the TZ. This method not only has the potential to guide clinical colposcopic examinations in resource-limited health care settings but also encourages the advancement of AI-guided digital colposcopic systems. By incorporating all colposcopic findings, such as TZ type and features of both minor and major lesions, AI-guided digital colposcopy could become a mature assistant for universal clinical colposcopy examinations.

Methods

Study Patients

This retrospective study included 3616 women who underwent colposcopy examinations across 6 multicenter hospitals in China between January 2019 and October 2021. These hospitals were

the Gansu Maternity and Child Healthcare Hospital, Second Hospital of Shanxi Medical University, Shenzhen Maternity and Child Healthcare Hospital, Jiangxi Maternity and Child Healthcare Hospital, The Affiliated Hospital of Qingdao University, and Chengdu Women's and Children's Central Hospital. The digital clinical records, including cytology, human papillomavirus infection status, colposcopy findings, and pathological results, were collected. All colposcopy images were captured using digital high-definition video colposcopes (Zonsun Healthcare Co Ltd, Edan Instruments). Colposcopy findings, including adequacy, SCJ visibility, TZ determination, and provisional diagnosis, were qualified by colposcopy experts from tertiary hospitals. The "ground truth" for TZ classification and SCJ visibility in this study was determined by an expert panel following IFCPC guidelines.

The inclusion criteria were as follows: (1) women with complete colposcopy findings and pathological diagnosis, (2) ages between 24 and 65 years, and (3) each record containing at least 5 satisfactory colposcopic images before and after acetic acid staining. The exclusion criteria were (1) all saline solution images (preacetic), since the TZ typically appears after acetic acid staining; (2) poor-quality images, such as overexposed images or those where the cervix was obscured or there was bleeding after the biopsy; (3) inadequacy colposcopic examination; and (4) records with missing TZ types or SCJ visibility. For external validation, an independent dataset was derived from The Affiliated Hospital of Qingdao University and Chengdu Women's and Children's Central Hospital. The training dataset was divided into training and test sets in an 8:2 ratio, with 10% of the training set used for validation during model tuning.

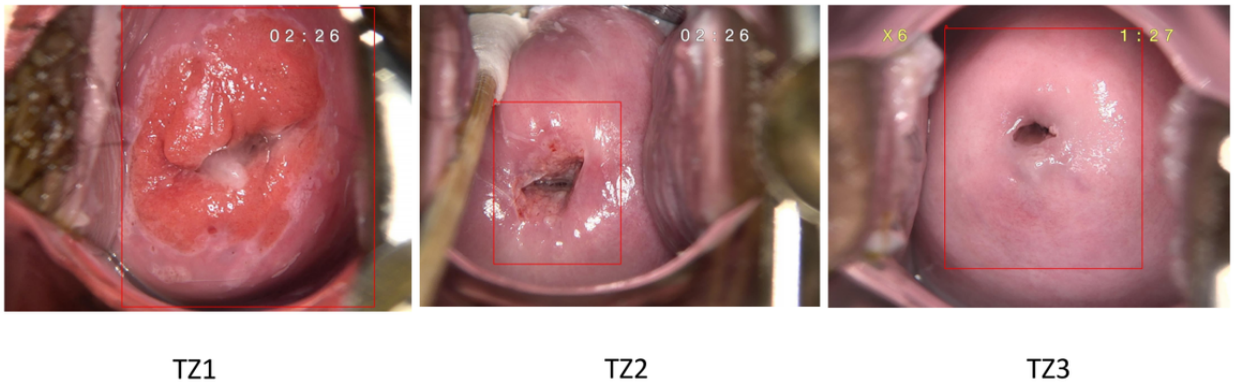
Transformation Zone

The colposcopy procedures adhered to the standard guidelines for *Colposcopy and Treatment of Cervical Precancer* [6]. Time-series images were captured for each case, including 1 original cervix image (saline solution) and at least 4 acetic acid–stained cervix images. The classification of TZ followed guideline criteria. TZ was classified as type 1 when it was entirely located on the ectocervix. In type 2, TZ was partially located within the endocervical canal, but its upper limit could be visualized using auxiliary instruments. Type 3 TZ lies partly or entirely inside the endocervical canal, with its upper limit being partially or completely invisible, even with the aid of auxiliary instruments.

Image Preprocessing

The colposcopy images were captured in high resolution. However, they contained irrelevant objects, such as the endocervical speculum, cotton swabs, or large regions of the vaginal wall, which could interfere with the classification model's ability to extract critical features of the cervix. To address this issue, we used the YOLOv5 network to detect the region of interest. Its single forward pass design enables real-time object detection with high efficiency and precision [16,17]. The integrated cervical region was automatically segmented and examined by a specialist (Figure 1). It divides an image into an S×S grid cell, predicting the bounding box locations and their associated categories. All segmented colposcopy images were resized to 224×224 pixels to align with the input specifications of the model.

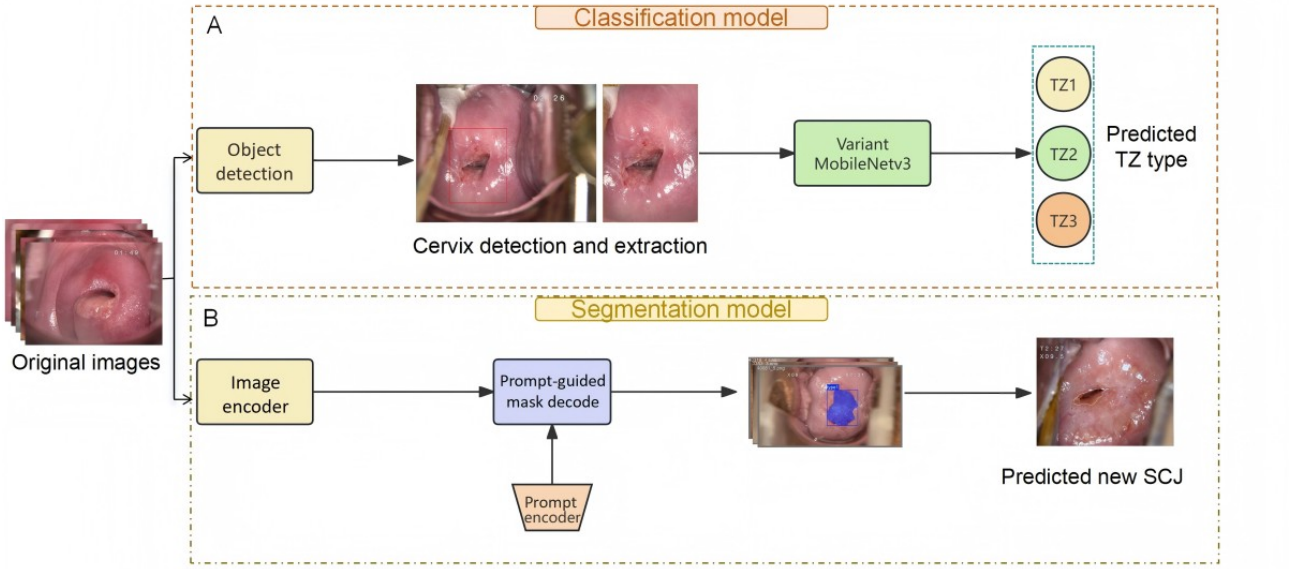
Figure 1. Cervix region of interest detection with bounding box examples in representative images. TZ: transformation zone.



Development of the TZ Classification Model

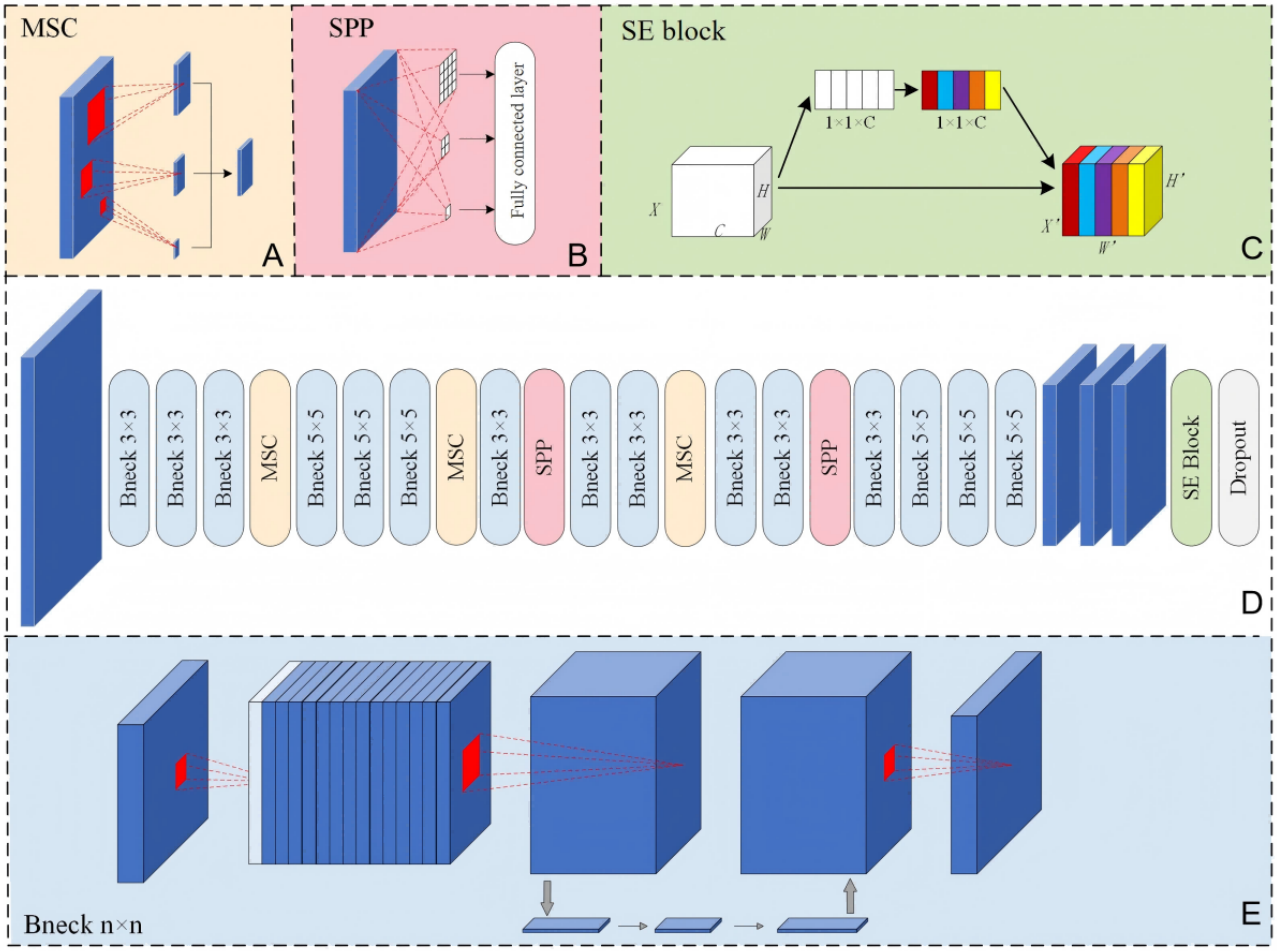
Initially, colposcopy images were input into a detection model to determine the cervical region, which was then preprocessed to enhance feature extraction. These data were then applied to a classification model to determine the types of TZ present (Figure 2A). In the second part of the method, the original images were annotated with new SCJ prompts to guide a general segmentation model in inferring the potential location of the TZ (Figure 2B).

Figure 2. The overall inference process of the TZ identification system consists of two stages: (A) detection and extraction of the entire cervical region from original images, followed by feature extraction using the variant MobileNetV3 and inference TZ1, TZ2, and TZ3 types; and (B) using the original images and images with polygon of the new SCJ as prompts to FastSAM, which then outputs the mask prediction of the out-of-TZ area to infer the new SCJ guidance line. SCJ: squamocolumnar junction; TZ: transformation zone.



We proposed a variant of the MobileNetV3 architecture for classifying TZ, a lightweight convolutional neural network (CNN) specifically designed for efficient operation on portable devices, which was particularly suitable for deployment in resource-limited settings. The overall model structure is presented in Figure 3. Our model retained the depthwise separable convolutions and the HardSwish (H-swish) activation function of MobileNetV3, which reduced computational demand and parameters without compromising accuracy. Multiscale convolution modules have been incorporated into the model to effectively extract features at multiple focal points (Figure 3A), from subtle acetowhite changes in the columnar epithelium to varying TZ-type scopes in the cervix. In addition, we introduced a spatial pyramid pooling module to address features at multiple scales while preserving spatial information in the input images to ensure the richness of feature representation (Figure 3B). The Squeeze-and-Excitation module is a lightweight attention mechanism designed to automatically prioritize the most diagnostically relevant visual patterns within the MobileNetV3 architecture (Figure 3C). In our model, the Squeeze-and-Excitation module was embedded to adjust global feature weights (Figure 3D), enhancing the overall effectiveness of feature extraction. It increased the sensitivity to important features, thereby improving classification performance [18].

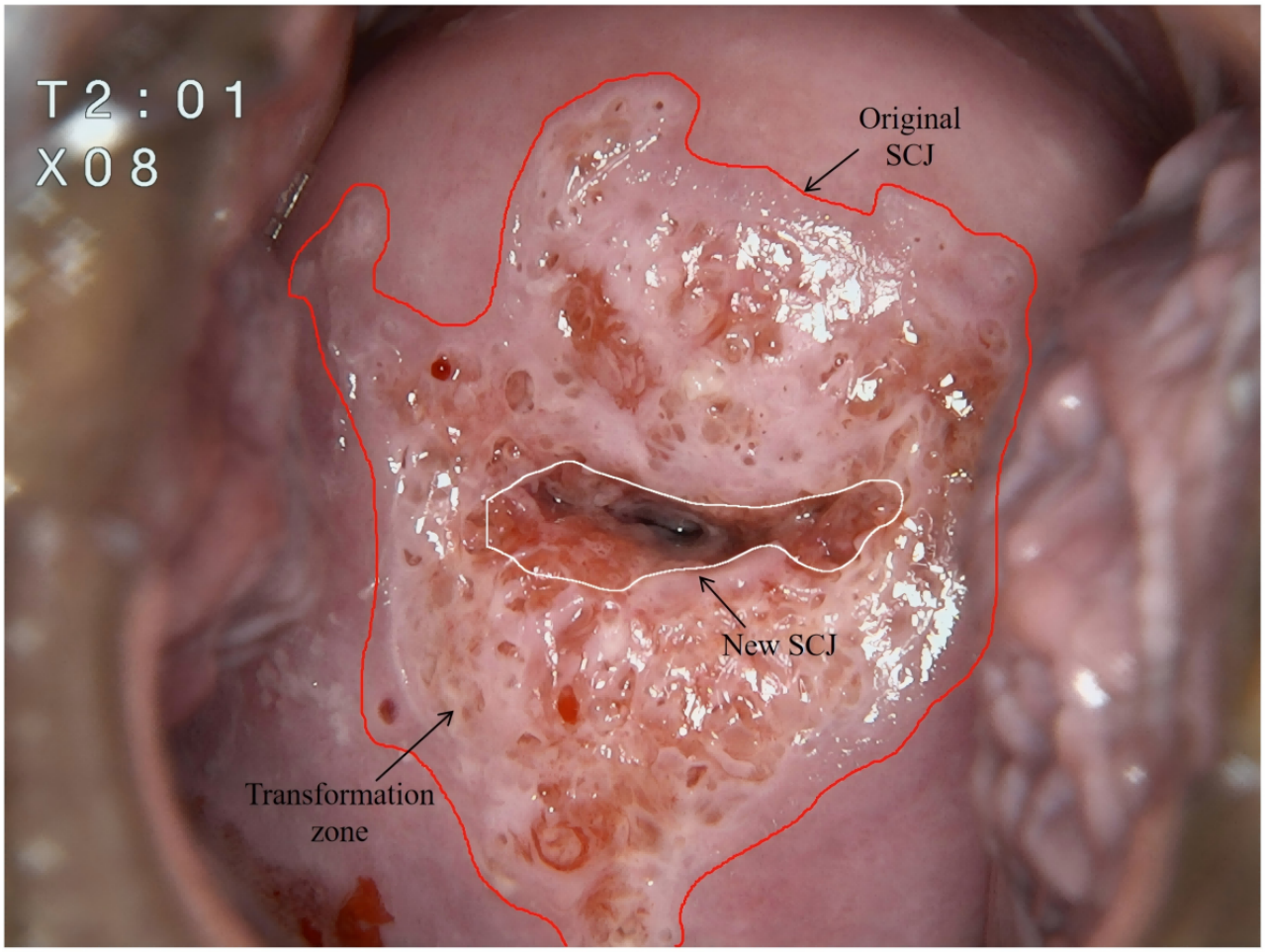
Figure 3. The overall architecture of the proposed model: (A) details of the multiscale convolution (MSC) module; (B) details of the Spatial Pyramid Pooling (SPP) module; (C) structure of the Squeeze-and-Excitation (SE) block module; (D) workflow of the model that included detailed layers of the network; and (E) structure of the inverted residuals module.



Development of the TZ Segmentation Model

The FastSAM model was used to segment the TZ from the entire cervix in colposcopy images (Figure 2B). FastSAM was fine-tuned for the specific dataset using a YOLO-based framework [19]. Colposcopic images were resized to 640×640 pixels, and the new SCJ was annotated using LabelMe (MIT CSAIL) [20]. By selectively freezing some of the backbone layers of the model, the deeper layers were fine-tuned to improve adaptation to the specific task, without compromising the extraction of general features. The model's performance was further enhanced by using coarse masks as spatial guides, which allowed the model to focus its attention on the most relevant image regions. The model was trained to delineate the new SCJ, thereby identifying possible TZ areas by this critical boundary. TZ is a dynamic region defined by the area between the original SCJ and the new SCJ. Typically, the original SCJ is considered a virtual line, and identifying the new SCJ is crucial for determining the location of the TZ (Figure 4).

Figure 4. Illustration of the squamocolumnar junction (SCJ) and the transformation zone.



Software and Hardware

The study was performed on a Windows 11 operating system with an NVIDIA GeForce RTX 3080 8GB graphics card. The models were implemented using Python (version 3.6.13; Python Software Foundation) and the PyTorch (Meta AI) Library (version 1.7.1).

Evaluation Metrics

Experiment performance of the cervical TZ classification model was evaluated using accuracy, precision, recall, F_1-score, and average precision, as defined in Equations 1-6.

Where TP, FP, and FN denote the true positive, false positive, and false negative predictions, respectively. $r(k)$ denotes the recall at threshold k. $p(k)$ denotes the precision at threshold k, and n denotes the number of thresholds. AP_i is the average precision for class i, and mAP denotes the mean average precision, where n is the number of classes.

Statistical Analysis

The performance of the classification model was evaluated by comparing it with selected state-of-the-art (SOTA) deep-learning models using metrics such as accuracy, precision, recall, and the area under the precision recall curve, represented by average precision. To evaluate the classification of TZ types, sensitivity, specificity, positive predictive value, and negative predictive value (NPV) were calculated, along with their corresponding 95% CIs. The evaluation metric was defined as the agreement with the expert-provided TZ classification. The demographic characteristics of the study participants were summarized using means and SDs for continuous variables and percentages for categorical variables. A P value of less than .05 (two-sided) was considered statistically significant. Statistical analyses were performed using SPSS (version 27.0; IBM Corp) and Python (version 3.7).

Ethical Considerations

This study was approved by the Institutional Review Board of the Chinese Academy of Medical Sciences and Peking Union Medical College (CAMS and PUMC-IEC-2022-022). Informed consent was not required due to the retrospective nature of the dataset, and all personal information and images were

completely anonymized. We commit that all research data will be used for academic research purposes.

Results

General Information of the Study Dataset

For the classification modeling, 8335 colposcopy images from 4 hospitals were selected for training and evaluation. These images included consisting of 2788 images of TZ1, 2663 images of TZ2, and 2884 images of TZ3. In the external validation study, 1335 cases were selected from 2 hospitals for model inference to predict the TZ types (Figure 5). The demographic characteristics of each participant, including age, colposcopic findings, and the distribution of each TZ type, are provided in Table 1. TZ types were found to be significantly associated with the participants' age group distribution (P<.001). The mean age of all participants at the time of colposcopy was 37.87 (SD 9.99) years, with the mean age of the TZ3 group being 45.04 (SD 11.69) years.

Figure 5. Flowchart of the study case collection. SCJ: squamocolumnar junction; TZ: transformation zone.

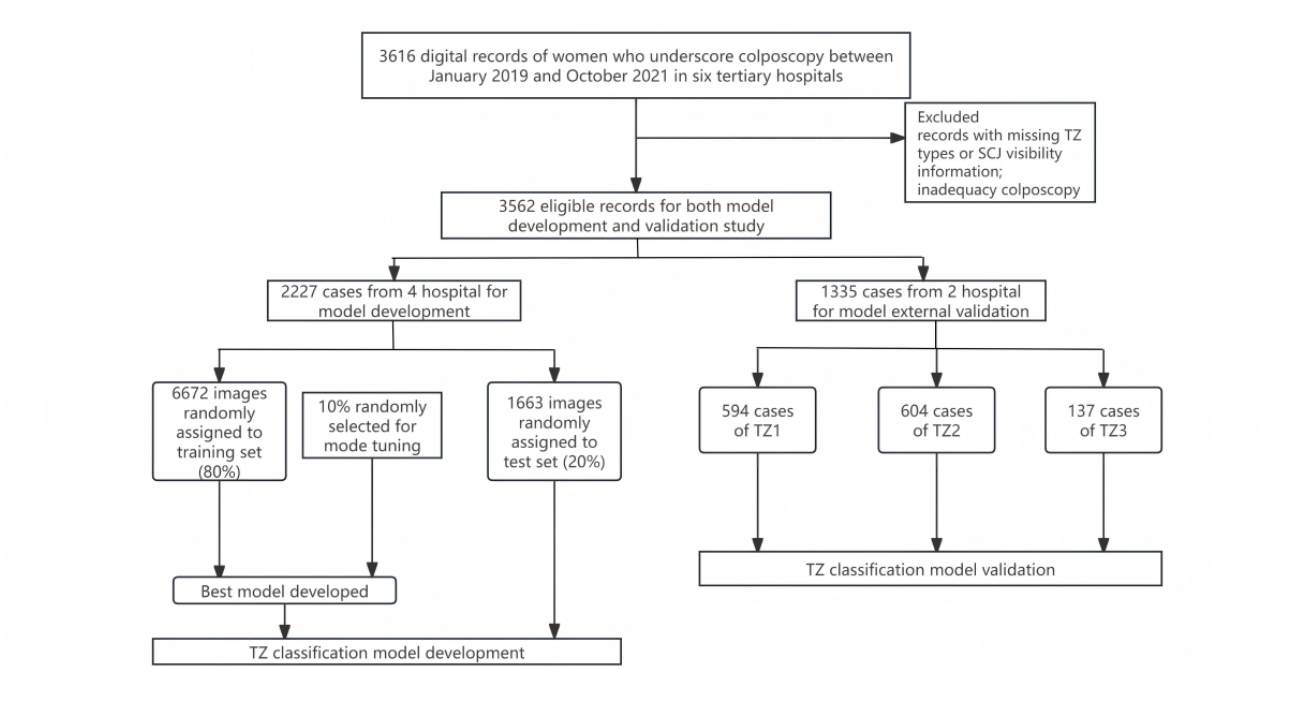


Table 1. Demographic characteristics and distribution of cervical TZ[a] types. Differences were analyzed using the chi-square test.

Characteristic	TZ1 (n=594)	TZ2 (n=604)	TZ3 (n=137)	*P* value
Age group (year), mean (SD)	34.44 (8.45)	39.62 (9.64)	45.04 (11.69)	<.001
24-29	140 (23.6)	62 (10.3)	10 (7.3)	
30-49	386 (65.0)	432 (71.5)	79 (57.7)	
50-65	68 (11.4)	110 (18.2)	48 (35.0)	
Parity, n (%)				<.001
0	159 (26.8)	95 (15.7)	18 (13.1)	
1-3	428 (72.1)	488 (80.8)	116 (84.7)	
>3	7 (1.2)	21 (3.5)	3 (2.2)	
Menstrual status, n (%)				<.001
Menopause	25 (4.2)	71 (11.8)	43 (31.4)	
No menopause	569 (95.7)	533 (88.2)	94 (68.6)	
SCJ[b] visibility, n (%)				<.001
Completely visible	594 (100.0)	73 (12.1)	0 (0)	
Partially visible	0 (0)	531 (87.9)	28 (20.4)	
Invisible	0 (0)	0 (0)	109 (79.6)	
Pathologic diagnosis, n (%)				<.001
Normal or benign	220 (37)	188 (31.1)	26 (19)	
CIN[c]1	220 (37)	213 (35.3)	51 (37.2)	
CIN2	124 (20.8)	131 (21.7)	23 (16.8)	
CIN3	24 (4)	63 (10.4)	13 (9.5)	
Cancer	6 (1)	9 (1.5)	24 (17.5)	

[a]TZ: transformation zone.

[b]SCJ: squamocolumnar junction.

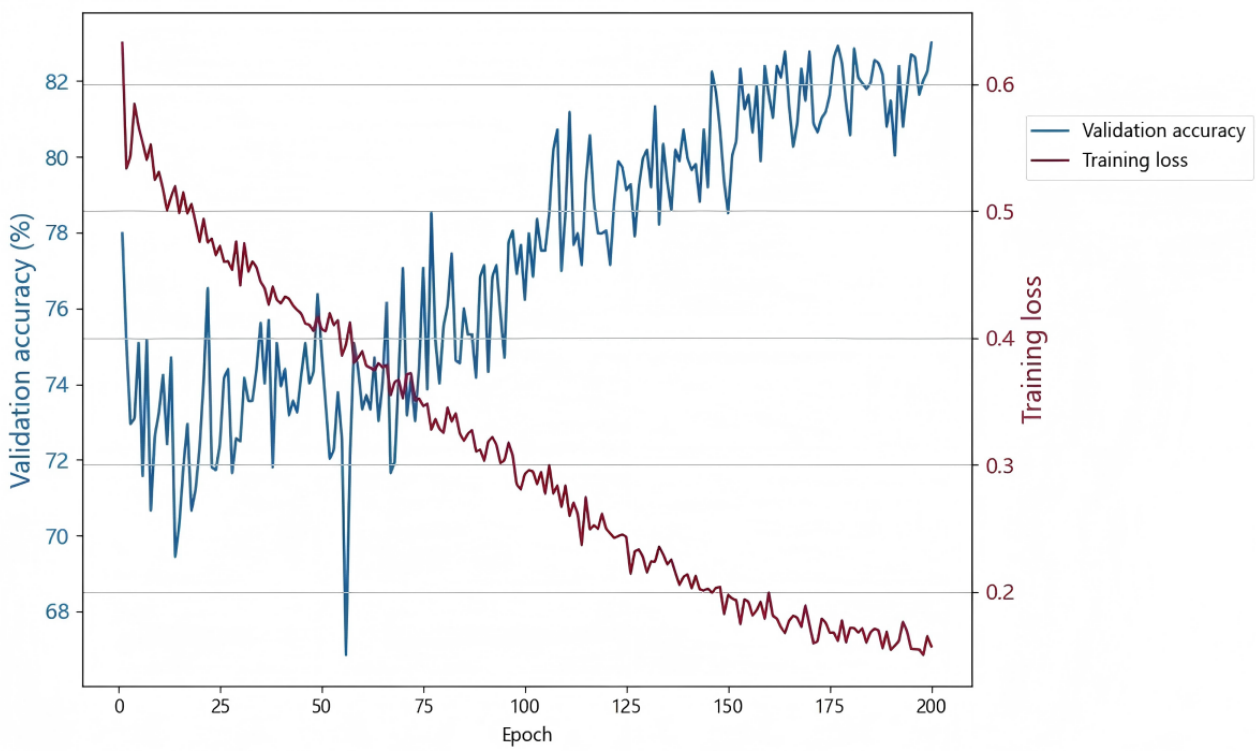
[c]CIN: cervical intraepithelial neoplasia.

TZ Classification Results

Cervix detection with a bounding box was used for feature engineering. Proper cervix extraction improved classification accuracy. A total of 8335 cervix images were investigated and resized to 224×224 pixels before being input into the classification model. Out of these, 1663 images were used to evaluate the model's performance.

Around 80% of the images were randomly selected as the training set, with the optimal weight parameter model selected during training being used to classify the images in the test set. The validation accuracy of the model gradually improved as the number of epochs increased, as shown in the validation plot (Figure 6). After approximately 100 epochs, both validation accuracy and training loss stabilized, reaching a peak validation accuracy of 83% at the 200th epoch. The validation accuracy fluctuated significantly during the first 25 epochs, it showed a general upward trend, while training loss rapidly decreased, indicating active learning and adjustment of the model. The model reached an optimal balance point and performed well after the 75th epoch when validation accuracy stabilized and training loss continued to decrease but with minimal changes. Overall, the model demonstrated rapid convergence in the early stages of training, with steady performance improvements in the later stages.

Figure 6. Validation accuracy and training loss curve for the proposed classification model.



Our proposed model accurately classified cervical TZ types, as shown in Table 2. The highest classification accuracy of 83.97% and precision of 83.93% were achieved on the test set. For the 3 TZ types, the sensitivity was 84.74% for TZ1, 78.95% for TZ2, and 87.87% for TZ3, while the specificity was 89.99% for TZ1, 91.98% for TZ2, and 94.03% for TZ3. The detailed values for sensitivity, specificity, positive predictive value, and NPV are presented in Table 3. According to the classification performance of TZ types, the sensitivity and NPV for TZ2 were significantly lower.

Table 2. Comparative performance results of the proposed classification model with other state-of-the-art models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F_1-score (%)	mAP[a] (%)
ResNet50	59.22	57.84	56.89	54.66	63.13
VGG16	70.33	69.76	69.21	69.19	74.64
ViT	59.46	57.64	57.39	56.49	63.76
EfficientNet	62.06	60.97	61.18	60.66	63.63
ShuffleNet	74.47	73.9	73.67	73.66	80.98
MobileNetV3	75.06	74.49	74.64	74.55	82.72
Proposed model	83.97	83.93	83.85	83.86	92.17

[a]mAP: mean average precision.

Table 3. Classification performance of the proposed classification model on the test set.

TZ[a] types	Sensitivity (%)	Specificity (%)	PPV[b] (%)	NPV[c], (%)	Average precision (%)
TZ1	84.74	89.99	80.96	92.15	91.84
TZ2	78.95	91.98	82.19	90.30	89.06
TZ3	87.87	94.03	88.64	93.60	95.62

[a]TZ: transformation zone.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

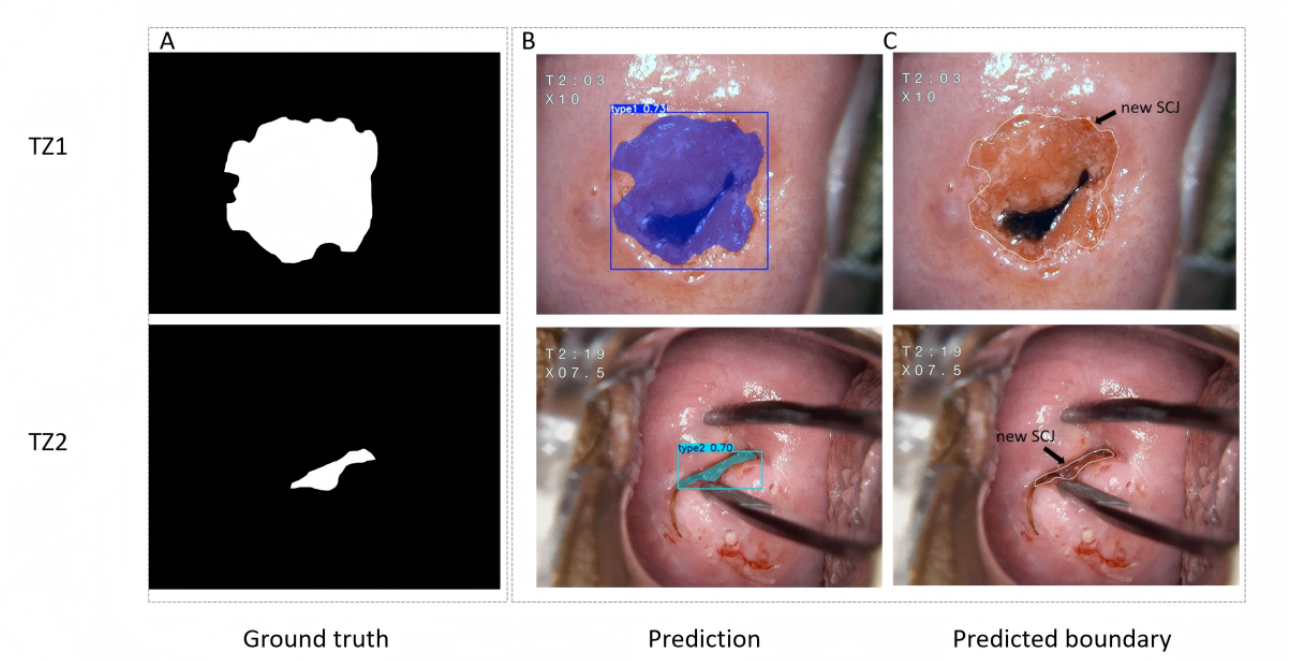
The SOTA deep-learning networks for interpreting colposcopy images were selected to train on our dataset for comparison with our model. Table 2 presents the experimental results of the test set based on various evaluation metrics. The traditional

ResNet50 and Vision Transformer models performed poorly in terms of accuracy and precision, achieving only 59.22% and 57.84%, and 59.46% and 57.64%, respectively. VGG16, with its relatively low-density network structure, performed better than ResNet50 in feature extraction. In addition, we selected 2 lightweight networks for comparison, EfficientNet and ShuffleNet, with higher accuracy of 62.06% and 74.47%. The proposed model achieved the highest accuracy of 83.97% among SOTA models.

TZ Segmentation Result

To enhance AI's clinical guidance, we used a pretrained FastSAM model to segment the TZ region by visualizing the new SCJ. A portion of the training data was used to annotate the new SCJ (Figure 7A), excluding TZ3, as the new SCJ is located in the endocervix. The segmentation model predicted the negative foreground region indicating the area outside the target TZ region (Figure 7B). The region boundary of the mask prediction was the predicted new SCJ (Figure 7C). In the test set, the overall recall and mAP50 (mean average precision at 50% Intersection over Union) of predicted mask were 0.78 and 0.75, respectively.

Figure 7. The segmentation model inference process. (A) Ground truth of the polygonal new SCJ outline. (B) Predictive segmentation of the out-of-TZ area. (C) Representative prediction of the new SCJ guidance line. SCJ: squamocolumnar junction; TZ: transformation zone.



Validation Results

An independent dataset of 1335 cases was used to evaluate the generalization of the TZ classification model, distributed as 594 cases of TZ1, 604 cases of TZ2, and 137 cases of TZ3. The overall classification accuracy was 79.33%. The model's predicted sensitivity for TZ1, TZ2, and TZ3 was 77.3% (95% CI 73.9%-80.6%), 81.1 (95% CI 78.0%-82.3%), and 80.3% (95% CI 73.7%-86.9%), respectively, while the specificity was 94.2% (95% CI 92.4%-95.8%), 83.3% (95% CI 80.6%-85.9%), and 90.7% (95% CI 89.1%-92.3%). The classification performance is presented in Figure 8 and Table 4.

Figure 8. Confusion matrix of the proposed classification model in validation. TZ: transformation zone.

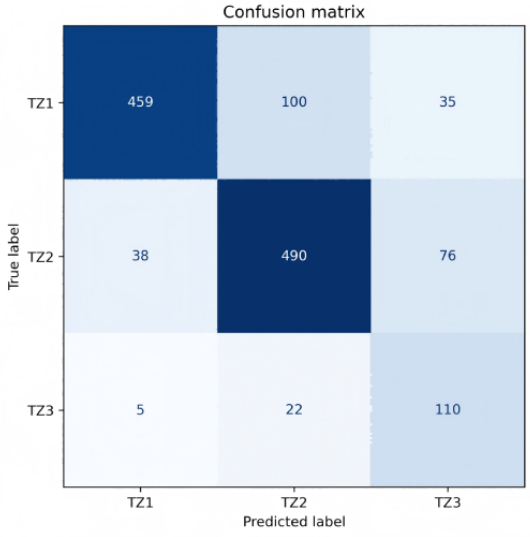


Table 4. Classification performance on the validation dataset.

TZ[a] types	Sensitivity (%; 95% CI)	Specificity (%; 95% CI)	PPV[b] (%; 95% CI)	NPV[c] (%; 95% CI)	F_1-score (%)
TZ1 (n=594)	77.3 (73.9-80.6)	94.2 (92.4-95.8)	91.4 (88.9-93.8)	83.8 (81.3-86.3)	83.76
TZ2 (n=604)	81.1 (78.0-82.3)	83.3 (80.6-85.9)	80.1 (76.8-83.2)	84.2 (81.5-86.9)	80.59
TZ3 (n=137)	80.3 (73.7-86.9)	90.7 (89.1-92.3)	49.8 (43.0-56.7)	97.6 (96.7-98.5)	61.45

[a]TZ: transformation zone.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

Discussion

Principal Findings

In this study, we proposed a classification and identification model that facilitates clinical colposcopy examinations of cervical TZ. Our AI model achieved an accuracy of 79.33% in the task of classifying the three types of TZ in the external dataset. The sensitivity of TZ3 was 80.3%, and the specificity was 90.7%, which were satisfactory. TZ3 could be accurately differentiated, which further decreased the chance of missed diagnoses of high-grade lesions [21], and it was valuable for guiding ECC and recommending appropriate treatments. The AI model had increased sensitivity on TZ2 and lower sensitivity on TZ1 compared to the performance on the test set, which might be attributed to the similarity between TZ1 and TZ2. According to the standard terminology of the IFCPC, TZ1 and TZ2 were both visible either with or without the assistance of an endocervical speculum [22]. TZ2 was nearly entirely visible in some cases, while the SCJ was only visible at the endocervical canal's border. Therefore, it was difficult to differentiate it from TZ1, which frequently required expert evaluation in strict accordance with established guidelines. In clinical practice, the treatment management for TZ1 and TZ2 is generally similar [6]. Our AI model demonstrated noninferior performance in classifying them and could be used to assist less experienced or inexperienced colposcopists in the colposcopic examination process. The overall model accuracy from the validation study was similar to the performance during model training, indicating that it is capable of generalizing fairly well.

Comparison With Other Studies

In terms of AI techniques, we proposed a method based on the variant MobileNetV3 architecture for cervical TZ classification and FastSAM for segmenting TZ in colposcopy. Currently, only a few deep learning–based models have been developed to classify cervical TZ from colposcopy images. Dash et al [23] conducted a TZ segmentation and classification model based on colposcopy images from the IARC image bank. Similarly, Cao et al [24] developed a high-performance, deep learning model based on retrospective image data collected from one hospital [24]. Comparatively, our method was based on a multicenter study with a more diverse dataset of colposcopy images. While this diversity added to the challenge of feature extraction, the classification model showed better accuracy. Second, the proposed segmentation model could precisely annotate the new SCJ and indicate the approximate location of TZ, which could greatly assist junior colposcopists in selecting biopsy sites. Effective biopsy prioritization focused on the most severe lesions, particularly those within the TZ. Compared with the previous study [23], our method delineated a new SCJ rather than the original SCJ, which can provide more effective assistance and insights for colposcopic clinical examination and treatment. Furthermore, conducted external validation for the TZ classification model for the first time, and our model demonstrated strong stability and generalization performance.

Clinical Implications

In resource-limited settings, colposcopists demonstrated significantly lower clinical skills and performance than those observed in our study. The results of a study evaluating the colposcopic abilities of colposcopists in underserved Chinese communities indicated that colposcopists had a mere 22% accuracy specifically for TZ3 [8]. Similarly, in another study, colposcopists' clinical diagnostic abilities were assessed before and after intensive training [9]. It was found that junior colposcopists had only 49.1% accuracy in classifying TZs. Despite the notable improvement, their accuracy remained below optimal levels at 68.6% following training. In addition, a comparable study conducted in Europe demonstrated that junior colposcopists were only able to detect TZs with a 55% accuracy rate [25]. According to these findings, colposcopists at less-experienced levels generally exhibited inadequate colposcopic performance. However, AI might offer a promising solution to enhance colposcopic capabilities and clinical decision-making confidence. Compared to these studies, our TZ classification model demonstrated strong performance with a high accuracy of 83.97% in the test set and achieved much higher sensitivity in predicting three TZ types in the validation study. Based on the results of previous studies, our model was more effective at stand-alone classification than that of junior colposcopists. Therefore, the method presented in this study accurately identifies TZ in colposcopy images, providing a valuable reference for colposcopists when making clinical decisions. The findings from this study supported the potential of the proposed AI-based TZ identification method as a promising adjunct tool for colposcopic examinations, particularly when integrated with AI colposcopic diagnostic systems and digital colposcopes. Dynamic digital imaging with AI assistance enhances the objectivity of colposcopic examinations and might address the diagnostic subjectivity of less experienced colposcopists. When AI-guided digital colposcopists are deployed in resource-limited health care settings, colposcopists will be able to receive intuitive and accessible guidance on clinical features from AI during the colposcopic examination in real time, supporting less-experienced colposcopists in improving their overall colposcopy skills. Combining AI results with colposcopist assessments helps reduce diagnostic bias, improve colposcopic examination capability, and narrow the gap with resource-rich areas. It addresses the minimization of missed diagnoses, the facilitation of early detection of lesions, the reduction of the risk of CIN progression, and the reduction of the burden of cervical cancer [26]. The application of AI will be integral to improving the quality of colposcopy services in low-resource settings.

The development and application research of AI-guided colposcopy models has emphasized their auxiliary value in clinical practice [14,27]. With the advent of innovations in AI algorithms, these models rely predominantly on mainstream CNNs coupled with transfer learning, as a method of unsupervised learning for image feature learning [28,29]. In some AI models, object detection is used as a metric for identifying lesion areas; however, it remains limited to binary classification tasks. Nevertheless, the multiclassification of normal cervix, low-grade squamous intraepithelial lesion, and high-grade squamous intraepithelial lesion plus poses significant challenges to these models. A clinical perspective suggested that acetowhite changes within TZ have a higher likelihood of association with CIN lesions than those outside of TZ. While the lesion extends across the new SCJ, a biopsy should be taken within the TZ or at the new SCJ. Therefore, by incorporating the TZ-type information as a weight in the model, the AI could be able to perform a more effective feature engineering process within the TZ. It is anticipated that the AI model will markedly enhance the capability of lesion detection, thus eliminating the current limitation in which precise identification is confined to CIN2+ lesions. Our proposed AI system delineated the new SCJ on high-resolution colposcopy images, and it enabled the approximate location of the TZ to be visualized more intuitively, thereby guiding more effective colposcopy examination and biopsy procedures.

For applying AI techniques in limited-resource settings, we proposed a model based on the MobileNetV3 architecture for cervical TZ classification in colposcopy images. Since Google introduced MobileNet in 2017 [30] as a lightweight CNN architecture, it has gained significant attention for its efficiency and accuracy. A lightweight AI model is more energy efficient and requires significantly fewer computational resources than a large-scale AI model, aligning with sustainable AI practices. It has been designed to operate efficiently on battery-operated devices, making it especially suitable for deployment in remote regions with limited power availability and internet access. By minimizing the reliance on high-cost hardware and extensive cloud infrastructure, lightweight AI models enable resource optimization in low-resource settings and facilitate equitable access to AI. Our AI model achieved efficient computation and robust classification performance on portable devices and can be applied actively in a variety of clinical settings to validate its generalization ability. A further economic evaluation is required to support the decision to adopt novel technologies in screening strategies.

Limitations and Future Directions

There are several limitations to this study. First, the scale of the dataset we included was limited, which restricted the ability to support training deeper or more complex deep neural networks. However, the quality and standardization of image acquisition were assured, and the appropriate network depth was chosen to maximize feature extraction. It is necessary to obtain more high-quality colposcopy images with endocervix expansion by auxiliary instruments per case to train more complex networks that can extract additional features and potentially improve classification accuracy in TZ2 and TZ3. To address the suboptimal performance in TZ2 classification, expanding the dataset for TZ2 in future studies could enable the model to capture more distinct texture features that differentiate it from TZ1 and TZ3. The model could be further enhanced by applying edge detection techniques in regions where it is difficult to distinguish TZ2 from TZ3 and to highlight subtle morphological changes at the new SCJ. Second, the external validation was limited to evaluating the AI results alone. Further research is needed to assess the impact of AI-aided colposcopists in the classification of TZ. A prospective clinical trial is further needed

to investigate whether the method can be applied to real-world colposcopy clinics. In addition, although we collected the validation datasets from two different hospitals, the scale of the external validation dataset is relatively limited, particularly when it comes to representing clinical scenarios that involve different colposcope vendors and lighting conditions. Future studies will focus on expanding both the size and scope of the dataset, incorporating data from more regions for more comprehensive validation. Furthermore, transfer learning or domain adaptation techniques may be used to improve the model's robustness and adaptability to various imaging conditions. Finally, the TZ classification lacks an absolute subjective gold standard but is guided by colposcopic examination criteria and expert consensus. However, the colposcopy expert panel from tertiary hospitals conducted the "ground truth" categories for this study in accordance with established guidelines from IFCPC. In light of the subjectivity inherent in TZ classification, using only single-image modality data for model development is limited. Model optimization in the future may implement multimodal learning, which incorporates image data with associated clinical factors (eg, age, menstrual status, gravidity, and parity) to reduce the subjectivity of the model and enhance TZ feature discrimination. Furthermore, the construction of a knowledge graph based on existing consensus and colposcopic examination guidelines will help the model adhere to various rules or logic during the learning process.

Conclusion

In this study, an accurate AI-based method was developed and evaluated for automatically identifying cervical TZ using colposcopy images. The proposed method was the first application of a lightweight CNN for cervical feature extraction and applied a general segmentation model for TZ delineation among multicenter images, achieving commendable classification accuracy on TZ. The proposed approach has the potential to adapt to various colposcopy clinical environments and improve AI-guided colposcopy practice.

Acknowledgments

We acknowledge Haiyan Hu (Shenzhen Maternity and Child Healthcare Hospital), Zhilian Wang (Second Hospital of Shanxi Medical University), Jinfeng Liu (Gansu Maternity and Child Healthcare Hospital), and Jun Lei (Jiangxi Maternity and Child Healthcare Hospital) for assisting with colposcopy image collection and manual labeling. We thank Ms Qianling Dai (Chengdu Women's and Children's Central Hospital), and Ms Haiyan Hu for quality control. We also thank Ms Qianling Dai and Ms Mingyang Chen (Chinese Academy of Medical Sciences and Peking Union Medical College) for their comments on the original manuscript. This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS 2021-I2M-1-004), the National Natural Science Foundation of China (82404286), the China Postdoctoral Science Foundation (2023M740323), and the Postdoctoral Fellowship Program of CPSF (GZB20230076).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to data-sharing policies set by the funding institution but are available from the corresponding author on reasonable request.

Authors' Contributions

TW and YQ contributed to conceptualization. TW was responsible for writing—original draft, methodology, software, and formal analysis. TW and YW conducted data analysis and visualization. YW and PX were responsible for data curation and writing—review and editing. XC and YQ contributed to supervision. YQ and PX handled funding acquisition.

Conflicts of Interest

None declared.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263 [FREE Full text] [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Ogilvie G, Nakisige C, Huh WK, Mehrotra R, Franco EL, Jeronimo J. Optimizing secondary prevention of cervical cancer: recent advances and future challenges. Int J Gynaecol Obstet 2017;138 Suppl 1:15-19. [doi: 10.1002/ijgo.12187] [Medline: 28691338]
3. Leeson SC, Alibegashvili T, Arbyn M, Bergeron C, Carriero C, Mergui JL, et al. The future role for colposcopy in Europe. J Low Genit Tract Dis 2014;18(1):70-78. [doi: 10.1097/LGT.0b013e318286b899] [Medline: 23774077]
4. Zhao F, Lang J, Qiao Y, Zhu L. How can China achieve WHO's 2030 targets for eliminating cervical cancer? BMJ 2024;386:e078641 [FREE Full text] [doi: 10.1136/bmj-2023-078641] [Medline: 39214542]
5. Deng H, Mondal S, Sur S, Woodworth CD. Establishment and optimization of epithelial cell cultures from human ectocervix, transformation zone, and endocervix optimization of epithelial cell cultures. J Cell Physiol 2019;234(6):7683-7694 [FREE Full text] [doi: 10.1002/jcp.28049] [Medline: 30609028]

6. Prendiville W, Sankaranarayanan R. Colposcopy and Treatment of Cervical Precancer. Lyon, France: International Agency for Research on Cancer; 2017.
7. Wei B, Li Q, Seery S, Qiao Y, Jiang Y. Endocervical curettage for diagnosing high-grade squamous intraepithelial lesions or worse in women with type 3 transformation zone lesions: a retrospective, observational study. BMC Womens Health 2023;23(1):245 [FREE Full text] [doi: 10.1186/s12905-023-02297-0] [Medline: 37161558]
8. Cui X, Wang H, Chen M, Seery S, Xue P, Qiao Y, et al. Assessing colposcopy competencies in medically underserved communities: a multi-center study in China. BMC Cancer 2024;24(1):349 [FREE Full text] [doi: 10.1186/s12885-024-12106-y] [Medline: 38504211]
9. Chen M, Xue P, Li Q, Shen Y, Ye Z, Wang H, et al. Enhancing colposcopy training using a widely accessible digital education tool in China. Am J Obstet Gynecol 2023;229(5):538.e1-538.e9. [doi: 10.1016/j.ajog.2023.07.043] [Medline: 37516400]
10. Kim S, An H, Cho HW, Min KJ, Hong JH, Lee S, et al. Pivotal clinical study to evaluate the efficacy and safety of assistive artificial intelligence-based software for cervical cancer diagnosis. J Clin Med 2023;12(12):4024 [FREE Full text] [doi: 10.3390/jcm12124024] [Medline: 37373717]
11. Wu A, Xue P, Abulizi G, Tuerxun D, Rezhake R, Qiao Y. Artificial intelligence in colposcopic examination: a promising tool to assist junior colposcopists. Front Med (Lausanne) 2023;10:1060451 [FREE Full text] [doi: 10.3389/fmed.2023.1060451] [Medline: 37056736]
12. Yuan C, Yao Y, Cheng B, Cheng Y, Li Y, Li Y, et al. The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images. Sci Rep 2020;10(1):11639 [FREE Full text] [doi: 10.1038/s41598-020-68252-3] [Medline: 32669565]
13. Peng G, Dong H, Liang T, Li L, Liu J. Diagnosis of cervical precancerous lesions based on multimodal feature changes. Comput Biol Med 2021;130:104209. [doi: 10.1016/j.compbiomed.2021.104209] [Medline: 33440316]
14. Xue P, Tang C, Li Q, Li Y, Shen Y, Zhao Y, et al. Development and validation of an artificial intelligence system for grading colposcopic impressions and guiding biopsies. BMC Med 2020;18(1):406 [FREE Full text] [doi: 10.1186/s12916-020-01860-y] [Medline: 33349257]
15. Li Y, Liu ZH, Xue P, Chen J, Ma K, Qian T, et al. GRAND: A large-scale dataset and benchmark for cervical intraepithelial neoplasia grading with fine-grained lesion description. Med Image Anal 2021;70:102006. [doi: 10.1016/j.media.2021.102006] [Medline: 33690025]
16. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. arXiv. Preprint posted online on June 8, 2015 . [doi: 10.48550/arXiv.1506.02640]
17. Zheng Z, Wang P, Liu W, Li J, Ye R, Ren D. Distance-IoU loss: faster and better learning for bounding box regression. 2020 Presented at: Proceedings of the AAAI Conference on Artificial Intelligence; February 7-12, 2020; New York, NY. [doi: 10.1609/aaai.v34i07.6999]
18. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. 2018 Presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT. [doi: 10.1109/cvpr.2018.00745]
19. Zhao X, Ding W, An Y, Du Y, Yu T, Li M, et al. Fast segment anything. arXiv. Preprint posted online on June 21, 2023 . [doi: 10.48550/arXiv.2306.12156]
20. Torralba A, Russell BC, Yuen J. LabelMe: online image annotation and applications. Proceedings of the IEEE 2010;98(8):1467-1484. [doi: 10.1109/jproc.2010.2050290]
21. Wei B, Zhang B, Xue P, Seery S, Wang J, Li Q, et al. Improving colposcopic accuracy for cervical precancer detection: a retrospective multicenter study in China. BMC Cancer 2022;22(1):388 [FREE Full text] [doi: 10.1186/s12885-022-09498-0] [Medline: 35399061]
22. Bornstein J, Bentley J, Bösze P, Girardi F, Haefner H, Menton M, et al. 2011 colposcopic terminology of the international federation for cervical pathology and colposcopy. Obstet Gynecol 2012;120(1):166-172. [doi: 10.1097/AOG.0b013e318254f90c] [Medline: 22914406]
23. Dash S, Sethy PK, Behera SK. Cervical transformation zone segmentation and classification based on improved inception-resNet-V2 using colposcopy images. Cancer Inform 2023;22:11769351231161477 [FREE Full text] [doi: 10.1177/11769351231161477] [Medline: 37008072]
24. Cao Y, Ma H, Fan Y, Liu Y, Zhang H, Cao C, et al. A deep learning-based method for cervical transformation zone classification in colposcopy images. Technol Health Care 2023;31(2):527-538. [doi: 10.3233/THC-220141] [Medline: 36093645]
25. Sabrina F, Ilkka K, Mervi HN, Charles R, Simon L, Ameli T, et al. Fostering prevention of cervical cancer by a correct diagnosis of precursors: a structured case-based colposcopy course in finland, Norway and UK. Cancers (Basel) 2020;12(11):3201 [FREE Full text] [doi: 10.3390/cancers12113201] [Medline: 33143157]
26. Chen H, Zhao X, Hu S, You T, Xia C, Gao M, et al. Estimation of economic burden throughout course of cervical squamous intraepithelial lesion and cervical cancer in China: a nationwide multicenter cross-sectional study. Chin J Cancer Res 2023;35(6):675-685 [FREE Full text] [doi: 10.21147/j.issn.1000-9604.2023.06.11] [Medline: 38204443]

27. Bao H, Sun X, Zhang Y, Pang B, Li H, Zhou L, et al. The artificial intelligence-assisted cytology diagnostic system in large-scale cervical cancer screening: a population-based cohort study of 0.7 million women. Cancer Med 2020;9(18):6896-6906 [FREE Full text] [doi: 10.1002/cam4.3296] [Medline: 32697872]
28. Chandran V, Sumithra MG, Karthick A, George T, Deivakani M, Elakkiya B, et al. Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images. Biomed Res Int 2021;2021:5584004 [FREE Full text] [doi: 10.1155/2021/5584004] [Medline: 33997017]
29. Luo YM, Zhang T, Li P, Liu PZ, Sun P, Dong B, et al. MDFI: Multi-CNN decision feature integration for diagnosis of cervical precancerous lesions. IEEE Access 2020;8:29616-29626. [doi: 10.1109/access.2020.2972610]
30. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv. Preprint posted online on April 17, 2017 . [doi: 10.48550/arXiv.1704.04861]

Abbreviations

AI: artificial intelligence
CIN: cervical intraepithelial neoplasia
CNN: convolutional neural network
ECC: endocervical curettage
IFCPC: International Federation for Cervical Pathology and Colposcopy
mAP50: mean average precision at 50% Intersection over Union
NPV: negative predictive value
SCJ: squamocolumnar junction
SOTA: state-of-the-art
TZ: transformation zone

Edited by N Cahill; submitted 06.12.24; peer-reviewed by J Li, Y Chu; comments to author 07.01.25; revised version received 16.01.25; accepted 19.02.25; published 31.03.25.

Please cite as:
Wu T, Wang Y, Cui X, Xue P, Qiao Y
AI-Based Identification Method for Cervical Transformation Zone Within Digital Colposcopy: Development and Multicenter Validation Study
JMIR Cancer 2025;11:e69672
URL: https://cancer.jmir.org/2025/1/e69672
doi:10.2196/69672
PMID:

Using ChatGPT to Improve the Presentation of Plain Language Summaries of Cochrane Systematic Reviews About Oncology Interventions: Cross-Sectional Study

Jelena Šuto Pavičić[1], MD; Ana Marušić[2], MD, PhD; Ivan Buljan[3], PhD

[1]Department of Oncology and Radiotherapy, University Hospital of Split, Spinciceva 1, Split, Croatia

[2]Department of Research in Biomedicine in Health, Centre for Evidence-based Medicine, University of Split School of Medicine, Split, Croatia

[3]Department of Psychology, Faculty of Humanities and Social Sciences, University of Split, Split, Croatia

Corresponding Author:
Jelena Šuto Pavičić, MD
Department of Oncology and Radiotherapy, University Hospital of Split, Spinciceva 1, Split, Croatia

Abstract

Background: Plain language summaries (PLSs) of Cochrane systematic reviews are a simple format for presenting medical information to the lay public. This is particularly important in oncology, where patients have a more active role in decision-making. However, current PLS formats often exceed the readability requirements for the general population. There is still a lack of cost-effective and more automated solutions to this problem.

Objective: This study assessed whether a large language model (eg, ChatGPT) can improve the readability and linguistic characteristics of Cochrane PLSs about oncology interventions, without changing evidence synthesis conclusions.

Methods: The dataset included 275 scientific abstracts and corresponding PLSs of Cochrane systematic reviews about oncology interventions. ChatGPT-4 was tasked to make each scientific abstract into a PLS using 3 prompts as follows: (1) rewrite this scientific abstract into a PLS to achieve a Simple Measure of Gobbledygook (SMOG) index of 6, (2) rewrite the PLS from prompt 1 so it is more emotional, and (3) rewrite this scientific abstract so it is easier to read and more appropriate for the lay audience. ChatGPT-generated PLSs were analyzed for word count, level of readability (SMOG index), and linguistic characteristics using Linguistic Inquiry and Word Count (LIWC) software and compared with the original PLSs. Two independent assessors reviewed the conclusiveness categories of ChatGPT-generated PLSs and compared them with original abstracts to evaluate consistency. The conclusion of each abstract about the efficacy and safety of the intervention was categorized as conclusive (positive/negative/equal), inconclusive, or unclear. Group comparisons were conducted using the Friedman nonparametric test.

Results: ChatGPT-generated PLSs using the first prompt (SMOG index 6) were the shortest and easiest to read, with a median SMOG score of 8.2 (95% CI 8 - 8.4), compared with the original PLSs (median SMOG score 13.1, 95% CI 12.9 - 13.4). These PLSs had a median word count of 240 (95% CI 232 - 248) compared with the original PLSs' median word count of 364 (95% CI 339 - 388). The second prompt (emotional tone) generated PLSs with a median SMOG score of 11.4 (95% CI 11.1 - 12), again lower than the original PLSs. PLSs produced with the third prompt (write simpler and easier) had a median SMOG score of 8.7 (95% CI 8.4 - 8.8). ChatGPT-generated PLSs across all prompts demonstrated reduced analytical tone and increased authenticity, clout, and emotional tone compared with the original PLSs. Importantly, the conclusiveness categorization of the original abstracts was unchanged in the ChatGPT-generated PLSs.

Conclusions: ChatGPT can be a valuable tool in simplifying PLSs as medically related formats for lay audiences. More research is needed, including oversight mechanisms to ensure that the information is accurate, reliable, and culturally relevant for different audiences.

(JMIR Cancer 2025;11:e63347) doi:10.2196/63347

KEYWORDS

health literacy; patient education; health communication; ChatGPT; neoplasms; Cochrane; oncology; plain language; medical information; decision-making; large language model; artificial intelligence; AI

Introduction

The significance of health literacy has been well established through numerous studies [1,2], demonstrating its importance not only for individual health care [3] but also within the public health system [4,5]. Various organizations, including the Center for Disease Control and Prevention and the National Institute of Health, have underscored the importance of using plain language in health communication to enhance understanding

of medical conditions and patients' engagement [6,7]. Health literacy is particularly important in oncology, where the advancement in cancer treatment beyond traditional methods has positioned patients and their families even more in the center of making care decisions [8]. Upon receiving a cancer diagnosis, patients often turn to various sources for more information, such as the internet, forums, social support groups, and literature [9,10]. Enhancing patients' understanding of their conditions has been shown to positively impact patient adherence and clinical outcomes in various chronic disease models by influencing patient behavior [11-13], whereas patients' struggle to understand complex medical information can adversely affect their adherence to medical advice [14]. However, nearly half of cancer patients struggle to understand the information about their treatment options from scientific literature [15]. Even though the official recommendation of the American Medical Association is that health information should be written at the reading level of the sixth grade in the US education system [16], the complexity of web-based cancer information often exceeds the reading and comprehension abilities of an average person, failing to meet the necessary standards for readability and understandability of health information [17].

To address the gap between the complexity of scientific evidence and the public, many organizations, including Cochrane, dedicate a lot of effort to enhancing the quality of health information available to the public [18]. Cochrane Database of Systematic Reviews is recognized as a highly reliable source for evaluating the effectiveness of health interventions [19]. Cochrane systematic reviews provide both a scientific abstract for professionals and a plain language summary (PLS) for the lay public [20]. Studies have consistently demonstrated that PLSs, which are authored by the same researchers who write the corresponding scientific abstracts, tend to exhibit readability levels that exceed those recommended for health information intended for a general audience [20-24]. In our previous study, we found that Cochrane PLSs for oncology interventions not only required a reading proficiency well above the average public level but also used language that lacked engagement, potentially reducing the reader's interest [21]. Additionally, the PLSs frequently contained ambiguous or insufficiently clear conclusions regarding the efficacy of the interventions assessed in the systematic review [21]. This lack of clarity can leave readers, especially patients and nonspecialists, uncertain about treatment benefits or outcomes, thus diminishing the utility of these summaries in supporting informed health decisions. In addition, we have to bear in mind that the language people use plays a vital role in processing and interpreting information in text, as well as in shaping psychological responses to information and influencing whether the reader will perceive the content as more relatable [25,26]. Given the significant challenges associated with the readability and clarity of PLSs, it is important to explore innovative solutions that can enhance the communication of health information to the lay public.

Artificial intelligence (AI) tools recently emerged as a tool to help generate textual outputs relevant to health care, with the potential to revolutionize the medical sector [27,28]. A standout example of AI in action is the Chat Generative Pretrained Transformer (ChatGPT), a chatbot that operates on the Generative Pretrained Transformer technology [29]. This technology is a type of large language model (LLM) with over 175 billion parameters, capable of understanding and generating text that mimics human conversation [30]. ChatGPT has been trained on a wide variety of internet content, such as books, articles, and websites [31,32]. Its fine-tuning process includes reinforcement learning from human feedback, enhancing its ability to grasp complex user intents and respond accurately to a wide array of tasks, including those related to medical inquiries [30,33]. The deployment of natural language processing models like ChatGPT in the health care field promises to significantly improve access to medical information for both professionals and patients [29,34,35].

Recognizing the challenge posed by the low readability of PLSs authored by researchers for Cochrane systematic reviews, particularly in the context of oncology interventions, we sought to investigate the potential of using AI LLMs, specifically ChatGPT as one of the most accessible tools today, to create PLSs that are more relatable to the public. Cochrane reviews are renowned for their rigorous methodologies and comprehensive analyses; however, this complexity often results in PLSs that may be difficult for a lay audience to understand [20,21]. By using ChatGPT, we aimed to improve the readability and the linguistic characteristics of these summaries, so that they effectively communicate critical findings, methodologies, and implications in a manner that is clear and engaging for nonexpert audiences.

Methods

Study Design and Data Sources

In this study, we analyzed the readability, linguistic characteristics, and conclusiveness of the PLSs generated by ChatGPT from corresponding scientific abstracts of Cochrane systematic reviews of oncology interventions. We then compared them to the readability and linguistic characteristics of the original PLSs, as well as with the conclusiveness of the corresponding scientific abstracts. The dataset included 275 PLSs and corresponding scientific abstracts of Cochrane systematic reviews about oncology interventions up to February 2019 from our previous study [21]. In that study, we assessed the language characteristics of PLSs of Cochrane systematic reviews of oncology interventions in comparison with corresponding Cochrane scientific abstracts. We used this dataset as it included the scientific abstracts for which the conclusiveness of the efficacy of interventions was already assessed, so that we could compare the conclusiveness of the AI-generated PLS with that of the original scientific abstracts from which it was created. Cochrane systematic reviews included in the dataset addressed came from Cochrane review groups focused solely on oncology: Breast Cancer; Childhood Cancer; Colorectal Cancer; Gynaecological, Neurooncology, and Orphan Cancer; Haematological Malignancies; and Lung Cancer. These groups served as representatives of different clinical cancer types. Systematic reviews that did not address intervention studies were excluded. Summaries from the dataset have been analyzed in terms of their Simple Measure of Gobbledygook (SMOG) index, linguistic characteristics (word

count and percentage of words related to different emotions), and the category of conclusiveness [21]. The full dataset is publicly accessible via the Open Science Framework [36].

We used the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist for reporting the results of this study.

Generation of PLSs by ChatGPT

We used a subscription chatbot, ChatGPT (version 4; Open AI) [37]. At the moment of our research, training data for ChatGPT included information up to April 2023.

We formed 3 prompts, and an author (JŠP) asked ChatGPT each prompt for each PLS separately. Before asking these questions, we asked ChatGPT to explain what a SMOG index is. ChatGPT correctly described the readability measure, its formula, and the interpretation of results, confirming that ChatGPT was adequately familiar with the SMOG index. We then used the following prompts (Figure 1):

1. Can you rewrite this Cochrane Scientific Abstract into a Cochrane Plain Language Summary so that your text has a SMOG index of 6?
2. Can you rewrite this Plain Language Summary so it is more emotional?
3. Can you rewrite this scientific abstract so it is simpler, easier to read, and more appropriate for the lay audience?

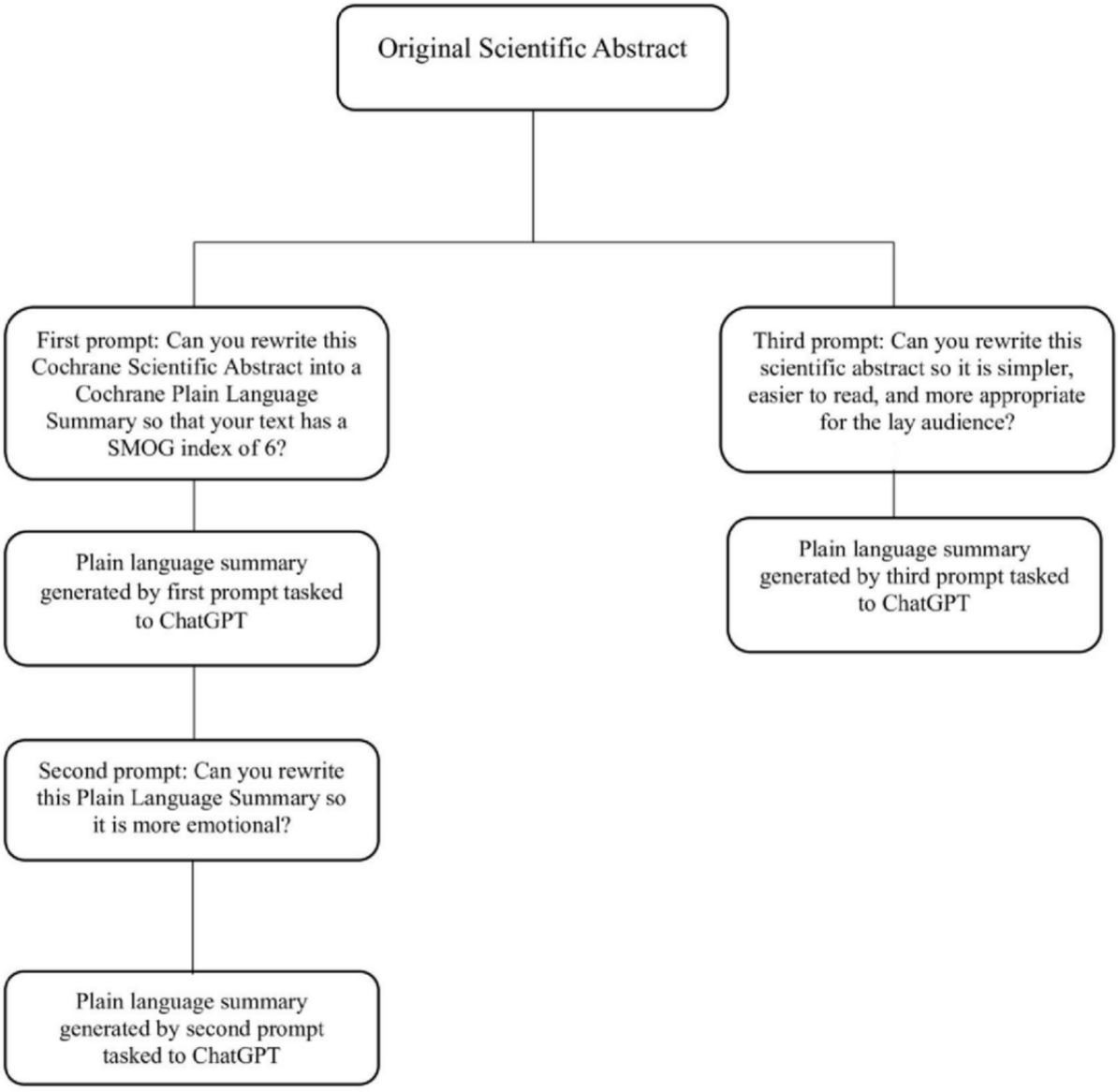
In the first prompt, we asked ChatGPT to rewrite the scientific abstract into PLS with a SMOG index of 6, because the official recommendation of the American Medical Association and National Health Institute is that the health information should be written at the reading level of the sixth grade in the US education system [16,38].

For the second prompt, we continued the conversation with ChatGPT that started in the first prompt. Since we found in our previous study that PLSs had language that was more emotional compared with the corresponding scientific abstracts and that PLSs that have a higher percentage of emotional tone have better readability [21], we used the PLS that was provided to ChatGPT in first prompt and asked ChatGPT to rewrite it again so that is more emotional, so this prompt was included to assess if adding emotional resonance could enhance reader engagement and relatability.

In prompt 3, similar to prompt 1, we asked ChatGPT to simplify the scientific abstract, but without defining the SMOG index. This prompt was designed to explore if using simpler language alone (without a specified readability index) would yield results with improved accessibility while retaining the essential information and nuance necessary for an accurate understanding of oncology topics. When tasking this prompt, we started a new chat with ChatGPT, so that it did not rely on the previous answer.

After that, for each PLS provided by ChatGPT, we measured its readability, expressed as SMOG index, and the following linguistic characteristics: word count and percentage of words related to authenticity, clout, emotional, and analytic tone. We analyzed only the first answers generated by ChatGPT for each of the 3 prompts and did not ask ChatGPT to revise the texts again. To be sure that the content of the PLS provided the same facts as the corresponding scientific abstract, we also checked the conclusiveness of each generated PLS, that is, checked whether there was any difference with the original conclusiveness category for the scientific abstract, determined in our previous study [21]. Multimedia Appendix 1 contains a supplementary table with examples of prompts.

Figure 1. Generation of Cochrane plain language summaries by ChatGPT. SMOG: Simple Measure of Gobbledygook.



SMOG Index

The readability of summary formats in English was assessed using the SMOG index [39]. SMOG index assesses the readability of certain content by counting polysyllabic words and the result is presented as the number of years of education required to understand a given text [39]. It is considered to be suitable for health information due to its consistent results, higher expected comprehension levels, application of recent validation criteria for estimating reading grade levels, and ease of use [39]. SMOG index for PLSs in English was calculated using a web-based tool "WebFX Readability Test Tool" [40]. The tool's reliability and accessibility made it a suitable choice for evaluating the readability of PLSs in our study. Regarding SMOG index interpretation, the official recommendation of the American Medical Association and National Health Institute is that health information should be written at the reading level of the sixth grade in the US education system [16,38].

Linguistic Characteristics

PLSs generated by ChatGPT were analyzed regarding their linguistic characteristics, using the Linguistic Inquiry and Word Count (LIWC) [41], a software tool designed to analyze a given text by comparing each word to its predefined dictionary. The tool categorizes text words into 4 main variables: analytical, clout, authenticity, and emotionality shared in the tone of the text, expressed as the percentage of words from the text in a particular category. The analytical thinking category is based on recognizing words associated with logic or connecting concepts and putting them into a relationship. Greater use of words related to analytical thinking is related to cognitive complexity and abstract thinking [26]. Clout speech is a variable that refers to the use of terms that denote self-confidence, leadership, or social status. A higher proportion of such words suggests that the author speaks from a position of expertise and certainty in what is stated, and a lower proportion suggests a style of presenting information that is humbler [42]. Authenticity is determined by the percentage of words related to personality,

such as the use of personal nouns in the first person (“I”, “my”, and “mine”), present tense, and relative adverbs (near, now). The use of these words is connected to writing that is more personal and honest [43]. Emotional share relates to how positive the tone is according to the words used. A score of 100 in emotional tone would mean the tone is maximally positive, while a score of 50 means an even balance of positive and negative emotion words [44].

Conclusiveness

The category of conclusiveness for each ChatGPT-generated PLS was analyzed by JŠP and checked by IB. Data extraction spreadsheet was tested by 2 authors (JŠP and IB). One author extracted the data, and the other one independently reviewed the data in a 10% random sample of PLSs and corresponding scientific abstracts. Then, it was checked whether the entry in the table was correct. Interobserver agreement was high (κ range 0.80 to 1.00, 95% CI 0.84 - 1.00). We resolved the differences in rating through the discussion with a third author (AM) before full data extraction.

The conclusiveness of statements about efficacy and safety was categorized into 3 categories [45]:

1. Conclusive: positive conclusive (there was moderate- or high-quality evidence indicating the effectiveness or safety of the intervention; ie, the drug was proven effective/safe); negative conclusive (there was moderate- or high-quality evidence indicating that the intervention is ineffective or harmful, or authors advised against the intervention/comparison or it is not recommended); or equal conclusive (the interventions analyzed were equally effective and safe).
2. Inconclusive: positive inconclusive (there was evidence suggesting effectiveness or safety, but it is of low quality or inconclusive, and the authors suggest that more research is needed); negative inconclusive (there was evidence of ineffectiveness or harm (evidence demonstrating that there was no effect or that the intervention was not safe) or authors urged against the intervention or comparison, or it is not recommended; however, the evidence is of low quality or inconclusive, or authors state that more research is needed); or equal inconclusive (the interventions appeared to be similarly effective and safe, but the evidence was of lower quality or inconclusive, and the authors suggest that more research is needed).
3. Unclear: no evidence (there was no evidence as the search did not retrieve any randomized controlled trials, ie, empty reviews); no opinion (the authors did not offer any opinion or judgment); and unclear (the authors did not give a clear conclusion or state that the more research is required).

Based on these criteria, we defined a category of conclusiveness for each of the derived PLS, and then the category of conclusiveness was compared with those from the original scientific abstracts to check whether they match and give the same conclusion about the effectiveness or safety of the intervention [21].

Statistical Analysis

Descriptive Statistics

The data on readability, word count, and linguistic characteristics were assessed as numeric variables. As the data deviated from normal distribution, the results were presented as medians and 95% CI and were presented on original PLS and across 3 prompt groups. The data on conclusiveness was assessed as frequencies and was presented on a bar chart across 3 different prompt groups.

Group Comparison

The results from the analysis of ChatGPT-generated PLSs were compared with the already published data for the original PLSs and scientific abstracts [21]. Since all versions were derived from the same PLS, the results were treated as within the subjects’ group under different conditions. Since the within-subjects ANOVA was not appropriate due to the deviations in the normality of data distribution, the comparison between groups was made using the Friedman nonparametric test for repeated measures, as nonparametric alternative and post hoc testing was made using the Conover post hoc test, since it is one of the recommended methods.

Statistical Software

Analyses were made using JASP (v.0.18.1.0; Jasp Team 2023) and R (v4.3.3; R core team, 2024) [46].

Ethical Considerations

The authors did not require ethical approval as the study is based solely on publicly available summaries of Cochrane systematic reviews. The research does not involve human participants or the use of animals. This is in accordance with the ethical code of the University of Split School of Medicine (April 2009).

Results

Overview

We generated a total of 275 PLSs for each of the 3 ChatGPT prompts. On average, all of them had statistically fewer words than the original PLSs (Table 1).

Table . Comparison of linguistic characteristics (median, 95% CI) between different plain language summary (PLS) groups.

Variable	Original[a], median (95% CI)	ChatGPT prompt, median (95% CI)			*P* value[b]
		First prompt: write from scientific abstract at SMOG[c] index 6	Second prompt: make PLS from first prompt more emotional	Third prompt: write simpler PLS from original PLS	
Word count	364 (339 - 388)	240 (232 - 248)	285 (278 - 292)	273 (266 - 278)	<.001
SMOG index[d]	13.1 (12.9 - 13.4)	8.2 (8.0 - 8.4)	11.4 (11.1 - 12)	8.7 (8.4 - 8.8)	<.001
Linguistic characteristics[e]					
Analytical tone	95.5 (95.0 - 95.8)	55.9 (53.6 - 57.9)[f]	85.7 (84.0 - 86.7)	60.9 (57.7 - 63.1)	<.001
Clout	50.0 (47.7 - 51.8)	67.2 (64.9 - 70.9)[f]	80.3 (77.8 - 83.6)	70.5 (68.0 - 72.8)	<.001
Authenticity	28.6 (26.2 - 31.3)	50.5 (47.0 - 53.5)[f]	38.0 (34.8 - 40.1)	49.4 (45.8 - 54.2)	<.001
Emotional tone	22.1 (18.5 - 26.2)	54.8 (51.3 - 58.7)[f,g]	63.9 (58.6 - 69.6)[f]	54.4 (51.9 - 56.4)	<.001

[a]The results for the original PLSs are from the previous study [21].

[b]Friedman nonparametric test. All post hoc differences were statistically significant except those labeled with symbols in superscript.

[c]SMOG: Simple Measure of Gobbledygook.

[d]Readability was measured as a SMOG index [39]. Higher scores indicated lower readability.

[e]Linguistic characteristics of the text were measured using dictionary-based text word categorizations by the Linguistic Inquiry and Word Count (LIWC) [26]. The variables are presented as the percentage of words from the text in a particular category.

[f]Not different from prompt 3.

[g]Not different from prompt 2.

SMOG Index and Linguistic Characteristics

PLSs generated by the first prompt (write at SMOG index 6) were the easiest to read, with the median SMOG index of 8.2 (95% CI 8 - 8.4), and the shortest. Regarding linguistic characteristics, these PLSs had less analytical tone and more authenticity, clout, and emotional tone when compared with the original PLSs written by authors (Table 1).

PLSs generated by the second prompt (make prompt 1 more emotional), had a median SMOG index of 11.4 (95% CI 11.1 - 12). Those PLSs also had more analytical tone and clout compared with the PLSs generated by the first prompt, but no difference in the emotional tone. They also used fewer words related to authenticity than the PLSs generated by the first prompt (Table 1).

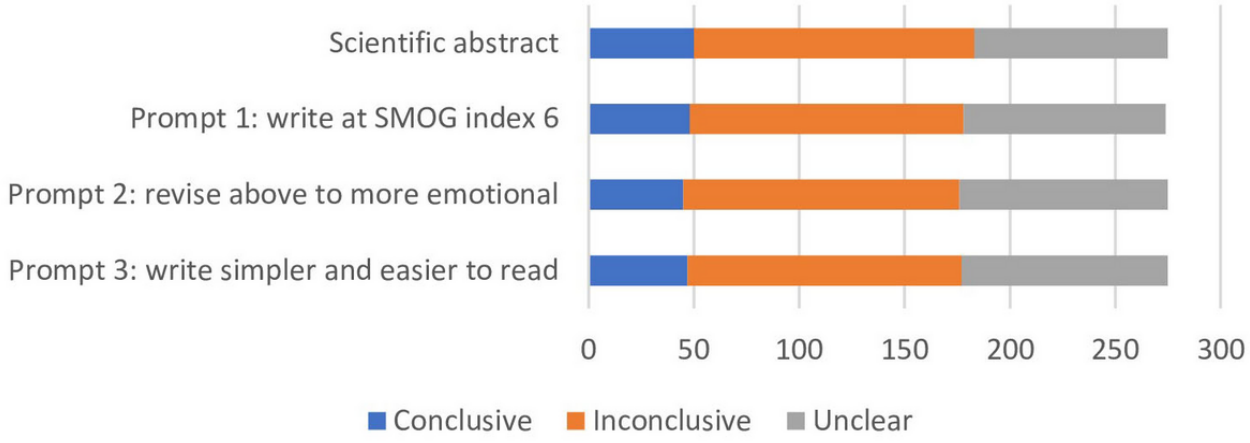
PLSs generated by the third prompt (write simpler PLS from original) had a median SMOG index of 8.7 (95% CI 8.4 - 8.8). Linguistic characteristics did not differ from the PLSs generated by the first prompt, but they had a less analytical and more authentic tone than the PLSs generated by the second prompt (Table 1).

Across the 3 GPT prompts, the results were consistent, without major outliers.

Conclusiveness

The category of conclusiveness of all 3 ChatGPT-generated PLSs did not differ from that of the original scientific abstract (Figure 2).

Figure 2. Distribution of plain language summaries according to the conclusiveness of the efficiency of interventions described in the systematic reviews across 4 different groups of writing prompts. SMOG: Simple Measure of Gobbledygook.



Discussion

Principal Findings

Our study investigated the potential of using an LLM-ChatGPT, to generate more readable and engaging PLSs for Cochrane systematic reviews in oncology. The results demonstrate that ChatGPT-generated PLSs were shorter and easier to read, as illustrated by the SMOG indices across different prompts. Across all prompts, ChatGPT-generated PLSs exhibited a decrease in analytical tone while showing higher levels of authenticity, clout, and emotional tone compared with the original PLSs. Notably, the categorization of conclusiveness in the original abstracts remained consistent in the ChatGPT-generated PLSs.

Specifically, PLSs created by the first prompt (targeting a SMOG index of 6) achieved the lowest SMOG index and the highest readability among the 3 approaches we used. The median SMOG index of the PLSs generated through prompt 1, while higher than the targeted level of 6, was still significantly lower than that of the original PLSs authored by researchers (median SMOG index 13.1).

The second prompt, which aimed to make the summaries contain text with more emotionally positive content, resulted in an increase in SMOG index and a notable shift toward analytical and clout tones, indicating that the addition of emotional language might inadvertently increase linguistic complexity. This observation is crucial because previous studies have shown that emotional resonance can improve reader engagement and comprehension [21], yet it must be balanced with readability to avoid overcomplicating the content.

The third prompt again managed to improve the readability of the summaries. Despite differences in readability and tone, the conclusiveness about the efficacy of the intervention in ChatGPT-generated PLSs remained consistent with that of the original scientific abstracts. This consistency is encouraging, as it suggests that ChatGPT can maintain the integrity of the original scientific conclusions while rephrasing content for a lay audience.

PLSs generated by ChatGPT had language characteristics more suitable for the lay population than the original PLSs from the published Cochrane systematic reviews [21]. They had fewer words and better readability—from the median of 8.2 to 11.2 compared with the original 13.1 SMOG index, bringing it closer to the recommended sixth-grade reading level for the health information intended for the lay public [16]. However, it is not clear whether we can expect at all for PLSs from oncology to be at the recommended reading level, as it is not sometimes possible to replace complex scientific expressions or names of the drugs (eg, trastuzumab-deruxtecan) without altering the meaning and jeopardizing the translation of the information to the reader in the correct way.

Regarding the linguistic content of the PLSs [16,42-44], those generated by ChatGPT had lower content with an analytical tone, meaning that they did not relate so much to abstract thinking or cognitive complexity as the original PLSs. They also had a higher positive emotional tone (a score over 50 for all ChatGPT-generated PLSs) than the original PLSs, which had a predominantly negative emotional tone (median of 22 out of maximum 100). ChatGPT-generated PLSs also had higher clout, meaning that the information came from the position of expertise and certainty, and they used more personal nouns, present tense, and relative adverbs, increasing authenticity tone, that is, making it more personal and honest. In this way, ChatGPT-generated PLSs could influence the subjective experience of people and their engagement in the given text when they process the information in the PLSs with respect to their opinion about the truthfulness of the information or their confidence in the information [25]. These cognitive processes are very important in reacting to health information, as it has been shown that patients adhere more to advice from doctors that contains more positive emotions [47].

Comparison to Prior Studies

ChatGPT-generated PLSs were closer in their characteristics to the press releases of Cochrane systematic reviews, written by professional writers. A study of Cochrane systematic reviews that had an official press release showed that these press releases were written in a more conversational and emotional language than the scientific abstracts or PLSs in different languages,

making them more engaging [22]. ChatGPT-generated PLSs had similar qualities, without losing the conclusiveness of the message, making them more suitable for health evidence translation to the patients and the general public.

Strength and Limitations

Regarding the strengths of the study, this study is among the first to evaluate the use of AI for generating PLSs in oncology, focusing on readability, linguistic characteristics, and consistency with original scientific conclusions. The use of multiple prompts provided a nuanced understanding of how prompt design influences AI-generated content.

These results must be considered in light of several limitations. First, we used PLSs from a single source, the Cochrane Library, and these PLSs were written by different authors. However, summaries from the Cochrane Library have the same format of presenting health information and specific guidance for writing PLSs [48], making them comparable. Second, we analyzed the PLSs in English only since it was the only common language for the summaries in oncology systematic reviews. The focus on English ensured uniformity in linguistic analysis, avoiding inconsistencies in translation processes. Third, a notable limitation of this study is that it required ChatGPT to generate PLSs from scientific abstracts and not from the complete texts of Cochrane systematic reviews. Typically, PLSs are derived from the full content of the reviews [48], which provides a more thorough understanding of the study's findings, methodologies, and contextual factors. Reliance only on scientific abstracts may lead to PLSs lacking depth and detail. Fourth, ChatGPT-4 was developed using a diverse dataset of publicly available information spanning multiple domains and at the time of our study, its information is limited to publicly available sources up to 2021 [49]. OpenAI does not specify the exact content of medical information included or the precise time frame of the dataset [49]. We used ChatGPT-4, the subscription version (unlike its predecessor ChatGPT-3.5), the most advanced and widely available version of ChatGPT during the study period, ensuring access to its latest capabilities, but it is available only to people willing to pay a monthly subscription. For the second prompt in our study, we relied on the answers from ChatGPT provided in the first prompt. For all 3 prompts, we did not ask the system to further rephrase the text but analyzed only the first output. Different AI models vary in training, algorithms, and capabilities, making the use and results of one model not universally applicable to others [50]. We did not use other AI tools such as Microsoft Bing AI, Bard, Jasper, or ChatSonic, which could have given different results. Microsoft Bing AI can only process up to 2000 characters [51], which is not suitable for summaries. Bard was not available in Croatia at the time of conducting the study [52], and Jasper and ChatSonic offered only paid subscriptions for which we did not have sufficient resources [53,54]. Additionally, we did not use specialized AI tools designed specifically for creating PLS, such as Sorcero's solution [55] or Putnam Associates' generative AI approach for PLSs [56]. Our decision not to use these specialized tools stemmed from a focus on general AI models that are more accessible to a wider range of users and our intent to evaluate the performance of widely available, nonspecialized AI for PLS generation. Fifth, it has to be kept in mind that, while our study assessed linguistic characteristics such as clout, analytical tone, authenticity, and emotional tone, it is equally important to consider the cultural and emotional sensitivities of the target audience [57]. AI models like ChatGPT-4 are trained on extensive datasets that may not fully capture the nuances of various cultural backgrounds. Consequently, the generated PLSs might lack the cultural relevance or sensitivity necessary to effectively communicate with all segments of the lay public. Further research should include PLSs from multiple sources to assess the generalizability of AI-generated PLSs across diverse formats and writing styles and explore the potential of AI tools in generating PLSs in languages other than English, to support Cochrane efforts to provide health information in 20 different languages [58].

Future Directions

What conclusions and recommendations can be drawn from our study? Our study is one of the first tests of AI language tools in creating health information from complex health evidence synthesis that is suitable for the lay public. The real-world implementation of medical AI interventions generally lacks high-quality evidence, as a recent systematic review identified 65 randomized controlled trials evaluating AI interventions, but only 7 with chatbots as an intervention [59]. We do not think that, at the current level of development, ChatGPT can replace evidence synthesis in real time, as it does not get updates in real time [60] and there is a potential for bias in the training data, which can result in biased or inaccurate responses [61]. In addition, text generated from the training data can have several other issues besides bias, such as plagiarism [62], lack of context, as well as underestimation of novelties in medicine that are important but may be less represented in web sources [63].

On the other hand, ChatGPT and other generative AI tools may be useful in ensuring the quality and appropriateness of the summary information for health evidence synthesis, such as Cochrane systematic reviews. Although Cochrane has clear guidance on writing PLSs [48], the evidence shows that they are not adequately implemented and published PLSs are not at the desired level of clarity and quality [64]. It seems that the authors of Cochrane systematic reviews have difficulties in translating their results into a language that is suitable for the lay public. This affects not only the usefulness of the PLSs but also their translations into a number of languages, where it is not clear whether there is a further loss to the clarity and understandability of the message to the lay public [22]. Readability metrics like the SMOG index provide an indication of text complexity but do not guarantee comprehension by the intended audience. This gap highlights the need for future research to evaluate the effectiveness of AI-generated PLSs in real-world settings and to determine their actual understandability among diverse patient populations.

Conclusions

Having all that in mind, ChatGPT may be a valuable tool in helping create content designed for the lay public. Cochrane should further explore the use of ChatGPT in generating PLSs, either as a tool for the authors or as an independent, systemic tool to generate high-quality, high-fidelity PLSs, also ensuring

that the main message of health information is unchanged and accurate.

Acknowledgments

This study was funded by the Croatian Science Foundation ("Professionalism in Health - Decision making in practice and research, ProDeM") under grant agreement number IP-2019-04-4882. The funder had no role in the design of this study, its execution, analyses, interpretation of the data, or decision to submit results.

Data Availability

The full summarized data generated or analyzed during this study are included in this published article and its supplementary information files. The datasets extraction sheets generated or analyzed during this study are available in the Open Science Framework repository [36].

Authors' Contributions

AM gave the idea for the study, and all authors contributed to the study plan. JŠP collected the data. IB analyzed the data and interpreted the results. JŠP wrote the first draft of the manuscript, and all authors critically revised the manuscript. All authors agreed to the version submitted for publication and take responsibility for the accountability of the study.

Conflicts of Interest

The authors declare no conflicts of interest, except for their membership in Cochrane.

Multimedia Appendix 1
Example sheet.
[XLSX File, 15 KB - cancer_v11i1e63347_app1.xlsx]

Checklist 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.
[DOCX File, 34 KB - cancer_v11i1e63347_app2.docx]

References

1. Institute of Medicine (US) Committee on Health Literacy. In: Nielsen-Bohlman L, Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Health Literacy: A Prescription to End Confusion: National Academies Press (US); 2004. [doi: 10.17226/10883]
2. van der Gaag M, Heijmans M, Spoiala C, Rademakers J. The importance of health literacy for self-management: a scoping review of reviews. Chronic Illn 2022 Jun;18(2):234-254. [doi: 10.1177/17423953211035472] [Medline: 34402309]
3. Berkman ND, Sheridan SL, Donahue KE, et al. Health literacy interventions and outcomes: an updated systematic review. Evid Rep Technol Assess (Full Rep) 2011 Mar(199):1-941. [Medline: 23126607]
4. Nutbeam D, Kickbusch I. Advancing health literacy: a global challenge for the 21st century. Health Promot Int 2000 Sep 1;15(3):183-184. [doi: 10.1093/heapro/15.3.183]
5. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80. [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
6. Plain writing at CDC. Centers for Disease Control and Prevention (CDC). 2024. URL: https://www.cdc.gov/other/plainwriting.html [accessed 2024-07-15]
7. Plain language at NIH. National Institutes of Health (NIH). URL: https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication/plain-language [accessed 2024-07-15]
8. Koay K, Schofield P, Jefford M. Importance of health literacy in oncology. Asia Pac J Clin Oncol 2012 Mar;8(1):14-23. [doi: 10.1111/j.1743-7563.2012.01522.x] [Medline: 22369440]
9. Chua GP, Tan HK, Gandhi M. Information sources and online information seeking behaviours of cancer patients in Singapore. Ecancermedicalscience 2018;12:880. [doi: 10.3332/ecancer.2018.880] [Medline: 30483360]
10. Carlsson M. Cancer patients seeking information from sources outside the health care system. Support Care Cancer 2000 Nov;8(6):453-457. [doi: 10.1007/s005200000166] [Medline: 11094989]
11. Bala S, Keniston A, Burden M. Patient perception of plain-language medical notes generated using artificial intelligence software: pilot mixed-methods study. JMIR Form Res 2020 Jun 5;4(6):e16670. [doi: 10.2196/16670] [Medline: 32442148]
12. Gold DT, McClung B. Approaches to patient education: emphasizing the long-term value of compliance and persistence. Am J Med 2006 Apr;119(4 Suppl 1):S32-S37. [doi: 10.1016/j.amjmed.2005.12.021] [Medline: 16563940]
13. Cho YI, Lee SYD, Arozullah AM, Crittenden KS. Effects of health literacy on health status and health service utilization amongst the elderly. Soc Sci Med 2008 Apr;66(8):1809-1816. [doi: 10.1016/j.socscimed.2008.01.003] [Medline: 18295949]

14. Murray MD, Tu W, Wu J, Morrow D, Smith F, Brater DC. Factors associated with exacerbation of heart failure include treatment adherence and health literacy skills. Clin Pharmacol Ther 2009 Jun;85(6):651-658. [doi: 10.1038/clpt.2009.7] [Medline: 19262464]
15. Kieffer Campbell J. Health literacy in adult oncology: an integrative review. Oncol Nurs Forum 2020 Jan 1;47(1):18-32. [doi: 10.1188/20.ONF.18-32] [Medline: 31845912]
16. Eltorai AEM, Ghanian S, Adams CA Jr, Born CT, Daniels AH. Readability of patient education materials on the American Association for Surgery of Trauma website. Arch Trauma Res 2014 Jun;3(2):e18161. [doi: 10.5812/atr.18161] [Medline: 25147778]
17. Friedman DB, Hoffman-Goetz L, Arocha JF. Readability of cancer information on the internet. J Cancer Educ 2004;19(2):117-122. [doi: 10.1207/s15430154jce1902_13] [Medline: 15456669]
18. Scholten R, Clarke M, Hetherington J. The Cochrane Collaboration. Eur J Clin Nutr 2005 Aug;59 Suppl 1:S147-S149. [doi: 10.1038/sj.ejcn.1602188] [Medline: 16052183]
19. Henderson LK, Craig JC, Willis NS, Tovey D, Webster AC. How to write a Cochrane systematic review. Nephrology (Carlton) 2010 Sep;15(6):617-624. [doi: 10.1111/j.1440-1797.2010.01380.x] [Medline: 20883282]
20. Banić A, Fidahić M, Šuto J, et al. Conclusiveness, linguistic characteristics and readability of Cochrane plain language summaries of intervention reviews: a cross-sectional study. BMC Med Res Methodol 2022 Sep 10;22(1):240. [doi: 10.1186/s12874-022-01721-7] [Medline: 36088293]
21. Šuto J, Marušić A, Buljan I. Linguistic analysis of plain language summaries and corresponding scientific summaries of Cochrane systematic reviews about oncology interventions. Cancer Med 2023 May;12(9):10950-10960. [doi: 10.1002/cam4.5825] [Medline: 36951519]
22. Karačić J, Dondio P, Buljan I, Hren D, Marušić A. Languages for different health information readers: multitrait-multimethod content analysis of Cochrane systematic reviews textual summary formats. BMC Med Res Methodol 2019 Apr 5;19(1):75. [doi: 10.1186/s12874-019-0716-x] [Medline: 30953453]
23. Langendam MW, Akl EA, Dahm P, Glasziou P, Guyatt G, Schünemann HJ. Assessing and presenting summaries of evidence in Cochrane reviews. Syst Rev 2013 Sep 23;2:81. [doi: 10.1186/2046-4053-2-81] [Medline: 24059250]
24. Glenton C, Santesso N, Rosenbaum S, et al. Presenting the results of Cochrane Systematic Reviews to a consumer audience: a qualitative study. Med Decis Making 2010;30(5):566-577. [doi: 10.1177/0272989X10375853] [Medline: 20643912]
25. Alter AL, Oppenheimer DM. Uniting the tribes of fluency to form a metacognitive nation. Pers Soc Psychol Rev 2009 Aug;13(3):219-235. [doi: 10.1177/1088868309341564] [Medline: 19638628]
26. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language. use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
27. Totlis T, Natsis K, Filos D, et al. The potential role of ChatGPT and artificial intelligence in anatomy education: a conversation with ChatGPT. Surg Radiol Anat 2023 Oct;45(10):1321-1329. [doi: 10.1007/s00276-023-03229-1] [Medline: 37584720]
28. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595. [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]
29. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98. [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
30. Suleimenov IE, Vitulyova YS, Bakirov AS, Gabrielyan OA. Artificial intelligence: what is it? Presented at: Proceedings of the 2020 6th International Conference on Computer and Technology Applications; Apr 14, 2020; Antalya Turkey p. 22-25. [doi: 10.1145/3397125.3397141]
31. Liu S, Wright AP, Patterson BL, et al. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. J Am Med Inform Assoc 2023 Jun 20;30(7):1237-1245. [doi: 10.1093/jamia/ocad072] [Medline: 37087108]
32. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
33. Carlbring P, Hadjistavropoulos H, Kleiboer A, Andersson G. A new era in internet interventions: the advent of Chat-GPT and AI-assisted therapist guidance. Internet Interv 2023 Apr;32:100621. [doi: 10.1016/j.invent.2023.100621] [Medline: 37273936]
34. The Lancet Digital Health. Large language models: a new chapter in digital health. Lancet Digit Health 2024 Jan;6(1):e1. [doi: 10.1016/S2589-7500(23)00254-6] [Medline: 38123249]
35. Ayre J, Mac O, McCaffery K, et al. New frontiers in health literacy: using ChatGPT to simplify health information for people in the community. J Gen Intern Med 2024 Mar;39(4):573-577. [doi: 10.1007/s11606-023-08469-w] [Medline: 37940756]
36. Suto Pavicic J, Marusic A, Buljan I. Using chatgpt to improve the presentation of plain language summaries of cochrane systematic reviews about oncology interventions: cross-sectional study [Data set]. : Open Science Framework; 2024. [doi: 10.17605/OSF.IO/92TNV]
37. ChatGPT. URL: https://chatgpt.com/ [accessed 2024-11-04]

38. Clear communication. National Institutes of Health (NIH).: U.S. Department of Health and Human Services, National Institutes of Health URL: https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication [accessed 2024-07-15]
39. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. Res Social Adm Pharm 2013;9(5):503-516. [doi: 10.1016/j.sapharm.2012.05.009] [Medline: 22835706]
40. Readability test tool. WebFX. URL: https://www.webfx.com/tools/read-able/ [accessed 2024-02-24]
41. Introducing LIWC-22. LIWC. URL: https://www.liwc.app/ [accessed 2024-01-25]
42. Moore RL, Yen CJ, Powers FE. Exploring the relationship between clout and cognitive processing in MOOC discussion forums. Brit J Educational Tech 2021 Jan;52(1):482-497. [doi: 10.1111/bjet.13033]
43. Lehman DW, O'Connor K, Kovács B, Newman GE. Authenticity. ANNALS 2019 Jan;13(1):1-42. [doi: 10.5465/annals.2017.0047]
44. Cohn MA, Mehl MR, Pennebaker JW. Linguistic markers of psychological change surrounding September 11, 2001. Psychol Sci 2004 Oct;15(10):687-693. [doi: 10.1111/j.0956-7976.2004.00741.x] [Medline: 15447640]
45. Bralić N, Mijatović A, Marušić A, Buljan I. Conclusiveness, readability and textual characteristics of plain language summaries from medical and non-medical organizations: a cross-sectional study. Sci Rep 2024 Mar 12;14(1):6016. [doi: 10.1038/s41598-024-56727-6] [Medline: 38472285]
46. A fresh way to do statistics. JASP (version 017). 2023. URL: https://jasp-stats.org/ [accessed 2023-10-10]
47. Falkenstein A, Tran B, Ludi D, et al. Characteristics and correlates of word use in physician-patient communication. Ann Behav Med 2016 Oct;50(5):664-677. [doi: 10.1007/s12160-016-9792-x] [Medline: 26968166]
48. New standards for plain language summaries. Cochrane Consumer Network. URL: https://consumers.cochrane.org/PLEACS [accessed 2023-06-12]
49. Chelli M, Descamps J, Lavoué V, et al. Hallucination rates and reference accuracy of ChatGPT and bard for systematic reviews: comparative analysis. J Med Internet Res 2024 May 22;26:e53164. [doi: 10.2196/53164] [Medline: 38776130]
50. Zhou H, Liu F, Gu B, et al. A survey of large language models in medicine: progress, application, and challenge. arXiv. Preprint posted online on 2023 URL: https://arxiv.org/abs/2311.05112 [accessed 2024-07-25]
51. Okemwa K. Microsoft bumps Bing Chat's character limit to 4,000. OnMSFT.com. 2023. URL: https://onmsft.com/news/bing-chat-character-limit-bumped-to-4000/ [accessed 2024-03-29]
52. Manyika J. Being bold on AI means being responsible from the start. Google AI. 2023. URL: https://blog.google/technology/ai/google-responsible-ai-io-2023/ [accessed 2024-03-29]
53. Paid subscription jasper AI losing momentum to freemium AI tools like chatgpt. Earnest Analytics. 2023. URL: https://www.earnestanalytics.com/insights/paid-subscription-jasper-ai-losing-momentum-to-freemium-ai-tools-like-chatgpt [accessed 2024-02-22]
54. Cetojevic V. Chatsonic review: can it compete with chatgpt in 2025. MSPoweruser. 2024. URL: https://mspoweruser.com/chatsonic-review/ [accessed 2024-09-15]
55. Illuminate actionable insights from complex medical data. Sorcero. URL: https://www.sorcero.com [accessed 2024-06-12]
56. AI at putnam. Putnam Associates. URL: https://www.putassoc.com/our-expertise/ai-at-putnam [accessed 2024-10-12]
57. Schouten BC, Meeuwesen L. Cultural differences in medical communication: a review of the literature. Patient Educ Couns 2006 Dec;64(1-3):21-34. [doi: 10.1016/j.pec.2005.11.014] [Medline: 16427760]
58. Cochrane evidence in different languages. Cochrane. URL: https://www.cochrane.org/news/cochrane-evidence-different-languages [accessed 2023-12-01]
59. Martindale APL, Llewellyn CD, de Visser RO, et al. Concordance of randomised controlled trials for artificial intelligence interventions with the CONSORT-AI reporting guidelines. Nat Commun 2024;15(1):1619. [doi: 10.1038/s41467-024-45355-3]
60. Sallam M, Salim NA, Barakat M, Al-Tammemi AB. ChatGPT applications in medical, dental, pharmacy, and public health education: a descriptive study highlighting the advantages and limitations. Narra J 2023 Apr;3(1):e103. [doi: 10.52225/narra.v3i1.103] [Medline: 38450035]
61. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 4;47(1):33. [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
62. Huang J, Tan M. The role of ChatGPT in scientific communication: writing better scientific review articles. Am J Cancer Res 2023;13(4):1148-1154. [Medline: 37168339]
63. Dahmen J, Kayaalp ME, Ollivier M, et al. Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. Knee Surg Sports Traumatol Arthrosc 2023 Apr;31(4):1187-1189. [doi: 10.1007/s00167-023-07355-6]
64. Jelicic Kadic A, Fidahic M, Vujcic M, et al. Cochrane plain language summaries are highly heterogeneous with low adherence to the standards. BMC Med Res Methodol 2016 May 23;16:61. [doi: 10.1186/s12874-016-0162-y] [Medline: 27216616]

Abbreviations:

AI: artificial intelligence
LIWC: Linguistic Inquiry and Word Count
LLM: large language model
PLS: plain language summary
SMOG: Simple Measure of Gobbledygook
STOBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by N Cahill; submitted 17.06.24; peer-reviewed by B Rasuli, K Gainey; revised version received 22.01.25; accepted 27.01.25; published 19.03.25.

Please cite as:
Šuto Pavičić J, Marušić A, Buljan I
Using ChatGPT to Improve the Presentation of Plain Language Summaries of Cochrane Systematic Reviews About Oncology Interventions: Cross-Sectional Study
JMIR Cancer 2025;11:e63347
URL: https://cancer.jmir.org/2025/1/e63347
doi:10.2196/63347

Assessing the Quality and Reliability of ChatGPT's Responses to Radiotherapy-Related Patient Queries: Comparative Study With GPT-3.5 and GPT-4

Ana Grilo[1], PhD, MSc; Catarina Marques[2], BSc; Maria Corte-Real[2], BSc; Elisabete Carolino[1], PhD; Marco Caetano[2], MSc

[1]Research Center for Psychological Science of the Faculty of Psychology, University of Lisbon to CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Av. D. João II, Lote 4.69.01, Parque das Nações, Lisboa, Portugal

[2]Escola Superior de Tecnologia da Saúde, Instituto Politécnico de Lisboa, Lisboa, Portugal

Corresponding Author:
Ana Grilo, PhD, MSc
Research Center for Psychological Science of the Faculty of Psychology, University of Lisbon to CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Av. D. João II, Lote 4.69.01, Parque das Nações, Lisboa, Portugal

Abstract

Background: Patients frequently resort to the internet to access information about cancer. However, these websites often lack content accuracy and readability. Recently, ChatGPT, an artificial intelligence–powered chatbot, has signified a potential paradigm shift in how patients with cancer can access vast amounts of medical information, including insights into radiotherapy. However, the quality of the information provided by ChatGPT remains unclear. This is particularly significant given the general public's limited knowledge of this treatment and concerns about its possible side effects. Furthermore, evaluating the quality of responses is crucial, as misinformation can foster a false sense of knowledge and security, lead to noncompliance, and result in delays in receiving appropriate treatment.

Objective: This study aims to evaluate the quality and reliability of ChatGPT's responses to common patient queries about radiotherapy, comparing the performance of ChatGPT's two versions: GPT-3.5 and GPT-4.

Methods: We selected 40 commonly asked radiotherapy questions and entered the queries in both versions of ChatGPT. Response quality and reliability were evaluated by 16 radiotherapy experts using the General Quality Score (GQS), a 5-point Likert scale, with the median GQS determined based on the experts' ratings. Consistency and similarity of responses were assessed using the cosine similarity score, which ranges from 0 (complete dissimilarity) to 1 (complete similarity). Readability was analyzed using the Flesch Reading Ease Score, ranging from 0 to 100, and the Flesch-Kincaid Grade Level, reflecting the average number of years of education required for comprehension. Statistical analyses were performed using the Mann-Whitney test and effect size, with results deemed significant at a 5% level (P=.05). To assess agreement between experts, Krippendorff α and Fleiss κ were used.

Results: GPT-4 demonstrated superior performance, with a higher GQS and a lower number of scores of 1 and 2, compared to GPT-3.5. The Mann-Whitney test revealed statistically significant differences in some questions, with GPT-4 generally receiving higher ratings. The median (IQR) cosine similarity score indicated substantial similarity (0.81, IQR 0.05) and consistency in the responses of both versions (GPT-3.5: 0.85, IQR 0.04; GPT-4: 0.83, IQR 0.04). Readability scores for both versions were considered college level, with GPT-4 scoring slightly better in the Flesch Reading Ease Score (34.61) and Flesch-Kincaid Grade Level (12.32) compared to GPT-3.5 (32.98 and 13.32, respectively). Responses by both versions were deemed challenging for the general public.

Conclusions: Both GPT-3.5 and GPT-4 demonstrated having the capability to address radiotherapy concepts, with GPT-4 showing superior performance. However, both models present readability challenges for the general population. Although ChatGPT demonstrates potential as a valuable resource for addressing common patient queries related to radiotherapy, it is imperative to acknowledge its limitations, including the risks of misinformation and readability issues. In addition, its implementation should be supported by strategies to enhance accessibility and readability.

(JMIR Cancer 2025;11:e63677) doi:10.2196/63677

KEYWORDS

artificial intelligence; ChatGPT; large language model; radiotherapy; patient information; quality; internet access; health information; cancer awareness; accuracy; readability; chatbot; patient query; chat generative pretrained transformer; OpenAI; natural language processing; patients with cancer

Introduction

In an increasingly digitized society, patients frequently resort to the internet to access information about cancer [1-3]. However, despite being one of the most favored informational modalities, websites often require more content accuracy and better readability [1].

Recently, artificial intelligence (AI)–powered chatbots such as ChatGPT have signified a potential paradigm shift in how patients with cancer can access a vast amount of medical information [1,3,4]. The rise of these AI platforms, accessible to the general public, has escalated notably since OpenAI released version 3.5 of ChatGPT (GPT-3.5) on November 30, 2022 [4-13], which amassed over 1 billion users in March 2023 [4].

ChatGPT, a large language model (LLM) [6,14-17], uses natural language processing to offer varied responses to the same query considering the context of the conversation and individual user preferences [18]. Through text-to-text communication, ChatGPT can engage with humans [12] and aims to deliver responses resembling human interactions [6,14,18]. This model has undergone extensive training on a diverse corpus of text data encompassing a broad spectrum of sources, including books, scholarly articles, and web pages, enabling it to effectively comprehend and respond to natural language queries across a broad range of topics [19,20]. Moreover, the model's performance is enhanced through reinforcement learning from human feedback, which enables it to produce more coherent and contextually relevant responses [21]. Additionally, ChatGPT can compose emails, essays, and medical reports, as well as solve problems and provide clarification [10,13,22,23].

On March 14, 2023, OpenAI announced the release ChatGPT-4 (GPT-4), which became available through a subscription-based model [9,12,16]. This new version demonstrated outstanding performance across numerous academic and professional benchmarks, providing more refined and varied responses than GPT-3.5 [24].

In this context, ChatGPT has emerged as a contender for traditional search engines, such as Google, because of its capacity to filter vast quantities of data and provide easily comprehensible responses [4,6]. Consequently, ChatGPT is a potentially reliable source of medical information to both the public and patients with cancer, and it is capable of offering insights regarding radiotherapy [4,25]. This is particularly significant given the general public's limited knowledge of this treatment [15,26] and concerns regarding its possible side effects [27].

Radiotherapy is a well-established treatment that delivers targeted ionizing radiation with precision with the aim of destroying cancer cells while minimizing damage to healthy tissues. Approximately half of all patients diagnosed with cancer undergo radiotherapy as a part of their care. Advances in radiotherapy have increased its complexity, requiring greater preparation and support for patients who may face physical and psychological challenges [28]. Considering that approximately 80% of patients have limited knowledge regarding radiotherapy and associated expectations regarding treatment, many have significant misconceptions. These commonly include concerns about radiation burns or the possibility of becoming radioactive as a result of the treatment [29,30]. Such misunderstandings, coupled with the unfamiliarity of radiotherapy for most patients and the inherent invisibility of the treatment, further complicate their ability to fully comprehend the process [31,32]. Therefore, providing clear and accessible information is essential for reducing patients' fear of treatment [31]. Previous studies have explored radiotherapy educational resources, such as videos, and tested group education in radiotherapy settings. However, these studies did not specifically address individual patient education and support needs at key time points [33]. Alternatively, written documentation has proven effective for patients who may feel overwhelmed by excessive verbal information, as it allows them to process the material at their own pace and share it with family and friends [34]. Therefore, ChatGPT offers a convenient and accessible method for patients to obtain written information and support [19].

Given that patient education is particularly crucial for patients with cancer because of the complexity of their treatment pathways [20], providing them with comprehensive information about radiotherapy at appropriate stages may enhance adherence to the treatment plan, because inadequate information can lead to increased uncertainty, unnecessary anxiety, and distress among patients and their families [27,35,36]. Additionally, poorly informed patients are likely to be dissatisfied with their care, have difficulty coping [35], and have many follow-up questions regarding the treatment process. Moreover, patients with cancer often feel uncomfortable discussing their body image and sexual health with their clinicians. Consequently, patient communication with ChatGPT may lower these barriers [36].

However, given that ChatGPT was not explicitly trained for oncology-related inquiries, the quality of the information it provides remains unverified [7,14,36]. Evaluating the quality of responses is crucial, as misinformation can foster a false sense of knowledge and security, lead to noncompliance, and result in delays in receiving appropriate treatment [4,14,15]. Nevertheless, various limitations of ChatGPT have been identified. It has been observed to fall below the expected educational level [4], as health-related materials intended for patient consumption are typically recommended to have a reading level equivalent to fifth and sixth grades [4,37]. Furthermore, the training data for GPT-3.5 are outdated, limited to the information available up until September 2021, and lack access to newer knowledge beyond that date [5,38,39]. To address this constraint, GPT-4 introduces a novel feature that allows the use of external plug-ins [25]. However, this new version is available exclusively through paid subscription [9,12,16]. Additionally, ChatGPT tends to provide unreliable or inaccurate information, potentially generating incorrect or misleading responses [14,16]. This issue often arises from the dependence of models on their training data, which may not always be up-to-date or fully comprehensive [40].

To date, limited research has been conducted on the application of language models in the medical domain and the effectiveness of ChatGPT in patient education remains indeterminate [14].

Although the literature addressing ChatGPT's capabilities has proliferated in recent months, there remains a lack of data regarding the quality and reliability of the responses it provides [11,18]. This gap underscores the necessity for more comprehensive studies to evaluate the performance of language models, including ChatGPT, in the medical context. Ensuring that these models are equipped with the most current and comprehensive data is essential for their effective application in radiotherapy health care.

This study aimed to evaluate the quality and reliability of ChatGPT responses to common patient queries regarding radiotherapy to ascertain its potential as a reliable source of patient information. Additionally, it aimed to compare the performance of GPT-3.5 with GPT-4 in generating responses to the same radiotherapy queries.

Methods

Prompt Generation

To determine the most common patient queries regarding radiotherapy, an assessment was conducted using articles that addressed topics related to the most relevant patient concerns. These served as the foundation for the development of 128 questions, 90 of which were derived from the studies by Halkett et al [27,35,41] and Zeguers et al [32], whereas the remaining 38 were sourced from the National Cancer Institute [42]. The questions were then organized into a table to facilitate the identification of duplicates and the selection of the most pertinent ones. There were 36 questions identified as duplicates, and 43 were deemed specific to certain pathologies or specialized treatments, leaving a total of 49 questions. Four authors (AG, CM, MC-R, and MC) excluded 9 additional questions upon agreement, resulting in a final set of 40 queries to be input into ChatGPT. This exclusion aimed to ensure that the responses could be applied to all patients receiving radiotherapy, thereby reflecting their primary concerns and doubts. The questions were intentionally phrased in the first person to mirror the way patients might typically frame their queries when interacting with ChatGPT [43] and were structured to address the informational needs of patients at various stages of radiotherapy [27]. The final set of questions was categorized into three dimensions: general information (n=14), planning and treatment (n=16), and side effects (n=10) (Textbox 1). These dimensions were selected based on previous studies [27,31,32,44], which assessed the most critical information needs for patients receiving radiotherapy, and they were further chosen to evaluate the strengths and limitations of responses across various topics in radiotherapy.

Textbox 1. Common patient queries regarding radiotherapy by dimension inserted in ChatGPT.

General information

1. Why is radiotherapy recommended?
2. What does radiotherapy involve?
3. When should radiotherapy and chemotherapy be combined?
4. What's the cost of radiotherapy treatment?
5. Who will be providing my radiotherapy treatment?
6. How does the radiotherapy treatment machine work?
7. What impact will radiotherapy treatment have on my life?
8. What impact will radiotherapy treatment have on my health in the future?
9. During the period of radiotherapy, will I have to follow a particular diet?
10. Will radiotherapy make me radioactive?
11. What does radiotherapy do to healthy cells?
12. How long does radiotherapy take to work?
13. Can I be cured of my disease through radiotherapy treatments?
14. What will happen after the radiotherapy treatment is finished?

Planning and treatment

1. Can I maintain my daily routine and activities during radiotherapy?
2. Can I keep working while undergoing radiotherapy treatments?
3. Are complementary medicines recommended while undergoing radiotherapy treatments?
4. What's the planning appointment in radiotherapy and what does it involve?
5. Why is computed tomography (CT) planning necessary in radiotherapy?
6. Why are tattoos useful in radiotherapy CT planning?
7. What happens on the first day of radiotherapy treatment?
8. Will the radiotherapy treatment schedule be adjusted to my availability?
9. What am I expected to do during the radiotherapy treatment?
10. Does the radiotherapy machine make noise?
11. How close is the radiotherapy treatment machine going to get?
12. What happens during radiotherapy treatment?
13. Is there a possibility of experiencing pain due to the radiotherapy treatment?
14. How long does a radiotherapy session last?
15. What should I wear for radiotherapy treatment?
16. Will there be follow-up after the end of radiotherapy treatments?

Side effects

1. What are the side effects of radiotherapy?
2. What skin care should I have during and after radiotherapy?
3. Am I going to feel tired after the radiotherapy treatments?
4. What hygiene care should be taken after radiotherapy treatments?
5. Which steps should be taken to reduce radiotherapy side effects?
6. Will the radiotherapy treatment be interrupted if I experience adverse side effects?
7. Who can I go to if the radiotherapy side effects become too burdensome?
8. Will radiotherapy affect my fertility?
9. Will radiotherapy cause hair loss?
10. Will radiotherapy cause permanent damage?

Data Collection

Responses were collected from ChatGPT between April 6, 2024 and April 9, 2024. Each question was queried on both versions of ChatGPT in English. Each query was entered separately using the "New Chat" function, acknowledging that ChatGPT considers the context of the conversation, which can influence responses. Therefore, the memory retention option was disabled when the questions were introduced into ChatGPT to ensure independence of the responses. The queries were then regenerated in each version of ChatGPT, and both responses were documented to analyze consistency.

Various methods were then used, as described in later sections, to assess the quality and reliability of the response content, response consistency, response readability, and similarity between responses from GPT-3.5 and GPT-4.

Outcomes

Quality and Reliability

To evaluate the quality and reliability of the information provided by ChatGPT, we used a 5-point Likert scale, known as the General Quality Score (GQS), which has been used in previous studies [14,45]. The assessment criteria included accuracy, lay-language use, information flow, usefulness, and empathy. The 5-point Likert scale was defined as follows: (1) inaccurate information, poorly organized text, missing important details, and not helpful for patients; (2) limited accuracy, some relevant information is present, but still not easily understandable for patients; (3) adequately accurate information and some important details are explained in plain language; (4) accurate information, well-organized text, and most relevant details are presented in a patient-friendly manner; and (5) extremely accurate information, well-structured text, and all relevant details are presented in a compassionate and patient-friendly manner [14].

The median GQS was calculated by averaging the ratings provided by 16 independent radiotherapy experts with substantial experience in managing oncology patients undergoing radiotherapy. The experts were randomly assigned to evaluate either GPT-3.5 or GPT-4, assuring that each expert evaluated only the responses from one of ChatGPT's versions to reduce potential bias during the evaluation process, thereby decreasing the likelihood of altering assessments and enhancing their credibility [46]. All the experts received detailed instructions on the evaluation guidelines to promote a uniform understanding of the assessment process. Furthermore, the responses from ChatGPT were provided to the experts in paper format, and their evaluation was conducted in real time without internet access and without knowledge of which version the responses corresponded to, thereby ensuring a blinding effect. Moreover, the authors (CM and MC-R) who analyzed the obtained results were unaware of the identity of the radiotherapy experts.

Consistency and Similarity

The consistency and similarity of the responses were evaluated using the cosine similarity score. This method involves transforming the text information provided by ChatGPT into vectors, then calculating the cosine of the angle between the two vectors, indicating how similar the responses are to each other. Scores were calculated using an online tool. The cosine similarity score ranges from 0 to 1, where a score of 0 indicates complete dissimilarity between the texts, and a score of 1 indicates complete similarity [14,36].

To assess the similarity between the responses generated by GPT-3.5 and GPT-4, the initial responses to the same question provided by both versions were inserted into the web-based tool to determine the cosine similarity score between them.

The consistency of the responses generated by ChatGPT was assessed by entering the same question into both versions and calculating the cosine similarity score between the two responses to the same question. By regenerating the same question, we aim to assess whether ChatGPT can provide consistent information or if its responses vary widely.

Readability

To evaluate readability, responses from both versions were assessed using a web-based Flesch Reading Ease Score (FRES) calculator. This calculator determined the responses' readability using two different indices: the FRES and the Flesch-Kincaid Grade Level (FKGL). These readability tests use mathematical formulas that consider factors such as sentence length and word count. The FRES is a numerical score ranging from 0 to 100, with higher numbers indicating better readability, meaning the content is easier to read and understand [8,47,48] and corresponds to a lower grade level [4,47,48] (Multimedia Appendix 1). The FKGL score indicates the average number of years of education needed to comprehend a text, with lower scores suggesting better readability [8,47,48] and correlating to the equivalent school level [4,47] (Multimedia Appendix 2).

Statistical Analysis

The data were analyzed using SPSS statistical software (version 29.0; IBM Corp). The results were considered statistically significant at the 5% level (P=.05). Exploratory data analysis was carried out using frequency analysis (n, %) for categorical variables and median and interquartile range (IQR = $Q_3 - Q_1$) for continuous variables. To test the normality of the data, the Shapiro-Wilk test was used. The Mann-Whitney test (since the normality assumption was not verified) and effect size were used to compare the evaluations between the 2 versions of ChatGPT. To analyze the question evaluations, scores were calculated for each question, considering the 8 experts assigned to each version. Krippendorff α and Fleiss κ were used to assess the agreement between experts. For this analysis, the experts' assessments were considered for all questions in each dimension in each version of ChatGPT.

Ethical Considerations

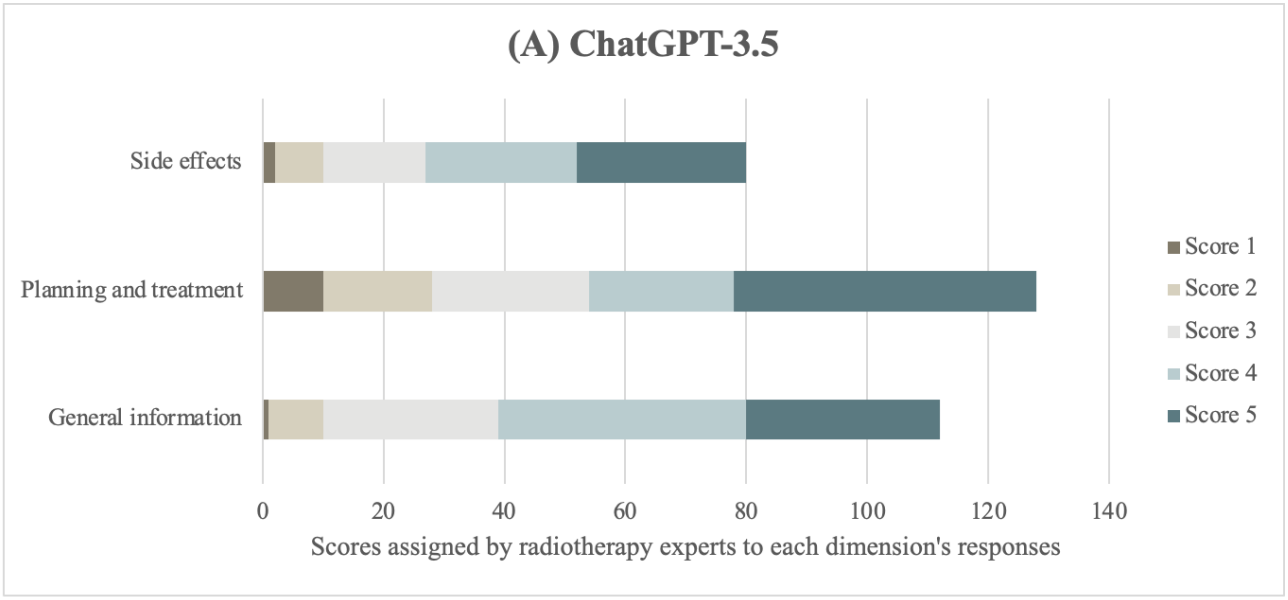
This study did not qualify as human subjects research due to the lack of patient involvement and identifying data for the health professionals involved; therefore, it was deemed exempt from institutional review board approval. Additionally, the use of ChatGPT, a public platform accessible to all, meant no permission was required to use the information generated in this study.

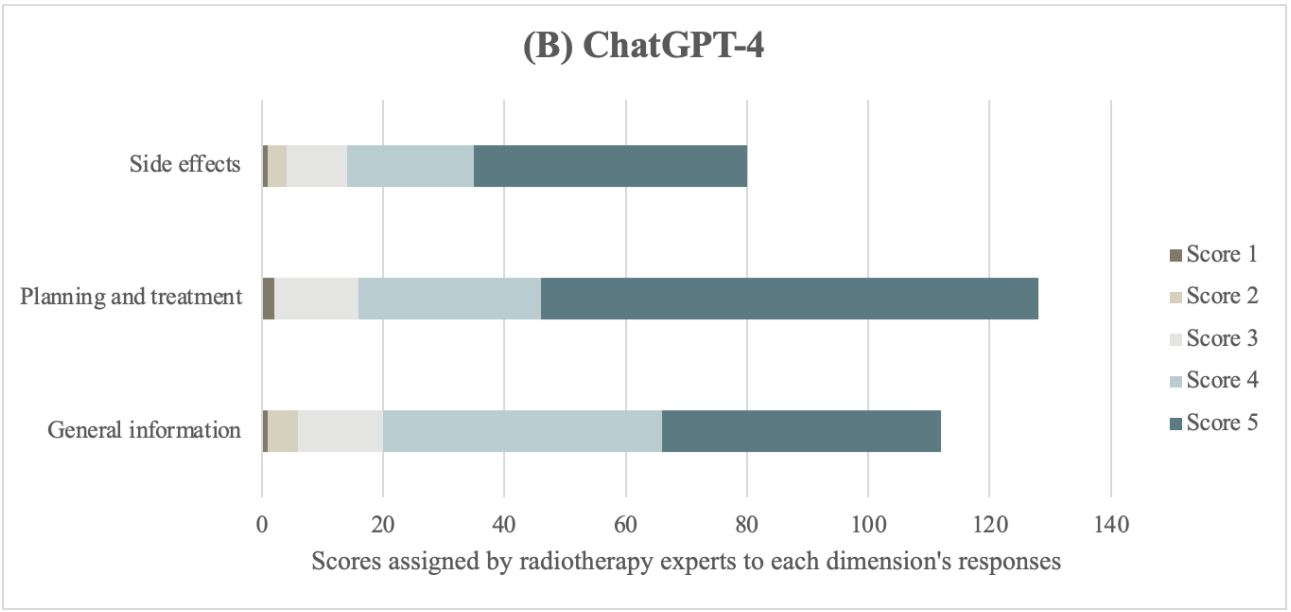
Results

Quality and Reliability

GPT-3.5 received primarily midrange scores, with most evaluations at levels 3 (n=72) and 4 (n=90), indicating generally accurate and comprehensible responses. Notably, many responses had the highest rating of 5 (n=110), providing extremely accurate and well-structured information. However, they also received low scores of 1 (n=13) and 2 (n=35), suggesting inaccurate or limited information, respectively. Conversely, GPT-4 received the highest score of 5 (n=173), indicating a superior ability to provide accurate and well-structured information. A significant number of responses were assigned a score of 4 (n=97), while a smaller proportion received a score of 3 (n=38), demonstrating that it consistently provided responses that were accurate, well-organized, and accessible to patients. Remarkably, GPT-4 exhibited a lower number of low scores of 1 (n=4) and 2 (n=8) than GPT-3.5. The score breakdown by the question dimension is shown in Figure 1.

Figure 1. Number of scores assigned by radiotherapy experts to the total number of responses in each dimension from (**A**) ChatGPT-3.5 and (**B**) ChatGPT-4. The Likert scale was defined as follows: score 1=inaccurate information; score 2=limited accuracy; score 3=adequately accurate information; score 4=accurate information; and score 5=extremely accurate information.





Considering the general information dimension, statistically significant differences were detected between the 2 versions of ChatGPT regarding questions 3 (P=.03, effect size=0.6) and 10 (P=.04, effect size=0.5). Regarding planning and treatment, statistically significant differences were detected for questions 5 (P=.046, effect size=0.5), 7 (P=.002, effect size=0.8), 9

(*P*=.003, effect size=0.7), and 11 (*P*=.02, effect size=0.6). Finally, regarding side effects, statistically significant differences were detected for question 9 (*P*=.04, effect size=0.5). In either situation, GPT-4 showed higher ratings (Table 1). The high effect size values revealed a weak overlap in the response distributions between the 2 versions of ChatGPT. However, in the results of the comparison of the evaluation of the 2 versions of ChatGPT, for the other questions, the effect size was low, revealing overlapping distributions of responses, which is why no statistically significant differences were detected. It can also be seen that, although not significant, version 4 of ChatGPT presents higher evaluation scores.

Table . Comparison of responses to questions about general information; planning and treatment; and side effects between the 2 versions of ChatGPT, with Mann-Whitney test and effect size results.

Dimension and questions	Number	Mean rank		*P* value	Effect size
		GPT-3.5	GPT-4		
General information					
Q1: Why is radiotherapy recommended?	8	7.56	9.44	.40	0.2
Q2: What does radiotherapy involve?	8	7.13	9.88	.23	0.3
Q3: When should radiotherapy and chemotherapy be combined?	8	6.00	11.00	.03	0.6
Q4: What's the cost of radiotherapy treatment?	8	9.19	7.81	.53	0.2
Q5: Who will be providing my radiotherapy treatment?	8	8.13	8.88	.72	0.1
Q6: How does the radiotherapy treatment machine work?	8	8.25	8.75	.82	0.1
Q7: What impact will radiotherapy treatment have on my life?	8	7.13	9.88	.20	0.3
Q8: What impact will radiotherapy treatment have on my health in the future?	8	7.13	9.88	.20	0.3
Q9: During the period of radiotherapy, will I have to follow a particular diet?	8	7.75	9.25	.44	0.2
Q10: Will radiotherapy make me radioactive?	8	6.25	10.75	.04	0.5
Q11: What does radiotherapy do to healthy cells?	8	8.75	8.25	.82	0.1
Q12: How long does radiotherapy take to work?	8	7.13	9.88	.21	0.3
Q13: Can I be cured of my disease through radiotherapy treatments?	8	9.69	7.31	.30	0.3
Q14: What will happen after the radiotherapy treatment is finished?	8	7.25	9.75	.22	0.3
Planning and treatment					

Dimension and questions	Number	Mean rank		*P* value	Effect size
		GPT-3.5	GPT-4		
Q1: Can I maintain my daily routine and activities during radiotherapy?	8	7.44	9.56	.24	0.3
Q2: Can I keep working while undergoing radiotherapy treatments?	8	7.25	9.75	.22	0.3
Q3: Are complementary medicines recommended while undergoing radiotherapy treatments?	8	9.50	7.50	.26	0.3
Q4: What's the planning appointment in radiotherapy and what does it involve?	8	7.00	10.00	.18	0.3
Q5: Why is computed tomography (CT) planning necessary in radiotherapy?	8	6.25	10.75	.046	0.5
Q6: Why are tattoos useful in radiotherapy CT planning?	8	7.13	9.88	.23	0.3
Q7: What happens on the first day of radiotherapy treatment?	8	5.13	11.88	.002	0.8
Q8: Will the radiotherapy treatment schedule be adjusted to my availability?	8	7.88	9.13	.44	0.2
Q9: What am I expected to do during the radiotherapy treatment?	8	5.19	11.81	.003	0.7
Q10: Does the radiotherapy machine make noise?	8	8.00	9.00	.54	0.2
Q11: How close is the radiotherapy treatment machine going to get?	8	5.75	11.25	.02	0.6
Q12: What happens during radiotherapy treatment?	8	6.88	10.13	.15	0.4
Q13: Is there a possibility of experiencing pain due to the radiotherapy treatment?	8	7.56	9.44	.41	0.2
	8	7.44	9.56	.34	0.2

Dimension and questions		Number	Mean rank		*P* value	Effect size
			GPT-3.5	GPT-4		
	Q14: How long does a radiotherapy session last?					
	Q15: What should I wear for radiotherapy treatment?	8	7.00	10.00	.06	0.5
	Q16: Will there be follow-up after the end of radiotherapy treatments?	8	7.38	9.63	.27	0.3
Side effects						
	Q1: What are the side effects of radiotherapy?	8	6.75	10.25	.12	0.4
	Q2: What skin care should I have during and after radiotherapy?	8	9.31	7.69	.48	0.2
	Q3: Am I going to feel tired after the radiotherapy treatments?	8	7.50	9.50	.26	0.3
	Q4: What hygiene care should be taken after radiotherapy treatments?	8	6.75	10.25	.10	0.4
	Q5: Which steps should be taken to reduce radiotherapy side effects?	8	7.94	9.06	.63	0.1
	Q6: Will the radiotherapy treatment be interrupted if I experience adverse side effects?	8	8.38	8.63	.91	0.0
	Q7: Who can I go to if the radiotherapy side effects become too burdensome?	8	6.81	10.19	.14	0.4
	Q8: Will radiotherapy affect my fertility?	8	7.63	9.38	.41	0.2
	Q9: Will radiotherapy cause hair loss?	8	6.38	10.63	.04	0.5
	Q10: Will radiotherapy cause permanent damage?	8	6.75	10.25	.12	0.4

Based on the analysis of Krippendorff α and Fleiss κ coefficients across the 3 dimensions (general information; planning and treatment; and side effects), the results indicated a low level of agreement in the classification of questions for both GPT-3.5 and GPT-4. This trend of weak agreement was consistent across the overall set of queries in Multimedia Appendix 3.

Consistency and Similarity

Regarding similarity and consistency, a cosine similarity score ranging from 0 to 1 was calculated, as previously described. Concerning similarity, the median (IQR) cosine similarity between GPT-3.5 and GPT-4 responses was 0.81 (IQR 0.05), indicating a reasonably good similarity between the 2 versions of ChatGPT. Notably, question 11 in the planning and treatment

dimension exhibited the lowest similarity, with a value of 0.68. With respect to consistency, the cosine similarity median (IQR) for GPT-3.5 and GPT-4 responses were 0.85 (IQR 0.04) and 0.83 (IQR 0.04), respectively. In both versions, consistency was demonstrated to be good or very good, with values ranging between 0.74 and 0.92.

Readability

The word count, sentence count, FRES, and FKGL score for both versions are summarized in Table 2. A relevant disparity was observed in the median (IQR) word count between GPT-3.5 and GPT-4 (299.00, IQR 176.5 versus 344.50, IQR 74.75). Additionally, the sentence count was higher in GPT-4 compared to GPT-3.5 (20.00, IQR 10.5 versus 18.00, IQR 17).

Table . Word count, sentence count, Flesch Reading Ease Score, and Flesch-Kincaid grade level score of responses from GPT-3.5 and GPT-4.

Dimension and questions[a]	GPT-3.5				GPT-4			
	Word count	Sentence count	FRES[b]	FKGL[c]	Word count	Sentence count	FRES	FKGL
General information								
Q1	332	22	35.31	12.08	378	25	32.36	12.50
Q2	414	27	35.97	12.05	453	28	35.97	12.26
Q3	304	18	24.11	14.09	340	17	25.80	14.63
Q4	188	7	13.53	18	268	15	25.81	14.10
Q5	246	18	28.58	12.67	305	21	21.78	13.83
Q6	378	27	35.96	11.72	431	27	36.55	12.13
Q7	422	27	35	12.26	358	27	41.90	10.71
Q8	389	22	32.74	13.09	351	16	30.79	14.41
Q9	332	25	41.23	10.81	311	25	57.92	8.27
Q10	84	5	17.56	14.98	223	16	21.59	13.71
Q11	304	26	36.90	11.02	352	21	37.45	12.2
Q12	178	7	33.21	14.94	298	15	47.28	11.60
Q13	177	8	27.61	14.91	231	9	23.30	16.39
Q14	348	22	35.92	12.18	410	13	19.24	18
Planning and treatment								
Q1	374	28	49.19	9.72	402	30	52.44	9.27
Q2	229	8	21.88	17.32	369	22	52.94	10.04
Q3	165	8	15.25	16.99	359	20	16.19	10.93
Q4	361	21	26.51	13.83	433	25	39.01	12.12
Q5	316	16	16.79	15.82	378	24	26.80	13.43
Q6	214	12	15.19	15.57	332	21	30.51	12.93
Q7	358	22	34.35	12.51	472	27	48.93	10.78
Q8	140	5	20.70	17.33	232	15	33.24	12.47
Q9	361	27	40.94	10.87	416	31	43.54	10.52
Q10	76	4	30.59	13.71	106	5	31.28	14.16
Q11	164	6	0	18	314	16	33.07	13.52
Q12	335	24	37.10	11.55	388	28	39.71	11.16
Q13	247	16	37.38	11.88	315	19	37.73	12.12
Q14	144	5	0	18	264	13	28.24	14.37
Q15	327	24	52.01	9.39	337	22	62.50	8.35
Q16	183	7	19.42	17.05	339	20	37.90	12.18
Side effects								
Q1	340	26	48	9.81	324	11	26.28	16.91
Q2	300	26	57.23	8.14	354	33	52.80	8.56
Q3	150	7	36.19	13.54	270	9	32.57	16.17
Q4	411	23	43.17	11.68	361	28	48.22	9.74
Q5	397	21	17.81	15.47	418	17	13.49	17.49
Q6	137	5	32.05	15.60	349	20	37.38	12.38
Q7	298	21	45.09	10.50	371	27	38.28	11.33

Dimension and questions[a]	GPT-3.5				GPT-4			
	Word count	Sentence count	FRES[b]	FKGL[c]	Word count	Sentence count	FRES	FKGL
Q8	164	8	23.53	15.07	277	10	17.15	17.75
Q9	108	5	43.91	12.50	214	15	57.94	8.72
Q10	212	10	29.29	14.44	330	12	26.13	16.45

[a]Please refer to Table 1 for the full questions.

[b]FRES: Flesch Reading Ease Score.

[c]FKGL: Flesch-Kincaid Grade Level.

The FRES median (IQR) for GPT-3.5 and GPT-4 responses were 32.98 (15.59) and 34.61 (16.07), respectively. This indicates that the responses generated by the two versions were considered college-level and difficult to read. The FKGL median (IQR) for GPT-3.5 and GPT-4 responses were 13.32 (3.79) and 12.32 (3.32), respectively. This suggests that at least 13 years of education (college-level) are required to understand the responses generated by GPT-3.5, whereas the responses from GPT-4 require at least of 12 years of education (college-level) for comprehension.

Discussion

Principal Findings

The power and utility of AI platforms in health care, such as ChatGPT, are rapidly evolving and improving and have the potential to significantly improve patient education [5,49]. This study sought to assess the quality and reliability of ChatGPT responses to common patient queries regarding radiotherapy with the aim of determining its potential as a reliable source of information for patients. We also aimed to compare the performances of GPT-3.5 and GPT-4 in generating responses to the same radiotherapy-related queries.

Although most responses were correct or close to correct, upon comparing the accuracy of responses between GPT-4 and GPT-3.5 in the 3 dimensions, it became evident that GPT-4 consistently offered improved elucidation of specific concepts relevant to radiotherapy treatment. In question 10 of the general information dimension, GPT-4 specifically delineated that patients are nonradioactive and may safely interact with others posttreatment (“You can safely be around others, including children and pregnant women, without any risk of exposing them to radiation”). However, this aspect was not as clearly articulated in GPT-3.5, which failed to mention that patients may come into contact with others after treatment. Additionally, within the side effects dimension, in questions 2 and 3, GPT-4 emphasized that the intended creams to use throughout radiotherapy treatment should only be those recommended by the health care provider (“Apply a fragrance-free moisturizer recommended by your healthcare provider”) and specified strategies to mitigate fatigue, a treatment-related side effect. However, this advice was not as detailed in the responses from GPT-3.5. Within the planning and treatment dimensions, GPT-3.5 demonstrated a propensity to diverge from directly addressing the queried issue in certain responses in contrast to GPT-4. In question 7, the response did not describe the first day of treatment but rather outlined the entire course of the patient’s radiotherapy. Question 11 failed to specify the distance between the equipment and the patient, a detail that was thoroughly addressed by GPT-4. In response to question 12, GPT-3.5 did not describe what occurs during treatment, instead reiterating the patient’s overall course. This indicates that GPT-3.5 exhibits reduced accuracy when responding to queries related to planning and treatment, as Valentini et al demonstrated [50].

However, in GPT-4’s response to the 13th planning and treatment question, specific information was inaccurately presented as it erroneously stated that radiotherapy induces direct pain (“Direct Pain from Treatment Site: Radiotherapy can cause localized pain at the site of treatment”). This error may have occurred because not all web-based sources are reliable, and because the model is trained on a diverse array of internet texts, it may incorporate biased or outdated information. Consequently, misinformation regarding cancer continues to pose a significant concern in online communication, which could result in responses or recommendations that do not consider the most current, evidence-based medical practices [21].

Moreover, there were a few occasions in both versions in which a lack of information was demonstrated. For instance, in question 7 of the side effects dimension, neither version mentioned that radiation therapists, who are team members that assist the patient daily throughout their treatment [31], could serve as advisers for patients experiencing severe side effects.

In summary, both GPT-3.5 and GPT-4 demonstrated the ability to address concepts related to radiotherapy. However, GPT-4 provided more targeted and detailed responses, thereby exhibiting superior performance compared to GPT-3.5, as corroborated by several studies [9,24,25,43,47]. The reduced number of scores of 1 and 2 assigned by radiotherapy experts to GPT-4 responses indicated a substantial improvement in response quality and reliability.

Comparison With Prior Work

In most responses, ChatGPT used a typical structure characterized by a succinct introductory paragraph, followed by 5 or 6 bullet points delineating the responses, culminating in a short concluding paragraph. Additionally, in a fair number of responses generated by GPT-3.5 (n=25) and GPT-4 (n=28), a statement was included advising that the information provided should always be discussed with health care providers, consistent with prior studies [43,51,52]. This recommendation is significant because the use of ChatGPT in health care must be carefully

monitored and should not be viewed as a substitute for human judgment. Its performance, safety, and associated risks require thorough evaluation by experts before integration into mainstream practice [53]. Moreover, it is essential that the model be trained on a substantial dataset validated by experts. This rigorous validation process could enhance the reliability and trustworthiness of ChatGPT responses, ultimately benefiting patient care [54].

The cosine similarity score indicated a reasonably substantial similarity and consistency, and while subtle changes in sentence structure were noted, most answers remained consistent, implying accuracy [3].

A key feature influencing consistency is the temperature parameter, a value ranging from 0 to 2, which adjusts the randomness of each subsequent word in the chat output. A value of 0 results in minimal variability, whereas values approaching 1 introduce greater randomness and creativity into the responses. Creativity is a powerful tool in communication, as it simplifies complex concepts, fosters critical thinking, and enhances the accessibility of intricate information, making it especially valuable for developing patient education materials. However, using ChatGPT with high creativity settings in clinical contexts may present challenges. By lowering the creativity level, we ensure that the summarized information remains faithful to the training data, thereby prioritizing accuracy and reliability over creative expression. Although this feature is not currently available for modification in ChatGPT, it may be included in future iterations of the tool's web interface [55].

Therefore, ensuring high reliability in ChatGPT's outputs is essential for users to trust its data-driven conclusions. Although advances in ChatGPT's performance can be attributed to key developments in its underlying technology, it is crucial that patients approach the information provided by AI tools such as ChatGPT with caution [56]. This is especially important given that ChatGPT does not disclose the bibliography used to generate responses [7,21,22]. This issue was observed during our study, as the bibliography was not disclosed in either version, indicating ChatGPT's inability to inform users of the contentious nature of certain information [7,10,21,22]. This lack of transparency is particularly significant, given the ethical concerns that arise regarding its application in patient care. Its implementation may lead to unintended or undesirable issues such as risks of bias and transparency, challenges related to interpretability, and generation of inaccurate content, all of which can have serious negative consequences for patient health [53].

Moreover, patient accessibility to AI technology varies significantly according to socioeconomic status, education, age, and geography. Individuals in higher socioeconomic groups or urban areas have better access to the necessary infrastructure, whereas those in lower socioeconomic conditions or rural regions face significant barriers. The effective use of AI also requires digital literacy and computational skills, making the understanding of technology crucial [57]. Although AI can revolutionize education and research, GPT-4 may widen the gap between the wealthy and poor [58]. Conversely, the free availability of GPT-3.5 helps reduce socioeconomic inequalities in cancer treatment by providing fast medical information to all, regardless of financial circumstances [21].

Concerning readability, all responses were considered more difficult to read than the sixth grade reading level recommended for patient consumption, a concern highlighted in prior studies [4,37]. This finding suggests that although the content was predominantly accurate, it was presented at a level too advanced for the public, particularly for individuals with lower health literacy [37,59]. Health literacy, defined as a patient's ability to read and understand health care information and make effective decisions, is crucial for quality patient engagement in health care options. An important aspect of health literacy is readability, which measures the ease with which text can be read and understood and is particularly relevant in radiotherapy due to the complexity of the field and its evolving nature. Patients with lower health literacy may have limited knowledge of radiotherapy; struggle to understand their conditions, treatments, and potential side effects; and often confuse different treatment modalities [60]. Owing to the heightened challenges faced by these patients, this bears particular significance [36].

Various studies have been conducted to assess ChatGPT's ability to enhance readability and simplify responses [36,61-63], considering that, when providing patient education in a written form, it is important to ensure that it is tailored to the reading level of the target population [62]. To address this issue, it was suggested that direct prompts such as "Explain this to me like I am in fifth grade" could assist in generating simplified responses [36,61-63]. This indicates the potential of ChatGPT to tailor responses to varying literacy levels and customize them based on an individual's educational background [61].

Strengths and Limitations

This investigation revealed that both GPT-3.5 and GPT-4 demonstrated proficiency in addressing radiotherapy-related concepts, with GPT-4 exhibiting notably higher performance. Although GPT-4 achieved marginally better readability scores, the content generated by both models remains complex for a general audience. Therefore, their use should be complemented by strategies to improve their accessibility and readability. Moreover, ChatGPT holds significant potential in promoting health behavior changes among patients with cancer by enhancing health literacy and supporting the self-management of radiotherapy-related conditions [64]. However, ChatGPT's responses must be validated by experts before they are integrated into a health care system to serve as a reliable source of information [53,54]. Therefore, ChatGPT shows promise in providing clinical guidance, suggesting treatment options, and serving as a valuable resource for medical education, facilitating a more effective shared decision-making process [6,47,59,65]. Hence, it can potentially serve as an alternative to current web-based resources [36].

This study had some limitations. First, the formulation and phrasing of queries in both versions may have influenced the performance of ChatGPT. Second, the queries were exclusively written in English, which restricted the responses to a single language. Third, although the total number of questions was comparable to other studies [1,8,11,18,47,50], the optimal number of queries needed to effectively evaluate the model

remains undetermined, and the sample may not capture the full diversity of patient concerns about radiotherapy. Fourth, the scoring process inherently involves subjectivity, particularly with the GQS, as different raters may interpret and prioritize quality aspects differently. Fifth, the potential bias introduced by using a 5-point Likert scale may lead to a central tendency bias, as respondents tend to avoid extreme options and cluster their answers around the midpoint, which can limit the granularity of the evaluations and distort the data [66,67]. Sixth, this study was conducted within a specific time frame (April 2024), and ChatGPT is expected to improve continuously over time. Repeating this study at a later time could improve response quality.

Another limitation of our study was the limitations of ChatGPT itself. First, it should be noted that the information provided by GPT-3.5 is available only up to September 2021. Second, GPT-4 has a limited number of questions that can be posed within a specific time frame and it is exclusively accessible through paid subscriptions, potentially constraining the public's access to more accurate information. Finally, ChatGPT is one of the many AI models available, making it uncertain whether the responses obtained represent the general characteristics of all LLMs.

Future Directions

Further research is essential to fully comprehend ChatGPT's role in patient education, including comparative studies with other AI models or traditional information sources, to better contextualize the findings. Additionally, future work should incorporate patient feedback into their understanding and satisfaction, providing valuable insights into the effectiveness of ChatGPT as an educational tool in real-world settings.

Conclusions

Both GPT-3.5 and GPT-4 demonstrated the ability to address concepts related to radiotherapy, with GPT-4 exhibiting superior performance. Although GPT-4 achieved slightly better readability scores, the content produced by both versions remains challenging for the general public. This highlights the need for caution regarding potential misinformation and readability. Furthermore, the paid subscription model for GPT-4 could exacerbate existing health care disparities by limiting access to certain patient populations. Despite these limitations, ChatGPT shows promise as a valuable tool for addressing common patient queries regarding radiotherapy. However, its use should be complemented by strategies to improve accessibility and readability.

Acknowledgments

This work received Portuguese national funding from FCT – Fundação para a Ciência e a Tecnologia, IP, through the Research Center for Psychological Science of the Faculty of Psychology, University of Lisbon (UID/04527). The funding source was not involved in the study design, analyses, or interpretation of the data; writing of the report; or decision to submit the paper for publication. The authors used AI to assist in rephrasing sentences and enhancing their clarity and effectiveness. After using the tool, the authors reviewed and edited the content as necessary and take full responsibility for the final publication.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

AG and MC contributed to the conceptualization, methodology, writing/review and editing, supervision, and project administration. CM and MC-R were responsible for the conceptualization, methodology, investigation, resources, data curation, and writing the original draft. EC conducted the statistical analysis. All authors approved the final version of the manuscript and had full responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flesch Reading Ease Score.
[DOCX File, 15 KB - cancer_v11i1e63677_app1.docx]

Multimedia Appendix 2
Flesch-Kincaid Grade Level score.
[DOCX File, 14 KB - cancer_v11i1e63677_app2.docx]

Multimedia Appendix 3
Analysis of Krippendorff α and Fleiss κ coefficients across the 3 dimensions.
[DOCX File, 16 KB - cancer_v11i1e63677_app3.docx]

References

1. Moazzam Z, Cloyd J, Lima HA, Pawlik TM. Quality of ChatGPT responses to questions related to pancreatic cancer and its surgical care. Ann Surg Oncol 2023 Oct;30(11):6284-6286. [doi: 10.1245/s10434-023-13777-w] [Medline: 37349615]
2. Haver HL, Ambinder EB, Bahl M, Oluyemi ET, Jeudy J, Yi PH. Appropriateness of breast cancer prevention and screening recommendations provided by ChatGPT. Radiology 2023 May;307(4):e230424. [doi: 10.1148/radiol.230424] [Medline: 37014239]
3. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. JNCI Cancer Spectr 2023 Mar 1;7(2):pkad015. [doi: 10.1093/jncics/pkad015] [Medline: 36929393]
4. Davis RJ, Ayo-Ajibola O, Lin ME, et al. Evaluation of oropharyngeal cancer information from revolutionary artificial intelligence chatbot. Laryngoscope 2024 May;134(5):2252-2257. [doi: 10.1002/lary.31191] [Medline: 37983846]
5. Gabriel J, Shafik L, Alanbuki A, Larner T. The utility of the ChatGPT artificial intelligence tool for patient education and enquiry in robotic radical prostatectomy. Int Urol Nephrol 2023 Nov;55(11):2717-2732. [doi: 10.1007/s11255-023-03729-4] [Medline: 37528247]
6. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol 2023 Jul;29(3):721-732. [doi: 10.3350/cmh.2023.0089] [Medline: 36946005]
7. Walker HL, Ghani S, Kuemmerli C, et al. Reliability of medical information provided by ChatGPT: assessment against clinical guidelines and patient information quality instrument. J Med Internet Res 2023 Jun 30;25:e47479. [doi: 10.2196/47479] [Medline: 37389908]
8. Wei K, Fritz C, Rajasekaran K. Answering head and neck cancer questions: an assessment of ChatGPT responses. Am J Otolaryngol 2024;45(1):104085. [doi: 10.1016/j.amjoto.2023.104085] [Medline: 37844413]
9. Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT, GPT-4, and Google Bard on a neurosurgery oral boards preparation question bank. Neurosurgery 2023 Nov 1;93(5):1090-1098. [doi: 10.1227/neu.0000000000002551] [Medline: 37306460]
10. Hopkins AM, Logan JM, Kichenadasse G, Sorich MJ. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm-shift. JNCI Cancer Spectr 2023 Mar 1;7(2):36808255. [doi: 10.1093/jncics/pkad010] [Medline: 36808255]
11. Mukherjee S, Durkin C, Pebenito AM, Ferrante ND, Umana IC, Kochman ML. Assessing ChatGPT's ability to reply to queries regarding colon cancer screening based on multisociety guidelines. Gastro Hep Adv 2023;2(8):1040-1043. [doi: 10.1016/j.gastha.2023.07.008] [Medline: 37974564]
12. Uprety D, Zhu D, West HJ. ChatGPT-a promising generative AI tool and its implications for cancer care. Cancer 2023 Aug 1;129(15):2284-2289. [doi: 10.1002/cncr.34827] [Medline: 37183438]
13. Guckenberger M, Andratschke N, Ahmadsei M, et al. Potential of ChatGPT in facilitating research in radiation oncology? Radiother Oncol 2023 Nov;188:109894. [doi: 10.1016/j.radonc.2023.109894] [Medline: 37659658]
14. Coskun B, Ocakoglu G, Yetemen M, Kaygisiz O. Can ChatGPT, an artificial intelligence language model, provide accurate and high-quality patient information on prostate cancer? Urology 2023 Oct;180(35–58):35-58. [doi: 10.1016/j.urology.2023.05.040] [Medline: 37406864]
15. Chow JCL, Wong V, Sanders L, Li K. Developing an AI-assisted educational chatbot for radiotherapy using the IBM Watson assistant platform. Healthcare (Basel) 2023 Aug 29;11(17):2417. [doi: 10.3390/healthcare11172417] [Medline: 37685452]
16. Li J, Zhong J, Li Z, Xiao Y, Wang S. Ectopic pituitary neuroendocrine tumor: a case report written with the help of ChatGPT. Cureus 2023 Oct;15(10):e46999. [doi: 10.7759/cureus.46999] [Medline: 37965416]
17. Lai UH, Wu KS, Hsu TY, Kan JKC. Evaluating the performance of ChatGPT-4 on the United Kingdom Medical Licensing Assessment. Front Med (Lausanne) 2023;10:1240915. [doi: 10.3389/fmed.2023.1240915] [Medline: 37795422]
18. Trager MH, Queen D, Bordone LA, Geskin LJ, Samie FH. Assessing ChatGPT responses to common patient queries regarding basal cell carcinoma. Arch Dermatol Res 2023 Dec;315(10):2979-2981. [doi: 10.1007/s00403-023-02705-3] [Medline: 37668714]
19. Javaid M, Haleem A, Singh RP. ChatGPT for healthcare services: an emerging stage for an innovative perspective. BenchCouncil Transactions on Benchmarks, Standards and Evaluations 2023 Feb;3(1):100105. [doi: 10.1016/j.tbench.2023.100105]
20. Ugas M, Giuliani M, Papadakos J. When is good, good enough? On considerations of machine translation in patient education. J Cancer Educ 2024 Oct;39(5):474-476. [doi: 10.1007/s13187-024-02401-4] [Medline: 38263349]
21. Cè M, Chiarpenello V, Bubba A, et al. Exploring the role of ChatGPT in oncology: providing information and support for cancer patients. BioMedInformatics 2024;4(2):877-888. [doi: 10.3390/biomedinformatics4020049]
22. Szczesniewski JJ, Tellez Fouz C, Ramos Alba A, Diaz Goizueta FJ, García Tello A, Llanes González L. ChatGPT and most frequent urological diseases: analysing the quality of information and potential risks for patients. World J Urol 2023 Nov;41(11):3149-3153. [doi: 10.1007/s00345-023-04563-0] [Medline: 37632558]

23. Emile SH, Horesh N, Freund M, et al. How appropriate are answers of online chat-based artificial intelligence (ChatGPT) to common questions on colon cancer? Surgery 2023 Nov;174(5):1273-1275. [doi: 10.1016/j.surg.2023.06.005] [Medline: 37482439]
24. Holmes J, Liu Z, Zhang L, et al. Evaluating large language models on a highly-specialized topic, radiation oncology physics. Front Oncol 2023;13:1219326. [doi: 10.3389/fonc.2023.1219326] [Medline: 37529688]
25. Huang Y, Gomaa A, Semrau S, et al. Benchmarking ChatGPT-4 on a radiation oncology in-training exam and Red Journal Gray Zone cases: potentials and challenges for AI-assisted medical education and decision making in radiation oncology. Front Oncol 2023;13:1265024. [doi: 10.3389/fonc.2023.1265024] [Medline: 37790756]
26. Smith SK, Zhu Y, Dhillon HM, et al. Supporting patients with low health literacy: what role do radiation therapists play? Support Care Cancer 2013 Nov;21(11):3051-3061. [doi: 10.1007/s00520-013-1875-7] [Medline: 23812495]
27. Halkett GKB, Kristjanson LJ, Lobb E, O'Driscoll C, Taylor M, Spry N. Meeting breast cancer patients' information needs during radiotherapy: what can we do to improve the information and support that is currently provided? Eur J Cancer Care (Engl) 2010 Jul;19(4):538-547. [doi: 10.1111/j.1365-2354.2009.01090.x] [Medline: 19708930]
28. Corish S, Fulton BA, Galbraith L, Coltart K, Duffton A. Impact of patient information format on the experience of cancer patients treated with radiotherapy. Tech Innov Patient Support Radiat Oncol 2024 Jun;30:100252. [doi: 10.1016/j.tipsro.2024.100252] [Medline: 38779037]
29. Chen JJ, Brown AM, Garda AE, et al. Patient education practices and preferences of radiation oncologists and interprofessional radiation therapy care teams: a mixed-methods study exploring strategies for effective patient education delivery. Int J Radiat Oncol Biol Phys 2024 Aug 1;119(5):1357-1367. [doi: 10.1016/j.ijrobp.2024.02.023] [Medline: 38437924]
30. Kumar KA, Balazy KE, Gutkin PM, et al. Association between patient education videos and knowledge of radiation treatment. Int J Radiat Oncol Biol Phys 2021 Apr 1;109(5):1165-1175. [doi: 10.1016/j.ijrobp.2020.11.069] [Medline: 33301819]
31. Douma KFL, Koning CCE, Zandbelt LC, de Haes H, Smets EMA. Do patients' information needs decrease over the course of radiotherapy? Support Care Cancer 2012 Sep;20(9):2167-2176. [doi: 10.1007/s00520-011-1328-0] [Medline: 22081119]
32. Zeguers M, de Haes H, Zandbelt LC, et al. The information needs of new radiotherapy patients: how to measure? Do they want to know everything? And if not, why? Int J Radiat Oncol Biol Phys 2012 Jan 1;82(1):418-424. [doi: 10.1016/j.ijrobp.2010.09.032] [Medline: 21075556]
33. Halkett GKB, Merchant S, Smith SK, et al. Supporting and preparing patients for radiotherapy: patients' and radiation therapists' perspectives on their one-to-one consultations. Eur J Cancer Care (Engl) 2020 Nov;29(6):32656882. [doi: 10.1111/ecc.13284] [Medline: 32656882]
34. Kim C, Dillon R, Nica L, Berthelet E, Keyes M, Hamilton SN. Information needs of patients diagnosed with head and neck cancer undergoing radiation therapy: a survey of patient satisfaction. J Cancer Educ 2021 Oct;36(5):941-949. [doi: 10.1007/s13187-020-01719-z] [Medline: 32189219]
35. Halkett GKB, Kristjanson LJ, Lobb E, et al. Information needs and preferences of women as they proceed through radiotherapy for breast cancer. Patient Educ Couns 2012 Mar;86(3):396-404. [doi: 10.1016/j.pec.2011.05.010] [Medline: 21664788]
36. Yalamanchili A, Sengupta B, Song J, et al. Quality of large language model responses to radiation oncology patient care questions. JAMA Netw Open 2024 Apr 1;7(4):e244630. [doi: 10.1001/jamanetworkopen.2024.4630] [Medline: 38564215]
37. Young JN, Poplausky D, et al. The utility of ChatGPT in generating patient-facing and clinical responses for melanoma. J Am Acad Dermatol 2023 Sep;89(3):602-604. [doi: 10.1016/j.jaad.2023.05.024] [Medline: 37207953]
38. Talyshinskii A, Naik N, Hameed BMZ, et al. Expanding horizons and navigating challenges for enhanced clinical workflows: ChatGPT in urology. Front Surg 2023;10:1257191. [doi: 10.3389/fsurg.2023.1257191] [Medline: 37744723]
39. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res 2023 Jun 28;25:e48568. [doi: 10.2196/48568] [Medline: 37379067]
40. Hang CN, Yu PD, Tan CW. TrumorGPT: query optimization and semantic reasoning over networks for automated fact-checking. 2024 Presented at: 2024 58th Annual Conference on Information Sciences and Systems (CISS); Mar 13-15, 2024; Princeton, NJ. [doi: 10.1109/CISS59072.2024.10480162]
41. Halkett GKB, Kristjanson LJ. Validity and reliability testing of two instruments to measure breast cancer patients' concerns and information needs relating to radiation therapy. Radiat Oncol 2007 Nov 25;2(1):43. [doi: 10.1186/1748-717X-2-43] [Medline: 18036247]
42. Radiation therapy and you: support for people with cancer. National Cancer Institute. URL: https://www.cancer.gov/publications/patient-education/radiation-therapy-and-you [accessed 2025-04-08]
43. Hermann CE, Patel JM, Boyd L, Growdon WB, Aviki E, Stasenko M. Let's chat about cervical cancer: assessing the accuracy of ChatGPT responses to cervical cancer questions. Gynecol Oncol 2023 Dec;179(164–8):164-168. [doi: 10.1016/j.ygyno.2023.11.008] [Medline: 37988948]
44. Flockton A, Leong A, Gilfillan D, Larsen P. A multi-centre survey of New Zealand cancer patients' preferences for radiation treatment information. J Med Radiat Sci 2024 Mar;71(1):91-99. [doi: 10.1002/jmrs.745] [Medline: 38131264]

45. Onder CE, Koc G, Gokbulut P, Taskaldiran I, Kuskonmaz SM. Evaluation of the reliability and readability of ChatGPT-4 responses regarding hypothyroidism during pregnancy. Sci Rep 2024 Jan 2;14(1):243. [doi: 10.1038/s41598-023-50884-w] [Medline: 38167988]
46. Parker RA, Berman NG. Blinding in Observational Studies: Cambridge University Press; 2016:334-340. [doi: 10.1017/CBO9781139024716.029]
47. Chiarelli G, Stephens A, Finati M, et al. Adequacy of prostate cancer prevention and screening recommendations provided by an artificial intelligence-powered large language model. Int Urol Nephrol 2024 Aug;56(8):2589-2595. [doi: 10.1007/s11255-024-04009-5] [Medline: 38564079]
48. Thia I, Saluja M. ChatGPT: is this patient education tool for urological malignancies readable for the general population? Res Rep Urol 2024;16:31-37. [doi: 10.2147/RRU.S440633] [Medline: 38259300]
49. Kuşcu O, Pamuk AE, Sütay Süslü N, Hosal S. Is ChatGPT accurate and reliable in answering questions regarding head and neck cancer? Front Oncol 2023;13:1256459. [doi: 10.3389/fonc.2023.1256459] [Medline: 38107064]
50. Valentini M, Szkandera J, Smolle MA, Scheipl S, Leithner A, Andreou D. Artificial intelligence large language model ChatGPT: is it a trustworthy and reliable source of information for sarcoma patients? Front Public Health 2024;12:1303319. [doi: 10.3389/fpubh.2024.1303319] [Medline: 38584922]
51. Rogasch JMM, Metzger G, Preisler M, et al. ChatGPT: can you prepare my patients for [18F]FDG PET/CT and explain my reports? J Nucl Med 2023 Dec 1;64(12):1876-1879. [doi: 10.2967/jnumed.123.266114] [Medline: 37709536]
52. Braun EM, Juhasz-Böss I, Solomayer EF, et al. Will I soon be out of my job? Quality and guideline conformity of ChatGPT therapy suggestions to patient inquiries with gynecologic symptoms in a palliative setting. Arch Gynecol Obstet 2024 Apr;309(4):1543-1549. [doi: 10.1007/s00404-023-07272-6] [Medline: 37975899]
53. Kumaresan A, Uden L, Ashraf S. Potential role of ChatGPT in healthcare in the prevention and management of non-communicable diseases. In: Communications in Computer and Information Science 2023, Vol. 1825:430-442. [doi: 10.1007/978-3-031-34045-1_34]
54. Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical considerations of using ChatGPT in health care. J Med Internet Res 2023 Aug 11;25:e48009. [doi: 10.2196/48009] [Medline: 37566454]
55. Davis J, Van Bulck L, Durieux BN, Lindvall C. The temperature feature of ChatGPT: modifying creativity for clinical research. JMIR Hum Factors 2024 Mar 8;11:e53559. [doi: 10.2196/53559] [Medline: 38457221]
56. Funk PF, Hoch CC, Knoedler S, et al. ChatGPT's response consistency: a study on repeated queries of medical examination questions. Eur J Investig Health Psychol Educ 2024 Mar 8;14(3):657-668. [doi: 10.3390/ejihpe14030043] [Medline: 38534904]
57. Keskin T, Vermeulen F. Overcoming the digital divide in the era of artificial intelligence. Presented at: Hawaii International Conference on System Sciences; Jan 3-6, 2024; Waikiki, HI p. 1471-1474. [doi: 10.24251/HICSS.2024.185]
58. Meyer JG, Urbanowicz RJ, Martin PCN, et al. ChatGPT and large language models in academia: opportunities and challenges. BioData Min 2023 Jul 13;16(1):20. [doi: 10.1186/s13040-023-00339-9] [Medline: 37443040]
59. Braithwaite D, Karanth SD, Divaker J, et al. Evaluating ChatGPT's accuracy in providing screening mammography recommendations among older women: artificial intelligence and cancer communication. Res Sq 2024 Jan 31:rs.3.rs-3911155. [doi: 10.21203/rs.3.rs-3911155/v1] [Medline: 38352437]
60. Prabhu AV, Hansberry DR, Agarwal N, Clump DA, Heron DE. Radiation oncology and online patient education materials: deviating from NIH and AMA recommendations. Int J Radiat Oncol Biol Phys 2016 Nov 1;96(3):521-528. [doi: 10.1016/j.ijrobp.2016.06.2449] [Medline: 27681748]
61. Mohammadi SS, Khatri A, Jain T, et al. Evaluation of the appropriateness and readability of ChatGPT-4 responses to patient queries on uveitis. Ophthalmol Sci 2025;5(1):100594. [doi: 10.1016/j.xops.2024.100594] [Medline: 39435137]
62. Moons P, Van Bulck L. Using ChatGPT and Google Bard to improve the readability of written patient information: a proof of concept. Eur J Cardiovasc Nurs 2024 Mar 12;23(2):122-126. [doi: 10.1093/eurjcn/zvad087] [Medline: 37603843]
63. Vallurupalli M, Shah ND, Vyas RM. Validation of ChatGPT 3.5 as a tool to optimize readability of patient-facing craniofacial education materials. Plast Reconstr Surg Glob Open 2024 Feb;12(2):e5575. [doi: 10.1097/GOX.0000000000005575] [Medline: 38313589]
64. Alanezi F. Examining the role of ChatGPT in promoting health behaviors and lifestyle changes among cancer patients. Nutr Health 2024 Apr 3:2601060241244563. [doi: 10.1177/02601060241244563] [Medline: 38567408]
65. Choi J, Kim JW, Lee YS, et al. Availability of ChatGPT to provide medical information for patients with kidney cancer. Sci Rep 2024 Jan 17;14(1):1542. [doi: 10.1038/s41598-024-51531-8] [Medline: 38233511]
66. Chyung SYY, Roberts K, Swanson I, Hankinson A. Evidence-based survey design: the use of a midpoint on the Likert scale. Perf Improv 2017 Nov;56(10):15-23. [doi: 10.1002/pfi.21727]
67. Weijters B, Cabooter E, Schillewaert N. The effect of rating scale format on response styles: the number of response categories and response category labels. International Journal of Research in Marketing 2010 Sep;27(3):236-247. [doi: 10.1016/j.ijresmar.2010.02.004]

Abbreviations

AI: artificial intelligence
FKGL: Flesch-Kincaid Grade Level
FRES: Flesch Reading Ease Score
GQS: General Quality Score
LLM: large language model

Edited by N Cahill; submitted 27.06.24; peer-reviewed by CN Hang, D Chrimes, M Chatzimina, Z Hou; revised version received 30.01.25; accepted 27.02.25; published 16.04.25.

Please cite as:
Grilo A, Marques C, Corte-Real M, Carolino E, Caetano M
Assessing the Quality and Reliability of ChatGPT's Responses to Radiotherapy-Related Patient Queries: Comparative Study With GPT-3.5 and GPT-4
JMIR Cancer 2025;11:e63677
URL: https://cancer.jmir.org/2025/1/e63677
doi:10.2196/63677

Original Paper

Developing and Assessing a Scalable Digital Health Tool for Pretest Genetic Education in Patients With Early-Onset Colorectal Cancer: Mixed Methods Design

Jessica N Rivera Rivera[1], MPH, PhD; Moran Snir[2], MSc, MBA; Emilie Simmons[2], MS; Tara Schmidlen[2], MS, LCGC; Misha Sholeh[3], MS; Melinda Leigh Maconi[4], PhD; Carley Geiss[4], PhD; Hayden Fulton[4], MA; Laura Barton[5], MA, MS, LCGC; Brian D Gonzalez[6], PhD; Jennifer Permuth[7], PhD; Susan Vadaparampil[8], PhD, MPH

[1]Healthcare Delivery Research Network, MedStar Health Research Institute, Washington, DC, United States

[2]Nest Genomics, New York, NY, United States

[3]Non-Therapeutic Research Office, Moffitt Cancer Center, Tampa, FL, United States

[4]Participant Research, Interventions, and Measurement Core, Moffitt Cancer Center, Tampa, FL, United States

[5]Department of Pathology, Moffitt Cancer Center, Tampa, FL, United States

[6]Research Diversity and Workforce Development, Moffitt Cancer Center, Tampa, FL, United States

[7]Department of Gastrointestinal Oncology, Moffitt Cancer Center, Tampa, FL, United States

[8]Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, United States

Corresponding Author:
Jessica N Rivera Rivera, MPH, PhD
Healthcare Delivery Research Network
MedStar Health Research Institute
100 Irving Street NW
Washington, DC, 20010
United States
Phone: 1 443 692 1138
Email: jessica.n.riverarivera@medstar.net

Abstract

Background: National guidelines recommend germline genetic testing (GT) for all patients with early-onset colorectal cancer. With recent advances in targeted therapies and GT, these guidelines are expected to expand to include broader groups of patients with colorectal cancer. However, there is a shortage of genetic professionals to provide the necessary education and support for informed consent. As such, there is a pressing need to identify alternative approaches to facilitate and expedite access to GT.

Objective: This study describes the development of a pretest education intervention, Nest-CRC, to facilitate the uptake of germline GT among patients with early-onset colorectal cancer. Patients with early-onset colorectal cancer and health care providers reviewed Nest-CRC, and their reactions and recommendations were captured using a nested mixed methods approach.

Methods: Using the learner verification approach, we conducted 2 sequential phases of surveys and interviews with English- and Spanish-speaking patients with early-onset colorectal cancer and health care providers. The surveys assessed participants' experiences with genetic services and provided immediate feedback on the Nest-CRC genetic education modules. Semistructured interviews evaluated participants' perceptions of self-efficacy, attraction, comprehension, cultural acceptability, and usability of Nest-CRC. Survey data were analyzed using descriptive statistics (mean, median, and proportions), while interview data were analyzed through line-by-line coding of the transcribed interviews. After each phase, Nest-CRC was refined based on participants' recommendations.

Results: A total of 52 participants, including 39 patients with early-onset colorectal cancer and 13 providers, participated in the study. Of these, 19 patients and 6 providers participated in phase 1 (N=25), and 20 patients and 7 providers participated in phase 2 (N=27). Most participants (phase 1: 23/25, 92%, to 25/25, 100%; phase 2: 24/27, 89%, to 27/27, 100%) agreed that each of the 5 education modules was easy to understand and helpful; 13 patients reported no history of GT, with 11 (85%) expressing interest in GT and 2 (15%) remaining unsure after completing Nest-CRC. Participants reported that Nest-CRC provided sufficient information to help them decide about GT. The tool was deemed acceptable by individuals from diverse backgrounds, and participants found it visually attractive, easy to comprehend, and user-friendly.

Conclusions: The findings revealed that Nest-CRC is a promising strategy for facilitating pretest education and promoting GT. Nest-CRC has been refined based on participant recommendations and will be re-evaluated.

(JMIR Cancer 2025;11:e59464) doi:10.2196/59464

KEYWORDS

genetic education; genetic testing; genetic counseling; digital health; early-onset colorectal cancer

Introduction

Colorectal cancer (CRC) is the third most common cause of cancer incidence and mortality among men and women in the United States [1]. By 2030, it is projected to become the leading cause of death among patients diagnosed with early-onset cancer (under the age of 50 years) [2]. Approximately 14%-25% of early-onset CRCs are linked to hereditary factors, irrespective of family history [3,4]. Identifying germline variants in patients with CRC can help reduce morbidity and mortality by enabling guided treatment decisions, risk management to prevent and detect new primary cancers early, and cascade testing for at-risk relatives [5-9]. The National Comprehensive Cancer Network (NCCN) Guidelines recommend multigene panel testing (MGPT) for *all* individuals diagnosed with CRC before the age of 50 years and consider its use for all individuals diagnosed with CRC [10]. MGPT is recommended because it can simultaneously identify gene variants associated with various cancers and simplifies referrals for genetic testing (GT), as neither family history nor patient tumor characteristics are required.

Genetic services heavily depend on clinicians to identify and refer high-risk patients for genetic counseling (GC) and GT. However, approximately 40% of patients with early-onset CRC are not referred for GC [11,12]. Additionally, racial and ethnic disparities exist in germline studies and access to genetic services [11,13]. A study conducted between 2009 and 2017, involving patients with early-onset CRC treated at a tertiary-care referral center and a safety-net health system, found that Black patients were less likely to attend GC compared with Hispanic and non–Hispanic White patients [11]. Another retrospective study, using data from 2012 to 2016 across 4 academic medical centers, found that Black and Hispanic patients with CRC were referred to genetic specialists less often than non–Hispanic White patients [13]. However, among those referred, no racial or ethnic differences were observed in GC attendance [13], highlighting a missed opportunity for guideline-concordant genetic care. Therefore, systematic strategies are necessary to ensure GT services are offered to all patients with early-onset CRC.

With the expanding indications for genetic services, there is a shortage of genetic counselors and qualified genetic professionals [14,15], which can result in delays in GT. Although oncologists and other health care providers can order GT for at-risk patients, most lack the expertise or time to provide adequate genetic education [16,17]. Consequently, the American Society of Clinical Oncology (ASCO) acknowledges GC as the standard of care but also advocates for alternative approaches to delivering genetic services [18]. Studies examining the uptake of genetic services among patients with early-onset CRC have identified cost, limited availability of services, and racial and ethnic referral disparities as key barriers [11,19]. To address these known barriers to GT, the National Institute of Health Clinical Genome Resource's Consent and Disclosure Recommendations Working Group suggested reserving traditional, provider-delivered pre- and posttest GC for patients with greater clinical and genetic complexity, such as those with conditions lacking well-established testing and risk management criteria [20]. Alternative approaches to genetic education are needed to facilitate, expedite, and expand access to GT for patients at risk of cancer without overburdening GC resources [21]. To address this, we developed a digital health tool designed for patients with early-onset CRC from diverse racial and ethnic backgrounds. This tool systematically delivers pretest education and triages patients to GC and GT.

Previous studies involving patients with cancer suggest that digital genetic education is well-accepted and effective in improving knowledge, decisional satisfaction, and reducing decisional conflict [22-24]. However, most educational interventions addressing germline testing have focused on patients with breast cancer [25-27]. To our knowledge, only 2 studies have evaluated alternative strategies for genetic education in patients with CRC. These studies were not specific to patients with early-onset CRC, and their educational content focused on tumor testing [28] or GC [29]. Therefore, in this study, we propose a digital health tool designed for patients with early-onset CRC to promote autonomy by allowing them to access relevant germline information at their convenience, make informed decisions about GT, and opt-in to pretest GC if desired. This study outlines patient and provider feedback on the digital pretest genetic education tool and provides recommendations for its implementation, using a mixed methods approach.

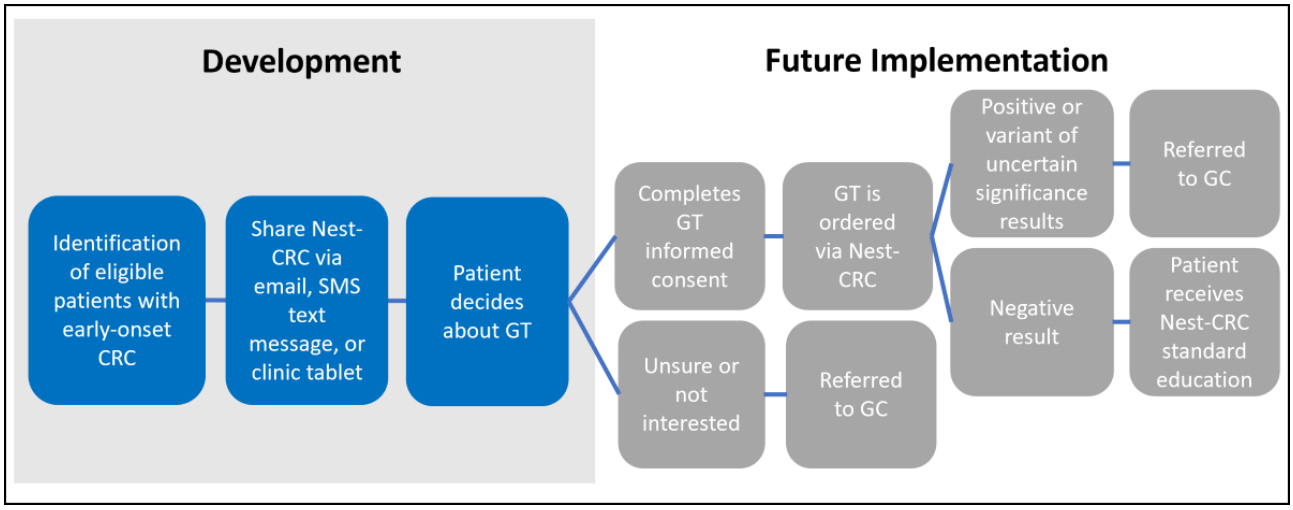
Methods

Intervention

Nest Genomics is a software company specializing in developing tools that help patients and providers scale the delivery and long-term implementation of genomic information. The Nest platform is a comprehensive, Health Insurance Portability and Accountability Act (HIPAA)–compliant solution designed to launch, implement, and scale longitudinal genomic programs, supporting both patients and clinicians throughout the care continuum—from patient identification and education to test ordering, result integration, and long-term management. Within Nest, our research team developed the Nest-CRC, a digital health tool designed to provide pretest genetic education for patients with early-onset CRC from diverse racial and ethnic groups (Figure 1). Nest-CRC is not publicly available at this time. The tool is divided into 5 brief modules. The modules

included in the first version of Nest-CRC are (1) hereditary CRC, (2) GT, (3) benefits and risks, (4) care recommendations, and (5) implications for family members. Nest-CRC takes approximately 10 minutes to complete. The information covered in Nest-CRC is supported by ASCO content recommendations for pretest genetic education [18], standard informed consent for GT, and feedback from genetic counselors and experts on the study team. Nest-CRC delivers education through text and images and is accessible on any personal device with internet access [30]. It includes images that are representative of different ages and races, with written content at a 5th-grade reading level in both English and Spanish.

Figure 1. Development and future implementation of the Nest-CRC tool. CRC: colorectal cancer; GC: genetic counseling; GT: genetic testing.

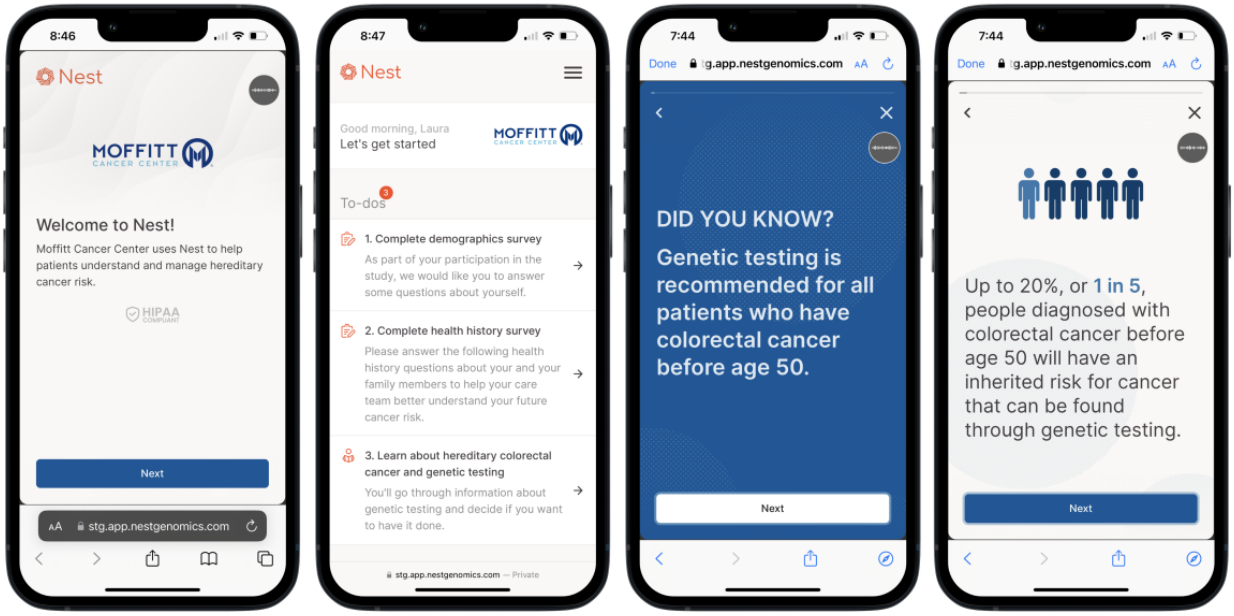


Procedure

We conducted a nested mixed methods study to develop and refine Nest-CRC for patients with early-onset CRC. Following the learner verification approach [31], we carried out 2 sequential phases of patient and provider surveys and interviews about Nest-CRC. Learner verification, which is useful for formative research, uses semistructured individual interviews to assess the appropriateness of materials for a target population. The quantitative data collected in each phase were used to describe the demographic characteristics of our sample population, their experiences with genetic services, and to obtain immediate feedback on their experience navigating each of the genetic education modules.

During phase 1, participants were emailed a link to a brief survey covering demographics, clinical characteristics, and experiences with genetic services, as well as the Nest-CRC educational modules, which included integrated questions about each module. Participants completed a semistructured interview after finishing the survey and Nest-CRC education. Patients could complete the survey and Nest-CRC on their personal device or a clinic tablet. After phase 1, we refined Nest-CRC based on participant recommendations (Figure 2). The revised version of Nest-CRC was then re-evaluated in phase 2 using the same procedure as in phase 1. However, in phase 2, participants also had the option to receive the link via SMS text message. Each version of Nest-CRC was reviewed by the entire study team for final edits.

Figure 2. Example of Nest-CRC for phase 2.



Ethical Considerations

This study was reviewed and approved by the Moffitt Cancer Center's Scientific Research Committee and Institutional Review Board (approval number 22176) on November 15, 2022. Before enrollment, all participants were provided with a copy of the informed consent form. All patients gave verbal informed consent in person or over the phone, and all providers consented to participate via email. Participants who agreed to take part in the study received a link to the survey and Nest-CRC education. The interviewer reviewed the informed consent form again with each participant before beginning the interview. All data collected for this study were deidentified using a unique ID number. Participants received US $25 upon completion of the interview.

Recruitment

From February to August 2023, English- and Spanish-speaking adult patients with early-onset CRC with upcoming medical appointments at the Moffitt Cancer Center Gastroenterology clinic were contacted in the clinic or by phone and invited to participate in the study. Recruitment flyers in both English and Spanish were posted in various waiting areas at Moffitt Cancer Center and distributed to community partners for sharing on their social media platforms (eg, Facebook). Interested potential participants responded to flyers and internet advertisements by calling or emailing the study team. The study team then contacted these individuals to screen them, obtain informed consent, and schedule their interview.

We purposely recruited at least 20% (n=4) Spanish-speaking and 20% (n=4) Black patients for each phase. These groups were specifically targeted because a lower proportion of Spanish-speaking and Black patients seen at the oncology clinic were eligible for the study. Gastroenterologists, oncologists, nurse practitioners, and genetic counselors with at least 2 years of experience working with patients with CRC were recruited from Moffitt Cancer Center (phases 1 and 2), MedStar Health (phase 2 only), and through referrals (phases 1 and 2). Different individuals participated in each phase.

Survey Measures

Before reviewing Nest-CRC, participants were asked to complete a brief survey capturing relevant sociodemographic, clinical, and epidemiologic characteristics adapted from the Health Information National Trends Survey (HINTS) 5 [32]. Patients were asked about their age, gender, race, ethnicity, country of birth, marital status, education level, employment status, household income, and insurance, while providers were asked about their age, gender, marital status, race, ethnicity, and professional degree. Additional information collected from patients included self-reported technology literacy (using the 3-item Digital Health Care Literacy Scale [33]), health literacy (using the 3-item Short Literacy Survey [34]), clinical details about their cancer diagnosis (including the type of CRC, cancer stage, age of diagnosis, and treatment history), family history of early-onset CRC, Ashkenazi Jewish ancestry (due to the known genetic risk for CRC and other cancers in this population), awareness of genetic services (ie, GC, GT, hereditary cancers, and Lynch syndrome), and history of genetic services (ie, referrals to genetic services, GC, and GT). We also evaluated patients' perceived importance of GT for cancer prevention and early detection (adapted from HINTS 5) [32]. Providers' self-reported practice characteristics included the frequency of communication with patients about genetic risk, referrals to GT, working with patients with early-onset CRC, and working with ethnic/racial minority patients. At the end of each module (n=5), participants were asked 2 questions: whether the information was easy to understand and whether it was helpful, using a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree. Responses of strongly agree and agree were recoded as agreed, while neither agree nor disagree, disagree, and strongly disagree were recoded as did not agree.

Upon completing Nest-CRC, participants were asked if they were interested in GT, with response options of yes, no, or unsure.

Interview Process

Interview guides were based on key elements of learner verification. Patients and providers gave feedback on the Nest-CRC tool's attractiveness, comprehension, self-efficacy, cultural acceptability, and usability (Table 1). The development of these questions was informed by prior studies using the learner verification approach [31,35,36] and refined by the study team. All interviews were recorded, and audio files were transcribed verbatim. Interviews conducted in Spanish were translated into English [37]. JRR conducted all English and Spanish interviews via Zoom (Zoom Communications/Qumu Corporation; phase 1: mean 26.96 minutes, range 17.26-37.36 minutes; phase 2: mean 26.04 minutes, range 15.07-36.02 minutes).

Table 1. Sample questions included in the interviews.

Key elements	Example of questions
Attractiveness	• What was the first thing that came to your mind when using the Nest-CRC tool? • How do you feel after going through this tool? • What attracted you or did not attract you about Nest-CRC?
Comprehension	• Overall, did you find the tool easy to understand?/Overall, did you find the tool easy to understand for patients that you typically see in the clinic?[a] • While completing Nest-CRC, can you describe any words, phrases, or sections that were difficult to understand? • What information do you think might be missing from the genetic education?
Self-efficacy	• Do you think this tool provides enough information to make an informed decision about getting or not genetic testing? • After completing the Nest-CRC tool, can you give me some examples of what happens after genetic testing?/After completing the Nest-CRC tool, would the patients have enough information for getting genetic testing?[a]
Cultural acceptability	• What are your thoughts about Nest-CRC being appealing to people from different backgrounds? • Were there any sections of the genetic education that made you feel uncomfortable?/Were there any sections of the genetic education that made you feel uncomfortable, or do you think patients might feel uncomfortable?[a]
Usability	• Overall, did you have any challenges using the tool? • Would you recommend this tool to other patients with colorectal cancer?/Would you recommend this tool to patients with early-onset colorectal cancer?[a] • If your health care provider had referred you to this tool, would you have completed it?/Would you think your patients will complete this tool?[a] • What do you think is the best way to share this tool with other patients with colorectal cancer diagnosed before the age of 50 years? • How should this tool be used in the clinic?[a]

[a]Provider-specific questions.

Data Analysis

Descriptive statistics (mean, median, and proportions) were calculated using IBM SPSS software (version 28). All qualitative data were transcribed into English and reviewed by team members. The research team met to develop the initial codebook, using deductive codes derived from the key elements of the interview guides. The themes and codebook, along with operational definitions for each code, were subsequently refined during the intercoder reliability process. Three research team members (CG, MLM, and HF) coded the transcripts using a direct content analysis approach. Intercoder reliability was assessed until Cohen κ reached 0.80, indicating substantial agreement [38]. Qualitative analysts performed line-by-line coding of all interview data using NVivo 12 software (Lumivero).

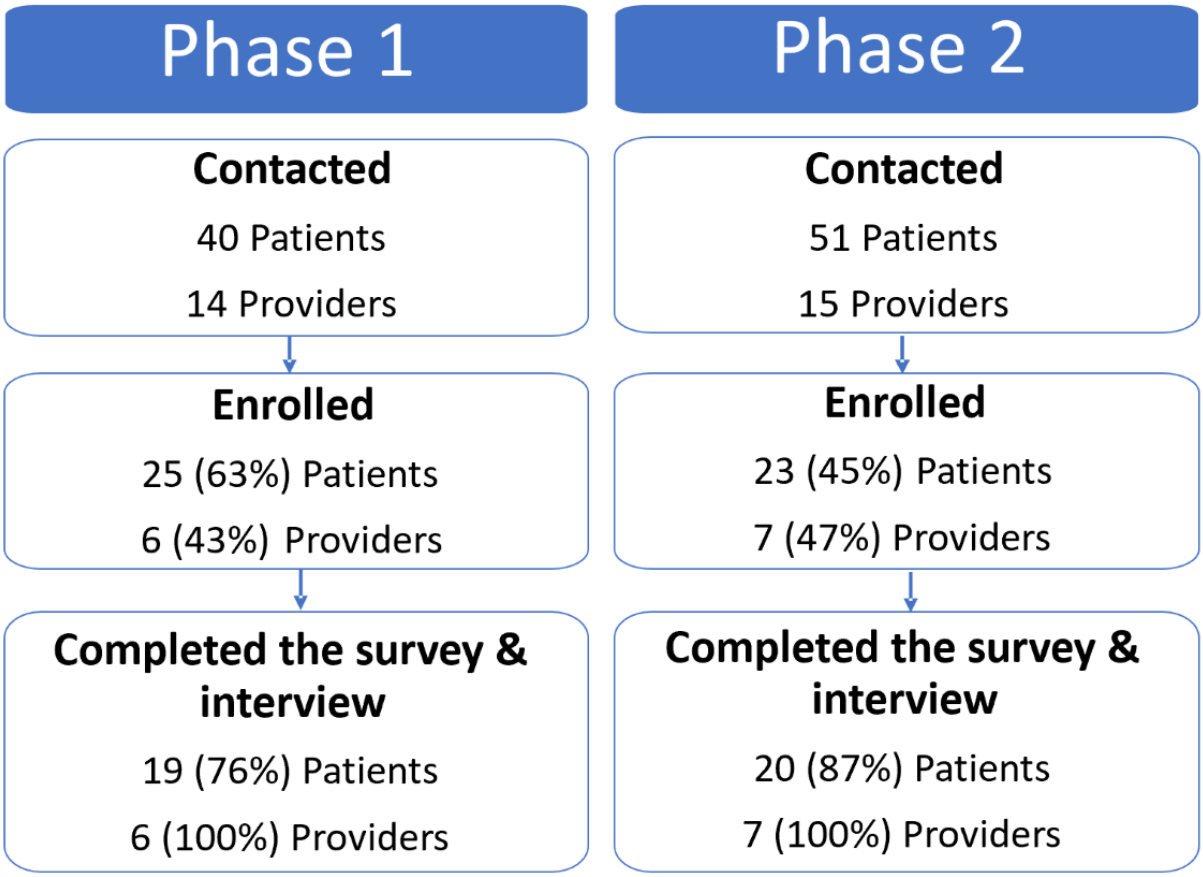
Results

Demographics and Clinical Characteristics

Phase 1

We contacted a total of 40 patients with early-onset CRC and 14 providers, of whom 19 patients and 6 providers completed the survey and interview (Figure 3). The most common reasons patients did not participate were a lack of interest or unsuitable timing due to their recent diagnosis and treatment. Providers generally declined participation passively.

Figure 3. Recruitment study flow for phases 1 and 2.



The median age of patients was 43 (range 27-51) years (Tables 2 and 3), with 10 out of 19 (53%) being female, 2 (11%) identifying as Black, 6 (32%) of Hispanic ethnicity, and 4 (21%) preferring Spanish (Table 1). About half (n=10) had at least some college education, and most were employed (n=11). One-third of the patients had stage 4 cancer (n=7), and half were undergoing active treatment (n=9). Most patients reported adequate health literacy (median 4, range 1-4) and technology literacy (median 4, range 1.6-4). Before completing Nest-CRC, 18 of the 19 (95%) patients indicated that GT was very important for cancer prevention and early detection, though only 10 (53%) and 15 (79%) reported awareness of GC and GT, respectively.

The median age of providers was 39 (range 31-46) years (Table 4). Half of the providers were female, all identified as White, and 1 was Hispanic and preferred Spanish. Two-thirds were medical or surgical oncologists, and 2 were genetic counselors. All providers had at least 2 years of experience with patients with CRC (median 11 years, range 2-13) and discussed genetic risk with patients at least 50% of the time.

Table 2. Patients' demographic characteristics by phase.

Demographic	Phase 1 (n=19)	Phase 2 (n=20)
Age (years), median (range)	43 (27-51)	47 (36-59)
Gender, n (%)		
Female	10 (53)	11 (55)
Male	9 (47)	9 (45)
Race, n (%)		
White only	14 (74)	12 (60)
Black only	2 (11)	4 (20)
More than one race	2 (11)	3 (15)
Other	1 (5)	1 (5)
Hispanic ethnicity, n (%)	6 (32)	7 (35)
Spanish-preferring, n (%)	4 (21)	4 (20)
Born in the US mainland, n (%)	15 (79)	14 (70)
Marital status, n (%)		
Married/partnered	13 (68)	14 (70)
Single	4 (21)	2 (10)
Divorced/separated	2 (11)	4 (20)
Education, n (%)		
<High school diploma	2 (11)	0 (0)
High school diploma or General Educational Development	6 (32)	1 (5)
Some college/vocational school	4 (21)	6 (30)
≥College graduate	7 (37)	13 (65)
Employment status, n (%)		
Employed[a]	12 (63)	14 (70)
Unemployed	3 (16)	2 (10)
Homemaker	1 (5)	1 (5)
Disable	3 (16)	3 (15)
Annual household income (US $), n (%)		
<19,999	3 (16)	3 (15)
20,000-49,999	5 (26)	0 (0)
50,000-99,999	2 (11)	4 (20)
≥100,000	8 (42)	10 (50)
Do not know	1 (5)	3 (15)
Insurance, n (%)		
No insurance	1 (5)	2 (10)
Private/commercial	15 (79)	15 (75)
Medicare/Medicaid	3 (16)	2 (10)
Other	0 (0)	1 (5)

[a]Includes self-employed.

Table 3. Patients' clinical characteristics by phase.

Clinical characteristics	Phase 1 (n=19)	Phase 2 (n=20)
Age at diagnosis (years), median (range)	41.5 (26-49)	44.5 (35-49)
Health literacy, median (range)	4 (1-4)	3.7 (1-4)
Technology literacy, median (range)	4 (1.6-4)	3.7 (2-4)
Type of cancer, n (%)		
Colon	11 (58)	11 (55)
Rectal	6 (32)	9 (45)
Do not know	2 (11)	0 (0)
Cancer stage, n (%)		
Stage 0	1 (5)	1 (5)
Stage 1	0 (0)	1 (5)
Stage 2	1 (5)	2 (10)
Stage 3	7 (37)	7 (35)
Stage 4	7 (37)	9 (45)
Do not know	3 (16)	0 (0)
Treatment history, n (%)		
No treatment	2 (11)	0 (0)
Received treatment[a]	17 (89)	20 (100)
Active cancer treatment, n (%)	9 (47)	7 (35)
Ashkenazi Jewish ancestry, n (%)	3 (16)	1 (5)
Family history of early-onset colorectal cancer, n (%)	2 (11)	0 (0)
Awareness of genetic counseling, n (%)	10 (53)	9 (45)
Awareness of genetic testing, n (%)	15 (79)	17 (85)
Awareness of hereditary cancers, n (%)	12 (63)	13 (65)
Awareness of Lynch syndrome, n (%)	4 (21)	7 (35)
Importance of genetic information for prevention , n (%)		
Very	18 (95)	15 (75)
Somewhat	1 (5)	4 (20)
A little	0 (0)	1 (5)
Not at all	0 (0)	0 (0)
Importance of genetics for early cancer detection , n (%)		
Very	18 (95)	18 (90)
Somewhat	1 (5)	2 (10)
A little/not at all	0 (0)	0 (0)
History of genetic counseling, n (%)	8 (42)	7 (35)
History of genetic testing, n (%)	10 (53)	16 (80)

[a]Treatment included surgery, chemotherapy, radiation, and immunotherapy.

Table 4. Providers' demographic and clinical characteristics.

Providers	Phase 1 (n=6)	Phase 2 (n=7)
Age (years), median (range)	39 (31-46)	40 (31-51)
Spanish-preferring, n (%)	1 (17)	0 (0)
Gender, n (%)		
Female	3 (50)	7 (100)
Male	3 (50)	0 (0)
Race, n (%)		
White only	6 (100)	3 (43)
Black	0 (0)	1 (14)
Asian only	0 (0)	2 (29)
Other	0 (0)	1 (14)
Hispanic ethnicity, n (%)	1 (17)	0 (0)
Marital status, n (%)		
Married/partnered	5 (83)	7 (100)
Single	1 (17)	0 (0)
Type of provider, n (%)		
Physician (MD)	4 (67)	2 (29)
Board-certified genetic counselor	2 (33)	2 (29)
Physician assistant	0 (0)	2 (29)
Nurse practitioner	0 (0)	1 (14)
Years working with patients with colorectal cancer, median (range)	11 (2-13)	10 (4-17)
Proportion of time communicating about the genetic risk to patients, n (%)	6 (100)	5 (71)
<10%	0 (0)	1 (14)
10%-29%	0 (0)	1 (14)
30%-49%	0 (0)	0 (0)
50%-69%	1 (17)	0 (0)
≥70%	5 (83)	5 (71)
Proportion of time referring patients to genetic services, n (%)		
<10%	0 (0)	2 (29)
10%-29%	2 (33)	1 (14)
30%-49%	0 (0)	0 (0)
50%-69%	3 (50)	1 (14)
≥70%	1 (17)	3 (43)
Proportion of time seeing patients with early-onset colorectal cancer , n (%)		
<10%	0 (0)	1 (14)
10%-29%	3 (50)	4 (57)
30%-49%	1 (17)	1 (14)
50%-69%	2 (33)	1 (14)
≥70%	0 (0)	0 (0)
Proportion of time working with racially/ethnically minority patients, n (%)		
<10%	0 (0)	0 (0)
10%-29%	3 (50)	2 (29)
30%-49%	1 (17)	3 (43)

Providers	Phase 1 (n=6)	Phase 2 (n=7)
50%-69%	2 (33)	2 (29)
≥70%	0 (0)	0 (0)

Phase 2

We contacted a total of 51 patients with early-onset CRC and 15 providers, of whom 20 patients and 7 providers completed the survey and interview (Figure 3). The most common reasons for nonparticipation were the same as in phase 1. The median age of phase 2 patients was 47 (range 36-59) years, with about half being female, 4 out of 20 (20%) identifying as Black, 7 (35%) as Hispanic, and 4 (20%) as Spanish-preferring (Tables 2 and 3). Most patients had at least some college education (n=18), were employed (n=13), and had health insurance (n=18). Patients also reported adequate health literacy (median 3.7, range 1-4) and technology literacy (median 3.7, range 2-4). Similar to phase 1, before completing Nest-CRC, 15 (75%) and 18 (90%) patients indicated that GT was very important for cancer prevention and early detection, respectively. However, only 9 (45%) and 17 (85%) reported awareness of GC and GT, respectively.

Nest-CRC Findings and Recommendations

Nest-CRC Quantitative Data

In phase 1, 9 patients reported no history of GT, and after completing the education, 7 (78%) were interested in GT, while 2 (22%) were unsure; none of the participants reported having no interest in GT. In phase 2, 4 patients reported no history of GT, and all of them indicated interest in GT after completing Nest-CRC. Across both phases, most participants reported that each of the Nest-CRC modules was useful and easy to use (phase 1: 23/25, 92%, to 25/25, 100%; phase 2: 24/27, 89%, to 27/27, 100%; Table 5). The average completion time for patients in phase 2 was 11 (range 5-26) minutes.

Table 5. Comprehension and usefulness of each Nest-CRC module.

Modules	Phase 1, n (%)		Phase 2, n (%)	
	Patients (n=19)	Providers (n=6)	Patients (n=20)	Providers (n=7)
Hereditary colorectal cancer, n (%)				
Easy to understand	19 (100)	5 (83)	18 (90)	7 (100)
Helpful	19 (100)	4 (67)	17 (85)	7 (100)
Genetic testing , n (%)				
Easy to understand	19 (100)	5 (83)	18 (90)	7 (100)
Helpful	19 (100)	6 (100)	19 (95)	7 (100)
Benefits and risks, n (%)				
Easy to understand	18 (95)	6 (100)	19 (95)	7 (100)
Helpful	19 (100)	5 (83)	19 (95)	7 (100)
Care recommendations, n (%)				
Easy to understand	19 (100)	5 (83)	18 (90)	7 (100)
Helpful	18 (95)	6 (100)	18 (90)	7 (100)
Family members implications, n (%)				
Easy to understand	19 (100)	6 (100)	19 (95)	7 (100)
Helpful	19 (100)	6 (100)	19 (95)	7 (100)

Qualitative Interviews

Attraction/Visual Appeal

Participants in both phases reported finding Nest-CRC visually appealing and well-suited to its goals. The layout was described by phase 1 participants as “straightforward,” “concise,” and “clean” (Table 6). Phase 1 participants appreciated how each slide presented information in “bite-sized” amounts, making it “easy to digest and read” and helping to prevent feelings of being overwhelmed, which echoed the quantitative findings. Recommendations for improvement included enhancing the “dark mode” to increase readability and incorporating more visual elements (eg, photos, animations, and diagrams) to maintain attention and simplify complex concepts.

Table 6. Interview excerpts by study phase.

Theme	Phase 1	Changes implemented for phase 2	Phase 2
Attraction/Visual Appeal	• "I really think it's concise. I think you'll lose people if you make them read through too much information even when it's important, so I thought it was – it was a good capture of the important information." [Participant #1219, patient] • "Maybe having something graphic might make it a little bit better. Because it's a lot of text. So, I don't know if maybe having either, like, a little video or animation, at least just for the introduction, that explains what genes are – and mutations or variants – are. That might be helpful." [Participant #1301, provider]	• Included more images and graphics • Simplified complex concepts • Used colors and font size to highlight important information • Improved readability for dark mode	• "Yes, I had an 8-year-old child at the time I was making that decision and wanted to know what would impact him the most [...] I thought it was pretty comprehensive [...] I found that very useful, especially having a child that I feel like, anything I needed – I wanted to know everything I could know." [Participant #2206, patient] • "I thought the technology was good, simple, you click, next, go to the next one, and when you finish number 1 it takes you to number 2. My question is, is that going to be the format you are going to use? Is it going to be that shape and color? It's a little bit boring." [Participant #2304, Spanish-speaking patient]
Comprehension	• "Yes, because I guess it said that you could be discriminated against. Obviously, that's a huge red flag in my opinion. So, that would be the only thing. It didn't really say much about it. So, that would be the only thing to deter me from getting it because, obviously, I'd have cancer. So, it's hard to get insurance period. So, if that makes it even harder to get insurance or my children hard to get insurance, then I wouldn't wanna get it. Or I would need it explained to me a little bit more so that I would know it's not really that big of a deal or it is a big deal." [Participant #1202, patient]	• Included introduction with summary information about the importance and benefits of GT[a], what to expect from the education, and the next steps for GT • Added more information about sporadic, familial, and hereditary cancer, GT versus somatic/tumor testing, incidental findings from GT, and insurance discrimination • Simplified information and provided examples about GT, patients who are at greater risk/in need of testing, and treatment recommendations based on results • Included additional optional information about genes, most common types of hereditary colorectal cancer syndromes, GC[b], and insurance discrimination • Added word definitions • Explained that cancer-causing gene variants are also known as gene mutations and referred to them as mutations throughout. • Added at the end a summary of the information covered	• "I thought that the explanations were really easy to understand for people like myself not in the medical field [...] So, it was really easy to understand. And I think kinda gave us a lot of information but not make it overwhelming." [Participant #2203, patient] • "I wanna suggest if there's any data about how minorities are hit pretty hard with colon cancer. If you could possibly put something like that in there because I know sometimes myself – I'm an African American – sometimes minorities feel a little bit afraid about doing the GT [...]" [Participant #2215, patient]
Cultural Acceptability	• "We know that this population [cancer patients under 50], they have different needs about treatment and other things and this is something that would be very helpful for this population." [Participant #1106, provider] • "I think the text was short enough that it was easy to read through, but I guess maybe having audio in case people, I don't know, can't read well." [Participant #1102, provider]	• Added voice-over feature	• "I think the language is quite basic, and concise, but very appropriate. I don't think some words are difficult to understand for a person from Peru or Venezuelan or Argentinean. The vocabulary is easy to understand. I didn't see any questionable vocabulary." [Participant #2304, Spanish-Speaking patient] • "[A language option] would be easier for [my mother and grandmother]. [...] Maybe having that little option might be better for them and more comfortable for them to participate." [Participant #2218, patient]

Theme	Phase 1	Changes implemented for phase 2	Phase 2
Self-Efficacy	• "It is easy to use, easy to navigate, bite-sized bits of information, which is completely different from all the other information you're getting in this process, quick, structured well, and it does take you towards the information you need to make a decision." [Participant #1213, patient] • "I think another reason I have felt more comfortable about it is because my provider had mentioned it to me and kinda talked about it, so I kind of understood it. So, it was just basically like an extra confirmation." [Participant #1215, patient] • "Well, I guess that's one comment then. That wasn't clear to me. [...] I don't remember seeing like if you don't have questions and want to proceed with testing, we can do it today versus seeing a counselor. And like, and then you see a counselor if you have any positives, or you could just see a counselor before testing because that wasn't clear to me. [...]" [Participant #1102, provider]	• Added information about GT cost, data security, GT/GC process and wait time, and behavioral and environmental cancer risk factors • Added and simplified information about genetic test results and care implications • Reordered the content so personal and family benefits will be presented before the risks • Insurance discrimination information was simplified	• "I mean, honestly overall it's a thing of – it's a positive thing overall. It's just – it's simple to be done and there's not – there won't be repercussions for finding out information [...] 'Cause that's what something that people worry about is – so, if I find out this is a hereditary thing are my kids gonna be denied whatever, insurance or whatever down the road because I did this now [...] So, you know – I think between that and how simple the test – knowing that there won't be those repercussions and knowing how simple the testing is I think that was a good message for me." [Participant #2208, patient] • "And that's the tricky part. If somebody gets diagnosed with colon cancer, the first they are thinking is am I going to live? Am I going to not live? And what's the treatment? They're not truly thinking about genetics. So, maybe once they met with the surgeon and oncologist and have had a treatment plan. At that time maybe a good way to bring in that. We have universal protocol anyways for colon cancer. Every colon cancer should be tested. So, it has to be brought up at some point. [...]" [Participant #2101, provider]
Usability/Utility	• "I think the app is pretty simple to navigate... I think overall, this does a pretty good job of making sure that somebody at 50 that doesn't have a whole lot of experience with technology and somebody at 20 who basically that's all he does is technology are able to use it. Again, I think it's pretty good across the board, but you're still gonna have some outliers." [Participant #1210, patient] • "Yes, because it was difficult for me because I needed my daughter's help and when my daughter wasn't there, I couldn't enter because I would copy the information and I would go to a link, and it wouldn't let me enter. Yes, it was difficult to me... Oh, that [being sent a text with a direct link] would be easier for those of us who don't know that much about technology. It would be easier where there is just a link, and you click on it and it takes you to the information." [Participant #1306, Spanish-speaking patient]	• The education was sent via SMS text message, and a direct link was provided via email or SMS text message • An audio option was implemented • Added an option to obtain additional information	• "I did like that it allows for audio and visual learning. That was kinda one of my favorite parts because a lot of times, essentially, some patients can hear what you have to say, but the ability to be able to kinda pause and listen or go back and listen again, I think that was a very smart use of the learning modality. I thought that the information was very clear and concise, and it wasn't very cumbersome. It wasn't just a lot of information on each slide." [Participant #2105, provider] • "A circle that came out there? I could see it moving but I didn't know what it was to read it for me. I thought they were listening to me. I kept reading and doing the up and down but I didn't use it to listen." [Participant #2301, Spanish-speaking patient] • "But I guess maybe what might be helpful is – I don't know how to say this. Sorta like an outline or something. You know what I mean? So that you can skip forward or skip back [...]." [Participant #2203, patient]

[a]GT: genetic testing.

[b]GC: genetic counseling.

The main factors most patients and providers emphasized in phase 2 were a desire to protect family members and trust in science and medical providers. Patients stated that a recommendation from a provider to review the education would influence their interest in it. The information provided in the education, along with its ease of use, was highlighted in phase 2 as the primary factor that attracted participants.

For the most part, participants did not have strong opinions about the visuals of the tool. Most described it as "clean" and "easy to navigate" and reported that the text was clear and easy to read. One participant found the aesthetic dull and "boring" and felt it needed more color. Another participant mentioned that the lack of personal interaction decreased their interest, preferring to speak with a person instead.

Comprehension

Participants in both phases found the information easy to comprehend and described the intervention as "straightforward," "simple," "succinct," "understandable for a layperson," and "super easy to understand."

Nearly all phase 1 patients stated that they understood GT could be used to determine genetic predisposition for their CRC and that learning whether they carry a particular gene could be helpful for family members. The most commonly cited point of confusion in phase 1 was the possibility of discrimination based on their GT results. This was surprising to many participants, with some requesting links or resources for further information. Other recommendations from phase 1 participants included adding information on the most common types of hereditary CRC syndromes (eg, Lynch syndrome, familial adenomatous polyposis), clarifying which patients are most at risk or in need of testing, distinguishing between GT and somatic/tumor testing, addressing incidental findings from GT (eg, beyond CRC, such as BRCA1), defining specific gene associations, discussing treatment options based on variants, reducing redundancies in "informed consent," and communicating clearly to lay audiences (eg, clarifying language and providing examples). When suggested by the interviewer, patients agreed that a dictionary tool to define unfamiliar terms would be useful.

Phase 2 participants felt the intervention provided just the right amount of information: educational, yet not overwhelming for a new patient. Most phase 2 providers found the intervention easy to understand, free of jargon, and containing an appropriate level of detail.

Although participants were overall very satisfied, they suggested some additional educational topics. One of the most commonly requested topics by both patients and providers was information on what other cancer and health condition risks can be detected via GT. The general consensus was that most patients would want GT to detect different cancers if it were covered by insurance or affordable. Some participants raised privacy concerns regarding who can access their results and what can be done with their DNA and results in the future. Many providers suggested that it could be beneficial to outline privacy policies within the intervention. Participants were also interested in the impact and next steps for family members based on their GT results. One participant requested more data on cancer prevalence by race/ethnicity.

Some participants wanted more information (eg, the likelihood of having Lynch syndrome, more details about specific genes, and variants of uncertain significance), while others felt the level of information provided was sufficient and that additional details should be discussed with a genetic counselor. In both phases 1 and 2, there was disagreement regarding the preferred term "variant" versus "mutation" and when to use each.

Cultural Acceptability

All participants in phase 1 and phase 2 indicated that they felt the information was acceptable to a wide audience of varied backgrounds. None reported concerns that the content of Nest-CRC was offensive or inappropriate for any groups. Instead, participants described the material as "neutral," "broad," and believed "it can help anyone."

Most of the concerns regarding the acceptability of Nest-CRC in phase 1 were related to accessibility. English-preferring participants inquired whether the education was available in Spanish and emphasized the need for patient-friendly language. Participants found the online delivery to be appropriate, given that the target audience is under 50 years of age, but recommended reevaluation if Nest-CRC is expanded to older patients. Participants also suggested adding an audio voice-over option or videos within Nest-CRC to improve accessibility for blind people/those with visual impairment, those not fully literate in English, or those who prefer listening over reading.

One phase 2 participant appreciated that the materials were inclusive of adopted patients who may be unaware of their family history. Phase 2 participants were also appreciative of the audio voice-over option.

Self-Efficacy

Both phase 1 and phase 2 participants described Nest-CRC as "very informative" and "comprehensive," and most felt that the intervention provided enough information to help them decide if they wanted GT without being overwhelming. Most patients had already undergone GT or were interested in it, evaluating Nest-CRC as a helpful way to be "proactive," with some wishing it had been available when they were first diagnosed.

Trust in their care team's recommendations was a major facilitator for using Nest-CRC. Participants responded positively when asked if they would complete the intervention if their provider recommended it, including patients who were undecided about GT. A few participants did not fully understand the benefits of GT (eg, therapeutic decision-making, risk management recommendations, and cascade testing), given that they already had cancer, with one commenting that it would be "pointless" now. Other potential barriers to GT included cost, information sharing/data security, and fear and anxiety around test results. Some participants also described fear of insurance discrimination as a potential barrier and felt this risk should be better clarified.

Self-efficacy for phase 2 participants was bolstered by their greater satisfaction with the information provided about insurance, costs, and the risks of discrimination, as well as the

available legal protections. Participants identified information on screenings and sample collection modalities (saliva vs blood) as important topics to support GT decision-making. Many providers and patients noted an assumption that GT was more invasive or complicated than it is and felt that providing information about the minimally invasive process could increase interest. The option for saliva testing was highlighted as important to make GT more appealing to patients afraid of blood and needles and more manageable for those undergoing multiple treatments or procedures for cancer care.

For phase 2, the remaining barriers to GT decision-making included unresolved concerns about discrimination and the ability to afford GT. Despite being informed that most insurance plans cover GT, participants remained concerned that their results could be positive and that cascade testing could be recommended for uninsured family members. One patient, who had previously experienced mishandling and misuse of medical results, expressed distrust in the efficacy of policies and institutions despite education about protections. Some newly diagnosed younger patients stated that being diagnosed with cancer was overwhelming and found it difficult to process information and make decisions, even with the necessary resources. With this in mind, some participants recommended that providers bring up or offer Nest-CRC again during the second or third visit to allow patients time to check insurance coverage for GC and GT.

Usability/Utility

Participants in both phase 1 and phase 2 described using the Nest-CRC tool as “user-friendly,” “very easy to use,” and “seamless.” They found the tool logically structured and easy to navigate, and appreciated that it is accessible on both desktop and mobile devices.

Phase 1 participants encountered minor technical issues, such as difficulty getting the link to work and not being able to return to where they left off. Participants felt usability could be improved by providing a direct link (rather than requiring copying and pasting into a browser), adding an audio option, allowing easier navigation of content out of sequence, and offering access to Nest-CRC via tablet in the clinic waiting room.

Phase 2 participants who noticed the audio option rated it positively, even if they preferred to read. However, this feature was not apparent to all users. For users with their phone set to “silent,” the audio option was also silenced, even if the phone volume was turned up. This led to confusion among many participants who were unable to hear the audio. One patient even thought the tool was recording audio because they could not hear it and noticed the sound wave animation within the audio button.

Summary and Next Steps

Most participants found each of the modules easy to understand and helpful. In general, patients described the content as straightforward, easy to comprehend, beneficial to anyone, informative, and useful for making decisions about GT. Following phase 1, usability was improved by providing participants with a direct link to Nest-CRC, sending links via SMS text message, and adding an audio option. The educational modules were also reordered to highlight the benefits first and end with the risks. Several additions were made to the educational content, including optional pathways and links for more detail on specific topics (eg, Lynch syndrome), definitions of key terms (eg, genes), graphs, images, and clarifying information about GT cost, insurance discrimination, and the GT process.

For the next phase, the audio voice-over will be improved, and additional optional pathways (eg, genes commonly tested in MGPT) will be incorporated. The revision will include more information about variants of uncertain significance, as well as content explaining the difference between GT and somatic testing, and addressing GT privacy concerns. Patients will also have the option to select their preferred language. Navigation within Nest-CRC will be simplified to allow easier return to core educational content from optional paths and modules.

Discussion

Principal Findings

Nest-CRC was developed as a user-friendly, scalable digital health tool designed to improve access to GT by facilitating pretest education at a lower cost. Findings indicated strong attraction, comprehension, cultural acceptability, self-efficacy, and usability of Nest-CRC in both phases. Endorsement of Nest-CRC was high, with participants recommending it be offered routinely and repeatedly to patients at different stages of the cancer journey, as some patients may prioritize GT immediately, while others may postpone it until survivorship care. Nest-CRC was also described as an acceptable alternative for empowering patients with information about GT and supporting their decision-making. Therefore, it was identified as a viable strategy for streamlining patients with early-onset CRC toward GT.

Comparison With Prior Work

NCCN guidelines recommend MGPT for all patients with early-onset CRC [10], but referrals to genetic services for these patients are inconsistent [11,12]. A prior study from 2 Texas health systems, evaluating data from 2009 to 2017, revealed that 58% of patients with early-onset CRC were referred for GC, and only 37% completed GT [11]. A more recent study, examining retrospective data from 2010 to 2019 at Cleveland Clinic, found that 62% of patients with early-onset CRC were referred to GC, 49% completed GC, and 48% completed GT [12]. In our study, less than half of the patients reported a history of GC, while two-thirds reported a history of GT. The proportion of patients reporting GT in our study was higher than previously reported rates for patients with early-onset CRC. However, interviews revealed that some patients were confusing their experiences with somatic testing and germline testing, which may have contributed to the elevated reporting of GT. Among those patients with early-onset CRC who denied GT at baseline, most expressed interest in GT after completing Nest-CRC. Additionally, the few patients who were unsure about GT showed interest in GC. This suggests that Nest-CRC can be used to streamline the triage of patients with early-onset CRC

for GT, while complex patients needing more support can be prioritized to maximize the efficiency of limited GC resources.

Prior studies examining barriers to genetic services among patients with early-onset CRC have identified cost, a shortage of qualified genetic professionals, and racial/ethnic referral disparities [11,19]. In our study, patients expressed concerns about the cost of GT, insurance coverage and discrimination, and the potential misuse of their DNA. While the cost of GT remains a concern, it has decreased substantially, making it more affordable for many individuals without health insurance. In the United States, patients who meet insurance criteria for GT typically pay between US $0 and US $100 out of pocket, while those without insurance may pay around US $300. Additionally, some laboratories offer financial assistance. Furthermore, the United States has federal and state laws that protect patients from insurance discrimination. For example, the Genetic Information Nondiscrimination Act (GINA) prevents nonmilitary employers and health insurance companies from using GT results against patients [39]. However, GINA does not apply to life, disability, or long-term insurance companies. Some states, like Florida, have enacted laws that protect patients from life, disability, or long-term insurance companies using GT results against individuals residing in the state [40,41]. Therefore, patients' commonly reported concerns about GT could be alleviated through education about existing resources and federal and state protections.

To address barriers to accessing GT information and services, the Clinical Genome Resource's Consent and Disclosure Recommendations working group recommended a brief pretest genetic education approach for more straightforward cases, reserving traditional GC for patients with greater clinical and genetic complexities or those without well-established testing recommendations [20]. Automated educational tools, such as videos and written materials, have proven effective in delivering genetic education, particularly for patients with high-risk cancer [21,22,25,42]. For example, a study involving patients with pancreatic cancer found that when oncology providers used an educational video to obtain GT informed consent, the rate of GT increased 6.5 times compared with previous years, when traditional GC referrals were used [21]. However, these advances still rely on clinic staff and providers to obtain informed consent and order GT.

Nest-CRC provides an alternative for GT education that can further alleviate the burden on patients, clinic staff, and institutions. Like other patient-driven digital tools [42,43], Nest-CRC can be completed conveniently from home, enhancing accessibility without incurring out-of-pocket costs for the patient. For institutions, the annual cost of using Nest-CRC can be tailored to the specific functionalities required, averaging a few thousand dollars per month—less than the cost of an entry-level GC assistant. Unlike a single assistant, Nest-CRC can scale the volume and capabilities of genetic services without being impacted by patient load, staff turnover, or the need for ongoing training. By automating workflows for education, consent, test ordering, and results return, tools such as Nest-CRC can significantly enhance clinic efficiency. While this study focuses on the perceived benefits of pretest genetic education, implementing a platform such as Nest has the potential to generate substantial cost savings and operational efficiencies across other routine tasks, such as family history collection, patient tracking, risk model calculations, and care plan management. As the demand for genetic services continues to grow, digital health tools such as Nest-CRC could be leveraged to identify high-risk patients and promote GT across various health care settings, such as primary care, oncology clinics, and even the general public. In this way, Nest-CRC offers an acceptable alternative strategy to expand equitable access to GT among high-risk patients.

Strengths and Limitations

The primary goal of this study was to evaluate the reactions and recommendations of English- and Spanish-speaking patients and providers to Nest-CRC, with preliminary data on patients' interest in GT as a secondary aim. This study is innovative in providing valuable insights from a diverse group of patients and providers, highlighting key considerations for developing pretest genetic education for patients with early-onset CRC. However, due to the small sample size, quantitative data were limited to descriptive purposes and were used primarily to describe the study sample and support the qualitative findings by incorporating patients' immediate feedback on the educational content. While we intentionally recruited Black and Spanish-speaking patients to ensure diverse representation, the findings should not be generalized to all patients with early-onset CRC. It is also important to note that during interviews, some participants expressed uncertainty about their GT history; therefore, the self-reported GT data should be interpreted with caution. This uncertainty underscores the potential benefit of a comprehensive digital health platform where GT history and related lifelong care recommendations can be easily accessed and shared with patients' clinicians and family members as needed.

Conclusions and Future Directions

Adults with early-onset CRC are at higher risk for having a hereditary cancer syndrome. GT to identify the causative variant can facilitate screening and risk reduction measures for both patients and their relatives. Despite GT being recommended for all patients with early-onset CRC, racial disparities persist in referrals for GT, access to GC, and uptake of both GC and GT. These issues are further compounded by a shortage of qualified genetic professionals and patients' concerns about the cost of genetic services. Our findings suggest that Nest-CRC is a promising strategy to scale genetic services by augmenting pretest genetic education and promoting GT uptake among patients with early-onset CRC from diverse racial and ethnic backgrounds. Future studies should implement digital GT platforms in clinical settings to evaluate their feasibility and acceptability among high-risk patients and their relatives from diverse racial and ethnic backgrounds, as well as assess their impact on lifelong care recommendations and survival outcomes.

Acknowledgments

This project was funded by the Society of Behavioral Medicine Bridging the Gap Research Award and the National Cancer Institute (grant T32 CA090314; principal investigators: Thomas Brandon and SV). We also want to acknowledge the support of Stefania Alastre Arcusa, MSPH, LCGC, for reviewing and providing feedback on the Nest-CRC education in Spanish for phases 1 and 2.

Data Availability

Deidentified data generated and analyzed during this study will be made available (as allowable according to institutional review board standards) by emailing the corresponding author.

Authors' Contributions

JRR wrote the first draft of the manuscript. JRR and M Sholeh were involved in data collection, data curation, and quantitative data analysis. MLM, CG, and HF conducted qualitative data analysis and wrote the first draft of the qualitative results. JRR and SV conceptualized the study and were responsible for funding acquisition, project administration, and supervision of the study team. M Snir, ES, and TS implemented Nest-CRC. JRR, M Snir, ES, TS, M Sholeh, LB, BDG, JP, and SV contributed to the development and revisions of Nest-CRC. All authors reviewed and edited the manuscript.

Conflicts of Interest

M Snir is the CEO of Nest Genomics, and ES and TS are employees of Nest Genomics. The remaining authors declare no potential conflicts of interest.

References

1. U.S. Cancer Statistics Working Group. Cancer statistics at a glance. U.S. Centers for Disease Control and Prevention (CDC). 2021. URL: https://www.cdc.gov/cancer/dataviz [accessed 2025-01-07]
2. Rahib L, Wehner MR, Matrisian LM, Nead KT. Estimated projection of US cancer incidence and death to 2040. JAMA Netw Open 2021 Apr 01;4(4):e214708 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.4708] [Medline: 33825840]
3. Stoffel EM, Koeppe E, Everett J, Ulintz P, Kiel M, Osborne J, et al. Germline genetic features of young individuals with colorectal cancer. Gastroenterology 2018 Mar;154(4):897-905.e1 [FREE Full text] [doi: 10.1053/j.gastro.2017.11.004] [Medline: 29146522]
4. Patel SG, Karlitz JJ, Yen T, Lieu CH, Boland CR. The rising tide of early-onset colorectal cancer: a comprehensive review of epidemiology, clinical features, biology, risk factors, prevention, and early detection. Lancet Gastroenterol Hepatol 2022 Mar;7(3):262-274. [doi: 10.1016/S2468-1253(21)00426-X] [Medline: 35090605]
5. Uson PLS, Riegert-Johnson D, Boardman L, Kisiel J, Mountjoy L, Patel N, et al. Germline cancer susceptibility gene testing in unselected patients with colorectal adenocarcinoma: a multicenter prospective study. Clin Gastroenterol Hepatol 2022 Mar;20(3):e508-e528 [FREE Full text] [doi: 10.1016/j.cgh.2021.04.013] [Medline: 33857637]
6. Mandelker D, Zhang L, Kemel Y, Stadler ZK, Joseph V, Zehir A, et al. Mutation detection in patients with advanced cancer by universal sequencing of cancer-related genes in tumor and normal DNA vs guideline-based germline testing. JAMA 2017 Sep 05;318(9):825-835 [FREE Full text] [doi: 10.1001/jama.2017.11137] [Medline: 28873162]
7. Gupta S, Provenzale D, Llor X, Halverson AL, Grady W, Chung DC, CGC, CGC, CGC, et al. NCCN guidelines insights: genetic/familial high-risk assessment: colorectal, version 2.2019. J Natl Compr Canc Netw 2019 Sep 01;17(9):1032-1041. [doi: 10.6004/jnccn.2019.0044] [Medline: 31487681]
8. Boardman LA, Vilar E, You YN, Samadder J. AGA clinical practice update on young adult-onset colorectal cancer diagnosis and management: expert review. Clin Gastroenterol Hepatol 2020 Oct;18(11):2415-2424. [doi: 10.1016/j.cgh.2020.05.058] [Medline: 32525015]
9. Valle L, Vilar E, Tavtigian SV, Stoffel EM. Genetic predisposition to colorectal cancer: syndromes, genes, classification of genetic variants and implications for precision medicine. J Pathol 2019 Apr;247(5):574-588 [FREE Full text] [doi: 10.1002/path.5229] [Medline: 30584801]
10. Gupta S, Weiss JM, Axel L, Burke CA, Chen LM, Chung DC, et al. NCCN guidelines: genetic/familial high-risk assessment: colorectal, endometrial, and gastric (version 3.2024). National Comprehensive Cancer Network (NCCN). 2023. URL: https://www.nccn.org/guidelines/guidelines-detail?category=2&id=1544 [accessed 2024-06-26]
11. Dharwadkar P, Greenan G, Stoffel EM, Burstein E, Pirzadeh-Miller S, Lahiri S, et al. Racial and ethnic disparities in germline genetic testing of patients with young-onset colorectal cancer. Clin Gastroenterol Hepatol 2022 Feb;20(2):353-361.e3 [FREE Full text] [doi: 10.1016/j.cgh.2020.12.025] [Medline: 33359728]
12. Syed H, Sommovilla J, Burke CA, McGee S, Macaron C, Heald B, et al. Referral, uptake, and outcome of genetic counseling and testing in patients with early-onset colorectal cancer. J Natl Compr Canc Netw 2023 Nov;21(11):1156-1163.e5. [doi: 10.6004/jnccn.2023.7057] [Medline: 37935108]

13. Muller C, Lee SM, Barge W, Siddique SM, Berera S, Wideroff G, et al. Low referral rate for genetic testing in racially and ethnically diverse patients despite universal colorectal cancer screening. Clin Gastroenterol Hepatol 2018 Dec;16(12):1911-1918.e2 [FREE Full text] [doi: 10.1016/j.cgh.2018.08.038] [Medline: 30130624]
14. Hoskovec JM, Bennett RL, Carey ME, DaVanzo JE, Dougherty M, Hahn SE, et al. Projecting the supply and demand for certified genetic counselors: a workforce study. J Genet Couns 2018 Feb;27(1):16-20. [doi: 10.1007/s10897-017-0158-8] [Medline: 29052810]
15. Jenkins BD, Fischer CG, Polito CA, Maiese DR, Keehn AS, Lyon M, et al. The 2019 US medical genetics workforce: a focus on clinical genetics. Genet Med 2021 Aug;23(8):1458-1464 [FREE Full text] [doi: 10.1038/s41436-021-01162-5] [Medline: 33941882]
16. Rodriguez NJ, Ricker C, Stoffel EM, Syngal S. Barriers and facilitators to genetic education, risk assessment, and testing: considerations on advancing equitable genetics care. Clin Gastroenterol Hepatol 2023 Jan;21(1):3-7 [FREE Full text] [doi: 10.1016/j.cgh.2022.10.025] [Medline: 36549838]
17. Sharma Y, Cox L, Kruger L, Channamsetty V, Haga SB. Evaluating primary care providers' readiness for delivering genetic and genomic services to underserved populations. Public Health Genomics 2021 Sep 07:1-10 [FREE Full text] [doi: 10.1159/000518415] [Medline: 34515218]
18. Robson ME, Bradbury AR, Arun B, Domchek SM, Ford JM, Hampel HL, et al. American Society of Clinical Oncology policy statement update: genetic and genomic testing for cancer susceptibility. J Clin Oncol 2015 Nov 01;33(31):3660-3667. [doi: 10.1200/JCO.2015.63.0996] [Medline: 26324357]
19. Noll A, J Parekh P, Zhou M, Weber TK, Ahnen D, Wu X, et al. Barriers to Lynch syndrome testing and preoperative result availability in early-onset colorectal cancer: a national physician survey study. Clin Transl Gastroenterol 2018 Sep 20;9(9):185 [FREE Full text] [doi: 10.1038/s41424-018-0047-y] [Medline: 30237431]
20. Hallquist MLG, Tricou EP, Ormond KE, Savatt JM, Coughlin CR, Faucett WA, et al. Application of a framework to guide genetic testing communication across clinical indications. Genome Med 2021 Apr 29;13(1):71 [FREE Full text] [doi: 10.1186/s13073-021-00887-x] [Medline: 33926532]
21. Ramsey ML, Tomlinson J, Pearlman R, Abushahin L, Aeilts A, Chen H, et al. Mainstreaming germline genetic testing for patients with pancreatic cancer increases uptake. Fam Cancer 2023 Jan;22(1):91-97 [FREE Full text] [doi: 10.1007/s10689-022-00300-5] [Medline: 35713757]
22. Russo J, McDougall C, Bowler N, Shimada A, Gross L, Hyatt C, et al. Pretest genetic education video versus genetic counseling for men considering prostate cancer germline testing: a patient-choice study to address urgent practice needs. JCO Precis Oncol 2021;5:1377-1386 [FREE Full text] [doi: 10.1200/PO.21.00238] [Medline: 34589662]
23. Birch PH. Interactive e-counselling for genetics pre-test decisions: where are we now? Clin Genet 2015 Mar;87(3):209-217. [doi: 10.1111/cge.12430] [Medline: 24828354]
24. Nazareth S, Hayward L, Simmons E, Snir M, Hatchell KE, Rojahn S, et al. Hereditary cancer risk using a genetic chatbot before routine care visits. Obstet Gynecol 2021 Dec 01;138(6):860-870 [FREE Full text] [doi: 10.1097/AOG.0000000000004596] [Medline: 34735417]
25. Quinn VF, Meiser B, Kirk J, Tucker KM, Watts KJ, Rahman B, et al. Streamlined genetic education is effective in preparing women newly diagnosed with breast cancer for decision making about treatment-focused genetic testing: a randomized controlled noninferiority trial. Genet Med 2017 Apr;19(4):448-456 [FREE Full text] [doi: 10.1038/gim.2016.130] [Medline: 27684037]
26. Conley CC, Castro-Figueroa EM, Moreno L, Dutil J, García JD, Burgos C, et al. A pilot randomized trial of an educational intervention to increase genetic counseling and genetic testing among Latina breast cancer survivors. J Genet Couns 2021 Apr;30(2):394-405 [FREE Full text] [doi: 10.1002/jgc4.1324] [Medline: 32936981]
27. Scheinberg T, Young A, Woo H, Goodwin A, Mahon KL, Horvath LG. Mainstream consent programs for genetic counseling in cancer patients: a systematic review. Asia Pac J Clin Oncol 2021 Jun;17(3):163-177. [doi: 10.1111/ajco.13334] [Medline: 32309911]
28. Manne SL, Meropol NJ, Weinberg DS, Vig H, Ali-Khan CZ, Manning C, et al. Facilitating informed decisions regarding microsatellite instability testing among high-risk individuals diagnosed with colorectal cancer. J Clin Oncol 2010 Mar 10;28(8):1366-1372 [FREE Full text] [doi: 10.1200/JCO.2009.25.0399] [Medline: 20142594]
29. Loader S, Shields C, Levenkron JC, Fishel R, Rowley PT. Patient vs. physician as the target of educational outreach about screening for an inherited susceptibility to colorectal cancer. Genet Test 2002;6(4):281-290. [doi: 10.1089/109065702760471813] [Medline: 12537652]
30. Snir M, Nazareth S, Simmons E, Hayward L, Ashcraft K, Bristow SL, et al. Democratizing genomics: leveraging software to make genetics an integral part of routine care. Am J Med Genet C Semin Med Genet 2021 Mar;187(1):14-27. [doi: 10.1002/ajmg.c.31866] [Medline: 33296144]
31. Rivera Rivera J, Fuzzell LN, Garcia J, Rathwell J, Robinson EJ, Chavez M, et al. Development of a patient activation toolkit for hepatitis C virus testing. J Cancer Educ 2023 Jun;38(3):931-939 [FREE Full text] [doi: 10.1007/s13187-022-02209-0] [Medline: 35971055]
32. National Cancer Institute. HINTS 5, Cycle 4 (2020) survey materials. HINTS. 2020. URL: https://hints.cancer.gov/data/survey-instruments.aspx#H5C4 [accessed 2025-01-07]

33. Nelson LA, Pennings JS, Sommer EC, Popescu F, Barkin SL. A 3-item measure of digital health care literacy: development and validation study. JMIR Form Res 2022 Apr 29;6(4):e36043 [FREE Full text] [doi: 10.2196/36043] [Medline: 35486413]
34. Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008 May;23(5):561-566 [FREE Full text] [doi: 10.1007/s11606-008-0520-5] [Medline: 18335281]
35. Chavarria EA, Christy SM, Simmons VN, Vadaparampil ST, Gwede CK, Meade CD. Learner verification: a methodology to create suitable education materials. Health Lit Res Pract 2021 Jan 11;5(1):e49-e59 [FREE Full text] [doi: 10.3928/24748307-20210201-02] [Medline: 34251862]
36. Scherr CL, Nam K, Augusto B, Kasting ML, Caldwell M, Lee MC, et al. A framework for pilot testing health risk video narratives. Health Commun 2020 Jun;35(7):832-841 [FREE Full text] [doi: 10.1080/10410236.2019.1598612] [Medline: 30999777]
37. Gwede CK, Jean-Francois E, Quinn GP, Wilson S, Tarver WL, Thomas KB, Tampa Bay Community Cancer Network partners. Perceptions of colorectal cancer among three ethnic subgroups of US blacks: a qualitative study. J Natl Med Assoc 2011 Aug;103(8):669-680 [FREE Full text] [doi: 10.1016/s0027-9684(15)30406-5] [Medline: 22046844]
38. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
39. Chapman CR, Mehta KS, Parent B, Caplan AL. Genetic discrimination: emerging ethical challenges in the context of advancing technology. J Law Biosci 2020;7(1):lsz016 [FREE Full text] [doi: 10.1093/jlb/lsz016] [Medline: 34221431]
40. Sprowls, Williamson, Aloupis, Andrade, Avila, Buchanan, Burton, Casello, Daniels, DiCeglie, Drake, Duggan, Fernandez-Barquin, Fischer, Gottlieb, Ingoglia, Latvala, Massullo, McClure, Perez, Plakon, Plasencia, Rommel, Sirois, Sullivan, Tomkow, Trumbull, Zika. HB 1189 — Genetic information for insurance purposes. The Florida Senate. 2020. URL: https://www.flsenate.gov/Committees/BillSummaries/2020/html/2232 [accessed 2025-01-07]
41. National Human Genome Research Institute. Genome Statute and Legislation Database. National Human Genome Research Institute. URL: https://www.genome.gov/about-genomics/policy-issues/Genome-Statute-Legislation-Database [accessed 2025-01-07]
42. Loeb S, Keith SW, Cheng HH, Leader AE, Gross L, Sanchez Nolasco T, et al. TARGET: a randomized, noninferiority trial of a pretest, patient-driven genetic education webtool versus genetic counseling for prostate cancer germline testing. JCO Precis Oncol 2024 Mar;8:e2300552 [FREE Full text] [doi: 10.1200/PO.23.00552] [Medline: 38452310]
43. Adam S, Birch PH, Coe RR, Bansback N, Jones AL, Connolly MB, et al. Assessing an interactive online tool to support parents' genomic testing decisions. J Genet Couns 2018 Jul 23;28(1):10-17. [doi: 10.1007/s10897-018-0281-1] [Medline: 30033481]

Abbreviations

ASCO: American Society of Clinical Oncology
CRC: colorectal cancer
GC: genetic counseling
GINA: Genetic Information Nondiscrimination Act
GT: genetic testing
HINTS: Health Information National Trends Survey
MGPT: multigene panel testing
NCCN: National Comprehensive Cancer Network

Edited by N Cahill; submitted 12.04.24; peer-reviewed by H Zierhut, K Le Du; comments to author 04.07.24; revised version received 29.08.24; accepted 16.12.24; published 17.01.25.

Please cite as:
Rivera Rivera JN, Snir M, Simmons E, Schmidlen T, Sholeh M, Maconi ML, Geiss C, Fulton H, Barton L, Gonzalez BD, Permuth J, Vadaparampil S
Developing and Assessing a Scalable Digital Health Tool for Pretest Genetic Education in Patients With Early-Onset Colorectal Cancer: Mixed Methods Design
JMIR Cancer 2025;11:e59464
URL: https://cancer.jmir.org/2025/1/e59464
doi:10.2196/59464
PMID:

Perception and Counseling for Cardiac Health in Breast Cancer Survivors Using the Health Belief Model: Qualitative Analysis

Sarah Tucker Marrison, MD, PhD; Nicholas Shungu, MPH, MD; Vanessa Diaz, MSCR, MD

Medical University of South Carolina, 135 Cannon Street, Suite 405, Charleston, SC, United States

Corresponding Author:
Sarah Tucker Marrison, MD, PhD
Medical University of South Carolina, 135 Cannon Street, Suite 405, Charleston, SC, United States

Abstract

Background: Breast cancer survivors have increased cardiovascular risk compared to those without cancer history. Cardiovascular disease is the most common cause of death in breast cancer survivors. Cardiovascular risk in breast cancer survivors is impacted by both cancer treatment–associated effects and in risk factors for breast cancer and cardiovascular disease overlap. Strategies to improve screening for and management of cardiovascular disease in breast cancer survivors are needed to improve the delivery of survivorship care.

Objective: This study aims to assess current cardiovascular risk counseling practices and perceived cardiovascular risk in breast cancer survivors.

Methods: Semistructured interviews were conducted from May to December 2021 with breast cancer survivors identified as having a primary care clinician within an academic family medicine center in Charleston, South Carolina. The interview guide and content were developed using the Health Belief Model with a focus on cardiovascular risk behaviors, risk perception, and barriers to risk reduction. Analysis of categorical data was conducted by frequency and quantitative variables by mean and SD. Template analysis was performed for qualitative analysis. Outcome measures included self-reported history of cardiovascular disease, risk perception, and risk behaviors.

Results: The average age of participants (n=19) was 54 (SD 7) years; 68% (13/19) were White and 32% (6/19) were Black or African American. Of the interviewed women, 90% (17/19) reported a personal history and 90% (17/19) reported a family history of cardiovascular disease. Only 53% (10/19) had previously reported receipt of cardiovascular counseling. Primary care most commonly provided counseling, followed by oncology. Among breast cancer survivors, 32% (6/19) reported being at increased cardiovascular risk, and 47% (9/19) were unsure of their relative cardiovascular risk. Factors affecting perceived cardiovascular risk included family history, cancer treatments, cardiovascular diagnoses, and lifestyle factors. Video (15/19, 79%) and SMS text messaging (13/19, 68%) were the most highly reported mechanisms through which breast cancer survivors requested to receive additional information and counseling on cardiovascular risk and risk reduction. Commonly reported barriers to risk reduction such as physical activity included time for meal planning and exercise, resources to support dietary and exercise changes, physical limitations, and competing responsibilities. Barriers specific to survivorship status included concerns for immune status during the COVID-19 pandemic, physical limitations associated with cancer treatment, and psychosocial aspects of cancer survivorship.

Conclusions: Breast cancer survivors identified that factors associated with their cancer diagnosis and treatment both impacted their cardiovascular risk and introduced additional barriers to risk reduction. Potential strategies to improve counseling and awareness around cardiovascular risk include video and messaging platforms. Further risk reduction strategies should consider the unique challenges of cancer survivorship in delivery and implementation.

(JMIR Cancer 2025;11:e71062) doi:10.2196/71062

KEYWORDS

cardiovascular health; cancer survivorship; lifestyle counseling; breast cancer; cancer survivors

Introduction

Breast cancer survivors are at increased risk of cardiovascular disease in part due to pre-existing comorbidities and cardiotoxicity associated with treatment [1]. Breast cancer survivorship has increased significantly, with over 3 million current breast cancer survivors. Breast cancer survivors are at increased risk of cardiovascular disease including heart failure, myocardial infarction, coronary artery disease, and vascular disease compared to those without a history of malignancy [2]. The etiology of the increased risk includes common risk factors, pre-existing comorbidities, and cardiotoxicity associated with breast cancer treatment. The likelihood of morbidity and mortality from cardiovascular disease exceeds that of breast cancer for many older women and those who have a personal history of cardiovascular disease [3-5]. In breast cancer survivors

older than 50 years of age, 35% of nonbreast cancer–associated mortality was attributed to cardiovascular disease [6]. In one cross-sectional study, over 60% of breast cancer survivors were diagnosed with hypertension, over 50% with hyperlipidemia, and 5% - 6% had a reported history of heart failure or stroke [7]. Cardiovascular comorbidity is common in breast cancer with significant implications for morbidity and mortality requiring screening, counseling, and management with a multidisciplinary health care team including primary care.

Comorbid risk factors for breast cancer include increasing age, alcohol consumption, dietary patterns, family history of cardiovascular disease, elevated BMI, and physical inactivity [2]. Breast cancer treatments can increase the risk of cardiovascular disease, including chemotherapy, endocrine therapy, and radiation therapy, although the specific cardiovascular risk profiles vary based on treatment history [2,8]. Management of comorbid medical conditions has also been shown to be suboptimal during active cancer treatment. Cancer disease processes themselves may cause subclinical myocardial damage [9]. Current strategies for risk factor reduction in breast cancer survivors include aggressive management of comorbid conditions, including hypertension, diabetes, and hyperlipidemia, and promotion of lifestyle factors such as smoking cessation, maintaining an appropriate body weight, and increased physical activity [1,10-12]. Survivorship guidelines for breast cancer recommend screening and management of cardiovascular risk among patients with breast cancer; however, current screening practices and management are inadequate [3,13]. Lifestyle interventions including nutrition, physical activity, and weight management are essential components to integrate into survivorship care to reduce cardiovascular risk [14].

When considering cardiovascular disease and breast cancer risk, it is advised to weigh factors both of breast cancer reoccurrence and cardiovascular disease risk [15]. However, previous studies have identified that women at increased risk of cardiovascular disease also have an increased risk of cancer reoccurrence [16]. Interventions targeting cardiovascular risk reduction also have the potential to impact breast cancer outcomes. Despite this, limited qualitative and interventional studies have evaluated strategies and interventions for cardiovascular risk reduction in breast cancer survivors. Observational studies have demonstrated that increasing physical activity and dietary changes improve all-cause survival and improve cardiovascular risk in breast cancer survivors [17,18]. Pilot studies using exercise-based interventions have been demonstrated to improve cardiovascular fitness in breast cancer survivors. A randomized trial by Lee et al [19] demonstrated an improved cardiovascular risk profile following participation in a 16-week exercise program [20]. Cardiovascular risk reduction consistent with guidelines and evidence-based interventions focuses on smoking cessation, increased physical activity, and dietary changes, although the specific dietary recommendations vary based on guidelines and interventions studied [21]. The purpose of this project is to identify multilevel barriers and facilitators to cardiovascular risk reduction among breast cancer survivors to best advise breast cancer survivors and to support the development of interventions focused on the unique demands of breast cancer survivors following treatment.

Methods

Qualitative analysis using semistructured interviews informed by the Health Belief Model was conducted to evaluate cardiovascular counseling and perceived risk in breast cancer survivors.

Inclusion Criteria

Inclusion criteria included female breast cancer survivors aged 40 - 65 years without metastatic disease seen by primary care at the Medical University of South Carolina in Charleston. The clinic population at the time of study initiation was predominately White and approximately 30% Black or African American. A majority of clinic patients were insured with either public or commercial insurance. For the purposes of this study, survivorship was defined as women who had completed their primary treatment for breast cancer. However, ongoing maintenance therapy, including endocrine therapy, was not considered an exclusion criterion.

Interviews

Participants were contacted using MyChart messaging within the electronic health record. Potentially eligible participants were identified through an electronic health record data inquiry based on the history of breast cancer and primary care appointments at the academic medical center within the last 3 years. Only individuals who had opted into a system-wide query about research contact were included in outreach. Of the individuals contacted, 11% (19/170) participated in the study. Outreach was stratified by age (<50 years old and >50 years old) and race (White, Black, or African American) with outreach equally to these groups to promote representation consistent with clinic demographics. A total of 19 semistructured interviews were conducted on Teams (Microsoft Corp) using audio recording only by a trained female, master-level study coordinator. The interview duration ranged from 11 to 17 minutes. Interviews were conducted until saturation of depth and diversity of themes occurred. Neither the interviewer nor the study team had previous interactions with the participants. Study objectives were provided to interview participants prior to informed consent. Interviews were transcribed and reviewed for accuracy. Participants were not contacted following the initial interview. Participants were provided with contact information for the study team with an opportunity to reach out about any additional questions or study results. Our research team is composed of primary care clinicians and public health researchers involved in primary care research, including researchers with expertise in onco-primary care. Our team values patient self-efficacy and the importance of interdisciplinary coordination of care guided by a patient-centered approach to care. To reduce potential bias, the study team member completing the interviews was not involved in data coding. Coding was completed independently and reviewed by consensus.

Conceptual Framework

The study was grounded in the Health Belief Model as a conceptual framework [22]. The Health Belief Model was selected based on its capacity to understand how individuals respond to communication of health-related information and their subsequent engagement in health-related behaviors. The Health Belief Model has been extensively used to assess cardiovascular risk perception, barriers to behavior change, and the development of interventions to address these [23-27]. More specifically, in this study, the focus on communication of information about cardiovascular risk in breast cancer survivors was predicted to inform and desire to and engagement in health behaviors to reduce cardiovascular risk. The interview guide was developed by the primary author after a review of previous literature, with revision following feedback from the research study team. Training was provided in the interview guide. Key constructs evaluated included perceived susceptibility to cardiovascular disease, perceived severity, self-efficacy, perceived barriers, and cues to action for engaging in cardiovascular risk reduction behaviors. Perceived susceptibility provides essential information on the participant's assessment of the probability of cardiovascular disease. Perceived severity focused on the understanding of the potential severity of health impacts of cardiovascular disease and its relation to breast cancer survivorship status. Participants identified their perceived barriers to engaging in health behaviors to promote improved cardiovascular health. Self-efficacy was assessed based on the patient's reported belief in their capacity to engage in health behaviors to reduce cardiovascular risk. Cues to action included experiences or actions within the participant's environment that prompted the decision to engage in cardiovascular risk reduction behaviors. Interview questions also included questions about adherence to cancer survivorship guidelines.

Template Analysis

Template analysis was conducted using an interview guide and codebook developed using the Health Belief Model [22,28-31]. Template analysis uses a structured template while still permitting a flexible coding process for the analysis of qualitative data. Since the interview guides in this study were based on the Health Belief Model framework, template analysis allowed researchers to use an initial template developed based on the constructs of the Health Belief Model framework, while also allowing codes to be updated in the coding process [30]. Coding was conducted by 3 independent coders with previous experience with qualitative analysis who each independently coded all transcripts. Following initial limited interview coding, further coding continued by consensus of themes which remained consistent across all coders. Nvivo (Lumivero) was used to assist in coding. Saturation of themes was reached after 19 interviews.

Ethical Considerations

The study was deemed by the Medical University of South Carolina institutional review board as exempt (Pro00073820) with a waiver for written informed consent. All participants provided verbal consent for participation. Participants self-reported current exercise participation, consumption of fruits and vegetables, and current cardiovascular health conditions. Interview transcripts were deidentified prior to coding and analysis. Interview participants received a US $35 gift card as compensation for their time for participation.

Statistical Analysis

Descriptive statistics including mean and SD were calculated for continuous variables and frequency data for qualitative variables.

Results

Participants

A total of 19 participants were enrolled in and completed interviews. The average participant age was 54 (SD 7; range 43 - 66) years. Participants averaged 3 primary care and 3 oncology visits annually. Over half (10/19, 52%) of participants reported receiving counseling on cardiovascular risk. Counseling was provided by primary care (8/10, 80%) and oncology (3/10, 30%). Demographic information is provided in Table 1. Themes and illustrative quotes identified are displayed in Table 2. Themes were consistent across participants based on age and race.

Table . Demographic information of participants (n=19).

Characteristics	Participants, n (%)
Race	
Black	6 (31)
White	13 (68)
Education	
High school or less	2 (10)
Some college	5 (26)
Bachelor's program	7 (37)
Master's program or graduate school	5 (26)
Employment	
Full time	13 (68)
Part time	2 (11)
No current employment	4 (21)
Marital status	
Married	11 (58)
Not currently married	8 (42)
Treatment history	
Surgery	19 (100)
Chemotherapy	12 (63)
Radiation	14 (74)
Endocrine therapy	14 (74)
General health (participant report)	
Poor	2 (11)
Fair or good	12 (63)
Very good or excellent	5 (26)
Adherence to survivorship guidelines	
>150 minutes exercise weekly	6 (32)
>5 servings fruits or vegetables daily	3 (16)
History of personal cardiovascular disease	17 (90)
Current cardiac medications	10 (50)
Family history of cardiovascular disease	17 (90)

Table . Health Belief Model themes.

	Quotation example	Theme
Perceived barriers	• "I'm almost 100 percent sure [my exercise limitations are] due to breast cancer treatment because I have some lung involvement. And so, you know, I've had issues with having a pleural effusion where it builds up the liquid in your lung." [INT7]	• General barriers identified included fatigue, limited time, and medications • Treatment-associated barriers included physical limitations and associated anxiety and stress
Overcoming barriers	• "But when I'm on the road, it's hard to find foods that fall in those categories ... When I know ahead of time that I'm going to have a pick up at a certain time. I'll make sure I either eat before I leave or I pack a lunch. But lots of times ... I only got like a 30 minute window to get them to pick it up." [INT17] • "So I think making the time about prioritizing ... prioritizing the time to exercise, and to shop appropriately, to eat healthy, so making the right decisions." [INT2] • "I think the main thing affecting me was I was kind of feeling because of what I went through, I was suffering a lot with PTSD and anxiety and all that. And I think because of that, I wasn't focusing too much on my health. I was stress eating and gained 20 pounds because I was like so stressed out from what happened and being on the hormonal meds and shot every month, it just really messed with my I guess, my whole, like, energy level and motivation." [INT4]	• Discussions with clinicians identified as factors promoting change • Scheduling considerations impact exercise and nutrition and mechanisms to prioritize change • Individual motivation important determinant for cardiovascular risk reduction
Perceived susceptibility	• "Genetics. On my father's side, everyone has died of heart disease, and I apparently got his genes." [INT3] • "I would say it's pretty high and I say that basically because when I had my radiation, they told me my tumor happened to be on my left side, so my radiation happened to be my heart. And so they told me that I was a high risk for congestive heart failure in the future." [INT12] • "I've never correlated the breast cancer to heart disease. I'm gonna need to think a little bit more about that. But I've never done that. So I would say no, because I just never thought about it in that way." [INT2]	• Individual factors impacting perceived risk include current personal cardiovascular disease, family history of cardiovascular disease, and lifestyle factors • Cancer and treatment-associated factors impact perceived risk of breast cancer in some individuals although there was uncertainty in others
Communication regarding cardiovascular risk and cancer	• "Possibly in the beginning ... My oncologist may have [discussed cardiovascular risk], but at that point, I think I was so overwhelmed with everything else that I probably didn't listen that well." [INT6] • "So, like, if I am at risk as being a breast cancer patient for heart issues, then that's something that I would want to stay on top of and have more information and have my primary care doctor or my oncologist talk to me about in ways that I could get on top of that before anything crazy happens." [INT7]	• Timing important factor in communication between health care team and patients about cardiovascular risk • Communication limited about cardiovascular risk and breast cancer.

	Quotation example	Theme
Self-efficacy	• "I stopped smoking was one [way I took control of my health]. Because the reality of possibly having cancer in another location or not healing from the procedure. I never realized that they actually kept you from healing. And diet improved. Everything had to improve." [INT5] • "I'm a survivor, thank God. And I think it has a lot to do with my healthy choices. But if you don't, if you're not educated on that stuff, most people don't know, you know. So it's important to talk to professionals and see what's the best thing for you to do." [INT8]	• Motivation to change and resources to support effective behavior change both reported as necessary by participants
Cues to action	• "You know, there's a publication that the [health system] puts out. I want to say four times a year ... There's a lot of just good little health tips, reminders that just kind of keep things at the forefront when you're starting to get busy." [INT1]	• Weight change and perception of body image motivation for lifestyle change • New diagnoses of cardiovascular disease in an individual or relative impacted motivation to seek care • Health promotion materials influence motivation for cardiovascular risk reduction

Perceived Barriers

Perceived barriers included both factors unrelated to cancer, as well as those associated with the late and long-term effects of breast cancer and treatment. Commonly reported barriers included fatigue, time limitations to participation in exercise or meal planning, motivation to change, and available resources to support behavioral changes to support cardiovascular risk reduction. Cancer-specific barriers included physical limitations, including those from previous cancer treatments that impacted the capacity for exercise. Anxiety and stress, especially associated with changes in health status associated secondary to cancer and cancer treatment, were reported and identified as a barrier to behavior changes in exercise and nutrition. Medication side effects following treatment were cited by one participant as a limitation to participation in exercise.

Overcoming Barriers

Engagement of the health care team and individual motivation for change were identified as factors promoting cardiovascular risk reduction behaviors. Participants reported that treatment of mental health conditions improved their capacity for behavior change. Participants highlighted the need for time management for meal preparation and exercise. The availability of group classes and discussions with clinicians in the health care team increased motivation for behavioral change.

Perceived Susceptibility

Participants identified breast cancer treatment–associated factors including radiation and chemotherapy as contributing to cardiovascular risk. Others expressed uncertainty about the potential impact of cancer treatment on their cardiovascular risk but desired more information on this from their health care team. Individual factors associated with risk perception included current diagnosis of cardiovascular disease or history of a cardiovascular event. A family history of cardiovascular disease was reported as increasing an individual's perceived personal risk of cardiovascular disease. Perceived protective factors included participant engagement in diet modification, regular exercise, and medication management of cardiovascular conditions.

Communication Regarding Cardiovascular Risk and Cancer

Participants reported limited communication regarding cardiovascular risk specifically related to their cancer history. However, they desired to receive information on cardiovascular risk. Timing of communication about cardiovascular risk further removed from the time of initial diagnosis and treatment was preferred.

Self-Efficacy

Participants identified self-efficacy as important in facilitating care and risk reduction. The ability to cook and resources for healthy food were identified as factors to support cardiovascular health. Motivation for change was identified as a necessary prerequisite for change.

Cues to Action

Participants reported inciting stimuli to engage in risk reduction behaviors as a new diagnosis of cardiovascular health conditions either in themselves or in family members or friends. Individuals who were experiencing symptoms reported a stronger motivation to seek medical care and engage in risk-reduction behaviors. Participants valued outreach from trusted organizations, including health care organizations, about behaviors to improve health and reduce cardiovascular risk as important in keeping their desire for change at the forefront of their priorities.

Sources of Information

The majority of individuals valued information from their health care team or other medical resources designed for patients. Specific sources of information on cardiovascular risk included clinicians, medical publications, and both medical and

nonmedical internet sites. The timing of communication impacted the reception of the discussion regarding risk. Some participants reported the initial stages of treatment as overwhelming in the amount of information delivered and reported being more treatment-focused at that time. Participants did desire to receive information about their cardiovascular risk both from their oncologists, as well as their primary care clinician. Desired mechanisms of delivery of information on cardiovascular risk in survivors of breast cancer included video (n=15, 79%), SMS text messaging (n=13, 68%), telephone (n=11, 58%), and mobile apps (n=10, 53%).

Summary

Based on the use of the Health Behavior Model, themes that promoted engagement in cardiovascular risk reduction were identified in each of the constructs assessed. Perceived barriers included both traditional barriers to cardiovascular risk reduction, as well as treatment-associated barriers derived from both mental and physical health. Factors promoting overcoming these barriers included clinician engagement with patients about cardiovascular risk reduction and the need for individual motivation for change to accomplish this. Although clinician communication was essential in promoting change, the timing of communication was identified as important. Immediately after diagnosis, breast cancer survivors preferred to focus discussion on treatment options. The availability of both financial resources and accessible resources supported the self-efficacy of participants with video and SMS text messaging as the preferred mediums for health information. Cues to action included changes in family and individual health. These interviews and themes have the capacity to inform further counseling and the development of interventions to support cardiovascular risk reduction both in breast cancer survivors and other cancer survivor populations.

Discussion

Principal Findings

Breast cancer survivors face additional cardiovascular risk following treatment, which impacts both overall and breast cancer–specific survival. Survivorship guidelines recommend screening and management strategies to reduce cardiovascular risk. Limited information is available on the barriers and facilitators of engaging breast cancer survivors in cardiovascular risk reduction behaviors. This qualitative study using semistructured interviews sought to evaluate facilitators and barriers to cardiovascular risk reduction in breast cancer survivors. Consistent with previous literature, uptake of survivorship guidelines including participating in physical activity and regular consumption of fruits and vegetables was limited [4]. The participants in the study overall received multiple treatment modalities that had the potential to increase individual cardiovascular risk.

The primary findings of the study included that there were opportunities to improve communication about cardiovascular risk in breast cancer survivors both within primary care and oncology. Breast cancer survivors valued communication about cardiovascular risk reduction, while also identifying the timing of delivering this information as important in engagement with subsequent risk reduction behaviors. Specifically, participants noted the importance of cardiovascular risk counseling as they progressed through treatment and into survivorship, rather than just at diagnosis. Many participants reported that they did not receive information on cardiovascular risk and were interested in learning more about their individual risks and the impact of their treatment on cardiovascular risk. These findings are in alignment with previous studies on both cancer survivors and health care professionals who identified that there was a knowledge gap among breast cancer survivors about cardiovascular risk. Cancer survivors consistently identified interest in discussions about heart disease and risk factors [32-34]. Although health care providers are generally aware of the increased cardiovascular risk in cancer survivors, system-based barriers include a lack of training, competing demands, and time limitations that limit the capacity for counseling and screening [34]. Furthermore, patient-level factors including socioeconomic factors and having a fatalistic outlook have been reported by health care clinicians as barriers to adopting risk reduction strategies [32,33]. In this study, the need for resources to promote self-efficacy was similarly identified by multiple participants even in a population with access to primary care [35].

Participants identified previously known barriers and facilitators to cardiovascular risk reduction, including motivation, time limitations, and resources to support efforts for change. Participants simultaneously identified breast cancer–specific factors as both significant components of their risk perception and as limitations to participation in risk-reducing behaviors. These same barriers were identified in previous formative research for intervention development, with the identification of cancer and other health conditions limiting the ability to increase physical activity. These studies simultaneously identified the need for individual motivation to facilitate change [34]. Participants in this study identified the need for multimodal changes, which aligns with previous literature supporting the greater efficacy of interventions that address factors beyond physical activity alone [30,31,34].

Breast cancer survivors desired additional communication about cancer-specific cardiovascular risk factors, preferably delivered after initial diagnosis and treatment have been established. Cardiovascular risk reduction strategies, including physical activity, have been demonstrated to reduce cardiovascular risk and improve quality of life, and therefore increased attention to cardiovascular risk reduction is essential to promoting improved survivorship care of breast cancer survivors [36,37]. Therefore, additional interventions that promote cardiovascular risk reduction and consider the elevated risks and unique barriers faced by breast cancer survivors are essential.

This study supports and aligns with existing literature regarding the need to support self-efficacy and consider the physical limitations of breast cancer survivors following treatment. Previous research identifies the critical role of clinicians in identifying and communicating the cumulative role of risk factors in cancer survivorship on cardiovascular risk and the importance of patient-clinician communication [38]. Engagement in communication about cardiovascular risk is limited and especially important in cancer survivors at increased

risk, including those with existing cardiovascular disease and obesity. Primary care clinicians can serve an important role in screening for cardiovascular risk, promoting risk reduction, and pharmacologic management when indicated [39]. Based on this study, primary care clinicians often initiate this discussion and have an opportunity to improve communication about risk and promotion of health behaviors. When primary care clinicians were surveyed about barriers and enablers of care, system-level barriers including lack of time and training were critical to address to promote improved cardiovascular survivorship care [32]. Clinical risk assessment and management tools have been identified as strategies to improve the delivery of cardiovascular disease care to reduce morbidity and mortality. This study identifies the need to consider mental health support and treatment in facilitating cardiovascular risk reduction. In addition, it identifies the importance of timing in the delivery of cardiovascular risk-reducing education during the breast cancer treatment-survivorship continuum in order to optimize the uptake of risk-reducing measures.

Strengths and Limitations

The proposed study is novel in its evaluation of breast cancer survivors–specific perceived risks and susceptibility to breast cancer, as well as potential survivorship-associated barriers to engagement in risk reduction. However, the study has a number of limitations. First, the study focuses on participants from a single academic medical center, which may limit its transferability. Further studies should include participants from a diversity of geographical locations and practice settings. A stratified recruitment approach was used to broaden the perspectives obtained as part of the interview process. Second, participants were selected based on having an existing primary care clinician, as the study included the role of primary care clinicians in survivorship. Future work would benefit from the inclusion of individuals with reduced access to care. Third, study findings may have been impacted by participation bias, where either the desire to engage in cardiovascular risk reduction behaviors or existing cardiovascular disease may have impacted the decision to participate. There are elevated rates of cardiovascular disease in breast cancer survivors, and this perspective is proportionally represented in this study. Another limitation is that the interview guide was not pretested before implementation. The interview guide was based strongly on the constructs of the Health Belief Model, which have been extensively studied. Future investigations could ensure the validity of the interview guide as related to capturing Health Belief Model constructs as related to cardiovascular risk in breast cancer survivors. Finally, the study is limited by the small sample size. This study focused on a limited number of individuals in order to obtain more breadth of information. While the achievement of theme saturation strengthens the study findings, future quantitative studies are needed to further evaluate the identified themes using a larger sample size to promote the development of interventions that improve communication between health care team members and breast cancer survivors about cardiovascular risk and engagement with risk reduction behaviors.

Conclusions

Breast cancer survivors face unique cardiovascular risks influenced by their diagnosis and treatment, which introduce barriers to effective risk reduction. Participants in this study expressed a desire for enhanced counseling on cardiovascular health, highlighting a significant knowledge gap regarding their risks and the impact of cancer treatments. Many reported insufficient communication from health care providers about cardiovascular risks, particularly after the initial treatment phases. To address these issues, health care systems should leverage electronic communication methods, such as video and SMS text messaging, to deliver tailored cardiovascular risk information. Interventions must consider the specific challenges of survivorship, including physical limitations and mental health concerns, to effectively engage survivors in risk-reducing behaviors. This study highlights the importance of improving adherence to guideline-based care for breast cancer survivors and emphasizes the need for further research to develop comprehensive strategies that facilitate cardiovascular health in this population. By prioritizing timely and accessible information, we can enhance self-efficacy and support better health outcomes for breast cancer survivors.

Acknowledgments

The authors would like to acknowledge Suzanne Perkins for her contributions to recruitment, data collection, and interviews. The authors would additionally like to thank Dr Emily Ramsayer for her contributions to data analysis and Nekia Robinson for her role in manuscript revision. This study was funded by pilot funding through the Cancer Control Group at Hollings Cancer Center at the Medical University of South Carolina. Generative artificial intelligence was not used in manuscript writing.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

STM contributed to conceptualization, formal analysis, funding acquisition, methodology, and writing including the original draft and review and editing. VD contributed to conceptualization, methodology, supervision, and writing including the original draft and review and editing. NS contributed to formal analysis, methodology, and writing in all stages of development.

Conflicts of Interest

None declared.

References

1. Montazeri K, Unitt C, Foody JM, Harris JR, Partridge AH, Moslehi J. ABCDE steps to prevent heart disease in breast cancer survivors. Circulation 2014 Oct 28;130(18):25462826. [doi: 10.1161/CIRCULATIONAHA.114.008820]
2. Sharma AV, Reddin G, Forrestal B, Barac A. Cardiovascular disease risk in survivors of breast cancer. Curr Treat Options Cardiovasc Med 2019 Dec 9;21(12):31820123. [doi: 10.1007/s11936-019-0788-2] [Medline: 31820123]
3. Snyder CF, Frick KD, Herbert RJ, et al. Comorbid condition care quality in cancer survivors: role of primary care and specialty providers and care coordination. J Cancer Surviv 2015 Dec;9(4):641-649. [doi: 10.1007/s11764-015-0440-4] [Medline: 25716644]
4. Mehta LS, Watson KE, Barac A, et al. Cardiovascular disease and breast cancer: where these entities intersect: a scientific statement from the American Heart Association. Circulation 2018 Feb 20;137(8):e30-e66. [doi: 10.1161/CIR.0000000000000556] [Medline: 29437116]
5. Jones DH, Nestore M, Henophy S, Cousin J, Comtois AS. Increased cardiovascular risk factors in breast cancer survivors identified by routine measurements of body composition, resting heart rate and arterial blood pressure. Springerplus 2014;3(150):150. [doi: 10.1186/2193-1801-3-150] [Medline: 24790808]
6. Bradshaw PT, Stevens J, Khankari N, Teitelbaum SL, Neugut AI, Gammon MD. Cardiovascular disease mortality among breast cancer survivors. Epidemiology 2016 Jan;27(1):6-13. [doi: 10.1097/EDE.0000000000000394] [Medline: 26414938]
7. Coughlin SS, Ayyala D, Majeed B, Cortes L, Kapuku G. Cardiovascular disease among breast cancer survivors. Cardiovasc Disord Med 2020;2(1):32661518. [doi: 10.31487/j.cdm.2020.01.01] [Medline: 32661518]
8. Florido R, Daya NR, Ndumele CE, et al. Cardiovascular disease risk among cancer survivors: the atherosclerosis risk in communities (ARIC) study. J Am Coll Cardiol 2022 Jul 5;80(1):22-32. [doi: 10.1016/j.jacc.2022.04.042] [Medline: 35772913]
9. Maas A, Ottevanger N, Atsma F, Cramer MJ, Leiner T, Poortmans P. Cardiovascular surveillance in breast cancer treatment: a more individualized approach is needed. Maturitas 2016 Jul;89(58-62):58-62. [doi: 10.1016/j.maturitas.2016.04.015] [Medline: 27180161]
10. Nathan PC, Amir E, Abdel-Qadir H. Cardiac outcomes in survivors of pediatric and adult cancers. Can J Cardiol 2016 Jul;32(7):871-880. [doi: 10.1016/j.cjca.2016.02.065] [Medline: 27179545]
11. Chen CL. Cardiovascular prevention in the cancer survivor. Curr Atheroscler Rep 2015 Mar;17(3):25637040. [doi: 10.1007/s11883-014-0484-3] [Medline: 25637040]
12. Kenyon M, Mayer DK, Owens AK. Late and long-term effects of breast cancer treatment and surveillance management for the general practitioner. J Obstet Gynecol Neonatal Nurs 2014;43(3):382-398. [doi: 10.1111/1552-6909.12300] [Medline: 24690072]
13. Lowenstein LM, Ouellet JA, Dale W, Fan L, Gupta Mohile S. Preventive care in older cancer survivors. J Geriatr Oncol 2015 Mar;6(2):85-92. [doi: 10.1016/j.jgo.2014.12.003] [Medline: 25547206]
14. Knobf MT. Clinical update: psychosocial responses in breast cancer survivors. Semin Oncol Nurs 2011 Aug;27(3):e1-e14. [doi: 10.1016/j.soncn.2011.05.001] [Medline: 21783006]
15. Singla A, Kumar G, Bardia A. Personalizing cardiovascular disease prevention among breast cancer survivors. Curr Opin Cardiol 2012 Sep;27(5):515-524. [doi: 10.1097/HCO.0b013e3283570040] [Medline: 22874128]
16. Feng X, Deng Z, McCullough MS, et al. The impact of cardiovascular risk factors on cancer progression: a prospective study in female breast cancer survivors. Breast Cancer Res Treat 2025 Apr;210(3):737-748. [doi: 10.1007/s10549-025-07611-y] [Medline: 39928264]
17. Knobf MT, Coviello J. Lifestyle interventions for cardiovascular risk reduction in women with breast cancer. Curr Cardiol Rev 2011 Nov;7(4):250-257. [doi: 10.2174/157340311799960627] [Medline: 22758626]
18. Jung W, Cho IY, Jung J, et al. Changes in physical activity and cardiovascular disease risk in cancer survivors: a nationwide cohort study. JACC CardioOncol 2024 Dec;6(6):879-889. [doi: 10.1016/j.jaccao.2024.09.013] [Medline: 39801643]
19. Lee K, Tripathy D, Demark-Wahnefried W, et al. Effect of aerobic and resistance exercise intervention on cardiovascular disease risk in women with early-stage breast cancer: a randomized clinical trial. JAMA Oncol 2019 May 1;5(5):710-714. [doi: 10.1001/jamaoncol.2019.0038] [Medline: 30920602]
20. Toohey K, Pumpa K, McKune A, et al. The impact of high-intensity interval training exercise on breast cancer survivors: a pilot study to explore fitness, cardiac regulation and biomarkers of the stress systems. BMC Cancer 2020 Aug 20;20(1):787. [doi: 10.1186/s12885-020-07295-1] [Medline: 32819304]
21. Khanji MY, van Waardhuizen CN, Bicalho VVS, Ferket BS, Hunink MGM, Petersen SE. Lifestyle advice and interventions for cardiovascular risk reduction: a systematic review of guidelines. Int J Cardiol 2018 Jul 15;263(142-51):142-151. [doi: 10.1016/j.ijcard.2018.02.094] [Medline: 29754910]
22. Champion VL, Skinner CS. Health Behavior and Health Education: Theory, Research, and Practice: Jossey-Bass/Wiley; 2008.

23. Korin MR, Chaplin WF, Shaffer JA, Butler MJ, Ojie MJ, Davidson KW. Men's and women's health beliefs differentially predict coronary heart disease incidence in a population-based sample. Health Educ Behav 2013 Apr;40(2):231-239. [doi: 10.1177/1090198112449461] [Medline: 22991050]
24. Ghorbani-Dehbalaei M, Loripoor M, Nasirzadeh M. The role of health beliefs and health literacy in women's health promoting behaviours based on the health belief model: a descriptive study. BMC Womens Health 2021 Dec 18;21(1):421. [doi: 10.1186/s12905-021-01564-2] [Medline: 34922505]
25. Komaç F, Duru P. The effect of education based on a health belief model and motivational interviews on cardiovascular disease risk factors and healthy lifestyle behaviour changes in patients with essential hypertension: a randomized controlled trial. Patient Educ Couns 2024 Mar;120(108126):108126. [doi: 10.1016/j.pec.2023.108126] [Medline: 38154390]
26. Ali NS. Prediction of coronary heart disease preventive behaviors in women: a test of the health belief model. Women Health 2002;35(1):83-96. [doi: 10.1300/J013v35n01_06] [Medline: 11942471]
27. Jones DE, Weaver MT, Grimley D, Appel SJ, Ard J. Health belief model perceptions, knowledge of heart disease, and its risk factors in educated African-American women: an exploration of the relationships of socioeconomic status and age. J Natl Black Nurses Assoc 2006 Dec;17(2):13-23. [Medline: 17410755]
28. Rosenstock IM. The health belief model: explaining health behavior through expectancies. In: Health Behavior and Health Education: Theory, Research, and Practice 1990:39-62.
29. Brooks J, McCluskey S, Turley E, King N. The utility of template analysis in qualitative psychology research. Qual Res Psychol 2015 Apr 3;12(2):202-222. [doi: 10.1080/14780887.2014.955224] [Medline: 27499705]
30. Crabtree BF, Miller WF. A template approach to text analysis: developing and using codebooks. In: Doing Qualitative Research: Sage Publications; 1992:93-109.
31. Crabtree BF, Miller WL. Using codes and code manuals: a template organizing style of interpretation. In: Doing Qualitative Research: Sage Publications; 1999:163-177.
32. Knowles R, Kemp E, Miller M, Koczwara B. Reducing the impact of cardiovascular disease in older people with cancer: a qualitative study of healthcare providers. J Cancer Surviv 2024 Jun;18(3):818-827. [doi: 10.1007/s11764-023-01331-2] [Medline: 36648743]
33. Knowles R, Kemp E, Miller M, Koczwara B. "There could be something going wrong and I wouldn't even know": a qualitative study of perceptions of people with cancer about cardiovascular disease (CVD) risk and its management. J Cancer Surviv 2025 Feb;19(1):319-325. [doi: 10.1007/s11764-023-01468-0] [Medline: 37775614]
34. Chang A, Boyd A, Leung I, et al. Formative research to adapt a cardiac rehabilitation program to breast cancer survivors: the HEART health after cancer treatment (HEART-ACT) study. Cardiooncology 2024 May 17;10(1):28. [doi: 10.1186/s40959-024-00228-y] [Medline: 38760873]
35. Koop Y, Dobbe L, Maas AHEM, et al. Oncology professionals' perspectives towards cardiac surveillance in breast cancer patients with high cardiotoxicity risk: a qualitative study. PLoS One 2021;16(3):e0249067. [doi: 10.1371/journal.pone.0249067] [Medline: 33788897]
36. Naaktgeboren WR, Groen WG, Jacobse JN, et al. Physical activity and cardiac function in long-term breast cancer survivors: a cross-sectional study. JACC CardioOncol 2022 Jun;4(2):183-191. [doi: 10.1016/j.jaccao.2022.02.007] [Medline: 35818555]
37. Montagnese C, Porciello G, Vitale S, et al. Quality of life in women diagnosed with breast cancer after a 12-month treatment of lifestyle modifications. Nutrients 2020 Dec 31;13(1):136. [doi: 10.3390/nu13010136] [Medline: 33396551]
38. Christian AH, O'Malley D, Barac A, Miller SM, Hudson SV. Cardiovascular risk and communication among early stage breast cancer survivors. Patient Educ Couns 2017 Jul;100(7):1360-1366. [doi: 10.1016/j.pec.2017.02.010] [Medline: 28215826]
39. Kidane RD, Ruddy KJ, Lin G, Sandhu NP. Cardiovascular health considerations for primary care physicians treating breast cancer survivors. Mayo Clin Proc 2025 Jan;100(1):124-140. [doi: 10.1016/j.mayocp.2024.08.014] [Medline: 39641716]

Edited by N Cahill; submitted 08.01.25; peer-reviewed by SJ Werts-Pelter, Y Zhu; revised version received 29.05.25; accepted 30.05.25; published 03.07.25.

Please cite as:
Marrison ST, Shungu N, Diaz V
Perception and Counseling for Cardiac Health in Breast Cancer Survivors Using the Health Belief Model: Qualitative Analysis
JMIR Cancer 2025;11:e71062
URL: https://cancer.jmir.org/2025/1/e71062
doi:10.2196/71062

Usability and Implementation Considerations of Fitbit and App Intervention for Diverse Cancer Survivors: Mixed Methods Study

Zakery Dabbagh[1], BS; Reem Najjar[1], BS; Ariana Kamberi[1], MBA; Ben S Gerber[1], MD, MPH; Aditi Singh[2], ScM; Apurv Soni[2], MD, PhD; Sarah L Cutrona[1,3], MD, MPH; David D McManus[2], MD; Jamie M Faro[1], PhD

[1]Department of Population and Quantitative Health Sciences, University of Massachusetts Chan Medical School, 55 N Lake Ave, Worcester, MA, United States

[2]Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[3]Center for Healthcare Organization and Implementation Research, Bedford VA Healthcare System, Bedford, MA, United States

Corresponding Author:
Jamie M Faro, PhD
Department of Population and Quantitative Health Sciences, University of Massachusetts Chan Medical School, 55 N Lake Ave, Worcester, MA, United States

Abstract

Background: Despite the known benefits of physical activity, cancer survivors remain insufficiently active. Prior trials have adopted digital health methods, although several have been pedometer-based and enrolled mainly female, non-Hispanic White, and more highly educated survivors of breast cancer.

Objective: The objective of this study was to test a previously developed mobile health system consisting of a Fitbit activity tracker and the MyDataHelps smartphone app for feasibility in a diverse group of cancer survivors, with the goal of refining the program and setting the stage for a larger future trial.

Methods: Participants were identified from one academic medical center's electronic health records, referred by a clinician, or self-referred to participate in the study. Participants were screened for eligibility, enrolled, provided a Fitbit activity tracker, and instructed to download the Fitbit: Health & Wellness and MyDataHelps apps. They completed usability surveys at 1 and 3 months. Interviews were conducted at the end of the 3-month intervention with participants and cancer care clinicians to assess the acceptability of the intervention and the implementation of the intervention into clinical practice, respectively. Descriptive statistics were calculated for demographics, usability surveys, and Fitbit adherence and step counts. Rapid qualitative analysis was used to identify key findings from interview transcriptions.

Results: Of the 100 patients screened for eligibility, 31 were enrolled in the trial (mean age 64.8, SD 11.1 years; female patients=17/31, 55%; Hispanic or Latino=7/31, 23%; non-White=11/31, 35%; less than a bachelor's degree=14/31, 45%; and household income <US $75,000=11/31, 35%). The mean (SD) years since diagnosis was 7.1 (8.2), and the two most frequent cancer diagnoses were prostate (9/31, 29%) and breast (4/31, 13%) cancer. Participants provided positive feedback on the MyDataHelps app usability; the overall app quality received a mean score of 3.79 (SD 0.82) on a 5-point Likert scale (1=worst, 5=best). Interviews with 10 patients yielded four themes: (1) Fitbit and app setup was easy but the research team provided assistance, when needed, which was helpful, (2) motivational messages within the app were not memorable, (3) step counts and Fitbit notifications were motivating, and (4) medical professionals viewing their data were acceptable. Interviews with 5 cancer care clinicians yielded four themes: (1) some patients used wearables but rarely discussed data with clinicians; (2) activity trackers can be helpful to motivate patients and keep them accountable; (3) objective activity measures—similar to BMI, weight, and blood pressure— that they can track over time and refer to afterward were preferred; and (4) training and systematic processes to view these data as part of active workflow were desired.

Conclusions: Implementing a remotely delivered, light-intensity physical activity program was feasible and acceptable in a sample of diverse cancer survivors. Future studies should consider registry-based methods and work with clinicians to engage hard-to-reach survivor populations who have low physical activity levels and disproportionately high adverse health outcomes.

Trial Registration: ClinicalTrials.gov NCT05417438; https://clinicaltrials.gov/study/NCT05417438

(JMIR Cancer 2025;11:e60034) doi:10.2196/60034

KEYWORDS

physical activity; cancer survivor; wearable device; smartphone app; diverse; Fitbit; wearable; feasibility; usability; digital health; digital health method; breast cancer; Hispanic; women; mobile health; activity tracker; mHealth

Introduction

The health benefits of physical activity for cancer survivors are widely known. Yet, few survivors are active during (<10%) and after cancer treatment (20% - 30%) [1,2]. Per the American Heart Association recommendation, cancer care clinicians should provide phsyical activity referrals to prevent cardiovascular disease in survivors [3]; however, clinicians do not always suggest these referrals [4]. Notably, acknowledging an individual's unique perspective and offering choices rather than referrals to one singular program have been shown to increase effectiveness [5,6].

There has been a recent increase in the number of physical activity programs that exist for survivors. Virtual group programs have appeared as an option that is preferred by some survivors, allowing them the social support they seek in a group format with the convenience of participating in a program from one's own home [7,8]. In a previous study, clinic-based referrals to group in-person and group virtual programs were acceptable to cancer survivors, but some also expressed a desire for a nongroup digital program option [8]. In cancer survivors, digital health programs can be effective in promoting physical activity and reducing participation barriers [9]. However, engaging diverse groups of survivors in digital programs continues to be a challenge [10], along with sustainably integrating digital programs into clinical survivorship care [11]. Ninety percent of US adults own a smartphone and 40% - 60% a wearable device [12,13]. As digital health programs are becoming increasingly more accessible, more strategies are needed to increase the reach of these programs, particularly as the number of cancer survivors grows annually [14].

This pilot study differs from past research in several ways. First, most prior digital health trials used pedometer-based interventions as opposed to Fitbit-based interventions [15], and many of these studies incorporated counseling and other forms of support in their interventions [15]. Additionally, the majority of trials using wearable physical activity trackers with cancer survivors have been done in homogenous populations, such as White, college-educated [10], and survivors of breast cancer [10,15]. Past evidence on the optimal frequency and timing of SMS text messages in this context is lacking and what evidence there is has been inconclusive [10]. Finally, very few studies have focused on older adults, who comprise the majority of cancer survivors in the United States [10].

Thus, the purpose of this study was to test a previously developed mobile health system in a diverse group of cancer survivors with the goal of refining the program and setting the stage for a larger future trial [16,17]. We tested the feasibility of using a wrist-based wearable activity tracker (Fitbit) and smartphone app dyad with survivors, including seeking feedback on survey data collection and message prompts within the app. We also sought feedback from our clinical partners to better understand the implementation of referrals to a digital health program and perspectives on accessing patient-generated health data from these programs.

Methods

Ethical Considerations

This study took place at the University of Massachusetts Chan Medical School with remote recruitment, allowing participants to partake in study activities outside of the institution. Procedures were approved by the University of Massachusetts Chan Medical School Institutional Review Board (#H00023545) and informed consent was obtained from each participant. All participant data were deidentified and study data were kept anonymous. Participants received US $25 compensation for participation in the pilot and US $25 for participation in a follow-up interview. The study was registered at ClinicalTrials.gov (NCT05417438).

Study Design

This study used a longitudinal, nonrandomized, multilevel mixed methods approach, involving both quantitative and qualitative data collection (explanatory sequential design) from enrolled patients and cancer care clinicians between January 2023 and July 2023. Cancer survivors were enrolled in a 3-month Fitbit or smartphone app intervention. The convenience sample size was not calculated, as this was a feasibility pilot study. Outcome data for the intervention included Fitbit adherence and step counts over 3 months, usability ratings of the MyDataHelps app at 1 month, and follow-up interviews. No data were collected regarding the usability of the Fitbit: Health & Wellness app, as the app was primarily used to continuously synchronize Fitbit data with the MyDataHelps app. Clinician semistructured interviews were conducted following the data collection period for the intervention (June-July 2023).

Recruitment

Eligibility criteria for the pilot study included having a past cancer diagnosis, owning a smartphone with internet access, and being deemed appropriate to participate by a medical professional as necessary. Recruitment was first conducted using paper flyers posted throughout clinics in the UMass Memorial Health network. The flyer provided interested participants with the study team's email address and telephone number to call and SMS text message to discuss the study further. Potential participants were additionally identified by extracting data directly from the electronic health record (EHR) data hosted in the UMass Chan Data Lake. Next, potential participants were sent a letter containing information about the study and instructions to call to opt out of further contact. Two weeks after being mailed, potential participants who did not opt out were called and asked if they were interested in joining the study. Potential participants who expressed interest were screened over the phone. Approval to participate in the study was sought from clinicians for potential participants based on responses in the Physical Activity Readiness Questionnaire (PAR-Q), a 7-step questionnaire screening for evidence of risk factors during moderate physical activity and reviewing for family history and disease severity [18]. Potential participants were then sent a secure email through REDCap containing a link directing them to an electronic consent form. After going through the electronic consent form with the study team over

the phone, eligible potential participants who were interested consented using their electronic signature.

Statistical Analysis

Healthy History, Demographics, and Physical Activity Readiness

Participants self-reported their age, sex, education level, race, ethnicity, date of last cancer diagnosis, and type of cancer via a telephone screening process with the study team. Physical activity readiness was assessed using the PAR-Q. Any "yes" response prompted approval to be sought from the participant's primary care provider or cancer care clinician. Once enrolled, patients completed additional health history baseline questionnaires through surveys delivered in the MyDataHelps app.

Fitbit Adherence and Step Count

Physical activity data were adjusted for a number of valid wear days. Consistent with prior studies [19-21], valid wear days were defined as those with a daily step count of 1500 or greater as measured by the Fitbit activity tracker.

App Usability

The Mobile App Rating Scale (MARS) was used to assess the acceptability of the MyDataHelps smartphone app at 1 month through a survey delivered in the MyDataHelps app. Specific assessments of app functionality and aesthetics included ease of use, navigation, visual appeal, performance, graphics, and layout, as well as overall app quality [22]. Items were rated by participants using a 5-point Likert scale from 1=inadequate to 5=excellent. Sample questions included "how easy is it to learn how to use the app?" and "how clear are the menu labels or icons and instructions?" The MARS was scored using a mean for each category. The MARS has demonstrated internal consistency (α=.9) and interrater reliability (intraclass coefficient=0.79) [22,23].

Participant and Clinician Acceptability

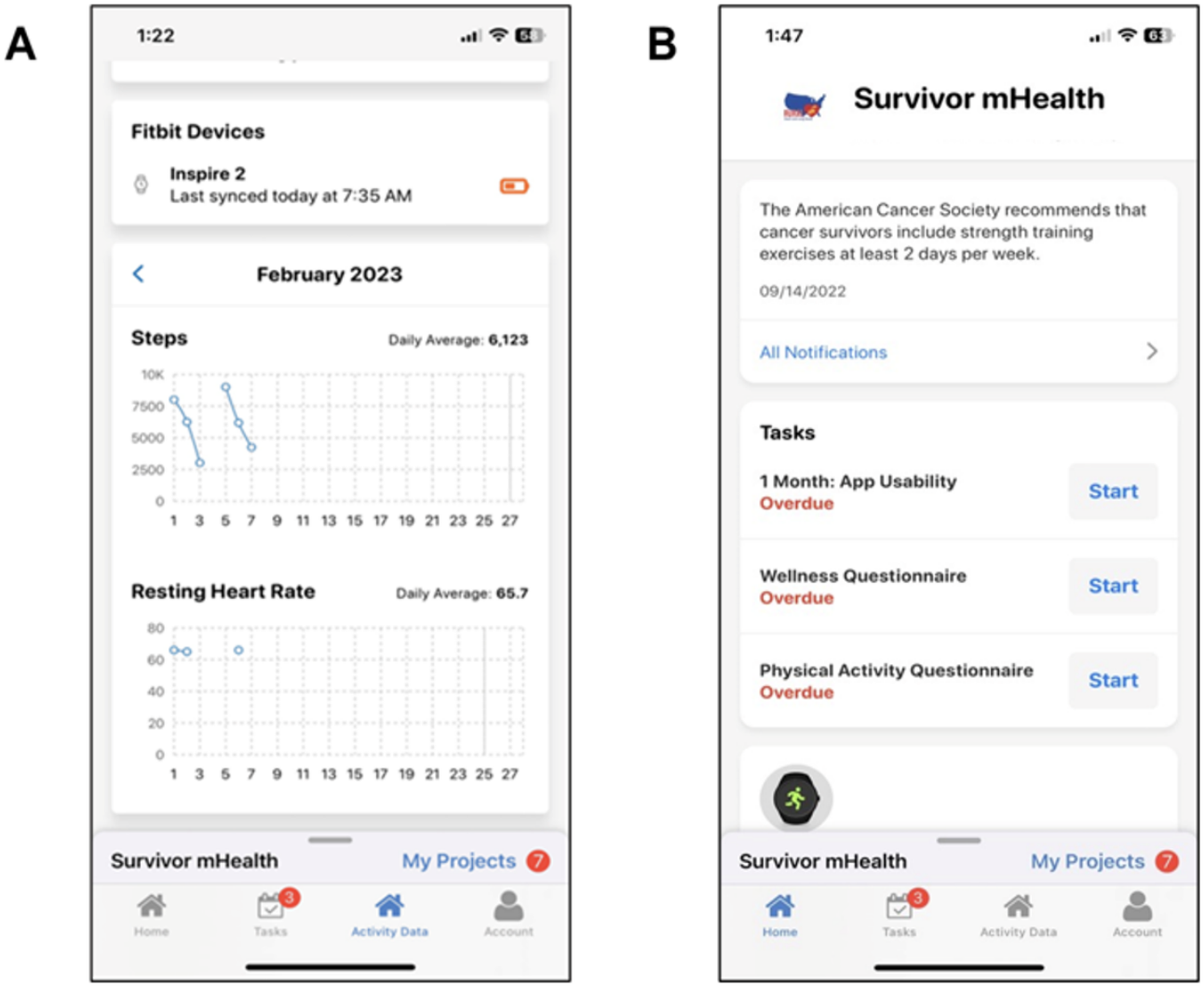
Enrolled participants completed semistructured interviews (Multimedia Appendix 1) after the 12-week intervention to assess their experiences during the study, including areas of improvement. Health care clinicians' perceptions on integrating this program into clinical practice were also assessed through semistructured interviews.

Protocol for Intervention

The intervention protocol, including the MyDataHelps app, was previously user-tested in a healthy cohort [24]. All participants received a Fitbit Charge device (versions 2 and 5) to keep as part of their compensation for participating. Following consent, participants were mailed their Fitbit along with paper instructions for device and app setup (including a unique study identifier) and were provided technical support over the phone by the study team. Study participants downloaded 2 apps onto their phones: Fitbit: Health & Fitness and MyDataHelps. Upon enrollment, participants were instructed that the standard goal steps/day set by Fitbit is 10,000 steps. They were told they could modify their step-count goal based on what their current activity level was and what they felt was achievable. Participants were instructed to wear their Fitbit for a total of 3 months, during which time they received standard Fitbit: Health & Fitness push notifications to their Fitbit devices, such as alerting them of achievement of goal steps per day, time spent in active heart rate zone minutes (if participants enter a high-intensity workout zone), and movement reminders (to get up and move if inactive for a period). These messages are preset by Fitbit to be delivered to participants based on their individual data.

Participants' smartphones also received push notifications from MyDataHelps to promote adherence to the study surveys, along with weekly push notifications of motivational messages (Figure 1). These messages were templated messages to inform participants about the benefits of exercise and strategies to incorporate exercise into their day, derived from the American Cancer Society (eg, "Research has shown that exercise is not only safe and possible during cancer treatment, but it can improve how well you function physically and your quality of life" and "End your exercise session with stretching or flexibility exercises. Hold a stretch for about 15 to 30 seconds and relax. Examples of stretching are reaching overhead, deep breathing, and bending over to touch your toes so that you relax all the muscle groups") [25].

Figure 1. MyDataHelps app displaying Fitbit data (A) and message notifications and survey tasks (B).



Participants who experienced technical challenges or difficulties adhering to study protocols (eg, syncing Fitbit, completing surveys) were contacted by phone to troubleshoot issues. Through the MyDataHelps app, participants completed surveys 24 hours, 1 month, and 3 months after enrollment. Participants also received push notifications alerting them that the surveys were available within the app (Figure 1). Participants' Fitbit devices were synced to the MyDataHelps app. Once participants registered for MyDataHelps and synced their Fitbit: Health & Wellness account, data collection from the Fitbit activity tracker began. The study team viewed activity and survey completion data weekly and conducted outreach to participants with missing data. We continued to reach out to participants, including giving them an opportunity to complete their 3-month survey. Those who did not complete their 3-month survey were considered lost to follow-up. After participants completed all the tasks required for the study, they were allowed to keep their Fitbit and were provided a US $25 Amazon gift card.

Protocol for Qualitative Feedback From Participants and Cancer Care Clinicians

Upon completion of the study tasks, participants were offered the option to take part in a 30-minute semistructured Zoom interview. Participants who took part were compensated with a US $25 Amazon gift card (in addition to the gift card provided for the completion of study tasks). Interview guides were developed by the study team using an iterative process of pretesting. The interview guide consisted of questions asking their perceptions of the program overall, as well as asking them ways to improve the current list of messages they received within the app (Multimedia Appendix 2). These guides were revisited after interviews to determine if any additional questions needed to be added, removed, or probed further. Cancer care clinicians, specifically advanced practice providers, were recruited via pre-existing relationships with the study team and through the survivorship coordinator. The survivorship coordinator emailed cancer care clinicians involved in survivorship care information about the study and about participating in interviews. Participating clinicians were asked to complete a 30-minute Zoom interview using a structured interview guide. The guide was designed to discuss implementing physical activity referrals to digital health programs into clinical workflow, along with preferences for reviewing those data. Again, the interview guide was revisited after each interview for revisions. The survivorship coordinator also provided feedback on the interview guides throughout the process. Zoom interviews were conducted by a trained member of the research team in conjunction with the study principal

investigator and recorded and transcribed electronically using an institutional review board–approved software.

Data Analysis

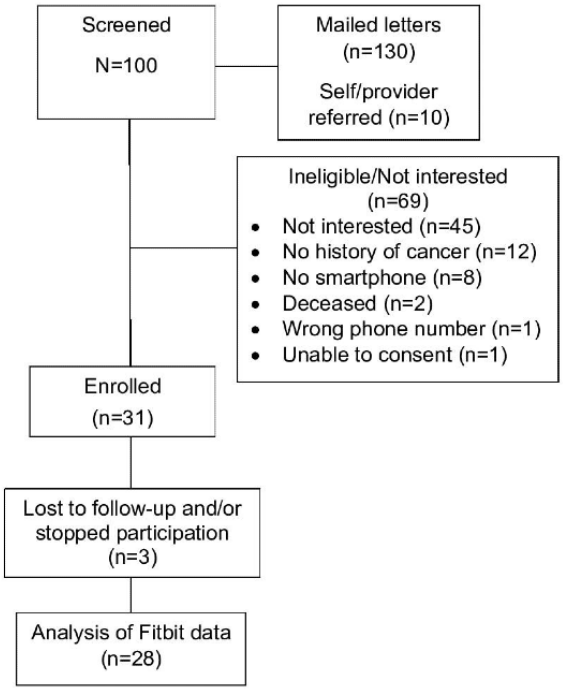
Quantitative analyses for demographics, usability surveys, and Fitbit data were completed using R 4.3.2 (The R Foundation) and Microsoft Excel. Analysis of Fitbit step count data (mean and SD of steps per day each week) was done on valid days, in which valid days were defined as those with a step count of 1500 or greater [20-22]. As such, the first valid day of Fitbit use was the first day the participant walked 1500 or more steps while wearing Fitbit, and data collection from Fitbit concluded 90 days following the first valid day. Any days with fewer than 1500 steps within the 90-day time frame of study participation were considered missing. Qualitative data were analyzed using rapid qualitative analysis [26]. This was done by first creating a matrix in Excel after the participants' interviews were completed. The matrix had a row for each participant and columns for each domain that corresponded with the interview guide. Domains included "experiences with patients and wearables devices," "when to integrate program into clinical practice," "preferences to clinical team viewing data," and "support needed for clinicians (eg, EHR staff) or patients." The study team (JMF, RN, AK, and ZD) then met to code the interview transcripts for 2 participants in each group, and then individually coded the remainder on their own. The study team met regularly to conduct data and coding checks and resolve any discrepancies in coding or domain assignment. The team noted the common themes between the participants' responses to each domain and identified key findings of the qualitative interviews. Each key finding discusses the main common themes within the qualitative interviews that align with the domains.

Results

Overall Results

Figure 2 presents the flow of participants from contact to the conclusion of the 12-week intervention. The study team mailed letters to 130 potential participants and had 10 potential participants either self-refer or were referred by their clinician. Of the 100 potential participants who were screened, 31 participants enrolled in the trial, achieving a response rate of 31%.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



Among the 31 participants enrolled in the intervention, the mean age was 64.8 (SD 11.1) years, 17 (55%) identified as female, 14 (45%) had less than a bachelor's degree level of education, 7 (23%) identified as Hispanic or Latino, and 11 (35%) identified as non-White (Table 1). The top two cancer diagnoses were prostate (9/31, 29%) and breast (4/31, 13%), and the mean years since diagnosis was 7.1 (SD 8.2).

Table . Enrolled patient baseline demographics (N=31).

Variables		Values
Age (years), mean (SD)		64.7 (11.1)
Sex, n (%)		
	Male	14 (45)
	Female	17 (55)
Education level, n (%)		
	High school or less	7 (23)
	Some college	7 (23)
	Bachelor's degree	6 (19)
	Advanced college degree	11(35)
Race, n (%)		
	White	20 (65)
	Black or African American	10 (32)
	Other	1 (3)
Ethnicity, n (%)		
	Hispanic or Latino	7 (23)
	Not Hispanic or Latino	24 (77)
Household income in US$, n (%) [a]		
	>75,000	18 (62)
	50,001-75,000	4 (14)
	25,000-50,000	5 (17)
	<25,000	2 (7)
Cancer type, n (%)[b]		
	Gynecologic	2 (6)
	Thyroid	2 (6)
	Skin	2 (6)
	Colon	1 (3)
	Prostate	9 (29)
	Breast	4 (13)
	Throat	2 (6)
	Kidney	1 (3)
	Blood	2 (6)
	Don't know	2 (6)
Years since diagnosis, mean (SD)		7.1 (8.2)

[a]Due to missing responses, n=29.

[b]Due to missing responses, n=27.

Fitbit Adherence and Activity

Physical activity adherence during the 12-week intervention was assessed using Fitbit data (Table 2). Based on this measure, the mean daily step count was 7219 (SD 4418) at baseline and 6687 (SD 3183) at 12 weeks.

Table . Mean (SD) number of valid days of adherence to wearing the Fitbit activity tracker and mean (SD) steps/day during the 12-week intervention period among cancer survivors (n=28).

Intervention week	Number of valid days/week, mean (SD)	Steps/day, mean (SD)
1	1.93 (2.61)	7218.61 (4417.71)
2	4.78 (2.83)	6447.97 (3405.71)
3	5.30 (2.52)	6200.9 (3382.68)
4	5.22 (2.61)	6645.24 (3607.05)
5	5.04 (2.89)	6782.65 (3901.27)
6	4.85 (2.74)	6639.79 (4072.6)
7	5.11 (2.53)	6570.96 (3950.37)
8	5.07 (2.89)	6628.75 (3745.22)
9	4.70 (3.01)	6871.6 (4134.65)
10	4.69 (3.17)	6598.36 (3481.8)
11	4.77 (2.96)	7384.56 (4157.78)
12	4.08 (2.62)	6686.95 (3183.27)

Mobile App Rating Scale

At 1 month, 25 participants completed the MARS usability scale responding to questions about the MyDataHelps app functionality, appearance, and overall rating on a scale of 1 to 5 (1=worst and 5=best; Table 3). The highest-rated category was “performance” with a mean score of 4.36 (SD 1.02), and the lowest-rated category was “ease of use” with a mean score of 3.79 (SD 1.22). The overall app quality received a mean score of 3.79 (SD 0.82).

Table . MARS (Mobile App Rating Scale) usability scores of the MyDataHelps smartphone app; ratings range from 1=worst to 5=best.

Variable and question	Values, mean (SD)
Performance: How accurately/fast do the app features (functions) and components (buttons/menus) work?	4.36 (1.02)
Ease of use: How easy is it to learn how to use the app? How clear are the menu labels/icons and instructions?	3.79 (1.22)
Navigation: Is moving between screens logical/accurate/appropriate/uninterrupted? Are all necessary screen links present?	4.21 (1.22)
Interactions: Are interactions (taps/swipes/pinches/scrolls) consistent and intuitive across all components/screens?	4.13 (1.05)
Layout: Is the arrangement and size of buttons/icons/menus/content on the screen appropriate or zoomable if needed?	4.04 (1.02)
Graphics: How high is the quality/resolution of graphics used for buttons/icons/menus/content?	3.80 (0.85)
Visual appeal: How good does the app look?	3.84 (0.78)
Overall rating: What is your overall star rating of the app?	3.79 (0.82)

Qualitative Feedback

Ten participants completed 3-month follow-up interviews assessing their overall experiences in the program. Four themes were identified from our analyses, and illustrative quotes can be seen in Table 4. Overall, participants felt the Fitbit activity tracker and the MyDataHelps app were easy to use but also benefited from the assistance of the research team. The weekly motivational messages within MyDataHelps were not very memorable and were often confused with the daily push notifications sent directly to the Fitbit activity tracker. For example, when asked about the content of the MyDataHelps weekly messages, a participant responded, “I didn’t pay a whole lot of attention to them. I know I got ‘em maybe daily or every other day.” However, participants noted overall that the messages and notifications motivated them to be more active and accrue additional steps. Lastly, most participants were comfortable with and liked the idea of sharing their physical activity data with their clinicians.

Table . Enrolled patient feedback from qualitative interviews.

Theme	Illustrative quotes
Fitbit and app setup were easy and assistance from the research team was helpful.	• "I didn't find it [Fitbit and MyDataHelps setup and navigation] to be a problem at all once you walked me through it." "The information provided was pretty self-explanatory, even though I'm not very computer-savvy and never had a Fitbit."
Motivational messages within the app were not memorable.	• "They weren't sent often. Not, not particularly unless I set it up wrong. I think if I had gotten something on a weekly or even a daily [basis], it would motivate me." "Well, the messages, if any, were, I believe...I forget how I was able to access them...I'm trying to think of a specific message."
The step counts and Fitbit notifications to be active were motivating.	• "Tracking my steps was helpful, especially given how I've changed my lifestyle from working full-time to finally being home most of the time. And I had to make sure I put some time into taking steps or helping my health progress." • "Before, I never actually paid attention to like, 'Okay, I'm gonna do this many steps today, or I'm gonna do this.' And that was pretty cool that it gave that option to keep track of all that. And if I beat my goal from the day before, it would let me know. It would tell me, 'Congratulations!' which was kind of cool."
Patients were comfortable with and liked the idea of medical professionals seeing their data.	• "I did tell my oncologist that I was doing this study and I also spoke with my pulmonologist...And I told my primary care, and she said, "Well, that's interesting." I said, "You know the thing is, you never know what [information] is going to come along that might be helpful to you in the long run." • "I would like that because it's able to connect some of the dots, right? They're only seeing, you know, the medical side of it, but they can't see your everyday activity."

Five advanced practice providers specializing in oncology completed interviews to assess their perceptions of the digital health program and workflow integration. All 5 clinicians were female, had been practicing in their clinic for >1 year, and treated either disease-specific patients (breast, gastrointestinal, or genitourinary) or patients of all diagnoses. Clinicians were from one academic hospital (n=4) and one community hospital (n=1). Four themes were identified with illustrative quotes represented in Table 5. Clinicians noted some patients use wearables but rarely bring them up during their clinic visits. They also expressed that goal setting may be helpful for their patients and thought it may be helpful to have objective metrics of physical activity to track over time in lieu of standard weight and BMI clinical measures. Most clinicians wanted more education and training on physical activity and wearable device programs and better standard operating procedures integrated into their workflows.

Table . Clinician feedback from qualitative interviews.

Theme	Illustrative quotes
Some patients use wearables but rarely discuss data with clinicians.	"I don't know that it comes up all that often. Things have changed since Covid. I used to have tons of patients who would do this kind of thing like the Silver Sneakers-type program...Not too much beyond the generic, like, 'I shoot for X number of steps a day and I usually [get] X.'"
Activity trackers can be helpful to motivate patients and keep them accountable.	"Some people are motivated by their internal motivations. And I think Fitbit still gives you that accountability and that trackability. I've had patients that have used pedometer-esque, Fitbit-like, applications, whether it's their iPhone or something like that, and they'll say, 'Okay, I got this many steps.' And I think that it does motivate them to be more active."
Clinicians prefer objective measures similar to BMI, weight, and blood pressure that they can track over time and refer to afterward (eg, vital signs).	"Typically, what I use is their BMI...to say, 'Oh, congratulations! Your BMI last visit was this and look what it is now. I can see that you have better energy, and your wellbeing seems to be improved as well.' It could be helpful for that sort of reinforcing and motivating and monitoring portion."
Clinicians want more training and systematic processes to view these data as part of active workflow.	"With training and guidance, if that was considered part of our scope, if we felt like patients were benefiting...yeah, absolutely...if it's something that would help patients and is clinically appropriate depending on whatever training we got."

Discussion

Principal Results

Our study found that deploying a Fitbit or app dyad remotely was feasible and acceptable in this convenience sample of diverse participants with histories of several types of cancers. Participant engagement, as indicated by mean valid wear days and mean daily steps each week, was highly consistent over the course of the 12-week intervention. This finding was similar to those of past trials in cancer survivors [27-29]. The main exception to this was that on average, participants had fewer valid wear days in the first week of the intervention, which, similar to other trials, can be attributed to an initial adjustment period to wearing a Fitbit daily [30]. Though engagement was high, step counts remained relatively unchanged from baseline to 12 weeks. Lastly, several cancer care clinicians were interested in the ability to deploy such an intervention into their clinical practice for physical activity surveillance and interventions for their patients with cancer histories.

The MARS app ratings were similar to other physical activity monitoring app studies in that functionality components were rated higher than aesthetic components [24,31]. The high functionality ratings were supported in our qualitative feedback, with participants specifically noting the app being easy to use even if they themselves were not very tech-savvy. This is critical in the cancer survivor population, as 67% of US cancer survivors are over the age of 65, and this proportion is expected to grow to 74% by 2040. There are noted disparities in the use of digital health tools in advanced age [32]. However, participant qualitative feedback revealed the messages sent within the app were not very memorable. This may have been due to the message content, or possibly that they did not know where to view the messages on the MyDataHelps app dashboard. Several participants made more mention of the daily Fitbit device notifications being helpful. As motivational messaging has been shown to be effective in promoting physical activity [33,34] and was memorable to our participants on the Fitbit device, it is possible the notifications, along with vibrations, going directly to a patient's device on their wrist are more noticeable than having to open an app to see the message. Despite notifications being reported as helpful, changes in step count over time were minimal. Though consistent with prior feasibility studies in this population [27,35], it will be critical to explore the effect of intervention components on step count over time in a large randomized trial. This includes exploring the addition of social support, a feature in Fitbit, and a feasible and acceptable intervention method for cancer survivors [35].

Qualitative feedback from participants yielded several important implications of recruiting, enrolling, and retaining patients throughout the study. The first important implication was that the MyDataHelps and Fitbit: Health and Wellness apps setup was straightforward and easy to follow. Some participants noted they were not tech-savvy, but they did not find it to be an issue. Participants also appreciated the help from the study team during the initial setup and to troubleshoot any issues that arose. To provide technical support, our study team used multiple methods to meet the needs of the participants. For most, this only entailed troubleshooting issues over the phone or SMS text message. Six participants additionally required videoconferencing and screen sharing to troubleshoot technical difficulties with a visual aid. Of the 6 participants, 4 chose to come in person for the study team to troubleshoot their issues. With additional assistance by video call or in person, these participants were also able to complete the study tasks quickly. Overall, the team found that participants greatly appreciated having options to connect and talk with them. The final theme from participants indicated they were comfortable sharing data with health care clinicians. Health care professionals can be critical in helping survivors engage in physical activity. Survivors reported a lack of physical activity guidance, prescriptions, and referrals from their care team as barriers to activity [36].

Clinicians revealed that patients rarely discuss wearable devices or data with them during their clinic visits. Oncology nurses have reported that patients lack interest in discussing physical activity with them. However, clinicians liked the notion of having long-term activity monitoring as a topic to discuss with patients in lieu of discussing weight or BMI. Physical activity independent of weight loss can improve health outcomes in individuals with obesity [37], and sedentary time, which is inversely correlated with physical activity levels, is an independent risk factor of cancer incidence [38]. Using a more neutral objective measure of physical activity rather than a more sensitive metric like weight or BMI may be more acceptable to discuss with patients who have weight concerns. To do this, clinicians expressed wanting more education and training on integrating physical activity digital health tools into their workflow as options for their patients. Lack of provider knowledge has been one of the most commonly reported barriers to the provision of physical activity promotion by cancer care clinicians to survivors [39]. Additionally, integrating wearable devices and subsequent data directly into clinicians' workflow poses challenges. Prior studies have identified barriers to integrating devices into the EHR, maintaining privacy and confidentiality of patient data, lack of system interoperability and connectivity of wearable devices and health systems, and patient information or data overload [40].

Limitations

This study has limitations to note. The small sample size is a statistical limitation. Similar to most studies done in the past, this study was limited to 3 months in duration; thus, there still remains little research regarding long-term outcomes from Fitbit-based interventions, including health outcomes [11,16]. As trials longer than 6 months have noted greater dropout rates over time [16], it will be important to explore this timeframe in future trials. Another limitation regarding the study design was the lack of a control group. Although this design has been used in similar feasibility trials, the lack of a control group limits the inferences we can make pertaining to the internal validity of the intervention. Though participant step counts were fairly consistent during the trial, this consistency could be attributed to having a highly motivated sample of individuals or to the Hawthorne effect and its impact on motivation and behavior. With regard to app usability, some participants reported confusing the SMS text messages delivered by the MyDataHelps app with push notifications sent by the Fitbit: Health & Wellness

app, which may have affected their MARS usability score evaluation. Additionally, familiarity and comfort with navigating smartphone apps associated with wearable technology were not assessed. Notably, we relied on recruiting patients from an EHR patient registry list. This method may have allowed us to capture patients who would not otherwise have reached out to join the study, but also may have introduced selection bias and is less pragmatic than a program that is implemented within routine clinical workflow. While we examined step counts only, it should be noted that Fitbit tracks additional metrics of minutes and intensity of activity and time spent being sedentary that should be considered. Lastly, we ascertained that most of the clinicians wanted training, guidance, or knowledge of how to integrate these into their workflow. Given the busy workload of clinicians, it is possible that in practice they would prefer better reports and objective data to be provided, rather than actively taking part in new tasks added into their workflow. This should be examined in future trials.

Conclusions

This study showed that a remotely delivered light-intensity physical activity program was feasible and acceptable in a sample of diverse cancer survivors. Future studies should consider registry-based methods and other strategies to engage hard-to-reach cancer survivor populations who are known to experience disproportionately high adverse health outcomes and low physical activity levels. These strategies should also be directed toward making improvements to recruit and engage larger numbers of participants from diverse sociodemographic backgrounds with consideration to technology access and use. On an individual scale, since some participants may have been more intrinsically motivated than others, future trials will benefit from assessing the underpinnings of participant motivation. Lastly, future trials should place emphasis on clinic implementation, including the quality of and method of delivery for reports, given the noted importance and use of the wearable device and app integration by cancer care clinicians for their survivors to be motivated to engage in and sustain physical activity.

Acknowledgments

This work was supported by the National Cancer Institute (5P50CA244693) and the National Heart, Lung, and Blood Institute (K01HL163254). The funders did not play a role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JMF, DDM, A Singh, SLC, A Soni
Data curation: JMF, ZD, RN, AK, BSG
Formal analysis: JMF, ZD, RN, AK, BSG
Funding acquisition: SLC, JMF
Investigation: A Singh, A Soni, JMF, ZD, RN
Project administration: A Singh, A Soni, JMF, ZD, RN
Software: A Singh, A Soni, DDM
Supervision: JMF
Visualization: ZD, RN, AK, BSG, A Singh, A Soni, SLC, DDM, JMF
Writing – original draft: ZD, RN, AK, BSG, A Singh, A Soni, SLC, DDM, JMF
Writing – review & editing: ZD, RN, AK, BSG, A Singh, A Soni, SLC, DDM, JMF

Conflicts of Interest

DDM reports receiving consulting fees from Avania, NAMSA, Heart Rhythm Society, and Fitbit; travel reimbursement from OHSU, Michigan, and UTSA; and grant funding and consulting from Flexcon. The authors have no further interests to declare.

Multimedia Appendix 1
Patient and provider interview guides.
[DOCX File, 16 KB - cancer_v11i1e60034_app1.docx]

Multimedia Appendix 2
MyDataHelps app push notifications.
[DOCX File, 16 KB - cancer_v11i1e60034_app2.docx]

References

1. Garcia DO, Thomson CA. Physical activity and cancer survivorship. Nutr Clin Pract 2014 Dec;29(6):768-779. [doi: 10.1177/0884533614551969] [Medline: 25335787]
2. Sturgeon KM, Deng L, Bluethmann SM, et al. A population-based study of cardiovascular disease mortality risk in US cancer patients. Eur Heart J 2019 Dec 21;40(48):3889-3897. [doi: 10.1093/eurheartj/ehz766] [Medline: 31761945]
3. Gilchrist SC, Barac A, Ades PA, et al. Cardio-oncology rehabilitation to manage cardiovascular outcomes in cancer patients and survivors: a scientific statement from the American Heart Association. Circulation 2019 May 21;139(21):e997-e1012. [doi: 10.1161/CIR.0000000000000679] [Medline: 30955352]
4. Stout NL, Sleight A, Pfeiffer D, Galantino ML, deSouza B. Promoting assessment and management of function through navigation: opportunities to bridge oncology and rehabilitation systems of care. Support Care Cancer 2019 Dec;27(12):4497-4505. [doi: 10.1007/s00520-019-04741-0]
5. Chua HF, Liberzon I, Welsh RC, Strecher VJ. Neural correlates of message tailoring and self-relatedness in smoking cessation programming. Biol Psychiatry 2009 Jan 15;65(2):165-168. [doi: 10.1016/j.biopsych.2008.08.030] [Medline: 18926523]
6. Patrick H, Canevello A. Methodological overview of a self-determination theory-based computerized intervention to promote leisure-time physical activity. Psychol Sport Exerc 2011 Jan;12(1):13-19. [doi: 10.1016/j.psychsport.2010.04.011] [Medline: 21103069]
7. Faro JM, Mattocks KM, Nagawa CS, et al. Physical activity, mental health, and technology preferences to support cancer survivors during the COVID-19 pandemic: cross-sectional study. JMIR Cancer 2021 Feb 3;7(1):e25317. [doi: 10.2196/25317] [Medline: 33471776]
8. Faro JM, Yue KL, Leach HJ, et al. Development and pilot testing of a clinic implementation program delivering physical activity electronic referrals to cancer survivors. Transl Behav Med 2023 Sep 28;13(10):794-803. [doi: 10.1093/tbm/ibad035] [Medline: 37318360]
9. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HWW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec;11(6):704-719. [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
10. Wang L, Langlais CS, Kenfield SA, et al. mHealth interventions to promote a healthy diet and physical activity among cancer survivors: a systematic review of randomized controlled trials. Cancers (Basel) 2022 Aug 6;14(15):3816. [doi: 10.3390/cancers14153816] [Medline: 35954479]
11. Ginsburg GS, Picard RW, Friend SH. Key issues as wearable digital health technologies enter clinical care. N Engl J Med 2024 Mar 21;390(12):1118-1127. [doi: 10.1056/NEJMra2307160] [Medline: 38507754]
12. Mobile fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/?menuItem=d40cde3f-c455-4f0e-9be0-0aefcdaeee00 [accessed 2022-06-13]
13. Shandhi MMH, Singh K, Janson N, et al. Assessment of ownership of smart devices and the acceptability of digital health data sharing. NPJ Digit Med 2024 Feb 22;7(1):44. [doi: 10.1038/s41746-024-01030-x] [Medline: 38388660]
14. Bluethmann SM, Mariotto AB, Rowland JH. Anticipating the "silver tsunami": prevalence trajectories and co-morbidity burden among older cancer survivors in the United States. Cancer Epidemiol Biomarkers Prev 2016 Jul;25(7):1029-1036. [doi: 10.1158/1055-9965.EPI-16-0133] [Medline: 27371756]
15. Singh B, Zopf EM, Howden EJ. Effect and feasibility of wearable physical activity trackers and pedometers for increasing physical activity and improving health outcomes in cancer survivors: a systematic review and meta-analysis. J Sport Health Sci 2022 Mar;11(2):184-193. [doi: 10.1016/j.jshs.2021.07.008] [Medline: 34314878]
16. McManus DD, Trinquart L, Benjamin EJ, et al. Design and preliminary findings from a new electronic cohort embedded in the Framingham Heart Study. J Med Internet Res 2019 Mar 1;21(3):e12143. [doi: 10.2196/12143] [Medline: 30821691]
17. Nuvvula S, Ding EY, Saleeba C, et al. NExUS-Heart: novel examinations using smart technologies for heart health—data sharing from commercial wearable devices and telehealth engagement in participants with or at risk of atrial fibrillation. Cardiovasc Digit Health J 2021 Oct;2(5):256-263. [doi: 10.1016/j.cvdhj.2021.08.001] [Medline: 35265917]
18. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
19. Tudor-Locke C, Barreira TV, Schuna JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. Med Sci Sports Exerc 2015 Apr;47(4):839-842. [doi: 10.1249/MSS.0000000000000476] [Medline: 25121517]
20. Mikkelsen MLK, Berg-Beckhoff G, Frederiksen P, et al. Estimating physical activity and sedentary behaviour in a free-living environment: a comparative study between Fitbit Charge 2 and Actigraph GT3X. PLoS ONE 2020;15(6):e0234426. [doi: 10.1371/journal.pone.0234426] [Medline: 32525912]
21. Algheryafi RA, Bevans KB, Hiremath SV, Lai JS, Tucker CA. Convergent validity of the Patient Reported Outcome Measurement Information System-Pediatric Physical Activity Instrument (PROMIS®-PA) with wearable devices in adolescents. Children (Basel) 2023 May 26;10(6):940. [doi: 10.3390/children10060940] [Medline: 37371172]
22. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]

23. Wilks CR, Chu C, Sim D, et al. User engagement and usability of suicide prevention apps: systematic search in app stores and content analysis. JMIR Form Res 2021 Jul 14;5(7):e27018. [doi: 10.2196/27018] [Medline: 34259163]
24. Ding EY, Pathiravasan CH, Schramm E, et al. Design, deployment, and usability of a mobile system for cardiovascular health monitoring within the electronic Framingham Heart Study. Cardiovasc Digit Health J 2021 Jun;2(3):171-178. [doi: 10.1016/j.cvdhj.2021.04.001] [Medline: 35265906]
25. Physical activity and the person with cancer. American Cancer Society. URL: https://www.cancer.org/cancer/survivorship/be-healthy-after-treatment/physical-activity-and-the-cancer-patient.html [accessed 2024-12-30]
26. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516. [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
27. Rossi A, Frechette L, Miller D, et al. Acceptability and feasibility of a Fitbit physical activity monitor for endometrial cancer survivors. Gynecol Oncol 2018 Jun;149(3):470-475. [doi: 10.1016/j.ygyno.2018.04.560] [Medline: 29692337]
28. Rahimy E, Usoz M, von Eyben R, et al. Phase II trial evaluating efficacy of a Fitbit program for improving the health of endometrial cancer survivors. Gynecol Oncol 2021 Apr;161(1):275-281. [doi: 10.1016/j.ygyno.2021.01.033] [Medline: 33551199]
29. Van Blarigan EL, Kenfield SA, Tantum L, Cadmus-Bertram LA, Carroll PR, Chan JM. The Fitbit One physical activity tracker in men with prostate cancer: validation study. JMIR Cancer 2017 Apr 18;3(1):e5. [doi: 10.2196/cancer.6935] [Medline: 28420602]
30. Hardcastle SJ, Jiménez-Castuera R, Maxwell-Smith C, Bulsara MK, Hince D. Fitbit wear-time and patterns of activity in cancer survivors throughout a physical activity intervention and follow-up: exploratory analysis from a randomised controlled trial. PLoS ONE 2020;15(10):e0240967. [doi: 10.1371/journal.pone.0240967] [Medline: 33075100]
31. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35. [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
32. Mubarak F, Suomi R. Elderly forgotten? Digital exclusion in the information age and the rising grey digital divide. Inquiry 2022;59:469580221096272. [doi: 10.1177/00469580221096272] [Medline: 35471138]
33. Hevel DJ, Amorose AJ, Lagally KM, Rinaldi-Miles A, Pierce S. Testing the effects of messaging on physical activity motivation in active and non-active adults. Psychol Sport Exerc 2019 Jul;43:333-342. [doi: 10.1016/j.psychsport.2019.04.010]
34. Koontz BF, Levine E, McSherry F, et al. Increasing physical activity in cancer survivors through a text-messaging exercise motivation program (ICanSTEP). Support Care Cancer 2021 Dec;29(12):7339-7349. [doi: 10.1007/s00520-021-06281-y] [Medline: 34050402]
35. Gell NM, Grover KW, Humble M, Sexton M, Dittus K. Efficacy, feasibility, and acceptability of a novel technology-based intervention to support physical activity in cancer survivors. Support Care Cancer 2017 Apr;25(4):1291-1300. [doi: 10.1007/s00520-016-3523-5] [Medline: 27957621]
36. Romero-Elías M, Beltrán-Carrillo VJ, González-Cutre D, Jiménez-Loaisa A. Barriers to physical activity participation in colorectal cancer patients during chemotherapy treatment: a qualitative study. Eur J Oncol Nurs 2020 May 13;46:101769. [doi: 10.1016/j.ejon.2020.101769] [Medline: 32506009]
37. Pojednic R, D'Arpino E, Halliday I, Bantham A. The benefits of physical activity for people with obesity, independent of weight loss: a systematic review. Int J Environ Res Public Health 2022 Apr 20;19(9):4981. [doi: 10.3390/ijerph19094981] [Medline: 35564376]
38. Friedenreich CM, Ryder-Burbidge C, McNeil J. Physical activity, obesity and sedentary behavior in cancer etiology: epidemiologic evidence and biologic mechanisms. Mol Oncol 2021 Mar;15(3):790-800. [doi: 10.1002/1878-0261.12772] [Medline: 32741068]
39. Karvinen KH, McGourty S, Parent T, Walker PR. Physical activity promotion among oncology nurses. Cancer Nurs 2012;35(3):E41-E48. [doi: 10.1097/NCC.0b013e31822d9081] [Medline: 21946904]
40. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861. [doi: 10.2196/12861] [Medline: 31512582]

Abbreviations

EHR: electronic health record
MARS: Mobile App Rating Scale
PAR-Q: Physical Activity Readiness Questionnaire

Edited by N Cahill; submitted 29.04.24; peer-reviewed by B Hasan, J Schilling; revised version received 25.11.24; accepted 29.11.24; published 24.02.25.

Please cite as:
Dabbagh Z, Najjar R, Kamberi A, Gerber BS, Singh A, Soni A, Cutrona SL, McManus DD, Faro JM
Usability and Implementation Considerations of Fitbit and App Intervention for Diverse Cancer Survivors: Mixed Methods Study
JMIR Cancer 2025;11:e60034
URL: https://cancer.jmir.org/2025/1/e60034
doi:10.2196/60034

Exploring Older Adult Cancer Survivors' Digital Information Needs: Qualitative Pilot Study

Lorelei Newton[1], PhD; Helen Monkman[2], PhD; Claire Fullerton[1], BScN

[1]School of Nursing, University of Victoria, PO Box 1700 STN CSC, Victoria, BC, Canada

[2]School of Health Information Science, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Lorelei Newton, PhD
School of Nursing, University of Victoria, PO Box 1700 STN CSC, Victoria, BC, Canada

Abstract

Background: Older adults (aged >65 years) are disproportionately affected by cancer at a time when Canadians are surviving cancer in an unprecedented fashion. Contrary to persistent ageist assumptions, not only do the majority of older adult cancer survivors use digital health technologies (DHTs) regularly, such technologies also serve as important sources of their health information. Although older adults' transition to cancer survivorship is connected to the availability and provision of relevant and reliable information, little evidence exists as to how they use DHTs to supplement their understanding of their unique situation to manage, and make decisions about, their ongoing cancer-related concerns.

Objective: This pilot study, which examined older adult cancer survivors' use of DHTs, was conducted to support a larger study designed to explore how digital health literacy dimensions might affect the management of cancer survivorship sequelae. Understanding DHT use is also an important consideration for digital health literacy. Thus, we sought to investigate older adult cancer survivors' perceptions of DHTs in the context of accessing information about their health, health care systems, and health care providers.

Methods: A qualitative pilot study, which involved semistructured interviews with older adult cancer survivors (N=5), was conducted to explore how participants interacted with, accessed, and searched for information, as well as how DHT use related to their cancer survivorship. Institutional ethics approval (#21 - 0421) was obtained. Interpretive description inquiry—a practice-based approach suitable for generating applied knowledge—supported exploration of the research question. Thematic analysis was used to examine the transcripts for patterns of meaning (themes).

Results: Assessing the credibility of digital information remains challenging for older adult cancer survivors. Identified benefits of DHTs included improved access to meet health information needs, older adult cancer survivors feeling empowered to make informed decisions regarding their health trajectory, and the ability to connect with interdisciplinary teams for care continuity. Additionally, participants described feeling disconnected when DHTs seemed to be used as substitutes for human interaction. The results of this pilot study were used to create 12 additional questions to supplement a digital health literacy survey, through which we will seek a more fulsome account of the relationship between digital health literacy and DHTs for older adult cancer survivors.

Conclusions: Overall, this pilot study confirmed the utility of DHTs in enhancing the connection of older adult cancer survivors to their health care needs. Importantly, this connection exists on a continuum, and providing greater access to technologies, in combination with human support, leads to feelings of empowerment. DHTs are an important aspect of contemporary health care; yet, these technologies must be seen as complementary and not as replacements for human interaction. Otherwise, we risk dehumanizing patients and disconnecting them from the care that they need and deserve.

(JMIR Cancer 2025;11:e59391) doi:10.2196/59391

KEYWORDS

older adults; cancer survivors; digital health literacy; digital health technologies; aging; qualitative; pilot study; semistructured interview

Introduction

Older adults (aged >65 years) are the fastest-growing segment of the Canadian population with unprecedented cancer survivorship [1]. Technology has become increasingly important, with older adult cancer survivors reporting that a significant proportion of their health care information is gathered from digital sources [2]. Despite such availability, finding and accessing information during transitions to survivorship are challenging [3]. Access to information plays a vital role in improving cancer survivors' health outcomes and quality of

life. Although older adult cancer survivors may have unique needs, there is little research examining the extent of digital health technology (DHT) use to support their health. DHTs "use computing platforms, connectivity, software, and sensors for health care and related uses" and include telemedicine, wearable devices, educational resources, remote patient monitoring, etc [4].

With the increasing reliance on DHTs [5], a consistent concern for older adult cancer survivors is interacting with health care systems digitally [6]. Contrary to persistent ageist assumptions, not only do 95% of this group use DHTs regularly, DHTs are substantial sources of their health information [7]. Understanding DHT use is also an important consideration for digital health literacy (DHL). DHL is the capacity "to acquire, process, communicate, and understand health information and services, make effective health decisions, and promote and improve individual and collective health in the context of the use of digital information and technologies" [7]. This pilot study, which examined older adult cancer survivors' use of DHTs, was conducted to support a larger study designed to explore how DHL dimensions might affect the management of cancer survivorship sequelae.

Methods

Ethical Considerations

The pilot study reporting was guided by the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Checklist 1) [8]. Institutional ethics approval (#21 - 0421) was obtained from the University of Victoria, and informed consent was secured in advance of conducting interviews. Confidentiality was maintained throughout the research process, and all transcripts were deidentified prior to analysis, with data stored in password-protected files. Participants were given a CAD $25 (US $17.53) gift card.

Recruitment

Participants were recruited through the social media platforms of local eldercare and retirement groups, and recruitment consisted of 3 recruitment posters that were posted weekly in attempts to elicit diverse responses. As only 2 participants responded to those efforts, an additional 3 participants were recruited, using the snowball method by way of the initial 2 participants' social networks. The inclusion criteria included being aged 65 years or older, speaking English, having completed cancer treatment within the past 5 years, and being a resident of British Columbia, Canada.

Data Collection

Semistructured, 90-minute, individual interviews were conducted over a secure videoconferencing platform. Interview questions investigated how participants interacted with, accessed, and searched for information, as well as DHT use related to their cancer survivorship (Textbox 1). The audio-recorded interviews were transcribed and deidentified by the research associate (CF), with two researchers (LN and HM) reviewing for accuracy, prior to analysis. Participants were offered an opportunity to review their transcript for accuracy.

Textbox 1. Interview questions. Research questions asked to the participants aimed to identify how older adults use digital health tools to interact with, access, and search for information related to their health care and cancer survivorship.

Interview questions

1. What information resource(s) did you use most often?
2. How did you manage multiple information resources?
3. How did you determine the information was credible?
4. Knowing what you know now, what information do you wish you had access to?
5. What questions did you have around your transition from treatment to survivorship?
6. What questions were the most difficult to find answers to?
7. What have you found difficult to understand about this transition?
8. What do you like/dislike about using digital health tools?
9. What makes it easy/difficult for you to use digital health tools?
10. Now that you have completed the eHealth literacy survey, and after thinking about the questions already asked, is there anything else you would like to tell us?

Reflexivity, Interpretive Description, and Thematic Analysis

All interviews were conducted by 2 PhD-prepared assistant professors (LN and HM) and an undergraduate student, who was included as a research associate (CF). The lead authors have experience with qualitative research; are from the Global North; and identify as cisgender, White, female individuals who endeavor to be reflexive on positionality and perspectives throughout the research process. Interpretive description inquiry—a practice-based approach suitable for generating applied knowledge—supported exploration of the research question [9]. Thematic analysis was used to examine the transcripts for patterns of meaning (themes) [10]. The researchers convened after independently reviewing the first transcript to ensure congruence. As the researchers became familiar with the data and generated initial codes, themes were reviewed together. Finally, the researchers agreed that the themes accurately reflected participants' responses.

Results

Participants

A total of 5 older adult cancer survivors (4 women and 1 man), with an average age of 69 (SD 2.06) years, completed interviews. All 5 had postsecondary education and stated that they accessed DHTs daily. All participants had completed cancer treatment within the past 5 years and represented different geographic areas of the province. Participants were not known to the researchers prior to the interviews; however, researcher information and an overview of the research project were included in the informed consent process. Time and an opportunity for participants to ask questions were provided prior to the interviews. Additionally, during the informed consent process, participants were assured that their data were private and were not to be shared publicly. Thus, supporting data are not available publicly.

Themes

Thematic analysis [10] highlighted the following three main themes: access, empowerment, and connection.

Access

All participants agreed that access to DHTs is crucial for health care purposes and for facilitating health care needs with minimal disruption to routines. Distinguishing between access and accessibility is essential; despite owning digital devices and feeling comfortable with using them, 3 participants had no interest in or did not enjoy using digital devices for health care. However, 1 participant expressed how DHTs "were very helpful" and that they "used them for follow up with therapists and counselors." Such access allowed health care needs, such as education and remote connection to health care professionals (HCPs), to be addressed conveniently.

Empowerment

All participants associated DHT with feelings of empowerment, that is, being involved in their health decisions and being in control of their information (including access). All participants discussed the autonomy provided by DHTs. Participants explained feeling empowered by being in control, by accessing education, and by the ability to book appointment times. One participant said, "You can get all the information you need…I never cease to be amazed at what technology can do." Another participant expressed their satisfaction with DHTs, stating "I don't have to track it down; everything is just right there." As such, participants found that independently booking appointments, retrieving educational materials, and reviewing available medical records (eg, laboratory results) improved their sense of empowerment.

Connection

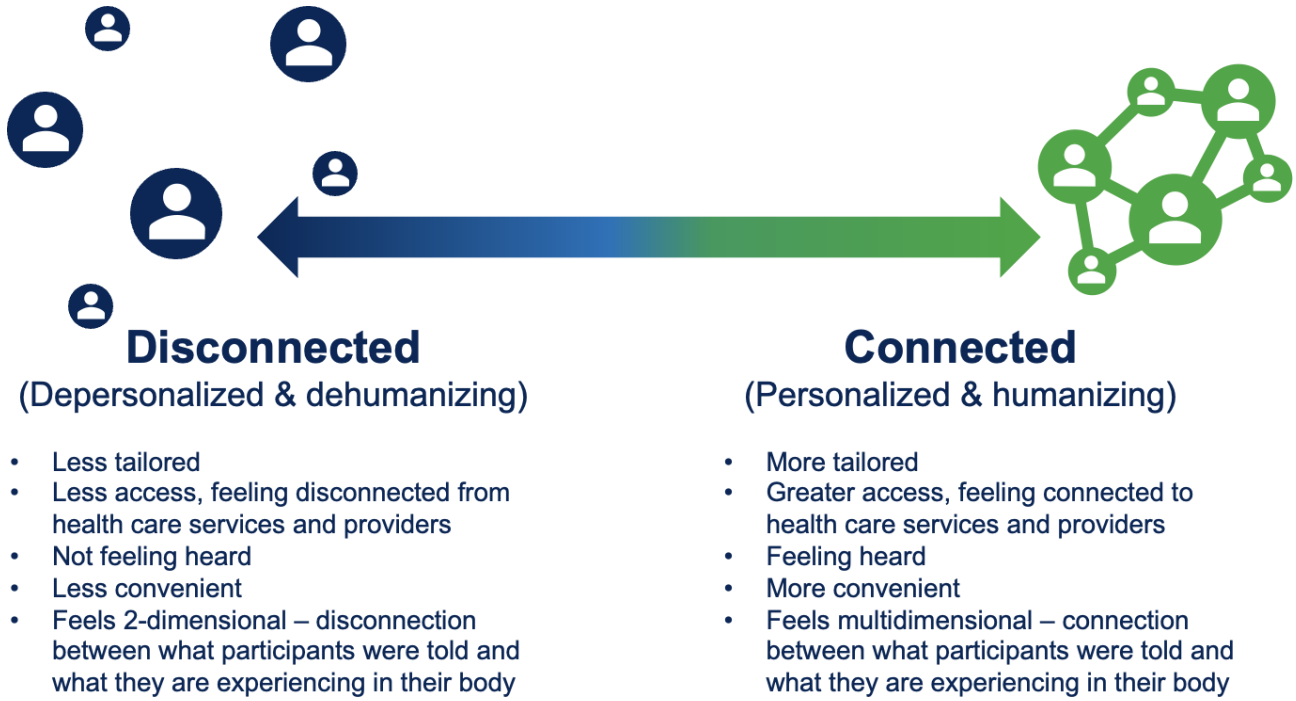
Participants described how DHTs enhanced connection. One participant stated that DHTs "allow for discussion that healthcare professionals wouldn't [otherwise] have time [to start]." A second participant echoed this, saying "When I was really sick, my doctor called me every day." This web-based connection was especially important in instances when participants were too ill to commute but not ill enough to be admitted to a hospital. In other cases, participants discovered community support through online groups. For example, one participant found videoconferencing with a survivorship group very helpful. Another participant stated, "I remember [medication] affected emotions; nothing could prepare you for that…My doctor and counselor didn't really talk about it...going online and finding that other people are having the same impact was very comforting." In this case, the patient found a sense of community through shared experiences on the web.

Paradoxically, participants also described experiences where human connection was absent. One participant shared an example of when an HCP noticed and addressed a physical concern as they were leaving an in-person appointment. This participant felt that this would never have been noticed in a web-based meeting. Another key observation was how intonation of voice, facial expressions, and body language of HCPs were difficult to discern and negatively impacted the participants' interpretations of interpersonal communication. One participant expressed frustration with "stereotypes placed on older people." Another stated, "People tend to treat everyone over the age of 65 like they are homogenous." Participants expressed annoyance with HCPs often overexplaining technology or assuming a lack of understanding of how to access DHTs when the participants felt that they could complete the task independently. The participants also pointed out instances in which HCPs lacked knowledge about the DHTs they expected patients to use.

Key Learnings to Inform Future Work

Participants reported that DHTs can facilitate access and empowerment. Participants' experiences also pointed to how DHTs can either foster connection or create barriers to human connection. That is, they described their feelings of connection as along a continuum from feeling disconnected from health care services to feeling connected with health care services (Figure 1). Participants outlined items that contribute to disconnection, such as the feeling of not being heard, services that are less tailored to their situation, and less accessibility. In contrast, participants further described feelings of greater connection with items such as greater access, the feeling of being heard by HCPs, and congruence between information received and what they are experiencing in their body. The results of this pilot study were used to create 12 additional questions to supplement a DHL questionnaire (Textbox 2) for the next phase of the project [11], in which we will seek a more fulsome account of the relationship between DHL and DHTs for older adult cancer survivors.

Figure 1. Connection continuum.



Textbox 2. Questions to be included in the eHealth literacy survey. After completing the interviews, the following questions, based on the insights from pilot study participants, were created to address additional areas of importance in the digital health literacy survey.

Questions to be included in the eHealth literacy survey

1. To support your survivorship, what information did you get using digital health tools?
2. Has the COVID-19 pandemic changed the way you think about digital health tools?
3. What digital health tools (apps, websites, wearables) did you use to support health and survivorship and how useful were they?
4. Did you join any support groups for people who have finished cancer treatment?
5. How often do you use digital health tools?
6. Is there anything you like about using digital health tools?
7. Is there anything you dislike about using digital health tools?
8. Is there anything you dislike about digital health tools as a part of health care?
9. If you are frustrated using digital health tools, how did you get through this frustration?
10. Where do you go if you need help with digital health tools?
11. In the past year, have you connected with a health care professional digitally?
12. What recommendations do you have about digital health tools for health professionals and organizations?

Discussion

Principal Findings

The participants in this pilot study confirmed some of the benefits of using DHTs for older adults' cancer survivorship; they can learn about their condition, connect with interdisciplinary teams for continuity of care, find connections and community support, and make educated and informed decisions regarding their health trajectory in survivorship. These findings are congruent with other studies, in which older adult cancer survivors expressed preferences for in-person visits and personalized telehealth visits [12], described how using DHTs could provide a sense of autonomy (ie, by allowing them to be actively involved with their health care) [2], and appreciated access to technology while maintaining a strong preference to be listened to with basic respect [13]. However, the propensity of health care systems to use DHTs as substitutes for human interactions to increase cost-effectiveness and efficiency can counteract those benefits. Indeed, given the existing structural ageism inherent in contemporary health care, there is a real danger of amplifying ageist processes, which can result in care that does not account for the intersection of normal aging and cancer survivorship [13]. Further, with the current lack of accounting for the burdens of navigating challenging cancer care systems, DHTs can either escalate feelings of disconnection or provide opportunities for connection and reconnection. Re-establishing the patient at the center of care and leveraging the humanity possible in DHTs are crucial; to continue to do

otherwise will ultimately lead to a sense of disengagement. DHTs are an important aspect of contemporary health care; yet, these technologies cannot replace HCP contact, or we risk dehumanizing patients and disconnecting them from the care that they need and deserve. By using DHTs compassionately and strategically for the ongoing care of older adult cancer survivors, HCPs can support this group along a continuum from feeling disconnected from health care services to feeling connected with health care services. Thus, it is imperative to determine the conditions under which DHTs complement health care and enhance rather than impair connection.

Limitations

Despite providing insights to augment the future survey, having only 5 participants inherently limits the scope and transferability of these findings. Further, although the participants represented geographic diversity, all identified as White, spoke English, and had postsecondary education; thus, this small, nonrepresentative sample may have reduced the richness of the data and the ability to achieve data saturation. Questions regarding diversity will also be added to the upcoming survey.

Conclusion

Overall, this pilot study confirmed the utility of DHTs in enhancing the connection of older adult cancer survivors to health care. Importantly, this connection exists on a continuum, and providing greater access to DHTs, in combination with human support, leads to feelings of empowerment. We are confident that applying these findings to further research will illuminate best practices for supporting older adult cancer survivors to optimize their cancer-free years.

Acknowledgments

The authors wish to acknowledge funding from the Internal Research Grant program through the Faculty of Human and Social Development, University of Victoria. Generative artificial intelligence tools were not used to create any aspect of this document.

Data Availability

In accordance with the informed consent process for this project, data are private and cannot be shared publicly.

Authors' Contributions

LN and HM conceptualized the overall research goals and aims and secured financial support for the project. All authors (LN, HM, and CF) participated in the data curation, formal analysis, investigation, and methodology. LN provided overall project administration and supervision. Although LN took the lead in writing original drafts and reviewing/editing, all authors participated in all stages of the manuscript through to publication.

Conflicts of Interest

None declared.

Checklist 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File, 1173 KB - cancer_v11i1e59391_app1.pdf]

References

1. Canadian cancer statistics 2023. Canadian Cancer Society. 2023 Nov. URL: https://cdn.cancer.ca/-/media/files/research/cancer-statistics/2023-statistics/2023_PDF_EN.pdf [accessed 2025-02-12]
2. Verma R, Saldanha C, Ellis U, Sattar S, Haase KR. eHealth literacy among older adults living with cancer and their caregivers: a scoping review. J Geriatr Oncol 2022 Jun;13(5):555-562. [doi: 10.1016/j.jgo.2021.11.008] [Medline: 34810146]
3. Fitch MI, Nicoll I, Newton L, Strohschein FJ. Challenges of survivorship for older adults diagnosed with cancer. Curr Oncol Rep 2022 Jun;24(6):763-773. [doi: 10.1007/s11912-022-01255-7] [Medline: 35286555]
4. What is digital health? U.S. Food and Drug Administration. 2020 Sep 22. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2025-02-12]
5. Holmes MM. Why people living with and beyond cancer use the internet. Integr Cancer Ther 2019;18:1534735419829830. [doi: 10.1177/1534735419829830] [Medline: 30741026]
6. Hall S, Sattar S, Ahmed S, Haase KR. Exploring perceptions of technology use to support self-management among older adults with cancer and multimorbidities. Semin Oncol Nurs 2021 Dec;37(6):151228. [doi: 10.1016/j.soncn.2021.151228] [Medline: 34753638]
7. Wang X, Luan W. Research progress on digital health literacy of older adults: a scoping review. Front Public Health 2022 Aug 5;10:906089. [doi: 10.3389/fpubh.2022.906089] [Medline: 35991040]
3. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

9. Thorne S. Interpretive Description: Qualitative Research for Applied Practice: Routledge; 2016. [doi: 10.4324/9781315545196]
10. Braun V, Clarke V. Thematic Analysis: A Practical Guide: SAGE Publications; 2021.
11. Kayser L, Karnoe A, Furstrand D, et al. A multidimensional tool based on the eHealth Literacy Framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36. [doi: 10.2196/jmir.8371] [Medline: 29434011]
12. Arthur EK, Pisegna J, Oliveri JM, Aker H, Krok-Schoen JL. Older cancer survivors' perspectives and use of telehealth in their cancer survivorship care in the United States: a ResearchMatch® sample. J Geriatr Oncol 2022 Nov;13(8):1223-1229. [doi: 10.1016/j.jgo.2022.08.004] [Medline: 35985929]
13. Fitch MI, Nicoll I, Lockwood G, Strohschein FJ, Newton L. Main challenges in survivorship transitions: perspectives of older adults with cancer. J Geriatr Oncol 2021 May;12(4):632-640. [doi: 10.1016/j.jgo.2020.09.024] [Medline: 33008768]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DHL: digital health literacy
DHT: digital health technology
HCP: health care professional

Edited by N Cahill; submitted 10.04.24; peer-reviewed by PH Liao, R Jin; revised version received 24.01.25; accepted 27.01.25; published 27.02.25.

Please cite as:
Newton L, Monkman H, Fullerton C
Exploring Older Adult Cancer Survivors' Digital Information Needs: Qualitative Pilot Study
JMIR Cancer 2025;11:e59391
URL: https://cancer.jmir.org/2025/1/e59391
doi:10.2196/59391

Analysis of the Relationship Between Rural-Urban Status and Use of Digital Health Technology Among Older Cancer Survivors Based on the Health Information National Trends Survey: Cross-Sectional Analysis

Samantha J Werts-Pelter[1], MPH, PhD; Zhao Chen[2], MPH, PhD; Jennifer W Bea[1], PhD; Amanda E Sokan[3], LLB, MHA, PhD; Cynthia A Thomson[1], RDN, PhD

[1]Department of Health Promotion Sciences, Mel and Enid Zuckerman College of Public Health, University of Arizona, 1295 N Martin Ave, Tucson, AZ, United States

[2]Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

[3]Department of Public Health Practice, Policy, & Translational Research, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

Corresponding Author:
Samantha J Werts-Pelter, MPH, PhD
Department of Health Promotion Sciences, Mel and Enid Zuckerman College of Public Health, University of Arizona, 1295 N Martin Ave, Tucson, AZ, United States

Abstract

Background: Though telehealth has been a promising avenue for engaging cancer survivors with health care and lifestyle programming, older and rural-dwelling cancer survivors may have additional challenges in accessing digital devices and tools that have not yet been described. This study aimed to use a robust, nationally representative sample collected in 2022 to provide an updated view of digital technology use and the use of technology for health in this population.

Objective: This study aimed to examine the prevalence of digital technology use for health-related activities among older cancer survivors in both rural and urban settings. The primary outcomes of interest included (1) internet access and use for health-related activities, (2) digital device ownership and use as a tool for health behaviors, (3) use of social media for health, and (4) use of telehealth.

Methods: A cross-sectional analysis of the National Cancer Institute's Health Information National Trends Survey Cycle 6 (HINTS 6) was completed to examine the prevalence of digital technology use among older cancer survivors. For analysis, the sample was restricted to cancer survivors over the age of 60 years (n=710). Unadjusted and adjusted logistic regression models were used to test the association between rurality and digital health tool use.

Results: Overall, 17% (125/710) of the sample lived in a rural area of the United States and the mean sample age was 73 (SD 8.2) years. Older cancer survivors, regardless of rural-urban status, reported a high prevalence of internet usage (n=553, 79.9%), digital device ownership (n=676, 94.9%), and social media use (n=448, 66.6%). In unadjusted models, rural survivors were less likely than urban survivors to report that they had used a health or wellness application in the previous year (odds ratio [OR] 0.56, 95% CI 0.32-0.97; *P*=.04). In adjusted models, rural survivors were more likely to report that they had shared personal health information on social media (OR 2.64, 95% CI 1.13-6.19; *P*=.03). There were no differences in the proportion of rural and urban respondents who reported receiving health services through telehealth in the previous year.

Conclusions: Regardless of the residential status, older cancer survivors report high internet and technology use for health-related activities. These results show promise for the feasibility of using digital technologies to implement supportive care and wellness programming with older cancer survivors.

(JMIR Cancer 2025;11:e66636) doi:10.2196/66636

KEYWORDS

cancer; non-metropolitan; disparities; digital divide; health research; aging; rural-urban; digital health technology; cross-sectional; health behaviors; mobile phone

Introduction

Improvement in cancer treatments and increasing life expectancy have led to a greater proportion of older, long-term survivors of cancer. As of 2022, 67% of cancer survivors in the United States were above the age of 65 years [1]. Though there are over 12 million older cancer survivors in the United States, their perspectives regarding survivorship and long-term care needs are not well described [1-4]. There are noted cancer care disparities among rural populations where the population tends to be older and to face a higher incidence and mortality rate from cancer than younger survivors. This is likely driven by barriers to accessing state-of-the-art cancer prevention, treatment, and survivorship services that support cancer-preventive behaviors [5-9].

Telehealth and other digital health tools offer an opportunity to bridge the gaps in care between rural and urban cancer survivors. Digital tools, such as the internet, electronic wearable devices, and social media, offer ways to disseminate health promotion materials and education remotely. Given the barriers that rural residents face in accessing in-person health behavior programming and resources, remote delivery with digital health tools is one strategy for delivering this information. In fact, a systematic review of lifestyle behavior change interventions for rural cancer survivors found that programming predominately relied on delivery using a hybrid or remote format by the use of digital technology, though the evidence of effectiveness was limited [10]. Remotely delivered programming to support psychosocial well-being and lifestyle behavior change has been shown to be effective for survivorship more generally, but the evidence of acceptability and usability for survivors living in rural areas is lacking [10]. While digital literacy and access disparities have long existed between rural and urban communities, commonly known as the digital divide [11-13], the COVID-19 pandemic has provided a push to address some of these barriers and has led to greater adoption of telehealth and other digital health tools [14-16]. For example, during the pandemic, providers pushed for the use of telehealth visits to complete routine check-ups with patients, especially immunocompromised cancer patients undergoing treatment. By providing information and instructions on how to complete telehealth visits remotely, many patients became familiar with how to access digital technology. Whether additional modifications in telemedicine and digital health delivery are needed to meet the supportive care needs of older survivors is not known. To address this gap in the literature, additional work is needed to understand the access and use of digital technology in this population.

Digital health technology is a promising tool to support lifestyle behavior change programming, but the acceptability and usability of digital technology for rural cancer survivor populations, particularly older survivors, has not been fully explored. This study aimed to use a robust survey-based dataset to (1) describe the use of digital technology among cancer survivors, (2) assess the use of telehealth and other digital technology to support health, and (3) examine the association between rurality and digital health tool use and access among older (>60 y) cancer survivors who responded to this nationally representative survey in 2022.

Methods

Study Design and Population

This cross-sectional analysis was derived from the National Cancer Institute (NCI) Health Information National Trends Survey (HINTS), a nationally representative sample of the adult, noninstitutionalized population in the United States. NCI has been collecting HINTS data every few years since 2003 to evaluate trends in health information access and attitudes toward digital health technology [17,18]. Detailed information about the methodology and publicly available deidentified datasets can be found on the NCI HINTS website [19]. This analysis was deemed exempt by the University of Arizona Institutional Review Board.

This analysis uses data from HINTS 6, which was collected between March and November 2022. As this was the first full HINTS survey completed after the beginning of the COVID-19 pandemic wherein telehealth use increased appreciably [15,20], this iteration was selected to best reflect the current prevalence of digital technology use. HINTS 6 is also the first to include stratification for residential status in the sampling strategy to ensure better representation of rural participants [21]. Out of the 6252 total respondents (response rate 28.1%), those without a history of cancer (n=4982), those under the age of 60 years (n=174), or those with missing data for cancer diagnosis or age (n=119) were excluded from analyses. Cancer diagnosis was determined with a self-report question asking if the respondent had ever been diagnosed with cancer. A cutoff of 60 years of age was used to define an “older cancer survivor” to maximize the analysis sample size and realizing that cancer survivors experience accelerated aging, meaning a 60-year-old cancer survivor may be more like a 65-year-old noncancer survivor. The final analytical sample included 710 cancer survivors 60 years of age or older. Most participants completed HINTS 6 using a mailed paper copy of the survey (435/710, 61.3%) compared with a web-based version of the survey (275/710, 38.7%).

Ethical Considerations

The HINTS 6 general population survey was designated “exempt research” and approved by the Westat Institutional Review Board (IRB) (Project # 6632.03.51). HINTS 6 was designated as “Not Human Subjects Research” from the National Institutes of Health Office of IRB Operations (iRIS reference number: 562715).

Outcome Variables

The primary measures of interest pertained to (1) internet access and use for health-related activities, (2) digital device ownership and use as a tool for health behaviors, (3) use of social media for health, and (4) use of telehealth.

Internet

Internet use was assessed with the question, “Do you ever go on-line to access the internet or World Wide Web, or to send and receive email?” Respondents who indicated “yes” were

then provided a follow-up question asking how they have used the internet for health-related needs in the previous 12 months with 4 statements to consider: "Look for health or medical information," "Send a message to a health care provider or a health care provider's office," "View medical test results," and "Make an appointment with a health care provider." Respondents answered each statement with a "yes" or "no" response. Those who reported using the internet were also asked about their level of satisfaction with their home internet connection using a 5-point scale ranging from "extremely satisfied" to "not at all satisfied."

Digital Devices

Survey respondents were asked if they own any of the following digital devices: tablet computer (eg, Apple iPad, Samsung Galaxy, or Kindle Fire), a smartphone (eg, Apple iPhone, Blackberry, or Android), or a basic mobile phone. These responses were then categorized to describe participants who owned these devices, those who owned multiple devices, and those who owned no digital devices. Those who indicated they own a tablet or smart phone were asked, "In the past 12 months, have you used a health or wellness app on your tablet or smartphone?" Ownership and use of wearable devices (eg, Fitbit, Apple Watch, or Garmin Vivofit) was assessed with the question, "In the past 12 months, have you used an electronic wearable device to monitor or track your health or activity?" To assess how respondents have used digital devices to support their health, participants were asked if they had shared any health information from either their wearable device or smartphone with a health professional in the previous year.

Social Media

Social media use was assessed with the question, "In the past 12 months, how often did you do the following?" The 5 statements included: "Visited a social media site," "Shared personal health information on social media," "Shared general health information on social media," "Interacted with people who have similar health or medical issues on social media or online forums," and "Watched a health-related video on a social media site (eg, YouTube)." The response options for each included 5 categorical frequency of use options ranging from "Never" to "Almost every day." These response types were dichotomized to capture if the respondent had used social media for the purpose at all in the last 12 months or not at all.

Telehealth

Telehealth use was assessed with the question, "In the past 12 months, did you receive care from a doctor or health professional using telehealth?" Those who indicated they had not used telehealth were then asked if telehealth had been offered by their provider if they had tried to schedule medical care. Those who did choose to use telehealth in the previous year were prompted with a set of statements and asked to indicate if they agreed that the statement reflected a reason they participated in a telehealth visit.

Exposure Variables

The exposure of interest for this analysis was residential status. This variable was dichotomized into rural and urban residency using the 2013 Rural-Urban Continuum (RUC) Codes set by the US Department of Agriculture Economic Research Service [22]. This code includes 9-values and is derived along a continuum based on population size and adjacency to a metropolitan area. Based on previous analyses, urban residencies were defined as RUC codes 1 - 3, while rural residencies were defined as RUC codes of 4 - 9 [11,23,24].

Sample Characteristics

The sociodemographic characteristics included in this analysis were age, sex, race and ethnicity, occupational status, annual household income, household size, marital status, education, and census region. To ensure adequate cell sample size for analysis, the race and ethnicity variable was condensed into non-Hispanic White, non-Hispanic Black or African American, Hispanic, and non-Hispanic other. Occupational status was organized into 4 categories including employed, retired, disabled, and other (ie, unemployed, homemaker, and student). Related to cancer history, the age of diagnosis, years since diagnosis, and cancer type were included in the summary of the sample.

Statistical Analysis

This is a cross-sectional analysis of the use of digital health tools and telehealth in older cancer survivors who completed the HINTS 6 survey collected in 2022. Provided survey weights were applied using the Jackknife repeated replication method for population level estimates.

Sociodemographic characteristics, health behaviors, and cancer history were stratified by residential status (rural vs urban) and differences were assessed using Wald design-based chi-square tests of independence or *t* tests. Rurality and associations with using digital health tools and telehealth were individually assessed using both unadjusted and adjusted multiple logistic regression models. Models were adjusted for age, race and ethnicity, annual household income, and education. Sex and marital status were not found to be confounders or effect modifiers and were thus not included in the adjusted model. An α level of .05 was considered statistically significant. A post hoc sensitivity analysis was conducted excluding respondents who reported diagnosis with nonmelanoma skin cancer as this is a population of people with a cancer diagnosis that experience very different treatment and survival outcomes from the general survivor population. All analyses were completed using STATA 17 (StataCorp LLC).

Results

Demographics

Out of the 710 older cancer survivors included in this analysis, 17% (n=125) were living in rural areas (see Table 1 for detailed sample characteristics). The average age of the respondents was 73 (SD 8.2) years. While most demographic characteristics did not differ between rural and urban survivors, rural cancer survivors were more likely to identify as non-Hispanic White (93% [n=95] vs 84% [n=415]; *P*=.007) and reported a lower annual household income (*P*=.01). When considering the US census region, a greater percentage of Midwest participants were living in rural areas compared to urban areas (37.3% vs 14.4%), while a lower percentage of participants in the West

and Northeast regions were living in rural areas compared to urban areas (*P*<.001). Most respondents were retired (n=497, 71.7%) and married or living with a romantic partner (n=356, 64.7%). Though there were no differences in age at diagnosis or years since diagnosis between groups, nonmelanoma skin cancer was the most prevalent cancer type (n=149, 31.2%); rural participants reported greater prevalence of breast and gynecological cancer than urban residents (*P*=.03). There were no differences in the survey form version completed between rural and urban residents, with 59.9% (n=350) of urban respondents and 68% (n=85) of rural respondents using a paper version of the survey (*P*=.09).

Table . Sociodemographic characteristics of cancer survivors participating in Health Information National Trends Survey 6 (HINTS 6) by residential status. Missing data <10%.

Characteristics	Total	Rural	Urban	*P* value
Participants, n (%)	710 (100)	125 (17.4)	585 (82.6)	
Sex, n (weighted %)				.94
Male	316 (46)	58 (46.5)	258 (45.9)	
Female	393 (54.1)	67 (53.5)	325 (54.1)	
Age, years, mean (SD)	73.0 (8.2)	72.3 (7.3)	73.2 (8.3)	.29
Race and ethnicity, n (weighted %)				.007
Non-Hispanic White	510 (85.6)	95 (92.8)	415 (84.1)	
Non-Hispanic Black or African American	73 (5.7)	7 (2.2)	66 (6.5)	
Hispanic	47 (5.8)	2 (0.9)	45 (6.7)	
Non-Hispanic Other	27 (2.9)	6 (4.2)	21 (2.7)	
Region, n (weighted %)				<.001
Northeast	92 (18.6)	12 (9.2)	80 (20.6)	
Midwest	114 (18.4)	36 (37.3)	78 (14.4)	
South	344 (42.3)	64 (44.9)	280 (41.8)	
West	160 (20.7)	13 (8.6)	147 (23.2)	
Education, n (weighted %)				.07
Less than high school	39 (5)	8 (5.5)	31 (4.9)	
High school graduate	130 (21.5)	32 (29.4)	98 (19.8)	
Some college	212 (43.1)	43 (44.4)	169 (42.8)	
College graduate or more	324 (30.5)	41 (20.8)	283 (32.5)	
Occupational status, n (weighted %)				.28
Employed	138 (20.4)	18 (14.7)	120 (21.6)	
Retired	497 (71.7)	94 (78.9)	403 (70.2)	
Disabled	32 (3.3)	6 (1.8)	26 (3.6)	
Other	35 (4.6)	6 (4.6)	29 (4.6)	
Annual household income, mean (SD)				.01
Less than $20,000	99 (8.6)	25 (11.8)	74 (7.9)	
$20,000 to <$35,000	92 (13)	25 (19.1)	67 (11.7)	
$35,000 to <$50,000	93 (14.8)	13 (13.9)	80 (15)	
$50,000 to <$75,000	128 (23.6)	25 (32)	103 (21.9)	
$75,000 or more	236 (40)	26 (23.2)	210 (43.4)	
Marital status, n (weighted %)				.70
Married or living with a romantic partner	356 (64.7)	60 (69.7)	289 (63.8)	
Divorced or separated	139 (10.4)	26 (9.7)	113 (10.6)	
Widowed	152 (15)	28 (15.4)	124 (15)	
Single, never married	59 (9.8)	4 (5.5)	55 (10.7)	
Age diagnosed, years	57.5 (14.7)	57.5 (14.8)	57.6 (14.7)	.94
Years since diagnosis, mean (SD)				.47

Characteristics	Total	Rural	Urban	*P* value
<1	61 (8.3)	11 (9.1)	50 (8.1)	
2 - 5	137 (20.2)	29 (22.9)	108 (19.7)	
6 - 10	127 (18.6)	21 (13.1)	106 (19.7)	
11+	371 (52.9)	62 (55)	309 (52.5)	
Cancer type, mean (SD)				.03
Breast	99 (16.5)	21 (24.7)	78 (14.8)	
Gynecological	51 (6.6)	14 (12.2)	37 (5.4)	
Colorectal	35 (5.4)	2 (2.1)	33 (6.1)	
Prostate	78 (14.2)	15 (15.9)	63 (13.9)	
Blood	32 (4.2)	3 (2.5)	29 (4.5)	
Skin, nonmelanoma	149 (31.2)	28 (33.5)	121 (30.8)	
Melanoma	36 (8.9)	5 (3.9)	31 (10)	
Other	77 (12.9)	7 (5.1)	70 (14.6)	

There were no differences in internet use or mode of access, digital device ownership, or social media use in the previous 12 months between older rural and urban survivors (Table 2). Overall, most older survivors were using the internet (n=553, 79.9%) and predominately used a high-speed service to connect (n=486, 89.8%). Though they shared the same prevalence of internet use, rural cancer survivors were more likely to report a lack of satisfaction with their internet connection than urban survivors (8.7%, n=7, rural vs 0.4%, n=3, urban; *P*<.001). Regardless of residential status, older survivors reported high rate of smartphone ownership (n=544, 78.5%) and only 5% (n=31) reported not owning any digital devices. About half of older survivors owned more than 1 digital device (n=345, 51.7%).

Table . Weighted prevalence of internet use, digital device ownership, and social media access in the past 12 months by residential status.

	Total, %	Rural, %	Urban, %	*P* value
Used the internet at all	79.9	78.1	80.1	.67
Mode of accessing the internet [a]				
Dial-up or telephone line	1.9	2	1.9	.95
High-speed service	89.8	84.7	90.8	.29
Cellular network	69.5	74.1	68.5	.45
Internet connection satisfaction				.006
Extremely satisfied	16.4	4.6	18.8	
Very satisfied	47.4	59.2	45	
Somewhat satisfied	28.8	22.2	30.1	
Not very satisfied	5.6	5.3	5.7	
Not at all satisfied	1.8	8.7	0.4	
Digital device ownership [a]				
58.9	48.6	61.1	.12	
78.5	76.1	79	.58	
9.2	13.6	8.2	.29	
51.7	41.4	53.9	.09	
5.1	3.2	5.5	.3	
Visited a social media site	66.6	71.6	65.6	.28

[a]Categories are not mutually exclusive.

Internet

In both unadjusted and adjusted logistic regression models, the use of the internet to support health did not differ between older rural and older urban cancer survivors. Regardless of residential status, most respondents who used the internet within the past 12 months indicated that they have used the internet to look for health information (n=463, 86.5%), they have sent a message

to their health care provider (n=362, 67.5%), and they have viewed their medical test results (n=405, 78.9%; Table 3).

Table . Association of rural versus urban residence and use of digital health tools in the previous 12 months.

Digital health tools	Weighted percent, %		Unadjusted OR (95% CI)	*P* value	Adjusted OR[a] (95% CI)	*P* value
	Rural	Urban				
Internet						
Used the internet to look for health or medical information	78.7	88.1	0.50 (0.22-1.11)	.09	0.73 (0.33-1.62)	.44
Used the internet to send a message to a health care provider or health care providers office	63.6	68.4	0.81 (0.40-1.61)	.54	0.80 (0.33-1.95)	.62
Used the internet to view medical test results	69.5	80.8	0.54 (0.26-1.11)	.1	0.63 (0.28-1.44)	.27
Used the internet to make an appointment with a health care provider	42.7	58.4	0.53 (0.27-1.07)	.07	0.52 (0.22-1.21)	.13
Digital devices						
Used a health or wellness app on a tablet or smartphone	36.9	51.3	0.56 (0.32-0.97)	.04	0.85 (0.47-1.53)	.58
Used an electronic wearable device to monitor or track health or activity	19.8	28	0.64 (0.33-.21)	.17	0.81 (0.38-1.73)	.58
Shared health information from an electronic monitoring device or smartphone with a health professional	21.8	24.4	0.87 (0.46-.63)	.65	1.25 (0.56-2.78)	.58
Social media						
Shared personal health information on social media	19	10.9	1.92 (0.82-4.51)	.13	2.64 (1.13- 6.19)	.03
Shared general health-related information on social media (ie, news article)	21.9	26.3	0.79 (0.35-1.79)	.56	0.55 (0.24-1.23)	.14
Interacted with people with similar health or medical issues on social media or online forums	19.7	16.5	1.24 (0.50-3.07)	.63	1.08 (0.46-2.58)	.85
Watched a health-related video on a social media site	40.2	42.3	0.91 (0.48-1.72)	.78	0.76 (0.37-1.55)	.44

[a]Logistic regression models were adjusted for age, race and ethnicity, annual household income, and level of education.

Digital Devices

Compared with urban survivors, rural respondents who owned a smartphone or tablet were less likely to report that they had used a health or wellness application in the previous year (36.9%, n=36, rural vs 51.3%, n=234, urban; unadjusted OR 0.56, 95% CI 0.32-0.97; *P*=.04). This difference was no longer significant in adjusted logistic regression models when age, annual household income, education, and race and ethnicity were considered (adjusted OR 0.85, 95% CI 0.47-1.53; *P*=.58). There were no differences in the use of wearable devices to track activity or to share health information with a health care provider. Overall, 26.5% (n=165) of older survivors reported using a wearable device to track their health or activity. While only 23.9% (n=166) of older survivors indicated they have shared data from a smartphone or wearable device with a health professional in the previous year, 81.3% (n=135) indicated that they would be willing to do so in the future.

Social Media

Two-thirds of older adults (n=448, 66.6%) visited a social media site in the previous year. Though social media use was similar for both urban and rural survivors, older cancer survivors living in rural areas were twice as likely to report that they had shared personal health information on social media in the previous year (adjusted OR 2.64, 95% CI 1.13-6.19; *P*=.03). Less than 20% (n=115) of older cancer survivors reported having used social media or a chat forum to interact with people who have similar health issues and 25.5% (n=173) indicated that they had shared general health information on social media in the previous year.

Telehealth

There were no differences in the proportion of rural and urban respondents who indicated they had received care from their health care provider using telehealth in the previous 12 months (Table 4). Rural and urban cancer survivors were equally likely to have been offered the option to have a telehealth visit by their providers. For respondents who indicated they received care using telehealth, the primary reason for choosing to participate in a telehealth visit was provider recommendation or requirement (n=215, 81.1%). A total of 54% (n=136) of older cancer survivors indicated that they chose to participate in telehealth because it was more convenient than going to a health care office. Compared with older urban survivors, rural survivors were less likely to indicate that one of their reasons for choosing telehealth was to seek advice about whether in-person care was needed (adjusted OR 0.21, 95% CI 0.05-0.94; *P*=.04).

Table . Association of rural versus urban residence with use and reasons for use of telehealth in the previous 12 months.

Statement	Weighted percent (%)		Unadjusted OR (95% CI)	*P* value	Adjusted OR[a] (95% CI)	*P* value
	Rural	Urban				
Received care from a doctor or health professional using telehealth	33.5	42.6	0.68 (0.40-1.15)	.15	0.79 (0.41-1.54)	.49
Offered the option to have a telehealth visit for any medical care	40.6	48.2	0.74 (0.36-1.49)	.39	0.87 (0.41-1.86)	.71
Reported technical problems with the telehealth visits	37.1	26.3	1.65 (0.64-4.24)	.29	1.80 (0.52-6.23)	.35
Reasons for choosing to participate in telehealth						
Health care provider recommended or required the visit use telehealth	74.6	82.1	0.64 (0.18-2.32)	.49	0.86 (0.12-6.07)	.87
Wanted advice about whether in-person medical care was needed	10.8	26.4	0.34 (0.12-1.00)	.05	0.21 (0.05-0.94)	.04
Wanted to avoid possible infection at the doctor's office or hospital	30.4	48.9	0.46 (0.19-1.10)	.08	0.43 (0.13-1.46)	.17
More convenient than going to the doctor	59.9	53.4	1.31 (0.52-3.29)	.56	1.94 (0.63-5.94)	.24
Could include family or other caregivers in the appointment	19.1	23.2	0.78 (0.31-1.98)	.59	0.44 (0.10-1.90)	.26

[a]Logistic regression models were adjusted for age, race and ethnicity, annual household income, and level of education.

Sensitivity Analysis

Exclusion of the 149 respondents diagnosed with nonmelanoma skin cancer did not materially change the study results (see Table S1 and S2 in Multimedia Appendix 1 and Multimedia Appendix 2, respectively). In general, no associations were identified between residential status and digital tool use in this smaller sample. There were significant differences in digital device ownership, with rural survivors being less likely to own a tablet computer (37.9%, n=37 vs 56.7%, n=265; *P*=.02) and to own multiple digital devices (28.7%, n=28 vs 51.5%, n=233; *P*=.003; Table S1 in Multimedia Appendix 1).

Discussion

Principal Findings

This study is among the first to examine the use of technology among older cancer survivors with a specific focus on the use of technology that became more common overall after the start of the COVID-19 pandemic. While findings from iterations of HINTS completed before the start of the pandemic indicated that rural survivors were less likely to access the internet than urban survivors [25-27], digital health tool access and use were similar between groups in this post–COVID-19 analysis. Regardless of residential status, older survivors of cancer reported a high prevalence of internet use (n=553, 79.9%), digital device ownership (n=676, 94.9%), and social media use (n=448, 66.6%). These post–COVID-19 pandemic prevalence results are similar to those found in an analysis of older cancer survivors who responded to the National Health and Aging Trends Study. In that analysis, a rise in digital health technology use was seen after the pandemic (52% in 2021), compared with before (45% in 2019), though they did not compare rural and urban populations [28]. As this is the first analysis to compare digital technology use between rural and urban older survivors after the pandemic, our results indicate that the pandemic may have enhanced uptake of digital technology across residential areas, potentially reducing the digital divide between rural and urban survivors.

Comparison With Previous Work

One of the differences noted between groups in this sample was that rural survivors, as compared with urban survivors, were

more likely to report that they had shared personal health information on social media in the previous year. There are several hypotheses for why this difference may arise. First, rural survivors experience greater barriers to accessing in-person social support, such as travel distance and access to transportation [5,29,30], which may leave phone or computer-based support as the more feasible option. Second, rural survivors tend to experience greater symptoms of anxiety and depression and poorer health-related quality of life than urban survivors, yet have limited access to mental health professionals [5,8,31,32]. Sharing on social media may be an avenue to garner support, elicit shared experience, or share positive outcomes. There is limited information available on rural cancer survivorship and social media use. As follow-up information was not collected in the HINTS survey about what type of personal health information was shared or who it was shared with, future research should explore how rural survivors use social media and how it is used as a tool for social support or information gathering.

The access and usability results from this study show promise for remote treatment and care for older cancer survivors, particularly those living in rural areas. However, adoption of wearable devices was modest (n=165, 26.5%) and there remains a lack of evidence exploring the barriers to digital technology adoption for those who do not access these tools [33,34]. Wearable devices are one tool to promote self-monitoring of healthy lifestyle behaviors that may be especially beneficial for rural survivors who have do not have access to in-person health coaching. With only 1 in 4 older cancer survivors using a wearable device, additional work is justified to explore the acceptability of wearable devices and to classify the barriers to use. In addition, while these results show promise for increasing acceptability and usability of telemedicine for health care, there is still a question of whether remote delivery of survivorship care and lifestyle behavior change programming is acceptable, feasible, and efficacious for older, rural cancer survivors [10,35]. A recently completed pilot trial examined this question using implementation of a remotely delivered, evidence-based group exercise program for older cancer survivors living in rural areas. This trial, the tele-EnhanceFitness program, incorporated remotely monitored Zoom (Zoom Video Communications) fitness classes 3 days a week for 16 weeks and found low attrition (5%) and high class attendance rate (87%) [36]. Additional research has identified that rural survivors are interested in the incorporation of remote lifestyle programs to their care, but additional study is needed to assess the efficacy of remote programming and barriers to implementation [10,35,37,38].

Strengths and Limitations

A strength of this study is the use of the HINTS dataset. HINTS is a nationally representative survey, which includes rigorous probability sampling of the US population. The HINTS 6 sampling strategy also introduced stratification for residential status to ensure better representation of rural participants [21]. In addition, the jackknife weighting strategy used for HINTS data analysis allows for population-level comparisons and estimates. Participants used a paper copy to complete the HINTS 6 survey regardless of rural-urban status, limiting potential bias that may have arisen from digital collection of the survey. As with all cross-sectional studies, a limitation of this analysis is the inability to determine cause and effect relationship. The sample size included for analysis was small given the exclusion of any respondents without a cancer history and relied on self-report, which may be affected by recall or response bias. It also lacked diversity reflecting the larger population of cancer survivors, limiting the ability to stratify analysis by race and ethnicity. To maximize sample size and to report results for the overall survivor population, cancer survivors of nonmelanoma skin cancer were included in analysis. This may have introduced confounding by favorable diagnosis, as survivors of nonmelanoma skin cancer generally do not receive chemotherapy or radiation. Another limitation is that the questions regarding digital health technology were general with limited follow-up about satisfaction with the technology or desire for future use. Finally, all data were self-reported, although bias related to reporting of social media use has not been previously reported.

Future Directions

While digital technology use was found to be similar between older urban and rural survivors, additional research is needed to explore the barriers and facilitators to digital technology adoption for this population and the acceptability of using digital tools for remote intervention delivery. Quantitative work is warranted to examine patterns of digital technology use over time to determine any trends that have emerged since the COVID-19 pandemic. Future work should aim to expand the generalizability by recruiting a diverse sample that better reflects the overall population of cancer survivors living in rural areas.

Conclusion

These findings provide valuable insight into the acceptability and usability of digital health technology for older cancer survivors. Regardless of rural-urban status, digital health technology use was found to be high among cancer survivors. This is the first analysis of digital health technology use among rural and urban residents after the start of the pandemic, indicating the digital divide may be narrowing as use and access to technology changes over time. For cancer survivors, these results indicate digital technology is a feasible method for delivering health information. Implementation of digital technology-based survivorship and lifestyle programming shows promise as a feasible solution to overcome barriers to high-quality cancer care for older and rural-dwelling survivors, yet additional work in this area is needed.

Acknowledgments

Health Information National Trends Survey (HINTS) is funded by the National Cancer Institute with support from the Health Communication and Informatics Research Branch of the Division of Cancer Control and Population Sciences. This work was

supported by the University of Arizona Cancer Center Support Grant (P30 CA023074) and by the Mel and Enid Zuckerman College of Public Health Dean's Student Dissertation Fund.

Disclaimer

Generative AI was not used in any portion of the manuscript writing process.

Data Availability

The datasets generated and analyzed for this study are publicly available in [19].

Authors' Contributions

SJW-P and CAT contibuted to conceptualization. SJW-P contributed to data curation. SJW-P, ZC, and CAT contributed to methodology. SJW-P and ZC performed formal analysis. SJW-P manged project administration. ZC, JWB, AES, and CAT performed supervision. SJW-P contributed to writing–original draft. SJW-P, ZC, JWB, AES, and CAT contibuted to writing–review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Weighted prevalence of internet use, digital device ownership, and social media access in the past 12 months by residential status, excluding those with nonmelanoma skin cancer.
[DOCX File, 18 KB - cancer_v11i1e66636_app1.docx]

Multimedia Appendix 2
Association of rural versus urban residence and use of digital health tools in the previous 12 months, excluding those with nonmelanoma skin cancer.
[DOCX File, 20 KB - cancer_v11i1e66636_app2.docx]

References

1. Miller KD, Nogueira L, Devasia T, et al. Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin 2022 Sep;72(5):409-436. [doi: 10.3322/caac.21731] [Medline: 35736631]
2. Fitch MI, Nicoll I, Lockwood G, Strohschein FJ, Newton L. Main challenges in survivorship transitions: Perspectives of older adults with cancer. J Geriatr Oncol 2021 May;12(4):632-640. [doi: 10.1016/j.jgo.2020.09.024] [Medline: 33008768]
3. Guerard EJ, Nightingale G, Bellizzi K, et al. Survivorship care for older adults with cancer: U13 conference report. J Geriatr Oncol 2016 Jul;7(4):305-312. [doi: 10.1016/j.jgo.2016.06.005] [Medline: 27424802]
4. Fletcher C, Flight I, Chapman J, Fennell K, Wilson C. The information needs of adult cancer survivors across the cancer continuum: A scoping review. Patient Educ Couns 2017 Mar;100(3):383-410. [doi: 10.1016/j.pec.2016.10.008] [Medline: 27765377]
5. Bhatia S, Landier W, Paskett ED, et al. Rural-urban disparities in cancer outcomes: opportunities for future research. J Natl Cancer Inst 2022 Jul 11;114(7):940-952. [doi: 10.1093/jnci/djac030] [Medline: 35148389]
6. Blake KD, Moss JL, Gaysynsky A, Srinivasan S, Croyle RT. Making the case for investment in rural cancer control: an analysis of rural cancer incidence, mortality, and funding trends. Cancer Epidemiol Biomarkers Prev 2017 Jul;26(7):992-997. [doi: 10.1158/1055-9965.EPI-17-0092] [Medline: 28600296]
7. Hirko KA, Xu H, Rogers LQ, et al. Cancer disparities in the context of rurality: risk factors and screening across various U.S. rural classification codes. Cancer Causes Control 2022 Aug;33(8):1095-1105. [doi: 10.1007/s10552-022-01599-2] [Medline: 35773504]
8. Moss JL, Pinto CN, Mama SK, et al. Rural-urban differences in health-related quality of life: patterns for cancer survivors compared to other older adults. Qual Life Res 2021 Apr;30(4):1131-1143. [doi: 10.1007/s11136-020-02683-3] [Medline: 33136241]
9. Weaver KE, Palmer N, Lu L, Case LD, Geiger AM. Rural-urban differences in health behaviors and implications for health status among US cancer survivors. Cancer Causes Control 2013 Aug;24(8):1481-1490. [doi: 10.1007/s10552-013-0225-x] [Medline: 23677333]
10. Werts SJ, Robles-Morales R, Bea JW, Thomson CA. Characterization and efficacy of lifestyle behavior change interventions among adult rural cancer survivors: a systematic review. J Cancer Surviv 2023 Sep 28. [doi: 10.1007/s11764-023-01464-4] [Medline: 37768420]

11. Okobi E, Adigun AO, Ozobokeme OE, et al. Examining disparities in ownership and use of digital health technology between rural and urban adults in the US: an analysis of the 2019 Health Information National Trends Survey. Cureus 2023 May;15(5):e38417. [doi: 10.7759/cureus.38417] [Medline: 37273368]
12. Greenberg AJ, Haney D, Blake KD, Moser RP, Hesse BW. Differences in access to and use of electronic personal health information between rural and urban residents in the United States. J Rural Health 2018 Feb;34 Suppl 1(Suppl 1):s30-s38. [doi: 10.1111/jrh.12228] [Medline: 28075508]
13. Hindman DB. The rural-urban digital divide. Journal Mass Commun Q 2000 Sep;77(3):549-560. [doi: 10.1177/107769900007700306]
14. Greenberg-Worisek AJ, Kurani S, Finney Rutten LJ, Blake KD, Moser RP, Hesse BW. Tracking healthy people 2020 internet, broadband, and mobile device access goals: an update using data from the Health Information National Trends Survey. J Med Internet Res 2019 Jun 24;21(6):e13300. [doi: 10.2196/13300] [Medline: 31237238]
15. Murphy A, Kirby A, Lawlor A, Drummond FJ, Heavin C. Mitigating the impact of the COVID-19 pandemic on adult cancer patients through telehealth adoption: a systematic review. Sensors (Basel) 2022 May 9;22(9):3598. [doi: 10.3390/s22093598] [Medline: 35591287]
16. Mackwood M, Butcher R, Vaclavik D, et al. Adoption of telemedicine in a rural US cancer center amid the COVID-19 pandemic: qualitative study. JMIR Cancer 2022 Aug 16;8(3):e33768. [doi: 10.2196/33768] [Medline: 35895904]
17. Hesse BW, Moser RP, Rutten LJF, Kreps GL. The health information national trends survey: research from the baseline. J Health Commun 2006;11 Suppl 1:vii-xvi. [doi: 10.1080/10810730600692553] [Medline: 16641070]
18. Nelson DE, Kreps GL, Hesse BW, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004;9(5):443-460. [doi: 10.1080/10810730490504233] [Medline: 15513791]
19. National Cancer Institute. Health Information National Trends Survey. URL: https://hints.cancer.gov [accessed 2023-12-11]
20. Americans' use of mobile technology and home broadband. : Pew Research Center; 2024 Jan 31 URL: https://www.pewresearch.org/internet/2024/01/31/americans-use-of-mobile-technology-and-home-broadband/ [accessed 2025-01-28]
21. Health Information National Trends Survey 6 (HINTS 6) methodology report. : WeStat; 2023 URL: https://hints.cancer.gov/docs/methodologyreports/HINTS_6_MethodologyReport.pdf [accessed 2025-01-28]
22. US Department of Agriculture. Rural-urban continuum codes. Economic Research Service. 2024. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx#.U0VBhleG-Hs [accessed 2024-01-31]
23. Zahnd WE, Davis MM, Rotter JS, et al. Rural-urban differences in financial burden among cancer survivors: an analysis of a nationally representative survey. Support Care Cancer 2019 Dec;27(12):4779-4786. [doi: 10.1007/s00520-019-04742-z] [Medline: 30972645]
24. Robertson MC, Song J, Taylor WC, Durand CP, Basen-Engquist KM. Urban-rural differences in aerobic physical activity, muscle strengthening exercise, and screen-time sedentary behavior. J Rural Health 2018 Sep;34(4):401-410. [doi: 10.1111/jrh.12295] [Medline: 29451333]
25. Chou WYS, Liu B, Post S, Hesse B. Health-related internet use among cancer survivors: data from the Health Information National Trends Survey, 2003-2008. J Cancer Surviv 2011 Sep;5(3):263-270. [doi: 10.1007/s11764-011-0179-5] [Medline: 21505861]
26. Fareed N, Swoboda CM, Jonnalagadda P, Huerta TR. Persistent digital divide in health-related internet use among cancer survivors: findings from the Health Information National Trends Survey, 2003-2018. J Cancer Surviv 2021 Feb;15(1):87-98. [doi: 10.1007/s11764-020-00913-8] [Medline: 32671557]
27. Greenberg-Worisek A, Ferede L, Balls-Berry J, et al. Differences in electronic personal health information tool use between rural and urban cancer patients in the United States: secondary data analysis. JMIR Cancer 2020 Aug 10;6(2):e17352. [doi: 10.2196/17352] [Medline: 32773369]
28. Zhou W, Cho Y, Shang S, Jiang Y. Use of digital health technology among older adults with cancer in the United States: findings from a National Longitudinal Cohort Study (2015-2021). J Med Internet Res 2023 May 31;25:e46721. [doi: 10.2196/46721] [Medline: 37256672]
29. Prochaska JJ, Coughlin SS, Lyons EJ. Social media and mobile technology for cancer prevention and treatment. Am Soc Clin Oncol Educ Book 2017;37:128-137. [doi: 10.1200/EDBK_173841] [Medline: 28561647]
30. Yli-Uotila T, Rantanen A, Suominen T. Motives of cancer patients for using the internet to seek social support. Eur J Cancer Care (Engl) 2013 Mar;22(2):261-271. [doi: 10.1111/ecc.12025] [Medline: 23320398]
31. Charlton M, Schlichting J, Chioreso C, Ward M, Vikas P. Challenges of rural cancer care in the United States. Oncology (Williston Park) 2015 Sep;29(9):633-640. [Medline: 26384798]
32. Burris JL, Andrykowski M. Disparities in mental health between rural and nonrural cancer survivors: a preliminary study. Psychooncology 2010 Jun;19(6):637-645. [doi: 10.1002/pon.1600] [Medline: 19582800]
33. Pang NQ, Lau J, Fong SY, Wong CYH, Tan KK. Telemedicine acceptance among older adult patients with cancer: scoping review. J Med Internet Res 2022 Mar 29;24(3):e28724. [doi: 10.2196/28724] [Medline: 35348462]
34. Sinha Gregory N, Shukla AP, Noel JJ, et al. The feasibility, acceptability, and usability of telehealth visits. Front Med (Lausanne) 2023;10:1198096. [doi: 10.3389/fmed.2023.1198096] [Medline: 37538312]
35. Blackwood J, Rybicki K. Outcomes of telehealth-delivered physical activity programs in adult cancer survivors: a systematic review. Rehabil Oncol 2021;39(3):128-136. [doi: 10.1097/01.REO.0000000000000249]

36. Gell NM, Dittus K, Caefer J, Martin A, Bae M, Patel KV. Remotely delivered exercise to older rural cancer survivors: a randomized controlled pilot trial. J Cancer Surviv 2024 Apr;18(2):596-605. [doi: 10.1007/s11764-022-01292-y] [Medline: 36374436]
37. Salerno EA, Gao R, Fanning J, et al. Designing home-based physical activity programs for rural cancer survivors: A survey of technology access and preferences. Front Oncol 2023;13. [doi: 10.3389/fonc.2023.1061641] [Medline: 36761969]
38. Lopez AM, Lam K, Thota R. Barriers and facilitators to telemedicine: can you hear me now? Am Soc Clin Oncol Educ Book 2021 Mar;41:25-36. [doi: 10.1200/EDBK_320827] [Medline: 34010056]

Abbreviations

HINTS: Health Information National Trends Survey
IRB: Institutional Review Board
NCI: National Cancer Institute
OR: odds ratio
RUCC: Rural-Urban Continuum Codes

Edited by N Cahill; submitted 18.09.24; peer-reviewed by E Salerno, SK Sharma; revised version received 15.01.25; accepted 16.01.25; published 04.03.25.

Please cite as:
Werts-Pelter SJ, Chen Z, Bea JW, Sokan AE, Thomson CA
Analysis of the Relationship Between Rural-Urban Status and Use of Digital Health Technology Among Older Cancer Survivors Based on the Health Information National Trends Survey: Cross-Sectional Analysis
JMIR Cancer 2025;11:e66636
URL: https://cancer.jmir.org/2025/1/e66636
doi:10.2196/66636

Exploring the Impact of the Multimodal CAPABLE eHealth Intervention on Health-Related Quality of Life in Patients With Melanoma Undergoing Immune-Checkpoint Inhibition: Prospective Pilot Study

Itske Fraterman[1], MSc; Lucia Sacchi[2], PhD; Henk Mallo[3], MSc; Valentina Tibollo[4], MSc; Savannah Lucia Catherina Glaser[5], MSc; Stephanie Medlock[5,6], PhD; Ronald Cornet[5,6], PhD; Matteo Gabetta[2], PhD; Vitali Hisko[7], BA; Vadzim Khadakou[7], BA; Ella Barkan[8], MA; Laura Del Campo[9], PhD; David Glasspool[10], PhD; Alexandra Kogan[11], MSc; Giordano Lanzola[2], PhD; Roy Leizer[11], MSc; Manuel Ottaviano[12], PhD; Mor Peleg[11], PhD; Konrad Śniatała[13], MSc; Aneta Lisowska[13], PhD; Szymon Wilk[13], PhD; Enea Parimbelli[2], PhD; Silvana Quaglini[2], PhD; Mimma Rizzo[14], MD; Laura Deborah Locati[15,16], MD; Annelies Boekhout[1], PhD; Lonneke V van de Poll-Franse[1,17], PhD; Sofie Wilgenhof[3], MD, PhD

[1]Department of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam, Netherlands

[10]Deontics Ltd, London, United Kingdom

[11]Department of Information Systems, University of Haifa, Haifa, Israel

[12]Life Supporting Technologies, Universidad Politécnica de Madrid, Madrid, Spain

[13]Institute of Computing Science, Poznan University of Technology, Poznan, Poland

[14]Division of Medical Oncology, Azienda Ospedaliero Universitaria Consorziale Policlinico di Bari, Bari, Italy

[15]Department of Internal Medicine and Medical Therapy, University of Pavia, Pavia, Italy

[16]Medical Oncology Unit, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy

[17]Department of Medical and Clinical Psychology, Center of Research on Psychological and Somatic Disorders (CoRPS), Tilburg University, Tilburg, Netherlands

[2]Department of Electric, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

[3]Department of Medical Oncology, Antoni van Leeuwenhoek, Amsterdam, Netherlands

[4]Laboratory of Informatics and Systems Engineering for Clinical Research, Istituti Clinici Scientifici Maugeri SpA SB IRCCS, Pavia, Italy

[5]Medical Informatics, Amsterdam UMC - University of Amsterdam, Amsterdam, Netherlands

[6]Methodology and Digital Health, Amsterdam Public Health, Amsterdam, Netherlands

[7]BITSENS JSC, Vilnius, Lithuania

[8]Department of Artificial Intelligence for Accelerated Healthcare and Life Sciences Discovery, IBM Research, University of Haifa, Haifa, Israel

[9]Associazione Italiana Malati di Cancro, Rome, Italy

Corresponding Author:
Itske Fraterman, MSc
Department of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam, Netherlands

Abstract

Background: Patients with melanoma receiving immunotherapy with immune-checkpoint inhibitors often experience immune-related adverse events, cancer-related fatigue, and emotional distress, affecting health-related quality of life (HRQoL) and clinical outcome to immunotherapy. eHealth tools can aid patients with cancer in addressing issues, such as adverse events and psychosocial well-being, from various perspectives.

Objective: This study aimed to explore the effect of the Cancer Patients Better Life Experience (CAPABLE) system, accessed through a mobile app, on HRQoL compared with a matched historical control group receiving standard care. CAPABLE is an extensively tested eHealth app, including educational material, remote symptom monitoring, and well-being interventions.

Methods: This prospective pilot study compared an exploratory cohort that received the CAPABLE smartphone app and a multisensory smartwatch for 6 months (intervention) to a 2:1 individually matched historical prospective control group. HRQoL data were measured with the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core

30 at baseline (T0), 3 months (T1), and 6 months (T2) after start of treatment. Mixed effects linear regression models were used to compare HRQoL between the 2 groups over time.

Results: From the 59 eligible patients for the CAPABLE intervention, 31 (53%) signed informed consent to participate. Baseline HRQoL was on average 10 points higher in the intervention group compared with controls, although equally matched on baseline and clinical characteristics. When correcting for sex, age, disease stage, and baseline scores, an adjusted difference in fatigue of −5.09 (95% CI −15.20 to 5.02, *P*=.32) at month 3 was found. No significant nor clinically relevant adjusted differences on other HRQoL domains over time were found. However, information satisfaction was significantly higher in the CAPABLE group (β=8.71, 95% CI 1.54 - 15.88, *P*=.02).

Conclusions: The intervention showed a limited effect on HRQoL, although there was a small improvement in fatigue at 3 months, as well as information satisfaction. When aiming at personalized patient and survivorship care, further optimization and prospective investigation of eHealth tools is warranted.

Trial Registration: ClinicalTrials NCT05827289; https://clinicaltrials.gov/study/NCT05827289

International Registered Report Identifier (IRRID): RR2-10.2196/49252

(JMIR Cancer 2025;11:e58938) doi:10.2196/58938

KEYWORDS

eHealth; melanoma; cancer; fatigue; quality of life; intervention; pilot study; exploratory; health-related; interventions; symptom; monitoring; well-being; immunotherapy; immune-related; immune-checkpoint inhibitor; patient; feasibility; smartphone; app; smartwatch; linear regression model; mobile phone

Introduction

Immunotherapy with immune-checkpoint inhibitors (ICIs) and targeted therapies with BRAF/MEK inhibitors have significantly improved clinical outcomes for patients with melanoma and have become standard treatment for patients with high-risk and advanced disease [1-6]. Nevertheless, these novel systemic treatments are associated with short- and long-term (immune-related) adverse events (AEs) [7-10]. Furthermore, these AEs have shown to affect physical and psychosocial well-being of patients with melanoma [10-13]. Most prevalent in patients with melanoma undergoing immunotherapy with ICIs are cancer-related fatigue (CRF) and emotional distress, affecting both health-related quality of life (HRQoL) and clinical outcome to immunotherapy [8,14-16]. Efforts to address CRF include exercise recommendation, psychosocial support, mindfulness-based interventions, and yoga, showing positive effects on fatigue, emotional distress, and HRQoL [14,15,17-23].

Insufficient monitoring and reporting of AEs can exacerbate side effects, possibly leading to more frequent hospital visits and admissions [24-26]. Electronic symptom monitoring has shown to be associated with improved clinical outcomes such as survival and HRQoL in patients with cancer undergoing chemotherapy [27-30]. One way to improve patient care in immunotherapy could therefore involve regularly gathering patient-reported outcomes (PROs), such as symptom information, using patient-reported outcome measures (PROMs) through eHealth tools [31-34]. Furthermore, using biometric sensors could potentially detect symptoms and track physical activity in outpatient oncology settings [35,36]. To date, evidence of the effect of eHealth tools monitoring patients with melanoma on treatment with ICIs is scarce. One study showed that an electronic PROMs tool could not reduce the number of severe AEs, although it did increase HRQoL [34,37].

Health apps have also the potential to fulfill patients' requirements for information and support, especially concerning symptom control and supportive services [38,39]. Furthermore, web-based programs and eHealth apps have incorporated nonpharmacological well-being interventions, such as promoting physical exercise, providing psychoeducation, mindfulness-based interventions, and yoga, to address CRF, showing encouraging outcomes [40,41]. By providing a combination of remote symptom monitoring, nonpharmacological well-being interventions, and information provision through an eHealth tool, patients believe this will positively affect their HRQoL and symptom burden [39].

Based on these insights, we previously developed a Cancer Patients Better Life Experience (CAPABLE) mobile app. CAPABLE is an extensively tested eHealth app as part of the EU Horizon 2020 program, designed to offer educational material, supportive care, remote symptom monitoring, and well-being interventions [42], initially for patients during and after ICIs, but open to treatment changes to targeted therapies. Development involved a user (patient)-centered design process in order to improve system usability and user acceptance [43]. The aim of this study was to explore the effect of CAPABLE on patient-reported outcomes, specifically fatigue and other HRQoL domains, compared with a matched historical control group receiving standard care [44].

Methods

Setting

The CAPABLE study was a prospective, exploratory, pilot study in which a cohort that received the CAPABLE smartphone app and a multisensory smartwatch (intervention) was compared with a historical prospective cohort that did not receive the CAPABLE app and smartwatch (control group). CAPABLE was registered as a medical device trial according to the Medical Device Regulation, article 62. A detailed description of the design of the pilot study and the design and development of the CAPABLE app was published previously [44]. Development and content was frozen during the trial. The CAPABLE app

was used in its final operational state for the first time in a trial setting during this study, although preliminary prototype testing was done during system development. Briefly, participants included in the study were provided with the CAPABLE mobile app and a multisensorial smartwatch (ASUS VivoWatch 5 HC-B0) during the first 6 months of treatment with ICIs. The mobile app consisted of 3 main components; first, facilitating symptom and mental well-being monitoring, second, providing educational material, and finally, providing well-being interventions through goal setting and demonstrating the well-being intervention activity through a video or text and figures. The symptom monitoring functionality was used "as needed." When a patient experienced a symptom, they were able to enter this into the CAPABLE app, upon which the decision support system managed the symptom [44]. No regular or static symptom monitoring was prompted by the app; however, the health care professional (HCP) monitored the symptom input coming from the patients on a daily basis and the information was included and discussed in clinical encounters. The well-being interventions could be executed from the home environment, and include a 30-minute walk, deep breathing practice, imagery training, physical activity of stretching, and strengthening exercises, Hatha Yoga or Nidra Yoga videos, or Tai Chi practice videos. The smartwatch collected data on heart rate, sleep (stages, hours, and performance), and physical activity, although data from the smartwatch were treated as ancillary data and not used for real-time symptom monitoring, decision support, or diagnosis. Over the course of the pilot study, participants were asked to complete PROMs at 3 time points. Results of the intervention (CAPABLE) group were compared with participants of 2 previously collected control groups (patient-reported outcomes in high risk and advanced melanoma patients cohort [PRO-MEL]; NL75996.031.20 and PROMs collected in clinical practice), which were 2 similar prospectively collected cohorts with the same inclusion criteria, but treated following standard of care and following the same follow-up schedule. PROMs in clinical practice were collected starting August 2024 and is still ongoing at the time of study. In addition, the PRO-MEL cohort was a prospective cohort that started inclusion in May 2021 and collected additional PROMs, as also collected in the CAPABLE cohort.

Ethical Considerations

The Medical Ethical Committee NedMec (Amsterdam, the Netherlands) granted ethical approval (reference 22 - 981/NL81970.000.22). The trial was prospectively registered at ClinicalTrials.gov (NCT05827289). Ethical approval also included the (secondary) use of data collected in the PRO-MEL and PROMs in clinical practice cohorts. Privacy and confidentiality protection was covered in the Medical Ethical Committee approval by a large data protection impact assessment. The study has not been amended during the course of the trial. Compensation to participants was not provided, except for the temporarily use of the smartphone and smartwatch used in the study.

Recruitment

During a 6-month inclusion period, from April to October 2023, participants were recruited through their treating HCP and the CAPABLE research team recruited in an oncology-specialized center in Amsterdam, the Netherlands. The target sample for feasibility end points of this pilot was to include 36 patients, corresponding to 60 eligible patients and a 60% inclusion and compliance rate in the inclusion period [44]. Eligible participants had histologically confirmed stage III or IV melanoma and planned to start treatment with ICIs (anti–programmed-death 1 with or without anti–cytotoxic T-lymphocyte associated protein 4) according to standard clinical practice. Furthermore, participants had to be >18 years of age, had a sufficient understanding of the Dutch language, and were able to use a smartphone.

Data Collection

Included patients were asked to use the CAPABLE app and smartwatch for a minimum period of 3 months and a maximum period of 6 months after start of treatment with ICIs. CAPABLE installation on mobile phones and baseline measurements took place after signing informed consent and before or during first ICI infusion. Research data were collected at baseline (T0), 3 months (T1), and 6 months (T2) after start of treatment by providing the participants a set of questionnaires. Clinical data (eg, staging, treatment details, and demographics) were extracted from the medical record during the study. PROMs data were stored and managed in ALEA (FormVision) [45]. Data generated through the CAPABLE app were stored on an internal secured drive.

To investigate the primary end point of this study, fatigue, the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire-Core 30 (QLQ-C30) was used [46], a questionnaire developed to assess the quality of life of patients with cancer. Responses to this questionnaire range from 1 (not at all) to 4 (very much) and are linearly transformed into a functioning or symptom scale ranging from 0 to 100, with higher scores representing more experienced symptoms or a higher functioning, respectively. Primarily, the changes in fatigue over time in the intervention cohort were compared with the changes in fatigue over time in the control cohort. The validated fatigue scale of the QLQ-C30 is constructed out of 3 questions in the QLQ-C30 questionnaire, "Did you need to rest?", "Have you felt weak?", and "Were you tired?". To explore secondary outcomes of this study, other domains of the EORTC QLQ-C30 were investigated (functioning and other symptom scales), as well as changes between the CAPABLE and (matched) control group when looking at the EuroQol 5D (EQ-5D-5L) [47], Functional Assessment of Cancer Therapy-Melanoma (FACT-M) [48], and EORTC Quality of Life Questionnaire-Information 25 (QLQ-INFO25) [49].

Feasibility outcomes were investigated throughout the course of this pilot study by exploring the inclusion and compliance rate of the CAPABLE app users. Recruitment rate was calculated as the percentage of patients included in the study out of the patients screened for eligibility. Patient compliance was calculated as the percentage of patients completing the

questionnaires per follow-up moment. Finally, patient retention was calculated as the percentage of patients adhering to the CAPABLE mobile app for 6 months (ie, ≥1 interaction with any of the functionalities within the follow-up period). Patient engagement with the app was presented by descriptive data on the use of the symptom reporting and well-being intervention functionalities. Extensive data collection methods and corresponding figures and tables are described in the previously published study protocol [44].

Data Analysis

Patients that completed at least 1 PROM over the course of the study were included in the final analysis. Because of low inclusion in the (control cohort) PRO-MEL study, matching was done on a control group composed of patients from both the PRO-MEL and PROMs in clinical practice cohort who filled in the EORTC QLQ-C30, FACT-M, and EQ-5D-5L according to the same follow-up schedule. Patients in the CAPABLE cohort were individually matched 1:2 with patients in the control cohorts based on sex, age, and tumor staging. To compare information needs (QLQ-INFO25) between the CAPABLE cohort and controls, no matching was performed, and comparison consisted of the entire PRO-MEL cohort to increase statistical power and be able to interpret results.

Descriptive statistics were calculated to provide information about the patient population, feasibility, and engagement with the CAPABLE app. For the purpose of this study, mean scores for fatigue and other QLQ-C30 domains were calculated and presented according to current guidelines [50]. To compare the mean fatigue scores and other HRQoL outcomes between the group receiving the CAPABLE intervention and the control group at each individual time point, independent sample *t* tests were used. To analyze the differences in all outcomes on different time points between the CAPABLE cohort and matched controls over time, linear mixed effects models were used. Statistical models were adjusted for age, sex, tumor stage, time, and baseline scores (with an interaction term between time and cohort). A 2-tailed *P* value<.05 was considered statistically significant, although *P* values in this pilot setting were not powered to provide much information due to the small sample size. Therefore, this study mostly focused on clinically relevant differences according to Cocks et al [51]. Similar methods were used for analyzing the EQ-5D-5L, FACT-M, and QLQ-INFO25.

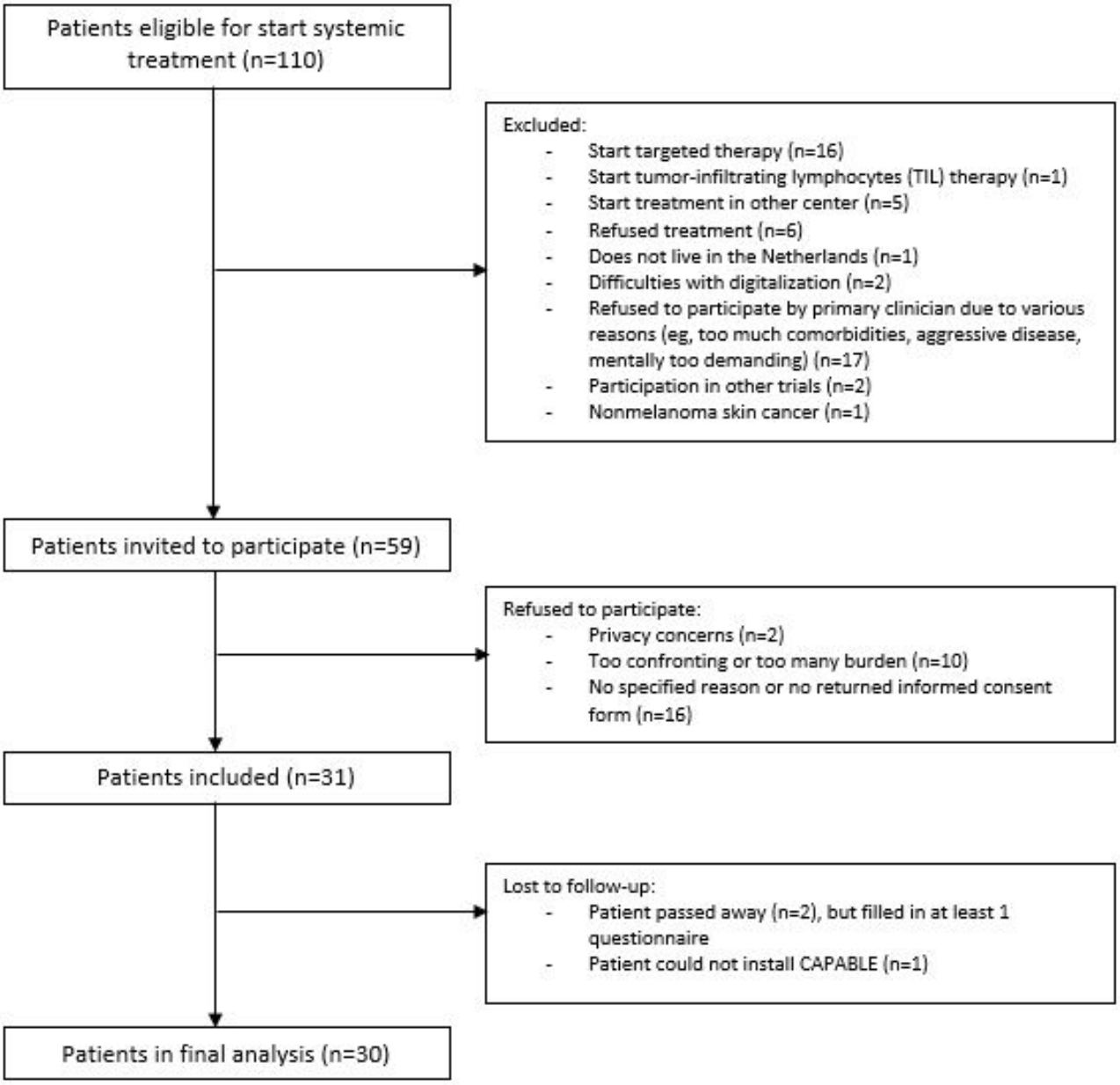
Missing items from the questionnaires were imputed according to corresponding EORTC guidelines [50]. The scale scores of the EORTC QLQ-C30 were set to missing if fewer than half of the items on a given scale were answered. Where at least 50% of the relevant scale scores were present, the missing values were replaced by the mean of the present values. We applied the same strategy to the other questionnaires as no other guidelines are available for those. Statistical analyses and matching procedures were done using Stata version 15 (StataCorp) [52].

Results

Overview

In total, 110 patients were screened for eligibility for the CAPABLE trial in the 6-month inclusion period (Figure 1). Main reasons for noneligibility were the start of targeted therapy instead of ICIs (n=16) or the patient not being invited for inclusion by the treating physician's decision (eg, aggressive disease, comorbidities, symptomatic brain metastasis, low health literacy, and mentally too demanding; n=17). Eventually, 59 out of 110 (54%) patients were contacted to participate. Most of the contacted patients who did not return a consent form did not specify a reason (n=16). Reasons specified for not participating in the pilot study included privacy concerns (n=2) and the expectation that the burden would be too high (n=10). Of the 31 included patients, 1 did not manage to install the CAPABLE app before T1. A total of 30 patients were taken into consideration for the statistical analysis, although 2 patients died due to progressive disease before reaching T2. In total, 297 patients (70 from the PRO-MEL cohort and 227 from the PROMs in clinical practice cohort) were eligible for individual matching. This yielded 56 patients that were matched 2:1 to the CAPABLE cohort. Thus, a group of 86 patients was included in the analysis of the EORTC QLQ-C30, FACT-M, and EQ-5D-5L. Since data on the EORTC QLQ-INFO25 were only available in the PRO-MEL cohort, the entire PRO-MEL cohort (n=70) was used in the comparison with the CAPABLE group (n=30) for these secondary end points.

Figure 1. Cancer Patients Better Life Experience (CAPABLE) study inclusion flowchart.



Both cohorts were equally balanced in terms of age, sex, and tumor stage, due to matching (Table 1). Median age of the CAPABLE cohort was 65 (IQR 55 - 72) years. Females represented 61% (19/31) of the included participants, and approximately half the participants had stage IV disease (17/31, 55%). Most patients received anti-PD1 monotherapy in the CAPABLE and matched control group (19/31, 61% and 37/56, 66%, respectively). In the matched control group, 21% (12/56) of patients received targeted therapy after therapy switch, compared with 7% (2/31) in the CAPABLE group, but this difference was not statistically significant (P=.07).

Table . Participants' clinical characteristics. Not all percentages add up to 100% as multiple patients received multiple treatments. Missing data were not taken into consideration when calculating the *P* values.

Characteristic		CAPABLE[a] cohort (n=31)	Control cohort (n=56)	*P* value[b]
Sex, n (%)				.96
	Male	12 (39)	22 (39)	
	Female	19 (61)	34 (61)	
Age (years), median (IQR)		65 (55-72)	64 (56-71)	.99
Tumor stage, n (%)				.96
	III	14 (45)	25 (45)	
	IV	17 (55)	31 (55)	
Treatment, n (%)				
	Anti-PD-1[c]	19 (61)	37 (66)	.66
	Anti-CTLA-4[d] + Anti-PD-1	14 (45)	23 (41)	.71
	Radiotherapy	6 (19)	12 (21)	.82
	Surgery before ICIs[e]	10 (32)	23 (41)	.42
	Targeted therapy	2 (7)	12 (21)	.07
Treatment line, n (%)				.90
	1	24 (77)	44 (79)	
	>1	7 (23)	12 (21)	

[a]CAPABLE: Cancer Patients Better Life Experience

[b]*P* values are based on χ^2 tests for group variables and Mann-Whitney test for age.

[c]PD-1: programmed-death 1.

[d]CTLA-4: cytotoxic T-lymphocyte associated protein 4.

[e]ICI: immune-checkpoint inhibitor.

Fatigue

The trend in unadjusted fatigue over time was similar between the CAPABLE group and the matched controls although fatigue seems to increase less in the CAPABLE cohort (Figure S1 in Multimedia Appendix 1). A 5-point difference was shown in baseline fatigue. The CAPABLE group had lower fatigue score on baseline compared with the matched controls (mean 18.4, SD 21.7) compared with the matched controls (mean 23.4, SD 19.4; *P*=.28), increasing to a difference of 8.2 points at month 3 (mean 23.0, SD 25.2 vs mean 31.2, SD 24.1; *P*=.17) (Table S1 in Multimedia Appendix 1). When correcting for sex, age, stage, time, and baseline scores, an adjusted difference in fatigue of −5.09 (95% CI −15.20 to 5.02; *P*=.32) for the CAPABLE group at month 3 was observed (Table 2). Although this result was not statistically significant, this difference was considered a small, clinically relevant difference. At month 6, a nonsignificant and nonclinically relevant difference was shown between the 2 groups (β=−2.32, 95% CI −12.81 to 8.16; *P*=.66).

Table . Adjusted mixed effects linear regression analysis on fatigue as measured by the Quality of Life Questionnaire-Core 30 (QLQ-C30) between Cancer Patients Better Life Experience (CAPABLE) group and matched controls over time.

Fatigue		β (95% CI)	*P* value	Clinical relevance
Cohort				
	Controls	Ref	—[a]	—
	CAPABLE	−1.43 (−9.00 to 6.13)	.71	Trivial
Sex				
	Male	Ref	—	—
	Female	1.31 (−4.06 to 6.68)	.63	Trivial
Age		0.03 (−0.22 to 0.27)	.83	Trivial
Stage				
	Stage III	Ref	—	—
	Stage IV	−0.35 (−5.36 to 4.66)	.89	Trivial
Time				
	Baseline	Ref	—	—
	Month 3	8.28 (2.38 to 14.18)	.01	Small
	Month 6	5.60 (−0.34 to 11.54)	.07	Small
Cohort×time				
	Controls × baseline	Ref	—	—
	CAPABLE × month 3	−5.09 (−15.20 to 5.02)	.32	Small
	CAPABLE × month 6	−2.32 (−12.81 to 8.16)	.66	Trivial
Baseline score		0.72 (0.58 to 0.85)	<.001	— (offset)

[a]Not applicable.

Health-Related and Melanoma-Specific Quality of Life

A significant difference in baseline scores was observed between the 2 cohorts for most of the HRQoL domains measured by the EORTC QLQ-C30 (Table S1 in Multimedia Appendix 1). On all functioning scales except cognitive functioning, the CAPABLE group reported better function in terms of both statistical significance and clinical relevance, with mean baseline differences ranging from 8.5 in social functioning (*P*=.11) to 12.0 in role functioning (*P*=.07) (Table S1 in Multimedia Appendix 1). After adjusting for covariates in multivariable regression analysis, no statistically significant nor clinically relevant changes on any of the HRQoL domains were observed between the CAPABLE group and matched controls at either month 3 or month 6 follow-up (Table 3).

Table . Adjusted mixed effects linear regression analyses on different health-related quality of life (HRQoL) outcomes as measured by the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire-Core 30 (QLQ-C30) between Cancer Patients Better Life Experience (CAPABLE) group and matched controls over time. Analyses are adjusted for age, sex, stage, and baseline scores.

HRQoL subscales		β (95% CI)[a]	*P* value	Clinical relevance
Physical functioning				
	CAPABLE × month 3	−4.45 (−11.93 to 3.03)	.24	Trivial
	CAPABLE × month 6	−6.31 (−14.09 to 1.47)	.11	Small
Role functioning				
	CAPABLE × month 3	0.55 (−11.88 to 12.99)	.93	Trivial
	CAPABLE × month 6	−3.82 (−16.74 to 9.10)	.56	Trivial
Emotional functioning				
	CAPABLE × month 3	−5.77 (−14.71 to 3.16)	.21	—[b]
	CAPABLE × month 6	−8.41 (−17.63 to 0.80)	.07	—
Social functioning				
	CAPABLE × month 3	−1.89 (−12.54 to 8.77)	.73	Trivial
	CAPABLE × month 6	1.09 (−9.95 to 12.13)	.85	Trivial
Cognitive functioning				
	CAPABLE × month 3	0.30 (−7.16 to 7.77)	.94	Trivial
	CAPABLE × month 6	1.06 (−6.64 to 8.77)	.79	Trivial
Insomnia				
	CAPABLE × month 3	3.65 (−1.68 to 13.93)	.58	Trivial
	CAPABLE × month 6	0.88 (−12.57 to 14.33)	.90	Trivial
Financial difficulties				
	CAPABLE × month 3	4.93 (−0.33 to 10.20)	.07	Small
	CAPABLE × month 6	7.47 (1.99 to 12.95)	.01	Small
Summary score				
	CAPABLE × month 3	−3.37 (−8.78 to 2.04)	.22	—
	CAPABLE × month 6	1.19 (−4.40 to 6.80)	.68	—

[a]Results presented are adjusted βs coming from interaction between cohort and time. Matched controls at baseline are reference group.

[b]Not applicable.

Similar baseline differences were observed in melanoma-specific quality of life as measured by the melanoma subscale (MS) and melanoma surgery subscale (MSS) of the FACT-M. Melanoma-specific quality of life showed a significant baseline difference between the CAPABLE group and matched controls when looking at both the MS (mean 57.5, SD 5.3 vs mean 50.4, SD 8.3; *P*<.001) and the MSS (mean 28.1, SD 4.4 vs mean 22.4, SD 6.3; *P*<.001). Mean scores did not change much over time in both cohorts (Table S3 in Multimedia Appendix 1). When adjusted for sex, age, tumor stage, time, and baseline scores, there were also no significant changes in melanoma-specific quality of life between both cohorts over time (Table 4). Utility scores of the EQ-5D-5L yielded similar results in terms of mean scores, and when adjusted for all covariates, no effect of CAPABLE was seen compared with matched controls (Table S3 in Multimedia Appendix 1; Table 4). However, HRQoL as measured by the visual analog scale, was significantly higher for the CAPABLE cohort compared with the matched controls on month 3 and month 6 when corrected for sex, age, and stage of disease (B=10.28, 95% CI 1.45 - 19.11, *P*=.02 and B=11.50, 95% CI 2.08 - 20.92, *P*=.017, respectively).

Table . Adjusted univariable mixed effects linear regression analyses on different patient reported outcome measures between CAPABLE group and matched controls. Analyses are adjusted for age, sex, stage, and baseline scores.

HRQoL[a] subscales			β (95% CI)[b]	*P* value
FACT-M[c]				
	MS[d] (range 0-64)			
		CAPABLE[e]×month 3	−1.54 (−5.37 to 2.29)	.43
		CAPABLE×month 6	−0.82 (−4.78 to 3.14)	.69
	MSS[f] (range 0-32)			
		CAPABLE×month 3	0.91 (−2.34 to 4.17)	.58
		CAPABLE×month 6	−0.68 (−4.00 to 2.64)	.69
EQ-5D-5L				
	Utility (range 0-1)			
		CAPABLE×month 3	−0.05 (−0.11 to 0.01)	.10
		CAPABLE×month 6	−0.03 (−0.10 to 0.03)	.33
	VAS[g] (range 0-100)			
		CAPABLE×month 3	10.28 (1.45 to 19.11)	.02
		CAPABLE×month 6	11.50 (2.08 to 20.92)	.02

[a]HRQoL: health-related quality of life.

[b]Results presented are adjusted βs coming from interaction between cohort and time. Matched controls at baseline are reference group.

[c]FACT-M: Functional Assessment of Cancer Therapy-Melanoma.

[d]MS: melanoma subscale.

[e]CAPABLE: Cancer Patients Better Life Experience.

[f]MSS: melanoma surgery subscale.

[g]VAS: visual analog scale.

Information Needs

Overall, information provision as reported by the EORTC QLQ-INFO25 was significantly lower in the control cohort on both baseline and month 6 (Table S4 in Multimedia Appendix 1). When adjusting for age, sex, and baseline scores, no separate domains showed significant improvements of the CAPABLE cohort. However, information satisfaction was significantly higher in the CAPABLE cohort (B=8.71, 95% CI 1.54 - 15.88; *P*=.02) (Table 5).

Table . Adjusted univariable mixed effects linear regression analyses on information domains between CAPABLE group and PRO-MEL controls. Analyses are adjusted for age, sex, stage, and baseline scores.

Information domains[a]	β (95% CI)	*P* value
Disease	3.38 (−5.49 to 12.25)	.46
Medical tests	9.35 (0.22 to 18.48)	.05
Treatment	1.59 (−6.89 to 10.06)	.71
Other services	2.46 (−7.28 to 12.19)	.62
Different places of care	−2.19 (−14.04 to 9.67)	.72
Things you can do to help yourself	−6.12 (−15.54 to 3.30)	.20
Satisfaction with information received	8.71 (1.54 to 15.88)	.02
Overall the information has been helpful	3.69 (−4.77 to 12.14)	.39

[a]Controls are reference group.

Feasibility

Because of project time constraints, only 31 of the planned 36 patients were included in this study. In total, 59 patients were eligible for study participation, resulting in a recruitment rate of 53% in the set period, whereas the planned recruitment rate for reaching the feasibility end point was 60%. However, patient compliance and patient retention remained high in the patients that were included. Patient compliance to the PROMs at baseline was 98% (30/31), at T1 was 90% (27/30), and at T2 was 79%

(22/28). Finally, 27 out of 31 (87%) patients adhered to using the CAPABLE app until at least T1, which dropped to 24 patients (77%) using the app at T2. Furthermore, 2 of those patients died because of rapidly progressive disease during the trial. Adherence to smartwatch use was lower with only 43% (13/30) usage at T2 due to smartwatch issues. In total, 27 individual problems with the CAPABLE app and smartwatch were reported during the trial, mostly in the first 3 months of usage. Almost half of patients (14/30, 47%) reported at least 1 problem with the CAPABLE app or one of its functionalities. The majority of reported problems were related to login issues (7/20, 35%), smartwatch problems (communication between app and discomfort of the smartwatch; 8/20, 41%), and problems with the symptom reporting workflow (3/20, 15%).

Engagement With the System

Concerning the symptom reporting functionality, 18 out of 30 patients have actively used the CAPABLE app and reported at least 1 distinct symptom or symptom episode (range 1 - 7) (Figure S2 in Multimedia Appendix 1). In total, 20 distinct immune-related AEs were reported through the CAPABLE app, with reports of 33 grade 1, 28 grade 2, 17 grade 3, and 3 grade 4 symptoms according to the mapped Common Terminology Criteria for AEs version 5 (Table S5 in Multimedia Appendix 1). Symptom episodes ranged from 1.5 days for headache to 149.6 days for muscle pain (Table S6 in Multimedia Appendix 1).

Engagement with the well-being interventions was on average lower than symptoms reporting (Table S7 in Multimedia Appendix 1). Interventions were not prescribed by HCPs and were free to use by the patients. The intervention related to taking a walk was the most often used, with 327 execution times, reported by 9 distinct patients. Some patients used the other interventions infrequently. Furthermore, 8 out of 30 patients used the interventions more than 5 times (range 8 - 213). Interventions that were used infrequently were mainly used during the first weeks of enrollment, suggesting that the interventions were tried out at the start of the pilot. The walking intervention was executed throughout the course of the pilot study, with more engagement in summer than in fall and winter (October-December).

Discussion

Principal Findings

The aim of this pilot study was to explore the effect of the CAPABLE mobile app on patient-reported outcomes, specifically fatigue, in patients with melanoma starting ICIs, compared with a historical control group receiving standard care. Our results showed no significant adjusted differences on fatigue and other HRQoL domains between the CAPABLE cohort and matched controls during the first 6 months of treatment. However, although not statistically significant, we did find that the CAPABLE cohort reported a smaller relevant increase in fatigue at month 3 follow-up compared with matched controls. Furthermore, information satisfaction was significantly higher in the CAPABLE users. The secondary goal of this study was to show patients' acceptance and feasibility of the CAPABLE app. With an observed recruitment rate of 53%, our feasibility end point of 60% was not met. A third of the patients refused to participate because study and questionnaires were expected to be too burdensome. Furthermore, HCPs did not feel comfortable including patients in the study because they expected it to be too burdensome for some patients (based on oral feedback). Furthermore, technology barriers might have played a role, as observed in other studies with eHealth apps [53].

We observed a significant baseline difference in almost all the HRQoL domains between the CAPABLE cohort and matched controls, suggesting a selection bias. Although matched on baseline characteristics, HRQoL domains in the CAPABLE group are clinically relevantly and significantly higher than in the matched controls. Therefore, results obtained in the CAPABLE group could have been influenced by the phenomenon "regression towards the mean," as improvement of HRQoL was almost impossible to achieve [54]. However, this phenomenon looks like it occurred in all HRQoL domains, except for fatigue, as we see an improvement only in fatigue for the CAPABLE cohort, although not statistically significant. This observation also underscores the importance of including patients with lower HRQoL in interventions designed to improve this outcome, for example, by minimizing the expected burden of participation [55].

Several studies have shown the benefits of eHealth on CRF. Supporting the small (nonsignificant) difference found in fatigue in our study, a meta-analysis done on 9 studies showed a statistically significant beneficial effect of eHealth interventions on CRF [23]. Furthermore, these eHealth tools were mostly designed to target CRF solely and did not have the multimodal aspects of our CAPABLE app. Furthermore, we were not able to disentangle which specific functionality was responsible for possible changes in fatigue, or if it was a combination of all functionalities. In addition, our sample size was not large enough to provide statistical significance; our pilot study was designed to provide descriptive statistics and focused largely on clinical relevance [51].

A similar study in Denmark, with electronic symptom monitoring carried out in patients with metastatic melanoma starting treatment with ICIs yielded improved HRQoL in the intervention group, as measured by the FACT-M and EQ-5D-5L [34,37], although the differences Tolstrup et al [37] obtained were also not clinically relevant. The Danish study was a randomized controlled trial (RCT) and had an active weekly symptom-monitoring component by their HCP; components of the trial that could have influenced the results compared with passive HCP monitoring in our real-world single-arm setting [56,57].

Information satisfaction was significantly higher in the CAPABLE cohort compared with the control group. Studies done on information provision through eHealth tools in the Dutch cancer care are still conflicting [58]. For example, in an RCT investigating a web-based eHealth app to support multiple cancer patient groups, improvement of knowledge was not reached [59]. However, another study showed that higher information satisfaction might contribute to patient knowledge and decision involvement [60]. Therefore, an eHealth tool, such

as CAPABLE, might still support patient knowledge and shared decision-making.

Results of our study might have been influenced by barriers when integrating and implementing eHealth into clinical practic. Although the CAPABLE system was developed using all relevant stakeholders, the CAPABLE app was not integrated in our electronic health record, causing our HCPs to work with 2 different systems, leading to resistance. In-depth results of user experience and usability research done in HCPs still need to be analyzed, but verbal feedback suggested this was a large barrier for monitoring patients with this app. To date, successful implementation and use of digital health interventions remains limited worldwide by integration into electronic health records, impairing the possible positive effect of such interventions [53]. Second, an existing challenge in digital health interventions research is the recruitment of target populations in need of such interventions [53,61,62]. Both not reaching our feasibility end point of 60% and high HRQoL baseline scores (probably because of selection bias) confirm this challenge. Consequently, results of this study need to be interpreted with caution and future research should make more effort into recruitment strategies including weaker populations, as well as considering health literacy. Furthermore, efforts are needed to reduce patient burden in this type of studies, both in terms of intervention as with research questionnaires, as it is essential to include all patients with serious health conditions [55].

Limitations

Several limitations of this study need to be considered when interpreting the results. Our sample size was too small to prove significant differences on HRQoL outcomes. Because of time constraints of our project, whose main focus was the design and development of the system, inclusion period was short and only 31 patients out of the anticipated 36 were included. Second, most of the control group was collected during the COVID-19 pandemic, which might have influenced HRQoL outcomes in these patients. However, a study done by van de Poll-Franse et al [63] showed that the crisis might have affected well-being of general population more than in that of cancer patients. In addition, another negative aspect of the COVID-19 pandemic was related to some delays in the software development for the CAPABLE eHealth app in an already-restricted project timeframe, which might have resulted in an increase in app issues reported by patients. While in-depth usability outcomes of this study have yet to be analyzed, we observed a relatively high proportion of technical problems during our study and we gathered multiple areas of improvement from verbal feedback from patients, which could have affected the results. Another limitation of the study was our decision to use a matched historical control cohort rather than an RCT. An RCT would have allowed better isolation of the effect of using the app. However, use of the historical cohort allowed us to recognize and characterize the participation bias in the intervention participants. This bias may have contributed to the lack of a clinically relevant effect in our main outcome.

However, a large strength of this study has been that CAPABLE was developed with co-design of patients and HCPs, following user-centered design principles, starting with explorative interviews and undergoing 3 testing rounds. Furthermore, our study was conducted in the real-world setting of the pilot trial; included patients were treated according to clinical practice. The CAPABLE app was added as a monitoring and coaching system tool and could intervene in management of patients after severe symptom reports. Due to the real-world setting of this trial, results on PROs, feasibility, and engagement with eHealth might be more generalizable to real-world patients compared with results found in clinical trials.

Future and larger studies could benefit from including patients more inclusively in terms of low health literacy and social economic status. Inclusion criteria should be broadened to reduce ceiling effects at baseline. Furthermore, the setting of this pilot study might have played a pivotal role, as we included patients in a dedicated cancer and melanoma center with a lot of ongoing clinical trials and other intervention studies and also standard care including easily accessible specialized nurse practitioners.

Conclusion

In our small, nonrandomized study we were unable to show that mobile-based coaching and follow-up affected HRQoL significantly, although results suggest a small clinical improvement of fatigue at 3 months follow-up in the app users. Ceiling effects due to large baseline differences might have caused the impact of CAPABLE to be negligible for patients with higher baseline HRQoL. CAPABLE resulted in significantly higher information satisfaction compared with controls. Although the feasibility end point of 60% was not met, adherence to the system was high. Further optimization of CAPABLE, taking into account patient-related and technology-related barriers is needed before future investigation in an RCT and might influence HRQoL end points. Furthermore, when aiming at personalized patient and survivorship care, further optimization and prospective investigation of eHealth tools is warranted.

Acknowledgments

This work was supported by the European Union (EU)'s Horizon 2020 research and innovation programme (grant agreement No 875052). The Cancer Patients Better Life Experience (CAPABLE) system was developed during the first 3 years of an EU project funded under the Horizon 2020 Research and Innovation Programme. CAPABLE is a 4-year project coordinated by the University of Pavia, Italy, and carried out by 12 partners in Europe and Israel. We like to thank all the partners and participants involved in this and previous CAPABLE-project studies.

Data Availability

The datasets generated during and analyzed during this study are not publicly available due to the sensitive information of patients but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization was contributed by all authors. Data curation and formal analysis was contributed by IF, LVvdPF, SW, and LS. Funding acquisition was managed by all authors. Investigation was handled by IF and SW. Methodology was contributed by IF, LVvdPF, SW, LDC, VT, RC, and SM. Project administration, resources, software supervision, and validation was handled by all authors. Visualization was contributed by IF, LVvdPF, SW, and LS. Writing – original draft was contributed by IF, SW, and LS. Writing – review and editing was contributed by all authors.

Conflicts of Interest

DG is an employee of and shareholder in Deontics Ltd. Deontics provided the Computer Decision Support platform component for the Cancer Patients Better Life Experience (CAPABLE) system used in the study. LDL reported the following: Conference honoraria/Advisory Board: EISAI, MSD, Merck Serono, Eli Lilly, Sanofi, Sunpharma, IPSEN, Bayer, Roche, Istituto Gentili Srl; New Bridge; Seagen; Novartis; Travel grant: Gilead.
All other authors report no conflict of interest.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File, 160 KB - cancer_v11i1e58938_app1.docx]

References

1. Larkin J, Del Vecchio M, Mandalá M, et al. Adjuvant nivolumab versus ipilimumab in resected stage III/IV melanoma: 5-year efficacy and biomarker results from CheckMate 238. Clin Cancer Res 2023 Sep 1;29(17):3352-3361. [doi: 10.1158/1078-0432.CCR-22-3145] [Medline: 37058595]
2. Robert C, Carlino MS, McNeil C, et al. Seven-year follow-up of the phase III KEYNOTE-006 study: pembrolizumab versus ipilimumab in advanced melanoma. J Clin Oncol 2023 Aug 20;41(24):3998-4003. [doi: 10.1200/JCO.22.01599] [Medline: 37348035]
3. Robert C, Grob JJ, Stroyakovskiy D, et al. Five-year outcomes with dabrafenib plus trametinib in metastatic melanoma. N Engl J Med 2019 Aug 15;381(7):626-636. [doi: 10.1056/NEJMoa1904059] [Medline: 31166680]
4. Wolchok JD, Chiarion-Sileni V, Gonzalez R, et al. Long-term outcomes with nivolumab plus ipilimumab or nivolumab alone versus ipilimumab in patients with advanced melanoma. J Clin Oncol 2022 Jan 10;40(2):127-137. [doi: 10.1200/JCO.21.02229] [Medline: 34818112]
5. Wolchok JD, Chiarion-Sileni V, Gonzalez R, et al. Overall survival with combined nivolumab and ipilimumab in advanced melanoma. N Engl J Med 2017 Oct 5;377(14):1345-1356. [doi: 10.1056/NEJMoa1709684] [Medline: 28889792]
6. Michielin O, van Akkooi ACJ, Ascierto PA, Dummer R, Keilholz U, ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org. Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Ann Oncol 2019 Dec 1;30(12):1884-1901. [doi: 10.1093/annonc/mdz411] [Medline: 31566661]
7. Haanen J, Obeid M, Spain L, et al. Management of toxicities from immunotherapy: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol 2022 Dec;33(12):1217-1238. [doi: 10.1016/j.annonc.2022.10.001] [Medline: 36270461]
8. Egeler MD, van Leeuwen M, Fraterman I, et al. Common toxicities associated with immune checkpoint inhibitors and targeted therapy in the treatment of melanoma: a systematic scoping review. Crit Rev Oncol Hematol 2023 Mar;183:103919. [doi: 10.1016/j.critrevonc.2023.103919] [Medline: 36736511]
9. Naidoo J, Page DB, Li BT, et al. Toxicities of the anti-PD-1 and anti-PD-L1 immune checkpoint antibodies. Ann Oncol 2015 Dec;26(12):2375-2391. [doi: 10.1093/annonc/mdv383] [Medline: 26371282]
10. O'Reilly A, Hughes P, Mann J, et al. An immunotherapy survivor population: health-related quality of life and toxicity in patients with metastatic melanoma treated with immune checkpoint inhibitors. Support Care Cancer 2020 Feb;28(2):561-570. [doi: 10.1007/s00520-019-04818-w] [Medline: 31089820]
11. Rogiers A, Boekhout A, Schwarze JK, Awada G, Blank CU, Neyns B. Long-term survival, quality of life, and psychosocial outcomes in advanced melanoma patients treated with immune checkpoint inhibitors. J Oncol 2019;2019:5269062. [doi: 10.1155/2019/5269062] [Medline: 31182961]
12. Rogiers A, Leys C, De Cremer J, et al. Health-related quality of life, emotional burden, and neurocognitive function in the first generation of metastatic melanoma survivors treated with pembrolizumab: a longitudinal pilot study. Support Care Cancer 2020 Jul;28(7):3267-3278. [doi: 10.1007/s00520-019-05168-3] [Medline: 31745697]

13. Saw RPM, Bartula I, Winstanley JB, Morton RL, Dieng M, Lai-Kwon J, et al. Melanoma and quality of life. In: Kassianos AP, editor. Handbook of Quality of Life in Cancer: Springer International Publishing; 2022:439-466. [doi: 10.1007/978-3-030-84702-9_26]
14. Bower JE. Cancer-related fatigue--mechanisms, risk factors, and treatments. Nat Rev Clin Oncol 2014 Oct;11(10):597-609. [doi: 10.1038/nrclinonc.2014.127] [Medline: 25113839]
15. Horneber M, Fischer I, Dimeo F, Rüffer JU, Weis J. Cancer-related fatigue: epidemiology, pathogenesis, diagnosis, and treatment. Dtsch Arztebl Int 2012 Mar;109(9):161-171. [doi: 10.3238/arztebl.2012.0161] [Medline: 22461866]
16. Fraterman I, Reijers ILM, Dimitriadis P, et al. Association between pretreatment emotional distress and neoadjuvant immune checkpoint blockade response in melanoma. Nat Med 2023 Dec;29(12):3090-3099. [doi: 10.1038/s41591-023-02631-x] [Medline: 37957378]
17. Paley CA, Boland JW, Santarelli M, Murtagh FEM, Ziegler L, Chapman EJ. Non-pharmacological interventions to manage psychological distress in patients living with cancer: a systematic review. BMC Palliat Care 2023 Jul 6;22(1):88. [doi: 10.1186/s12904-023-01202-8] [Medline: 37407974]
18. Bourmaud A, Anota A, Moncharmont C, et al. Cancer-related fatigue management: evaluation of a patient education program with a large-scale randomised controlled trial, the PEPs fatigue study. Br J Cancer 2017 Mar 28;116(7):849-858. [doi: 10.1038/bjc.2017.31] [Medline: 28196066]
19. Fabi A, Bhargava R, Fatigoni S, et al. Cancer-related fatigue: ESMO Clinical Practice Guidelines for diagnosis and treatment. Ann Oncol 2020 Jun;31(6):713-723. [doi: 10.1016/j.annonc.2020.02.016] [Medline: 32173483]
20. Puetz TW, Herring MP. Differential effects of exercise on cancer-related fatigue during and following treatment: a meta-analysis. Am J Prev Med 2012 Aug;43(2):e1-24. [doi: 10.1016/j.amepre.2012.04.027] [Medline: 22813691]
21. Song S, Yu J, Ruan Y, Liu X, Xiu L, Yue X. Ameliorative effects of Tai Chi on cancer-related fatigue: a meta-analysis of randomized controlled trials. Support Care Cancer 2018 Jul;26(7):2091-2102. [doi: 10.1007/s00520-018-4136-y] [Medline: 29564620]
22. Bruggeman-Everts FZ, Wolvers MDJ, van de Schoot R, Vollenbroek-Hutten MMR, Van der Lee ML. Effectiveness of two web-based interventions for chronic cancer-related fatigue compared to an active control condition: results of the "Fitter na kanker" randomized controlled trial. J Med Internet Res 2017 Oct 19;19(10):e336. [doi: 10.2196/jmir.7180] [Medline: 29051138]
23. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: a systematic review and meta-analysis. Psychooncology 2017 Sep;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
24. Champiat S, Lambotte O, Barreau E, et al. Management of immune checkpoint blockade dysimmune toxicities: a collaborative position paper. Ann Oncol 2016 Apr;27(4):559-574. [doi: 10.1093/annonc/mdv623] [Medline: 26715621]
25. Mayer DK, Travers D, Wyss A, Leak A, Waller A. Why do patients with cancer visit emergency departments? Results of a 2008 population study in North Carolina. J Clin Oncol 2011 Jul 1;29(19):2683-2688. [doi: 10.1200/JCO.2010.34.2816] [Medline: 21606431]
26. Vandyk AD, Harrison MB, Macartney G, Ross-White A, Stacey D. Emergency department visits for symptoms experienced by oncology patients: a systematic review. Support Care Cancer 2012 Aug;20(8):1589-1599. [doi: 10.1007/s00520-012-1459-y] [Medline: 22526151]
27. Denis F, Lethrosne C, Pourel N, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. J Natl Cancer Inst 2017 Sep 1;109(9). [doi: 10.1093/jnci/djx029] [Medline: 28423407]
28. Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
29. Bennett AV, Jensen RE, Basch E. Electronic patient‐reported outcome systems in oncology clinical practice. CA A Cancer J Clin 2012 Sep;62(5):336-347. [doi: 10.3322/caac.21150]
30. Jensen RE, Snyder CF, Abernethy AP, et al. Review of electronic patient-reported outcomes systems used in cancer clinical care. J Oncol Pract 2014 Jul;10(4):e215-e222. [doi: 10.1200/JOP.2013.001067] [Medline: 24301843]
31. Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res 2012 Oct;21(8):1305-1314. [doi: 10.1007/s11136-011-0054-x] [Medline: 22048932]
32. Iivanainen S, Alanko T, Peltola K, et al. ePROs in the follow-up of cancer patients treated with immune checkpoint inhibitors: a retrospective study. J Cancer Res Clin Oncol 2019 Mar;145(3):765-774. [doi: 10.1007/s00432-018-02835-6] [Medline: 30666409]
33. Iivanainen S, Alanko T, Vihinen P, et al. Follow-up of cancer patients receiving anti-PD-(L)1 therapy using an electronic patient-reported outcomes tool (KISS): prospective feasibility cohort study. JMIR Form Res 2020 Oct 28;4(10):e17898. [doi: 10.2196/17898] [Medline: 33112242]
34. Tolstrup LK, Bastholt L, Dieperink KB, Möller S, Zwisler AD, Pappot H. The use of patient-reported outcomes to detect adverse events in metastatic melanoma patients receiving immunotherapy: a randomized controlled pilot trial. J Patient Rep Outcomes 2020 Oct 30;4(1):88. [doi: 10.1186/s41687-020-00255-0] [Medline: 33125537]

35. Beauchamp UL, Pappot H, Holländer-Mieritz C. The use of wearables in clinical trials during cancer treatment: systematic review. JMIR Mhealth Uhealth 2020 Nov 11;8(11):e22006. [doi: 10.2196/22006] [Medline: 33174852]
36. Chung AE, Jensen RE, Basch EM. Leveraging emerging technologies and the “Internet of Things” to improve the quality of cancer care. J Oncol Pract 2016 Oct;12(10):863-866. [doi: 10.1200/JOP.2016.015784] [Medline: 27624946]
37. Tolstrup LK, Pappot H, Bastholt L, Möller S, Dieperink KB. Impact of patient-reported outcomes on symptom monitoring during treatment with checkpoint inhibitors: health-related quality of life among melanoma patients in a randomized controlled trial. J Patient Rep Outcomes 2022 Jan 21;6(1):8. [doi: 10.1186/s41687-022-00414-5] [Medline: 35061112]
38. Slev VN, Mistiaen P, Pasman HRW, Verdonck-de Leeuw IM, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: a meta-review. Int J Med Inform 2016 Mar;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]
39. Fraterman I, Glaser SLC, Wilgenhof S, et al. Exploring supportive care and information needs through a proposed eHealth application among melanoma patients undergoing systemic therapy: a qualitative study. Support Care Cancer 2022 Sep;30(9):7249-7260. [doi: 10.1007/s00520-022-07133-z] [Medline: 35589878]
40. Spahrkäs SS, Looijmans A, Sanderman R, Hagedoorn M. How does the Untire app alleviate cancer-related fatigue? A longitudinal mediation analysis. Psychooncology 2022 Jun;31(6):970-977. [doi: 10.1002/pon.5886] [Medline: 35060222]
41. Spahrkäs SS, Looijmans A, Sanderman R, Hagedoorn M. Beating cancer-related fatigue with the Untire mobile app: results from a waiting-list randomized controlled trial. Psychooncology 2020 Nov;29(11):1823-1834. [doi: 10.1002/pon.5492] [Medline: 33393199]
42. Lisowska A, Wilk S, Peleg M. SATO (IDEAS expAnded wiTh BCIO): Workflow for designers of patient-centered mobile health behaviour change intervention applications. J Biomed Inform 2023 Feb;138:104276. [doi: 10.1016/j.jbi.2022.104276] [Medline: 36586499]
43. Glaser SLC, Fraterman I, van Brummelen N, et al. Usability and usefulness testing of a symptom management and coaching system for cancer patients treated with immune checkpoint inhibitors: a comparative qualitative study. JMIR Preprints. Preprint posted online on Nov 13, 2024. [doi: 10.2196/preprints.57659]
44. Fraterman I, Wollersheim BM, Tibollo V, et al. An eHealth app (CAPABLE) providing symptom monitoring, well-being interventions, and educational material for patients with melanoma treated with immune checkpoint inhibitors: protocol for an exploratory intervention trial. JMIR Res Protoc 2023 Oct 11;12:e49252. [doi: 10.2196/49252] [Medline: 37819691]
45. ALEA Tools for Clincal Trials. URL: https://www.aleaclinical.eu/ [accessed 2025-01-21]
46. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 3;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]
47. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001 Jul;33(5):337-343. [doi: 10.3109/07853890109002087] [Medline: 11491192]
48. Winstanley JB, Saw R, Boyle F, Thompson J. The FACT-Melanoma quality-of-life instrument: comparison of a five-point and four-point response scale using the Rasch measurement model. Melanoma Res 2013 Feb;23(1):61-69. [doi: 10.1097/CMR.0b013e32835c7dd9] [Medline: 23262441]
49. Arraras JI, Greimel E, Sezer O, et al. An international validation study of the EORTC QLQ-INFO25 questionnaire: an instrument to assess the information given to cancer patients. Eur J Cancer 2010 Oct;46(15):2726-2738. [doi: 10.1016/j.ejca.2010.06.118] [Medline: 20674333]
50. Fayers P, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A. EORTC QLQ-C30 Scoring Manual, 3rd edition: European Organisation for Research and Treatment of Cancer; 2001.
51. Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. J Clin Oncol 2011 Jan 1;29(1):89-96. [doi: 10.1200/JCO.2010.28.0107] [Medline: 21098316]
52. New in stata 15. Stata. URL: https://www.stata.com/stata15/ [accessed 2025-01-21]
53. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023 Sep 18;6(1):161. [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
54. Carmona-Bayonas A, Jimenez-Fonseca P, Fernández-Somoano A, et al. Top ten errors of statistical analysis in observational studies for cancer research. Clin Transl Oncol 2018 Aug;20(8):954-965. [doi: 10.1007/s12094-017-1817-9] [Medline: 29218627]
55. Aiyegbusi OL, Roydhouse J, Rivera SC, et al. Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. Nat Commun 2022 Oct 12;13(1):6026. [doi: 10.1038/s41467-022-33826-4] [Medline: 36224187]
56. Chodankar D. Introduction to real-world evidence studies. Perspect Clin Res 2021;12(3):171-174. [doi: 10.4103/picr.picr_62_21] [Medline: 34386383]
57. van den Hurk CJG, Mols F, Eicher M, et al. A narrative review on the collection and use of electronic patient-reported outcomes in cancer survivorship care with emphasis on symptom monitoring. Curr Oncol 2022 Jun 17;29(6):4370-4385. [doi: 10.3390/curroncol29060349] [Medline: 35735458]

58. van Deursen L, Versluis A, van der Vaart R, et al. eHealth interventions for Dutch cancer care: systematic review using the Triple Aim Lens. JMIR Cancer 2022 Jun 14;8(2):e37093. [doi: 10.2196/37093] [Medline: 35699991]
59. van der Hout A, van Uden-Kraan CF, Holtmaat K, et al. Role of eHealth application Oncokompas in supporting self-management of symptoms and health-related quality of life in cancer survivors: a randomised, controlled trial. Lancet Oncol 2020 Jan;21(1):80-94. [doi: 10.1016/S1470-2045(19)30675-8] [Medline: 31838009]
60. Cuypers M, Lamers RED, Kil PJM, van de Poll-Franse LV, de Vries M. Impact of a web-based prostate cancer treatment decision aid on patient-reported decision process parameters: results from the Prostate Cancer Patient Centered Care trial. Support Care Cancer 2018 Nov;26(11):3739-3748. [doi: 10.1007/s00520-018-4236-8] [Medline: 29752528]
61. Mumtaz H, Riaz MH, Wajid H, et al. Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. Front Digit Health 2023;5:1203945. [doi: 10.3389/fdgth.2023.1203945] [Medline: 37840685]
62. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120. [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
63. van de Poll-Franse LV, de Rooij BH, Horevoorts NJE, et al. Perceived care and well-being of patients with cancer and matched norm participants in the COVID-19 crisis: results of a survey of participants in the Dutch PROFILES Registry. JAMA Oncol 2021 Feb 1;7(2):279-284. [doi: 10.1001/jamaoncol.2020.6093] [Medline: 33237294]

Abbreviations

AE: adverse event
CAPABLE: Cancer Patients Better Life Experience
CRF: cancer-related fatigue
EORTC: European Organization for Research and Treatment of Cancer
EQ-5D-5L: EuroQol-5D-5L
FACT-M: Functional Assessment of Cancer Therapy-Melanoma
HCP: health care professional
HRQoL: health-related quality of life
ICI: immune-checkpoint inhibitor
MS: melanoma subscale
MSS: melanoma surgery subscale
PRO: patient-reported outcome
PRO-MEL: patient-reported outcomes in high risk and advanced melanoma patients cohort
PROM: patient-reported outcome measure
QLQ-C30: Quality of Life Questionnaire-core 30
QLQ-INFO25: Quality of Life Questionnaire-Information 25
RCT: randomized controlled trial

Edited by N Cahill; submitted 28.03.24; peer-reviewed by E Nagele, L Kjær; revised version received 27.11.24; accepted 27.11.24; published 30.01.25.

Please cite as:
Fraterman I, Sacchi L, Mallo H, Tibollo V, Glaser SLC, Medlock S, Cornet R, Gabetta M, Hisko V, Khadakou V, Barkan E, Del Campo L, Glasspool D, Kogan A, Lanzola G, Leizer R, Ottaviano M, Peleg M, Śniatała K, Lisowska A, Wilk S, Parimbelli E, Quaglini S, Rizzo M, Locati LD, Boekhout A, van de Poll-Franse LV, Wilgenhof S
Exploring the Impact of the Multimodal CAPABLE eHealth Intervention on Health-Related Quality of Life in Patients With Melanoma Undergoing Immune-Checkpoint Inhibition: Prospective Pilot Study
JMIR Cancer 2025;11:e58938
URL: https://cancer.jmir.org/2025/1/e58938
doi:10.2196/58938

A Machine Learning Approach Using Topic Modeling to Identify and Assess Experiences of Patients With Colorectal Cancer: Explorative Study

Kelly Voigt[1], MD, PhD; Yingtao Sun[2], MSc; Ayush Patandin[3], MSc; Johanna Hendriks[4], MD, PhD; Richard Hendrik Goossens[2], PhD; Cornelis Verhoef[1], MD, PhD; Olga Husson[1,5], PhD; Dirk Grünhagen[1*], MD, PhD; Jiwon Jung[2,4*], PhD

[1]Department of Surgical Oncology and Gastrointestinal Surgery, Erasmus Medical Centre Cancer Institute, Rotterdam, Netherlands

[2]Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands

[3]Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, Netherlands

[4]Department of Surgery, Erasmus Medical Centre, Rotterdam, Netherlands

[5]Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Jiwon Jung, PhD
Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands

Abstract

Background: The rising number of cancer survivors and the shortage of health care professionals challenge the accessibility of cancer care. Health technologies are necessary for sustaining optimal patient journeys. To understand individuals' daily lives during their patient journey, qualitative studies are crucial. However, not all patients wish to share their stories with researchers.

Objective: This study aims to identify and assess patient experiences on a large scale using a novel machine learning–supported approach, leveraging data from patient forums.

Methods: Forum posts of patients with colorectal cancer (CRC) from the Cancer Survivors Network USA were used as the data source. Topic modeling, as a part of machine learning, was used to recognize the topic patterns in the posts. Researchers read the most relevant 50 posts on each topic, dividing them into "home" or "hospital" contexts. A patient community journey map, derived from patients stories, was developed to visually illustrate our findings. CRC medical doctors and a quality-of-life expert evaluated the identified topics of patient experience and the map.

Results: Based on 212,107 posts, 37 topics and 10 upper clusters were produced. Dominant clusters included "Daily activities while living with CRC" (38,782, 18.3%) and "Understanding treatment including alternatives and adjuvant therapy" (31,577, 14.9%). Topics related to the home context had more emotional content compared with the hospital context. The patient community journey map was constructed based on these findings.

Conclusions: Our study highlighted the diverse concerns and experiences of patients with CRC. The more emotional content in home context discussions underscores the personal impact of CRC beyond clinical settings. Based on our study, we found that a machine learning-supported approach is a promising solution to analyze patients' experiences. The innovative application of patient community journey mapping provides a unique perspective into the challenges in patients' daily lives, which is essential for delivering appropriate support at the right moment.

(JMIR Cancer 2025;11:e58834) doi:10.2196/58834

KEYWORDS

colorectal cancer; forum; topic modeling; patient journey; patient experience; AI; machine learning; cancer care; cancer survivor; United States; quality of life; post; topic; artificial intelligence

Introduction

Colorectal cancer (CRC), the third most common cancer in the Netherlands significantly impacts the health of individuals [1]. Supportive care involves adopting a person-centered care approach, intending to offer individuals affected by cancer the essential services required to address their informational, emotional, social, and physical needs and concerns throughout the entire patient journey [2,3]. Understanding the needs and concerns of those affected by this disease during their patient journey is crucial for improving patient outcomes and quality of care [4]. Several survey and qualitative studies involving

CRC survivors offer valuable information about the needs and perceptions of patients [4-7]. However, these survey- and qualitative studies are limited by sample size, labor-intensive processes, and could lead to socially desirable answers [8]. Efforts were made to include vulnerable populations who may not be accessible through focus groups, ensuring a more comprehensive understanding of diverse patient perspectives [9]. In the last decade, patient web forums have emerged as valuable platforms for individuals to openly share their experiences and thoughts related to CRC, providing unique insights into the social, physical, and emotional aspects of their patient journey [10,11]. These forums offer different opinions compared with traditional patient experience data collection methods, such as questionnaires and interviews.

While patient community forums contain a wealth of information, the analysis of the extensive unstructured data within these forums poses a considerable challenge [12]. Traditional manual qualitative analysis by human experts is time-consuming, labor-intensive, and lacks scalability, making it impractical to analyze the sheer volume of patient-generated unstructured content. Machine learning techniques offer a potential solution to address this challenge. This enables the automated processing and analysis of textual data, allowing for the efficient extraction and interpretation of large amounts of patient forum posts [13,14].

Our primary objective is to assess patient experiences using a novel machine learning–supported approach and data from patient forums. To achieve this, we used patient community journey mapping to better understand the experiences throughout the patient journey [15]. This is a machine learning-driven approach, that uses web-based patient forums as input data and is processed through topic modeling. By gaining insights into these patient experiences, we can shape future patient journeys such as remote monitoring systems to be aligned with current patients' needs.

Methods

Ethical Considerations

Ethical approval was given by the human research ethics committee of the Delft University of Technology (ID 2596). Informed consent was not required since the data used in this study were sourced from publicly available forums, in accordance with institutional guidelines. The data were accessed and analyzed in accordance with the terms of service and privacy policies outlined by the platform hosting the data. As the ethics committee suggested, all data used in this study were anonymized by removing all direct and indirect identifiers (eg, names, location, and user ID) to prevent potential privacy issues within our data set. Confidentiality was strictly maintained during data collection, analysis, and reporting to ensure that no identifiable information was disclosed.

In order to assess patient experiences using machine learning–driven analysis, this study used systematic data collection, advanced topic modeling techniqu and cocreation sessions with domain experts to interpret and validate identified clusters. The Consolidated Reporting Guidelines for prognostic and diagnostic machine learning modeling studies checklist can be found in Multimedia Appendix 1.

Data Source

To enable research, we used forum posts of patients with CRC scraped from the Cancer Survivors Network USA, an open-source patient community platform. The web-based platform provides support, education, and advocacy for patients affected by CRC that includes current patients, former patients, families, and caregivers. The posts are written with the intention of asking questions to peers and health care providers, as well as forsharing experiences among peers [16]. The initial CRC discussion thread (ie, main posts, comments, and replies) has remained active on the public platform since the year 2000. No distinctions were made between main thread posts and replies to posts.

The data collection involves 2 main steps: first involves using Selenium WebDriver to gather the URLs of discussion pages, and second, using BeautifulSoup to extract data from the HTML elements while keeping sensitive information secure.

Data Analysis (Topic Modeling)

Information about the personal (health) status of patients, which could be directly or indirectly identifiable, was excluded from the analysis. Topic modeling, a machine learning algorithm was used to recognize patterns in the platform data. Nonnegative matrix factorization (NMF) was used as a topic modeling technique to analyze the data set and identify topics with weighted keywords [17,18]. The number of topics was determined by evaluating topic coherence scores and model stability [19,20]. Model stability was assessed using Jaccard similarity to reduce the overlap of topics. Human evaluation was used to ensure that the output contains diverse and distinctive topics, without topics unrelated to patient experiences. This evaluation was conducted by 4 researchers to select the number of topics that yield the most diverse and distinctive clusters.

Interpretation of Data Analysis (Qualitative Analysis)

To comprehend each topic, researchers meticulously read and analyzed the top 50 most relevant posts and top 20 related keywords. The most relevant posts and keywords had the highest dominant topic score identified by the algorithm, indicating they were most closely associated with the topic. The algorithm efficiently identified the most relevant posts and keywords within each topic by using Term Frequency–Inverse Document Frequency to calculate the importance of each term in relation to each document [21].

Clusters

The identified topics were grouped into clusters for topics that revolved around similar concerns. Irrelevant topics (eg, those related to platform-specific issues) were excluded. The frequency of discussions on each topic was recorded by counting the number of posts associated with each topic.

Creation of Patient Community Journey Map

This study uses patient community journey mapping to visually present the identified patient experience topics. The creation of

the patient community journey map began with the conduction of 2 interviews with CRC medical doctors to outline the patient journey and its distinct phases such as diagnosis and follow-up of a patient with CRC.

The results of the interpreted topics were deliberated in an in-person cocreation session involving domain experts [22]. The cocreation session took place in the Erasmus Medical Centre in March 2023 and lasted 3 hours. The interdisciplinary team of domain experts reviewed and discussed the outputs comprehensively. The team comprised 2 oncological surgeons, 2 surgical oncology PhD candidates (MD), and an epidemiologist with an interest in oncological research. A total of 3 out of 5 participants have over 20 years of work experience, indicating a high level of expertise. During the cocreation session, clusters from the NMF results were interpreted, and feedback was provided on the generated interpretations. This process ensured a thorough examination and validation of the identified clusters within a clinically relevant context. During this session, the same domain experts also reviewed the patient community journey map. The identified topics were positioned on the journey map, with allocation to specific phases of the patient journey. Distinctions were made regarding whether each topic was taking place at a “hospital” or “home”setting.

Results

Data

A total of 294,166 posts were extracted from the patient forums of the Cancer Survivors Network USA. The posts on the website were created between the year 2000 and 2022. However, 212,107 posts were analyzed, as the remaining 82,059 posts were excluded through topic modeling.

Topic Modeling

Using NMF topic modeling, the topic coherence score, and model stability did not provide clear insights as the coherence score consistently declined. Therefore, human evaluation was required. Initially, 40 and 50 topics were considered the best amount of topics. Subsequently, the cut-off of 45 topics was also reviewed. Ultimately, 50 topics were chosen as the number that provided the most diverse and distinctive topics. This was confirmed by examining keywords and the most relevant posts for each cluster. Additionally, one topic bin identified 82,059 posts as unclassifiable," which could not be categorized into any of the 50 topics.

Identified Patient Experience Topics and Patient Community Journey Map

A total of 50 topics were identified with the use of NMF topic modeling. Thirteen topics were excluded as they were unrelated to patient experiences, such as platform use and expressions of gratitude. The data export of the relevant topics is shown in Table 1. These are the key patient experiences found in our study.

A total of 10 clusters were derived from 37 topics. The patient community journey map, shown in Figure 1, serves as a visual guide to navigate through the nuanced dynamics of patient experiences.

The topics in the home context have a more emotional content, as emotionally charged keywords such as “confused,” “bad,”and “worried” are more often discussed. Conversely, the hospital context is marked by the clinical terminologies “drug” and “node.” This reflects the distinct atmospheres characterizing discussions in different contexts.

The clusters commanding the highest share of posts are “Daily activities while living with CRC” (38,782, 18.3%) and, “Understanding treatment including alternatives and adjuvant therapy” (31,577, 14.9%) underscoring their important role in shaping the experiences of CRC survivors. Patients expressed significant concerns about test results, as indicated by the 14.1% of posts within cluster 4. As shown in Figure 1, “Understanding treatment including alternatives and adjuvant therapy” is a cluster that goes through almost the entire journey, while the “Daily activities while living with colorectal cancer,” is more in the home context and starts later in the follow-up.

Table . Overview of the clusters derived from the topics and their keywords.

Clusters and topics, (n, %)[a] (N=212,107)	Top 20 keywords	Number of posts[b], n
Cluster 1: Experience around medical professionals' opinion (9756, 4.6%)		
Doubts about treatment options from medical professionals	'doctor', 'oncologist', 'surgeon', 'told', 'patient', 'system', 'cell', 'wrong', 'reason', 'medical', 'asked', 'office', 'immune', 'trust', 'kill', 'testing', 'medicine', 'appointment', 'clinic', 'recommend'	5043
Suggestion to look for a second medical opinion	'care', 'second', 'opinion', 'ask', 'port', 'onc', 'put', 'taken', 'nurse', 'center', 'forget', 'top', 'first', 'team', 'comfortable', 'third', 'schedule', 'getting', 'question', 'hospital'	4713
Cluster 2: Understanding treatment including alternatives and adjuvant therapy (31,577, 14.9%)		
Patients share their research about alternative therapy options from websites and articles	'answer', 'information', 'might', 'research', 'alternative', 'therapy', 'study', 'perhaps', 'available', 'patient', 'consider', 'website', 'based', 'benefit', 'article', 'option', 'approach', 'internet', 'current', 'specific'	7854
Making treatment decisions for the future with regard to the best outcome and path	'best', 'wish', 'wishing', 'decision', 'health', 'possible', 'whatever', 'future', 'choice', 'outcome', 'decide', 'path', 'upcoming', 'regard', 'situation', 'advocate', 'choose', 'action', 'simply', 'circumstance'	4759
Listing type, side effect, regimen, and effectiveness of drugs	'side', 'effect', 'drug', 'folfox', 'round', 'cycle', 'dose', 'vitamin', 'week', 'oxaliplatin', 'avastin', 'reaction', 'irinotecan', 'affect', 'rash', 'onc', 'cell', 'regimen', 'session', 'effective'	8768
Sharing experience on using traditional Chinese medicine to manage health	'well', 'known', 'fairly', 'usual', 'responded', 'recall', 'crc', 'chinese', 'handling', 'version', 'correctly', 'content', 'deserved', 'treating', 'referring', 'nicely', 'handled', 'spelling', 'managing', 'tolerated'	3610
Share experiences and recommendations for supplements and medication	'congratulation', 'definitely', 'tried', 'experience', 'mentioned', 'suggestion', 'type', 'helped', 'mention', 'medication', 'interested', 'speak', 'form', 'help', 'suggest', 'using', 'supplement', 'suggested', 'detail', 'order'	2854
Sharing information about and experiences regarding clinical trials	'new', 'find', 'look', 'looking', 'awesome', 'forward', 'trial', 'start', 'working', 'step', 'move', 'hearing', 'hard', 'pic', 'yet', 'clinical', 'finding', 'meeting', 'hopeful', 'totally'	3732
Cluster 3: Surgery experience (9540, 4.5%)		
Sharing experience around radiofrequency ablation for the liver	'congrats', 'procedure', 'success', 'huge', 'liver', 'proud', 'shrink', 'treat', 'entire', 'tumor', 'ablation', 'option', 'rfa', 'candidate', 'location', 'status', 'shrinkage', 'inoperable', 'surgical', 'afterwards'	4979

Clusters and topics, (n, %)[a] (N=212,107)	Top 20 keywords	Number of posts[b], n
Sharing negative experience about resection in liver and lung	‘day’, ‘liver’, ‘resection’, ‘mets’, ‘every’, ‘met’, ‘pump’, ‘lung’, ‘ahead’, ‘week’, ‘couple’, ‘following’, ‘open’, ‘two’, ‘operation’, ‘followed’, ‘throughout’, ’single’, ‘complication’, ’section’	4561
Cluster 4: Experience regarding the test results (including being worried and confused) (29,984, 14.1%)		
Sharing negative emotions and experiences to live with colorectal cancer (CRC): overwhelmed, confused and scared, especially for the tests and screenings	‘like’, ’sound’, ‘wonderful’, ‘plan’, ‘place’, ‘picture’, ’show’, ‘meet’, ‘report’, ‘looked’, ‘absolutely’, ‘really’, ‘pick’, ‘button’, ‘familiar’, ‘machine’, ‘pretty’, ’smart’, ’silly’, ‘compare’	3873
Being worried about upcoming scans and results	‘great’, ‘result’, ‘idea’, ’scan’, ‘number’, ‘option’, ‘curious’, ‘present’, ‘follow’, ‘considering’, ‘treated’, ‘recurrence’, ‘appendix’, ‘initial’, ‘cea’, ‘mop’, ‘peritoneal’, ‘assuming’, ‘base’, ‘reliable’	3210
Share their outcomes (clear or not) from scanning and caring about how frequently they scan	‘year’, ‘month’, ‘ago’, ‘clear’, ‘last’, ‘three’, ‘two’, ‘free’, ’scan’, ‘colonoscopy’, ‘end’, ‘past’, ‘every’, ‘four’, ‘clean’, ‘date’, ’six’, ’safe’, ‘behind’, ‘later’	4293
Describing a stressful experience in a blood test; worry about the numbers in the result	‘everyone’, ‘word’, ‘blood’, ‘test’, ‘kind’, ‘wait’, ‘worry’, ‘fine’, ‘check’, ‘waiting’, ‘count’, ‘normal’, ‘wanted’, ‘level’, ‘concern’, ‘checked’, ‘else’, ‘low’, ‘high’, ‘appointment’	5048
Being worried and confused about odd scan results in liver, lung, and lymph	’scan’, ‘tumor’, ‘liver’, ‘node’, ‘lung’, ‘removed’, ‘radiation’, ’stage’, ‘lymph’, ’spot’, ’showed’, ’spread’, ’small’, ‘pet’, ‘biopsy’, ‘remove’, ‘mets’, ’surgeon’, ‘recurrence’, ‘week’	13,560
Cluster 5: Experience with side effects (26,152, 12.3%)		
Sharing their experience on managing the side effects of treatment	‘cold’, ‘nausea’, ‘taking’, ‘help’, ‘water’, ‘drink’, ‘warm’, ‘pill’, ‘infusion’, ’sleep’, ‘mouth’, ’sore’, ‘med’, ‘oxy’, ‘gave’, ‘fatigue’, ‘mum’, ‘nasty’, ‘appetite’, ’sleeping’	8341
Negative feelings of hair loss due to cancer treatments	’stuff’, ‘hair’, ‘funny’, ‘hate’, ‘bad’, ’suck’, ‘lose’, ‘cut’, ‘fall’, ‘wear’, ‘crap’, ‘lost’, ‘really’, ‘made’, ’stand’, ‘head’, ‘weird’, ‘losing’, ’strange’, ‘air’	5122
Feeling uncomfortable due to the obstipation	‘pain’, ‘control’, ‘nothing’, ‘relief’, ‘med’, ‘bowel’, ‘issue’, ‘breath’, ‘walking’, ‘walk’, ’scar’, ‘deep’, ‘kidney’, ‘causing’, ‘blockage’, ‘problem’, ‘intestine’, ‘tube’, ‘hospital’, ‘hernia’	6414

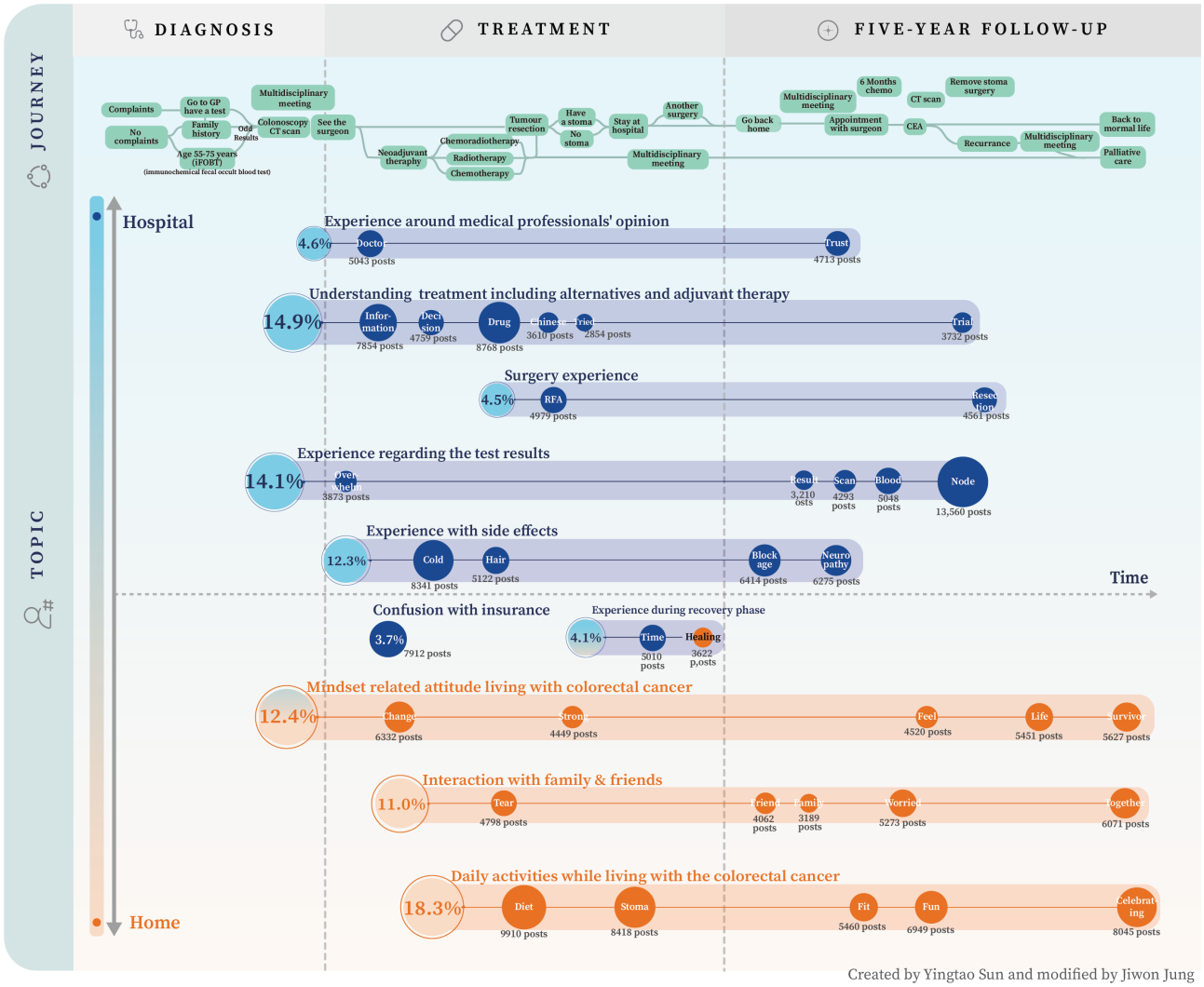
Clusters and topics, (n, %)[a] (N=212,107)	Top 20 keywords	Number of posts[b], n
Experiencing pain and numbness due to neuropathy	'hand', 'foot', 'hurt', 'neuropathy', 'bone', 'worse', 'arm', 'pain', 'caused', 'leg', 'left', 'painful', 'problem', 'experienced', 'damage', 'nerve', 'cause', 'related', 'shoulder', 'symptom'	6275
Cluster 6: Confusion with insurance (7912, 3.7%)		
Confusion about insurance coverage	'insurance', 'really', 'sent', 'email', 'needed', 'pay', 'company', 'money', 'state', 'phone', 'cost', 'card', 'mail', 'received', 'hospital', 'address', 'cover', 'attention', 'letter', 'service'	7912
Cluster 7: Experience during recovery phase (8632, 4.1%)		
Describing negative experience of repeated visits to the hospital	'time', 'took', 'full', 'week', 'short', 'last', 'shot', 'infection', 'heal', 'ended', 'went', 'due', 'drop', 'period', 'several', 'stopped', 'hospital', 'started', 'kept', 'taking'	5010
Staying positive and making life adjustments to their cancer circumstances during the recovery phase	'way', 'positive', 'sending', 'coming', 'along', 'vibe', 'healing', 'energy', 'half', 'begin', 'rid', 'improve', 'ton', 'complete', 'outlook', 'faster', 'headed', 'toward', 'improved', 'adjustment'	3622
Cluster 8: Mindset-related attitude living with CRC (26,379, 12.4%)		
Difficulties to adjust and adapt to their lives with CRC	'going', 'time', 'sometimes', 'enough', 'trying', 'probably', 'part', 'really', 'actually', 'course', 'change', 'either', 'pretty', 'different', 'body', 'point', 'never', 'hard', 'anyway', 'whole'	6332
Sharing how patients can be resilient and positive	'always', 'stay', 'remember', 'away', 'week', 'next', 'cry', 'laugh', 'sense', 'strong', 'forever', 'matter', 'extra', 'humor', 'staying', 'never', 'whenever', 'hero', 'sweetie', 'corner'	4449
Sharing their feelings: sick, tired, weak, and bad	'feel', 'better', 'feeling', 'soon', 'hopefully', 'getting', 'sick', 'felt', 'tired', 'real', 'making', 'really', 'starting', 'bad', 'stronger', 'gotten', 'start', 'expected', 'weak', 'normal'	4520
Sharing their positive philosophical thoughts about living with cancer	'life', 'live', 'mean', 'understand', 'people', 'important', 'living', 'world', 'fear', 'time', 'become', 'realize', 'death', 'die', 'quality', 'learn', 'focus', 'illness', 'save', 'allow'	5451
Survivors sharing their attitudes towards living with cancer along with survival rate	'story', 'old', 'survivor', 'disease', 'folk', 'site', 'year', 'member', 'cure', 'people', 'recently', 'personal', 'passed', 'woman', 'remission', 'given', 'survival', 'grateful', 'alive', 'lived'	5627
Cluster 9: Interaction with family and friends (23,393, 11%)		

Clusters and topics, (n, %)[a] (N=212,107)	Top 20 keywords	Number of posts[b], n
Family members' emotional struggle about having a cancer patient in their family	'husband', 'heart', 'face', 'eye', 'hold', 'head', 'tear', 'dog', 'stop', 'brought', 'hell', 'front', 'putting', 'child', 'soul', 'biggest', 'attack', 'mad', 'horse', 'men'	4798
Experience regarding relationships with friends while having cancer	'got', 'friend', 'heard', 'name', 'room', 'semi', 'never', 'bunch', 'forgot', 'hot', 'blue', 'chat', 'surprise', 'joke', 'facebook', 'picked', 'neighbor', 'girlfriend', 'Canadian', 'everywhere'	4062
Sharing changes in their family relationship due to cancer journey	'thinking', 'today', 'family', 'anyone', 'call', 'wondering', 'sister', 'called', 'brother', 'visit', 'yesterday', 'talk', 'close', 'else', 'sign', 'talking', 'law', 'wanted', 'concerned', 'appt'	3189
Being worried about their family members and seeking information on family history (eg, genetic testing)	'mom', 'dad', 'agree', 'mother', 'age', 'never', 'worried', 'father', 'stage', 'breast', 'symptom', 'died', 'diagnosed', 'yr', 'turned', 'happen', 'parent', 'told', 'tested', 'child'	5273
Arguing the importance to spend moments with family members during the cancer journey	'night', 'wife', 'daughter', 'kid', 'hour', 'beautiful', 'morning', 'together', 'friday', 'every', 'minute', 'day', 'moment', 'late', 'last', 'bed', 'home', 'thursday', 'gift', 'time'	6071
Cluster 10: Daily activities while living with the CRC (38,782, 18.3%)		
Sharing suggestions on diets focused on balanced meals and healthy alternatives	'eat', 'food', 'cheer', 'diet', 'eating', 'juice', 'sugar', 'juicing', 'favorite', 'fruit', 'glass', 'perfect', 'dream', 'green', 'drink', 'red', 'wine', 'meat', 'coffee', 'heck'	9910
Suggesting how to take good care of a stoma.	'bag', 'problem', 'add', 'colostomy', 'used', 'daily', 'ostomy', 'twice', 'radiation', 'reversal', 'bathroom', 'ileostomy', 'skin', 'diarrhea', 'prep', 'stool', 'stoma', 'rectum', 'careful', 'bowel'	8418
Sharing experience on being fit again, caring about weight control	'back', 'came', 'weight', 'exercise', 'return', 'set', 'running', 'self', 'track', 'pound', 'door', 'gain', 'lost', 'lb', 'fit', 'buzzard', 'sit', 'key', 'floor', 'put'	5460
Sharing ways to stay in a positive mood through planning for distractions	'going', 'enjoy', 'weekend', 'fun', 'trip', 'ready', 'tonight', 'vacation', 'game', 'watch', 'party', 'drive', 'house', 'weather', 'excited', 'car', 'town', 'near', 'planning', 'movie'	6949
Celebrating anniversaries and birthdays for patients as a big milestone of their lives	'happy', 'birthday', 'dance', 'holiday', 'healthy', 'yea', 'anniversary', 'thanksgiving', 'naked', 'celebrating', 'celebration', 'filled', 'happiness', 'raise', 'dancing', 'cake', 'spongebob', 'ending', 'celebrated', 'scouty'	8045

[a]Number of topics in each cluster.

[b]Number of posts related to the topic group.

Figure 1. Community journey map of patients with colorectal cancer. CEA: carcinoembryonic antigen; CT: computed tomography; GP: general practitioner; RFA: radiofrequency ablation.



Discussion

Principal Findings

This study delves into the daily experiences of patients using data from a web-based platform, applying topic modeling as a tool to efficiently analyze and interpret large volumes of patient-driven daily life experience data. We analyze patient experiences (eg, struggles, tips, and coping strategies) reported on a platform to better understand patient needs. This understanding of patients' needs has positive potential for future care path development such as information provision or psychosocial support in remote patient monitoring. The use of machine learning techniques in analyzing qualitative data holds significant promise to improve various aspects of health care. This includes optimizing patient care, by using advanced methods such as using symptom checkers to guide patients to the most suitable care journeys and support disease diagnosis. This approach will ultimately lead to patient-centered care to support and empower patients.

The identified clusters in this study align with findings from previous qualitative and survey research, emphasizing the potential to optimize the use of machine learning in qualitative studies of health care in the years to come [4,23]. Based on our analysis, it becomes evident that patients exhibit a notable interest in actively participating in discussions with other CRC survivors to exchange experiences regarding the daily challenges of living with CRC. This underscores the crucial role of social support and the dissemination of pragmatic, day-to-day coping strategies within the broader patient community. This theme aligns with observations from a systematic review detailing the experiences of CRC survivors, where the significance of coping and addressing functional limitations emerged as a central theme [23].

The cluster with the highest share of posts (cluster 10; 38,782, 18.3%) revolves around the challenges associated with resuming daily activities while living with CRC. This substantial volume of posts highlights the significant impact of this particular aspect on the experiences of CRC survivors, as well as the needed change in patients' mindset (cluster 8; 26,379/212,1087, 12.4%). The systematic review of survivorship experiences further supports these findings [23]. The alignment between these dominant themes in the posts and the insights from the qualitative literature underscores the robustness of our findings [24]. These findings emphasize the importance of integrating

comprehensive support mechanisms to facilitate the successful reintegration of patients into their daily lives.

On the platform, patients frequently engage in discussions centered around understanding treatment options (cluster 2; 31,577, 14.9%), highlighting its significance as a key topic. This observation aligns with the qualitative study conducted by van Deursen et al [4], where participants similarly emphasized the need for more information regarding their treatment. It is noteworthy that patients on the platform also frequently express concerns regarding test results (cluster 4; 29,984, 14.1%). This anxiety and fear of recurrence are recognized as common concerns among cancer survivors [25], highlighting the potential to offer valuable support in addressing this. Additionally, the positive outcomes associated with the interaction with family, as reported in the systematic review [23], further underscore the importance of familial support within the context of patient experiences in our findings [26].

A novel finding from our analysis was the open expression of patients' sentiments regarding the opinions of medical professionals (cluster 1; 9756, 4.6%). Not completely surprising as this is possibly due to the reluctance of patients to express negative feelings in a setting where researchers are present, such as focus groups or interviews. Furthermore, qualitative studies primarily concentrate on the follow-up phase, excluding discussions on the experiences of surgery and the subsequent recovery phase. These aspects might provide valuable insights into the comprehensive journey of patients.

The Added Value of Machine Learning to Support the Understanding of Patient Experiences

To reflect on our approach, topic modeling allows extraction from a vast pool of large-scale patient data. The alignment of our study with existing literature is indeed promising. Additionally, topic modeling offers a novel approach by sparing patients the burden of active participation in qualitative studies. Rather than focusing on a single aspect, as is often the case in conventional studies, this method presents a broader spectrum of patient experiences beyond specific questions from researchers. Notably, topic modeling relies on patient-driven data, in contrast to a physician-driven approach. This method's replicability allows seamless comparisons between various medical conditions. This not only contributes to a more extensive comprehension of diverse experiences but also facilitates the identification of commonalities and distinctions of different medical conditions. For example, we explored the present method with patients with pulmonary fibrosis and found comparable experiences to those with CRC, their increased focus on treatment options highlights the distinctive challenges inherent to their specific condition (unpublished data). Given the large data sets on forums, we can quantify patients' experiences with the implementation of topic modeling. This complements the challenges of focus groups and interviews, where the ability to quantify and prioritize findings is limited. As patients' experiences change throughout their journey, the simplicity of using topic modeling remains a flexible and efficient approach for understanding patients' experiences and priorities in the era of remote monitoring [27]. It offers a convenient and adaptable method, without burdening patients [28].

Limitations

Using stories from patients who prefer to share their experiences on a digital platform introduces a risk of selection bias. It is essential to address the potential overexpression of emotion in our findings as the platform primarily captures narratives from individuals who are willing to share their experiences on the web. However, it should be recognized that these individuals represent the intended user group for new care path developments such as remote monitoring support approaches. Patients who are not feeling well, patients who do not have access to the internet, or reserved patients may be less inclined to engage in web-based discussions and might be underrepresented [29]. To complement this, in previous research, we conducted interviews with vulnerable patient populations to ensure their perspectives were included in our understanding of patient experiences [9]. Similar themes such as the need for information provision, concerns about test results, and challenges in reintegrating into daily life were found in this study. Additionally, our approach demands a significant investment of human labor and time to interpret the meaning of topics after generating groups of topics. Thus, exploring another artificial intelligence technique to interpret patient experiences is a promising direction. As an example, a large language model, a pretrained model, can be used to interpret and summarize large quantity textual data [30]. The use of machine learning, even in the part where we used human experts' knowledge, is a promising area to labor-effectively understand the patients' experience while having the least safeguarding support of human validation. Another limitation is the lack of direct involvement of patients with CRC in the co-design or review of the map. We plan to address this in the future as this can provide valuable insights. However, the primary purpose of this method is to enhance the methodology while minimizing the burden on patients. To overcome this limitation, we sought input from highly experienced medical professionals' perspectives. Three of these experts possessed more than 20 years of medical and research experience, specifically with patients with CRC.

In future research, efforts should be directed towards further refining machine learning techniques with the ultimate aim of minimizing or eliminating the need for human involvement, as this will enable more frequent monitoring of patient experiences and thereby facilitate responsiveness to those experiences. Addressing the nuanced interpretation of emotions in patients remains a challenge for artificial intelligence systems. A promising direction is the automatic interpretation of patient-reported outcome measures using machine learning and NLP. Patient-reported outcome measures offer structured data that can improve AI's ability to understand emotional nuances. This approach aims to improve remote patient monitoring, reduce the burden on health care professionals, and identify those who face psychosocial challenges [31]. This approach not only empowers patients in managing their health but also leads to a proactive and personalized approach.

In conclusion, topic modeling and the use of patient forum data offer a robust and efficient approach to understanding patients' experiences in their daily lives. This approach reveals the challenges patients encounter in their daily life such as getting back to daily activities and the need for understanding their treatment. This study not only provides a comprehensive overview of patient experiences through web-based platforms but also highlights the potential for improving patient monitoring systems. This machine learning technique of identifying patient experiences contributes to a more efficient way of building value-based health care. By integrating these insights into the development of remote monitoring solutions, a patient-centered approach can be created that not only addresses medical concerns but also caters to the broader spectrum of challenges individuals face in their daily lives.

Acknowledgments

We extend our gratitude to all the individuals who shared their experiences on the Cancer Survivors Network USA, an open-source patient community platform. This provided valuable insights for this study. Generative artificial intelligence was not used in the writing process.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consolidated reporting guidelines for prognostic and diagnostic machine learning modeling studies
[PDF File, 200 KB - cancer_v11i1e58834_app1.pdf]

References

1. Cijfers darmkanker [numbers colorectal cancer]. IKNL [Netherlands comprehensive cancer organisation].: IKNL URL: https://iknl.nl/kankersoorten/darmkanker/registratie [accessed 2024-12-19]
2. Hui D. Definition of supportive care: does the semantic matter? Curr Opin Oncol 2014 Jul;26(4):372-379. [doi: 10.1097/CCO.0000000000000086] [Medline: 24825016]
3. Niedzwiedz CL, Knifton L, Robb KA, Katikireddi SV, Smith DJ. Depression and anxiety among people living with and beyond cancer: a growing clinical and research priority. BMC Cancer 2019 Oct 11;19(1):943. [doi: 10.1186/s12885-019-6181-4] [Medline: 31604468]
4. van Deursen L, Aardoom JJ, Alblas EE, Struijs JN, Chavannes NH, van der Vaart R. Exploring colorectal cancer survivors' perspectives on improving care delivery and the role of e-health technology: a qualitative study. Support Care Cancer 2023 Aug 31;31(9):544. [doi: 10.1007/s00520-023-08007-8] [Medline: 37650936]
5. Bahrami M, Masoumy M, Sadeghi A, Mosavizadeh R. The needs of colorectal cancer patients/survivors: a narrative review. J Educ Health Promot 2022;11(227):227. [doi: 10.4103/jehp.jehp_1093_21] [Medline: 36177430]
6. Rutherford C, Müller F, Faiz N, King MT, White K. Patient-reported outcomes and experiences from the perspective of colorectal cancer survivors: meta-synthesis of qualitative studies. J Patient Rep Outcomes 2020 Apr 25;4(1):27. [doi: 10.1186/s41687-020-00195-9] [Medline: 32335745]
7. Drury A, Payne S, Brady AM. Colorectal cancer survivors' quality of life: a qualitative study of unmet need. BMJ Support Palliat Care 2020 Jun 29:bmjspcare-2020-002190. [doi: 10.1136/bmjspcare-2020-002190] [Medline: 32601151]
8. Collier D, Mahoney J. Insights and pitfalls: selection bias in qualitative research. World Pol 1996 Oct;49(1):56-91. [doi: 10.1353/wp.1996.0023]
9. Voigt KR, de Bruijn EA, Wullaert L, et al. Assessing patients' needs in the follow-up after treatment for colorectal cancer-a mixed-method study. Support Care Cancer 2024 Feb 26;32(3):192. [doi: 10.1007/s00520-024-08401-w] [Medline: 38409637]
10. Pellino G, Simillis C, Qiu S, et al. Social media and colorectal cancer: a systematic review of available resources. PLoS One 2017;12(8):e0183031. [doi: 10.1371/journal.pone.0183031] [Medline: 28832603]
11. Dosani S, Harding C, Wilson S. Online groups and patient forums. Curr Psychiatry Rep 2014 Nov;16(11):507. [doi: 10.1007/s11920-014-0507-3] [Medline: 25273668]
12. Tanwar M, Duggal R, Khatri SK. Unravelling unstructured data: a wealth of information in big data. Presented at: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions); Sep 2-4, 2015; Noida, India. [doi: 10.1109/ICRITO.2015.7359270]
13. Jung J, Kim K, Peters T, Snelders D, Kleinsmann M. Advancing design approaches through data-driven techniques: patient community journey mapping using online stories and machine learning. Int J Design 2023;17(2):19-44.

14. Bahja M. Identifying patient experience from online resources via sentiment analysis and topic modelling approaches. Presented at: 39th International Conference on Information Systems; 2018; San Francisco. [doi: 10.1145/3006299.3006335]
15. Peters T, Snelders D, Kleinsmann M. Advancing design approaches through data-driven techniques: patient community journey mapping using online stories and machine learning. Int J Des 2023. [doi: 10.57698/v17i2.02]
16. Colorectal cancer. Cancer Survivors Network. URL: https://csn.cancer.org/categories/colorectal [accessed 2000-01-01]
17. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature New Biol 1999 Oct 21;401(6755):788-791. [doi: 10.1038/44565] [Medline: 10548103]
18. Lee D, Seung HS. Algorithms for non-negative matrix factorization. Presented at: Advances in Neural Information Processing Systems 13, Papers from Neural Information Processing Systems (NIPS); Dec 4-6, 2000; Denver, Colorado, USA URL: https://proceedings.neurips.cc/paper_files/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf [accessed 2024-09-01]
19. Bouma GJ. Normalized (pointwise) mutual information in collocation extraction. Proceedings of GSCL 2009 Sep 30;30:31-40.
20. Niwattanakul S, Singthongchai J, Naenudorn E, Wanapu S, editors. Using of Jaccard coefficient for keywords similarity. Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Mar 17;1(6):380-384.
21. TfidfVectorizer. Scikit learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html [accessed 2024-09-01]
22. Ind N, Coates N. The meanings of co‐creation. Eur Bus Rev 2013 Jan 4;25(1):86-95. [doi: 10.1108/09555341311287754]
23. Lim CYS, Laidsaar-Powell RC, Young JM, Kao SCH, Zhang Y, Butow P. Colorectal cancer survivorship: a systematic review and thematic synthesis of qualitative research. Eur J Cancer Care (Engl) 2021 Jul;30(4):e13421. [doi: 10.1111/ecc.13421] [Medline: 33733545]
24. Rivard SJ, Vitous CA, Bamdad MC, et al. "I Wish There had been Resources": a photo-elicitation study of rectal cancer survivorship care needs. Ann Surg Oncol 2023 Jun;30(6):3530-3537. [doi: 10.1245/s10434-022-13042-6] [Medline: 36847958]
25. Denlinger CS, Barsevick AM. The challenges of colorectal cancer survivorship. J Natl Compr Canc Netw 2009 Sep;7(8):883-893. [doi: 10.6004/jnccn.2009.0058] [Medline: 19755048]
26. Cotrim H, Pereira G. Impact of colorectal cancer on patient and family: implications for care. Eur J Oncol Nurs 2008 Jul;12(3):217-226. [doi: 10.1016/j.ejon.2007.11.005] [Medline: 18567538]
27. Teagle A, Gilbert DC. Remote follow-up strategies after cancer treatment: a lot of opportunities. Clin Oncol (R Coll Radiol) 2014 Oct;26(10):622-624. [doi: 10.1016/j.clon.2014.05.009] [Medline: 24969682]
28. Acocella I. The focus groups in social research: advantages and disadvantages. Qual Quant 2012 Jun;46(4):1125-1136. [doi: 10.1007/s11135-011-9600-4]
29. Ellenberg JH. Selection bias in observational and experimental studies. Stat Med 1994;13(5-7):557-567. [doi: 10.1002/sim.4780130518] [Medline: 8023035]
30. Van Veen D, Van Uden C, Blankemeier L, et al. Clinical text summarization: adapting large language models can outperform human experts. Res Sq 2023 Oct 30;PMID:37961377. [doi: 10.21203/rs.3.rs-3483777/v1] [Medline: 37961377]
31. Smith SK, Rowe K, Abernethy AP. Use of an electronic patient-reported outcome measurement system to improve distress management in oncology. Palliat Support Care 2014 Feb;12(1):69-73. [doi: 10.1017/S1478951513000345] [Medline: 24128592]

Abbreviations

CRC: colorectal cancer
NMF: negative matrix factorization

Edited by N Cahill; submitted 27.03.24; peer-reviewed by L He, M Fitch; revised version received 03.07.24; accepted 30.10.24; published 27.01.25.

Please cite as:
Voigt K, Sun Y, Patandin A, Hendriks J, Goossens RH, Verhoef C, Husson O, Grünhagen D, Jung J
A Machine Learning Approach Using Topic Modeling to Identify and Assess Experiences of Patients With Colorectal Cancer: Explorative Study
JMIR Cancer 2025;11:e58834
URL: https://cancer.jmir.org/2025/1/e58834
doi:10.2196/58834

Original Paper

Exploring Motives Behind Ideal Melanoma Survivorship Care Plans With Multiple Stakeholders: A Cocreation Study

Nadia Christina Willemina Kamminga[1], MD; Marjolein Lugtenberg[1,2], PhD; Julia Annabel Van den Broek[1], MD; Tamar Nijsten[1], MD, PhD; Marlies Wakkee[1], MD, PhD; Kasia Tabeau[3], PhD

[1]Department of Dermatology, Erasmus MC Cancer institute, University Medical Center Rotterdam, Rotterdam, Netherlands

[2]Department Tranzo, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands

[3]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, Netherlands

Corresponding Author:
Kasia Tabeau, PhD
Erasmus School of Health Policy & Management
Erasmus University Rotterdam
Burgemeester Oudlaan 50
Rotterdam, 3062 PA
Netherlands
Phone: 31 10 408 8555
Email: tabeau@eshpm.eur.nl

Abstract

Background: Survivorship care plans (SCPs), ie, personalized health care plans for cancer survivors, can be used to support the growing group of melanoma survivors throughout their disease trajectory. However, implementation and effectiveness of SCPs are suboptimal and could benefit from the involvement of stakeholders in developing a user-centered design.

Objective: The aim of this study was to identify the ideal SCP for patients with melanoma in terms of functions and features to be included according to different stakeholders and to explore their underlying motives.

Methods: In total, 3 cocreation sessions were organized with mixed samples of stakeholders, ie, patients with (a history of) melanoma (n=4), health care providers (HCPs) active in melanoma care (n=3), and IT specialists active in hospital IT departments (n=6). They were invited to compose their ideal melanoma SCP based on potential functions and features identified from prior qualitative research. These functions and features belonged to one of the four main categories of survivorship care (SSC): (1) information and education, (2) identification and treatment, (3) oncological follow-up, and (4) coordination. Participants were invited to explain their motives for including functions and features. Ideas were shared between stakeholders, and interaction was promoted. Descriptive statistics were used to determine the ideal SCP per stakeholder group. To analyze underlying motives, all cocreation sessions were audio-taped, transcribed verbatim, and analyzed in a thematic content analysis.

Results: With regard to their ideal SCPs, all stakeholders added functions from all 4 SSC categories. Patients assembled a rather compact SCP with category 2 on *identification and treatment* being most important. Both HCPs and IT professionals constructed a somewhat larger SCP, with category 3 on *oncological follow-up* being the most important aspect and HCPs also focusing on category 4 on *coordination*. As for the motives behind their ideal SCP compositions, patients predominantly added functions based on their personal experiences or experiences from fellow patients, whereas both HCPS and IT professionals based their compositions primarily on their respective areas of expertise: HCPs related their additions to their roles as medical practitioners; for example, in providing a complete treatment plan and obtaining informed consent, while IT professionals' contributions were mainly influenced by feasibility and privacy concerns.

Conclusions: This cocreation study provides insights into stakeholders' ideal melanoma SCP and the motivations behind them. Considering the diversity in both the preferences and underlying motives regarding SCP composition between patients, HCPs, and IT specialists, it is crucial to develop a broad SCP that extends beyond traditional SCP content, emphasizing personalization. In addition to continued stakeholder involvement, efforts should be focused on addressing potential feasibility and privacy issues to ensure the SCP meets both patients' and HCPs' needs.

(JMIR Cancer 2025;11:e55746) doi:10.2196/55746

KEYWORDS

cocreation; survivorship care; psycho-oncology; supportive care; motives; melanoma; cancer survivor; melanoma care

Introduction

In recent years, the prognosis of melanoma, one of the most aggressive forms of skin cancer, has significantly improved due to advancements in innovative treatments such as immunotherapy and targeted therapy [1]. With an estimated worldwide total of 325,000 new cases in 2022, increasing to an expected total of 510,000 new cases in 2040 [2], this results in an expanding cohort of melanoma survivors, ie, individuals living with or beyond melanoma [3]. To ensure that patients get the necessary support throughout their treatment trajectory and assist them in resuming life thereafter, it is important to provide them with effective survivorship care (SSC) [4].

SSC can be divided into four main categories [4,5], namely (1) information and education, (2) identification and treatment, (3) oncological follow-up, and (4) coordination (Textbox 1 [4,5]). Survivorship care plans (SCPs), personalized health care plans for cancer survivors, have an important role in the delivery of SSC, traditionally mainly regarding categories 1 and 3 of SSC (Textbox 1) [4]. However, notwithstanding their potential benefit for both patients and health care providers (HCPs) and the recommendation of their use in clinical guidelines, the present implementation and effectiveness of SCPs seem to be suboptimal [6-8]. Until now, most SCPs have been static, paper-based documents, while patients have shown a preference for dynamic, electronically accessible formats that permit alterations and accessibility for all stakeholders [8]. Personalization, an essential element of SCPs, has often been overlooked, despite evidence emphasizing the importance of tailoring SCPs to accommodate the diverse information needs among patients [9]. However, these findings are mostly based on evaluations of the SCP subsequent to its implementation and based on the feedback of 1 single type of stakeholder. Indeed, while stakeholder engagement seems critical for effective implementation [10], involvement of key stakeholders like patients and HCPs during SCP development has been limited until present [8,11]. The specific needs of patients with melanoma and their HCPs regarding the content of melanoma SCPs have been explored previously [12], which showed that while both stressed the importance of adequate information throughout the disease trajectory and personal oncological follow-up, different opinions existed regarding psychosocial support and coordination of care. However, the reasons why they consider these elements important and why their opinions differ remain unknown. Thus far, only the needs of users have been investigated, with developers' perspectives yet to be examined, even though this could provide valuable insights into the feasibility of desired content.

An approach to integrating the diverse perspectives of all stakeholder groups, ie, patients, HCPs, *and* IT professionals (future developers), is to engage them in a cocreation process [13,14] that encourages their direct involvement in SCP development. Cocreation allows stakeholders to be active partners in the development of innovations, as opposed to objects of study, resulting in products and services that people want and need [15,16]. Therefore, the objective of this study is to investigate the ideal SCP in terms of functions and features to be included per stakeholder group and to explore the motivations behind these preferences through a cocreation process. The findings of this study will serve as a basis to design a user-centered, practically feasible SCP that is tailored to the needs of stakeholders and thereby more easily integrated in clinical practice.

Textbox 1. The 4 main categories of survivorship care.

- *Information and education* about the disease, its treatment, and the possible early and late effects.
- *Identification and treatment* of the disease and therapy effects on all possible domains (ie, physical and psychosocial, including work- and insurance-related).
- *Oncologic follow-up* with surveillance for cancer progression, recurrences or second cancers.
- *Coordination* between all health care providers involved in the care process, to make sure the survivor's health needs are met.

Methods

Setting

This study was part of a regional project in which a digital and personalized melanoma SCP will be developed that will be linked to the patients' electronic health record and provided to patients from diagnosis onwards to help them deal with all disease and treatment-related impacts [17,18]. This project takes place in the region Rijnmond, the Netherlands, and forms a collaboration between 1 academic (Erasmus Medical Center) and 3 non-academic hospitals (Albert Schweitzer Hospital, Francicus Gasthuis & Vlietland, and Maasstad Hospital), in which, like internationally [7], SCPs are not yet routinely provided [6]. The project consists of multiple phases—from needs assessment to implementation—in which cocreation is used to develop an SCP that is adapted to all stakeholders' needs [13].

Study Design

In this study, qualitative research methods in terms of cocreation sessions were used to gain an in-depth understanding of the preferences of all stakeholders involved [19]. Conducting cocreation sessions allows both the end users, ie, patients and HCPs and developers, ie, IT specialists, to collaborate in the SCP design process to reduce the gap between research and implementation [13,14].

The COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [20] were used in reporting this qualitative study.

Participants and Recruitment

Eligible participants were patients with (a history of) cutaneous stage I-IV melanoma; HCPs involved in both primary, secondary, and tertiary melanoma care such as dermatologists, oncologists, surgeons, general practitioners, and nurse practitioners; and hospital-based IT professionals, for example, those working in organizational aspects of IT in health care and eHealth (future developers of the SCP). We aimed to recruit an equal number of patients, HCPs, and IT professionals to ensure a balanced cocreation session in terms of perspectives to be included and to explore the motives in these perspectives. All patients had to be treated in and therefore under follow-up in, and both HCPs and IT professionals had to be affiliated with one of the 4 participating hospitals. To select participants, first patients and HCPs that participated in prior qualitative research and/or had given consent to be contacted (again) for participation in a follow-up study were invited to participate. IT professionals as well as the remaining patients and HCPs were approached through the professional networks of the researchers. All potential participants received information about the study by email or by phone. Our aim was to recruit ±15 participants, in which we follow prior research (eg, Vandekerckhove et al [21]). In the end, 19 participants signed up, and based on their availability and eventual willingness to participate, a total of 3 mixed cocreation sessions with 4-5 participants were organized, with a total of 4 patients, 3 HCPs, and 6 IT specialists. No financial compensation was given for participation.

Cocreation Sessions

Input for the cocreation sessions was based on prior in-depth qualitative research, in which SSC needs of a total of 50 patients with stage I-IV melanoma and 24 HCPs were explored [6,17,18]. In total, 23 interview- and 8 focus group transcripts were re-analyzed for the purpose of this study using Nvivo version 12/R1, to identify potential functions and features of a melanoma SCP. *Functions* represent the overarching areas in which support can be provided, while *features* are tools to deliver that support. First, all transcripts were coded to identify patients' and HCPs' needs regarding SSC (including SCPs) by 2 researchers (JB, a female medical master student and a female health care management master student), which was then checked and complemented by 2 other researchers (NK, a female medical doctor, and KT, a female academic researcher in health care innovation and cocreation). The resulting SSC needs were reformulated as 44 potential features of an SCP by the research team (first reformulation by JB and a female health management student, under supervision of NK and KT, which was then discussed in the multidisciplinary research team including NK, ML, JB, KT, and a medical student until consensus was reached). Subsequently, an exploratory literature review was conducted to assess completeness, which did not reveal any new features. Subsequently, the features were structured according to the 4 main categories of SSC [4,5], as presented in Textbox 1, and further divided into 14 potential functions. Using this classification, the 44 potential features of a melanoma SCP were presented to the participants of the cocreation sessions (see Multimedia Appendix 1).

As a result of the national restrictions placed in response to the COVID-19 crisis, the 3 cocreation sessions were held online via Zoom in May 2021. All sessions took approximately 90 minutes and were moderated by JB and NK (experienced in moderating group discussions) and a health care management student, who were all not directly involved in melanoma care.

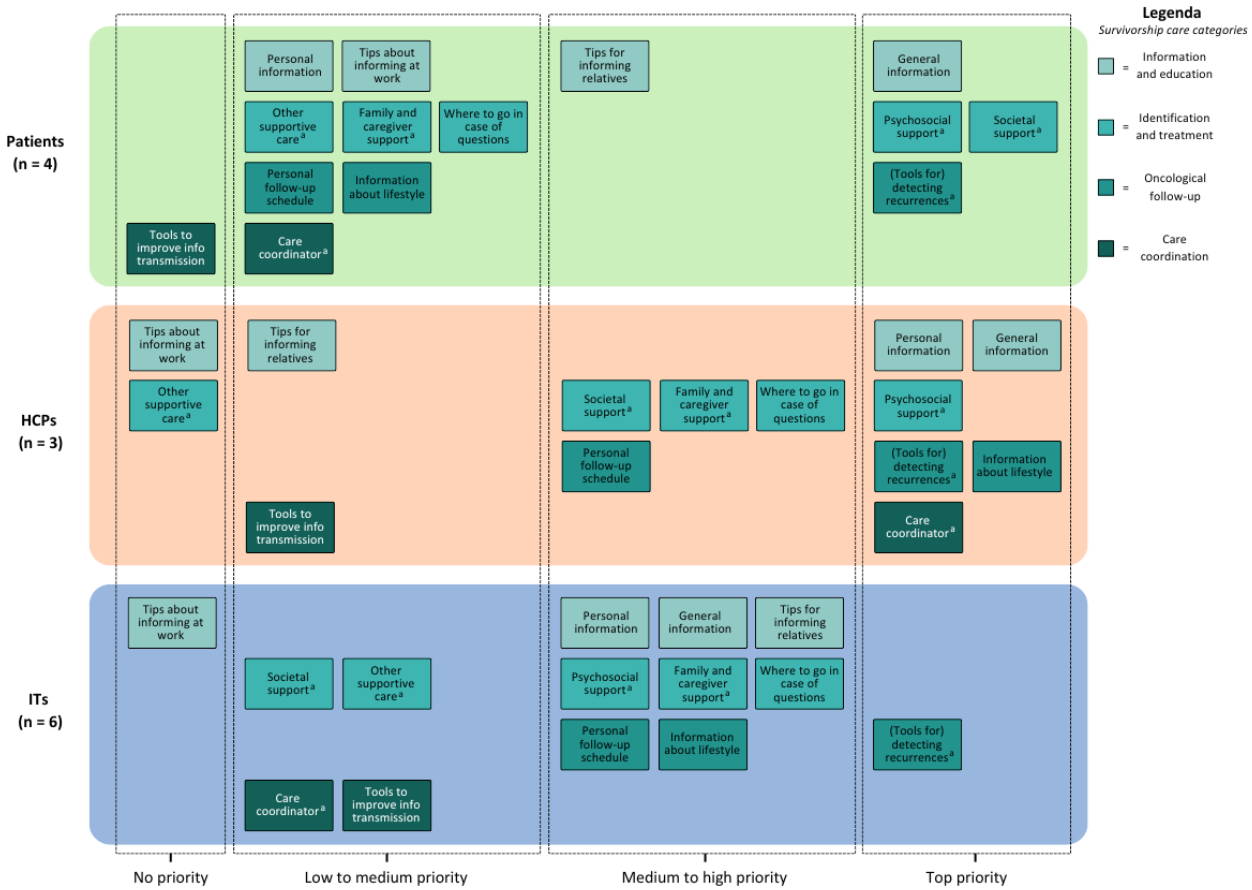
During the cocreation sessions, a PowerPoint presentation was used to present potential functions and features, and participants were invited to individually create their "ideal SCP" by placing their preferred features in a box. Subsequently, participants had a plenary, semistructured discussion in which they were encouraged to share their ideas about the ideal SCPs by sharing their screens. They were invited to elaborate on their motives for (not) including certain functions and features, and discussions arose. At the end of the sessions, the moderators questioned the participants about under-discussed features. All cocreation sessions were audio-taped, and participants were invited to hand in their filled-in sheets.

Data Analysis

All 3 audio-recordings were transcribed verbatim in anonymized form and were analyzed using Nvivo version 12/R1. Descriptive analyses were used, using Microsoft Excel, to determine the ideal SCPs of the different stakeholder groups based on the filled-in PowerPoint sheets and on comments they made during the sessions. For each function and feature, it was determined by how many percent of the participants they were included. Subsequently, all functions were categorized into "top," "medium to high," "low to medium," or "no priority" based on the following criteria: 100% of respondents adding the function to their ideal SCP was defined as "top priority," 50%-99% as "medium to high priority," 1%-49% as "low to medium priority," and 0% of respondents adding the function to their ideal SCP was defined as "no priority"; see also Figure 1. Based on this figure, we then described the ideal SCP par stakeholder group and the similarities and differences between them.

Second, underlying motivations of all participants for including functions and features in their ideal melanoma SCPs were analyzed in a thematic content analysis [22]. As a first step of the analysis, all transcripts were coded based on the functions and features within the 4 main categories of SSC, as previously mentioned (Textbox 1), by 2 researchers (JB and a health care management master student). This was checked and complemented by a third researcher (medical master student). Next, within these categories, underlying motives for composing the ideal SCPs were explored. Motives of each stakeholder group to include (or exclude) a specific function or feature in the SCP were openly coded by one researcher (JB), which was checked and complemented by a second researcher (NK). The next phase of analysis consisted of axial coding, in which the open codes were clustered in concept motives and links between motives and stakeholder groups were made in order to investigate the differences and resemblances between the ideal SCPs for patients, HCPs, and IT specialists. The resulting overview was discussed within the multidisciplinary research team until consensus was reached (JB, NK, ML, and KT).

Figure 1. Overview of composed ideal SCP per stakeholder group in terms of functions added per main category of SSC. For an overview of all functions, including their corresponding features, see Multimedia Appendix 1. [a]Information about and referral to reliable and up-to-date information and tools regarding this topic. HCP: health care provider; SCP: survivorship care plan; SSC: survivorship care.



Ethical Considerations

This cocreation study was part of a larger project, of which the study protocol was submitted to, and approved by, the Medical Ethics Committee Erasmus MC. After reviewing the protocol, the committee concluded that the Medical Research Involving Human Subjects Act (Dutch abbreviation: WMO) did not apply to this study (MEC-2020-0197). Written informed consent was obtained from all participants involved in the cocreation sessions and they were informed that they could withdraw at any point during the study. Participants did not receive any compensation for participation in this study.

Results

Participant Characteristics

The characteristics of the 13 participants in the cocreation sessions as well as the compositions of all 3 sessions can be found in Table 1.

Table 1. Characteristics of participants per cocreation session.

Participant	Gender	Stakeholder	Experience or expertise	Care setting
Session 1				
1	Female	Patient	Melanoma stage IV	Secondary or tertiary care
2	Female	IT professional	Assistant professor in organizing aspects of IT in health care	Tertiary care
3	Male	IT professional	Information management and IT advisor	Tertiary care
4	Female	IT professional	Clinical informatician and data protection officer[a]	Secondary care
Session 2				
5	Male	Health care professional	General practitioner	Primary care
6	Female	IT professional	Information manager in research and innovation	Secondary care
7	Female	IT professional	Clinical informatician: information advisor and architect, (application) implementation lead	Secondary care
8	Male	Patient	Melanoma stage I/II	Secondary care
Session 3				
9	Female	Health care professional	Oncological surgeon	Tertiary care
10	Female	Health care professional	Oncological nurse practitioner	Secondary care
11	Female	IT professional	Information manager	Tertiary care
12	Male	Patient	Melanoma stage IV	Secondary or tertiary care
13	Male	Patient	Melanoma stage IV	Secondary or tertiary care

[a]Having a partner with melanoma.

Ideal SCPs and Underlying Motives

In Figure 1, an overview is provided of the composed ideal SCPs per stakeholder group, followed by an in-depth description of their underlying motives for (not) including certain functions and features. All results are discussed per category of SSC (Textbox 1), and all motives are provided from the participants' perspectives.

Ideal SCPs

Overall, patients assembled the smallest SCP; for them, fewer functions were of medium to high, or top priority compared to the other 2 stakeholder groups. For patients, category 2 focusing on identification and treatment was most important. They primarily included general information and support for themselves, tips on how to inform their relatives, and information about (tools for) detecting recurrences. On the other hand, both HCPs and IT professionals constructed a somewhat larger SCP with more functions being of medium to high, or top priority, with category 3 focusing on oncological follow-up being the most important aspect. HCPs added more information about (personal) follow-up for the patient, including (tools for) detecting recurrences. Furthermore, they added both general and personal information, as well as support for both the patient and their family and caregivers. Additionally, they included information about a care coordinator. IT professionals, on the other hand, considered (tools for) detecting recurrences particularly important to add, and they included extensive information about the patient's (personal) follow-up, both personal and general information, along with support for the patient and their family and caregivers.

Underlying Motives Per Stakeholder Group

For each stakeholder group, motives for (not) including certain functions and features are discussed below per main category of SSC and can be found in Table 2. For readability, results are provided on the function level, which are **bold**. Submotives are in *italics.*

Table 2. Overview of motives for (not) including functions and features from the 4 main categories of SSC in the ideal SCP, per stakeholder group.

Motives and submotives	PT[a]	HCP[a]	IT[a]
Motives for including functions and features			
Informing patients			
To meet patients' information needs	1, 3	1, 2, 3	1, 2, 3
To provide understandable information	1	1	1
To provide reliable information	1	2	1
To help remember provided information	1	1	
To obtain informed consent		1	
Helping patients deal with psychosocial issues			
To identify psychosocial issues		2	
To alleviate patients' concerns	3	3, 4	3
To provide support for or treat psychosocial issues	2	2	2
To help patients deal with lack of understanding of others	1		1[b]
Improving patient empowerment			
To improve patients' self-management skills	3	3	1, 2, 3
To support patients in decision-making		4	
Preparing and providing structure for patients			
To prepare for or raise awareness about what to expect	1	1	
To prevent unnecessary consultations	3	2, 3	1
To provide structure for patients		3	
Empathizing with patients and their situation			
To empathize with (other) patients' needs	2, 3	2	1, 2, 3
To take the patient's context into account		1, 2, 4	
Meeting the relatives' needs			
To meet relatives' information needs		1, 2, 3	1
To meet relatives' support needs			2
Improving health and outcomes			
To improve melanoma outcomes		3	3
To improve health in general		3	
Providing integrated care			
To improve communication and collaboration between HCPs		4	4
To provide a complete treatment plan		2, 3	2
To provide an accessible contact point	2, 3, 4	2, 3, 4	2, 3, 4
Relating to experiences and expertise			
Based on own experiences	1, 2, 3, 4		1, 2, 3, 4[b]
Based on experiences of others	2, 3		1, 2[b]
Based on expertise		1, 2, 3	1, 2, 3, 4
Taking feasibility and privacy into account			
Function is feasible			1, 2, 3, 4
Function is important and this exceeds potential privacy issues			2, 3
Motives for not including functions and features			
Preventing patient distress			

Motives and submotives	PT[a]	HCP[a]	IT[a]
To prevent (irrelevant) information overload		1, 3	3
To prevent taking away hope	1		1[b]
Preventing unnecessary development			
SCP is not the right medium		2	4
Similar tools already exist	3	3	3
Function is applicable in general, not for melanoma specifically			1, 3
Lack patient need for function	3		
Taking feasibility and privacy into account			
Function is unfeasible			2, 3, 4
Privacy issues are more severe than the importance of including the function and its relevance for patients			3

[a]The numbers correspond with the main category of SSC to which the functions belonged and for which submotives were provided, namely (1) information and education, (2) identification and treatment, (3) oncological follow-up, and (4) coordination.

[b]Submotives for (not) including functions provided solely by an IT professional with a partner with melanoma.

Motives of Patients

The most important category for patients was category 2 (identification and treatment). Patients included **psychosocial support**, **societal support**, and **where to go in case of questions** because it would *help them* in *dealing with psychosocial issues*. Furthermore, they indicated that the SCP should contain information about and referral to (reliable and up-to-date information about) **psychosocial** and **societal support** because patients recognized the existence of such issues *based on their personal experiences* and because they *empathized with other patients' (diverse) needs.*

> *Of course it's all very individual.. One person might struggle with their mortgage, while another might not have financial issues due to their illness. […] And chatting with fellow patients for example, peer support, can be helpful for many people, although I personally don't have the need for it.* [Patient, male]

Patients included **general information** and **tips for informing relatives** from the second most important category, namely category 1 (information and education), because they valued being *informed* and *prepared* and to help them *deal with their psychosocial issues*. They indicated that the ideal SCP should provide **general information** that is *reliable* and *understandable* since much of what is currently included in, for example, the patient portal is too medical and written in "doctor's language." According to them, providing this information within the SCP could potentially prevent both patients and their relatives from searching for (incorrect) information on websites.

> *I did a lot of googling myself during the years I was under treatment, and at some point, you end up on really unpleasant websites where you're practically told that you'd be better off not living, and five years later, here I am. I found that very distressing, and it always made me very sad. So, I would like people to know how to find their way to the right information.* [Patient, female]

Tips or **information about informing relatives** could also *help them deal with psychosocial issues*, for example *with the (lack of) understanding of others.* A reason considered for not including functions from this category in their ideal SCP, such as specific parts of **general information**, was to *prevent getting distressed*; they emphasized not to add specific information, such as treatment effectiveness, because, *in their experience*, it *took away their hope.*

Being informed, prepared, and empowered were important motives to include **(tools for) detecting recurrences**, a **personal follow-up schedule**, and **information about a healthy lifestyle** from the third category (oncological follow-up), as was *dealing with their psychosocial issues*, which they did when *empathizing with other patients' needs*. Furthermore, patients stressed that adding information regarding **(tools for) detecting recurrences** could *improve their self-management skills, alleviate their concerns*, and (thereby) *prevent unnecessary consultations*. Motives not to include functions from this category (eg, **information about a healthy lifestyle**) were that similar functions or tools already exist so that *unnecessary development could be prevented*.

Functions in the fourth and last category (coordination), like information regarding a **care coordinator**, although less important in their ideal SCP, were added *based on their own experiences* since they missed an *easily accessible contact point* during their disease trajectory.

Motives of Health Care Providers

HCPs added functions from category 3 (oncological follow-up) to *inform*, *prepare*, and *empower patients*. More specifically, they indicated that the ideal SCP should include information about/referral to **(tools for) detecting recurrences** to *meet patients'* as well as *relatives' information needs*, *improve patients' self-management*, and *alleviate their concerns.* Altogether, this could *prevent unnecessary consultations. Meeting patients' information needs* and also *improving their melanoma outcomes and health in general* were reasons for adding **information about a healthy lifestyle**. Furthermore,

HCPs indicated that offering this information is needed in order *to provide* patients with *a complete treatment plan*, which they did *based on their expertise*. However, others mentioned that *similar tools already exist* and therefore did not add this function to the SCP. A **personal follow-up schedule** including background information was added by HCPs to *meet patients' information needs* and *to provide structure for them.* On the other hand, a reason for not including this background information was also mentioned, namely *to prevent (irrelevant) information overload.*

Identifying and providing support for or treating psychosocial issues were motives mentioned by HCPs for adding **psychosocial support** from category 2 (identification and treatment) to their ideal SCP. They did this *based on their expertise* and when *empathizing with patients' needs*. Other reasons mentioned for adding this function were to *provide reliable information* (preferably by linking it to existing trustworthy resources) that is up-to-date and to *provide an accessible contact point* for patients. Furthermore, HCPs indicated that information regarding and/or referral to **societal support** should be added to meet patients' *information needs* since questions regarding financial and work-related issues arise from diagnosis onwards. **Family and caregiver support** was added by the HCPs to *provide a complete treatment plan* for the patient and their loved ones and to *meet the relatives' information needs*, the latter being important as heredity is a topic that elicits many questions. HCPs indicated that by better informing relatives, *unnecessary consultations* could be *prevented*. Lastly, **where to go in case of questions** was added by HCPs to *provide structure for patients* and to help them find the right person to turn to among the many HCPs involved in their illness and treatment journey.

> *That the patient actually has some idea of where to go for which question, so they don't get lost in the maze of various professionals.* [Health care professional, male]

Similarly, regarding category 4 (coordination), they added information regarding a **care coordinator** to *provide an accessible contact point* for patients and their relatives. They also did this *to alleviate patients' concerns*, which is something that such a contact point could do. In addition, they considered this information important to ensure that the patient had a sense of having somewhere (or someone) to turn to for questions and uncertainties.

Functions from the first category (information and education) that were added to the ideal SCP of HCPs were both **personal** and **general information**, and they included them to *meet the patients' information needs, to provide understandable information, to help patients remember provided information,* and *to support them in (treatment) decision-making.* They mentioned that patients often forget important **personal information** about their received diagnosis and treatment options, but also **general information regarding** what certain treatments (effects) entail, and they considered it important to assist them in retaining this information. In general, HCPs *based this on their expertise* and indicated they should provide this information in order *to obtain informed consent.*

> *You're supposed to tell them what you're going to do, what the side effects are, what the chances are that it will work; make it clear what the patient can choose. Only then can they give informed consent. So, whether you're going to perform surgery or provide immunotherapy, this is the information you must share.* [Health care professional, female]

Motives of IT Professionals

Category 3 (oncological follow-up) was the most important category in the ideal SCP of IT professionals. They added information regarding **(tools for) detecting recurrences** based *on their own experiences* and when *empathizing with patients' needs.* Furthermore, *to improve patients' self-management skills, to alleviate their concerns*, and *to provide an accessible contact point* were also mentioned as motives for including this function. IT professionals indicated that they thought patients could have worries and uncertainties around potential recurrences and that support regarding these worries, which is also *feasible*, should be provided to them. While *empathizing with patients' needs,* they indicated that a **personal follow-up schedule** including background information should be added *to meet patients' information needs*. However, *taking feasibility and privacy into account,* they mentioned that while they considered adding background information feasible, they foresaw several *privacy issues* for including a **personal follow-up schedule** that *were more severe than the importance of including and the relevance of* this information.

> *For privacy regulations, I deliberately left those things out because otherwise, it makes it quite challenging when it comes to an appointment calendar or something similar. If you manage it yourself, it's fine, but if it has to come through the hospital information system, then you have to integrate with that, and it becomes really tricky from an IT and privacy standpoint.* [IT professional, female]

They added **information about a healthy lifestyle** for patients with melanoma when *empathizing with* and *to meet their needs*, as well as *to improve melanoma outcomes*. Reasons for not including this information were that *having a healthy lifestyle is applicable in general, to prevent (irrelevant) information overload* and because *similar tools already exist*.

IT professionals added functions belonging to category 2 (identification and treatment) *based on their expertise* and, more specifically, because they considered them *feasible.* In addition, they added **psychosocial support** to *provide support for or treat psychosocial issues* and *to improve patients' self-management*, which they *based on their own and others' experiences.* They added **family and caregiver support** to *meet patients' and relatives' information and support needs* and *to provide a complete treatment plan*. Some included information about **where to go in case of questions** to provide clarity for patients; although some of them questioned its feasibility, they thought *its importance and relevance for patients outweighed the potential privacy issues.*

> *I think that if you can have patients fill in information like, 'I have this general practitioner, I go to that hospital with that specialist, and these are the other health care providers involved,' and alongside that, a general guide saying, 'For these kinds of questions, contact your GP,' I think that can provide more value. But when you want to retrieve that information automatically, it becomes challenging.* [IT professional, female]

For others, these *potential issues* made them consider this function *unfeasible*, and they did not add this to their ideal SCP.

Informing, *preparing*, and *empoweringpatients* were reasons mentioned by IT professionals to include functions like **personal** and **general information** from category 1 (information and education). They added them *to meet* both *patients' and relatives' information needs, to provide understandable* as well as *reliable information*, and to, at the same time, *help patients deal with the lack of understanding of others*. **Tips to inform relatives** were added by the IT professionals to *meet relatives' information needs* and to *help patients deal with a lack of understanding of others*, which they did *based on the experiences* they gained *from others*. They considered adding these functions (**personal** and **general information** and **tips to inform relatives**) *feasible* as *based on their own expertise* and when *empathizing with patients' needs*. Reasons for not adding functions like **general information** included considering them *not applicable to melanoma specifically* and to prevent this general information from *taking away hope*.

IT professionals considered category 4 (coordination) least important. Nevertheless, they added functions (**care coordinator** and **tools to improve information transmission**) from this category *based on their own and others' experiences* and based on *their expertise*. They also provided motives for not including functions from this category; they indicated that according to them, the *SCP is not the right medium* to include **tools to improve the information transmission between HCPs**, a function that they also considered *unfeasible* to realize.

Discussion

Principal Findings and Comparison With Prior Work

The aim of this study was to gain insight into the ideal SCPs according to relevant stakeholder groups and to explore their motivations behind adding them. Patients composed a rather compact SCP, mainly focusing on category 2 on identification and treatment, including both information and support for themselves, with their motives being primarily based on their personal experiences and needs. HCPs and IT professionals constructed more comprehensive SCPs, with category 3 on oncological follow-up being the most important one, and HCPs additionally focusing on category 4 on *coordination*. When looking at their underlying motives, they all aligned with their respective areas of expertise: HCPs related their additions mainly to their roles as medical practitioners, such as providing a complete treatment plan and obtaining informed consent, while IT professionals' contributions were mostly influenced by feasibility and privacy concerns.

In light of our findings and prior research focusing on other types of cancer (eg, colorectal and gynecological [23-25]) than melanoma, it is evident that patients place considerable emphasis on *identification and treatment of disease and therapy effects* (category 2), perceiving the SCP mainly as an informational and supportive tool. Thereby, patients' unmet needs are related to their commonly faced challenges such as psychological distress, anxiety, depression, long-term and late effects, as well as the fear of recurrence. As a result, they express a need for a more comprehensive supportive care approach that includes nonmedical information, like peer support, financial guidance, return-to-work strategies, and psychological resource information [17,26-28]. Moreover, the emphasis on category 2 in our results underscores the importance of broadening traditional, current SCP content, which predominantly targets categories 1 and 3 on *information and education* and *oncological follow-up* [4]. Patients agreed that functions from category 2 (eg, psychosocial and societal support) are indeed essential for effective SSC, as it would both help them deal with certain issues and prevent searching for incorrect information online and related distress. Integrating (referrals to) this information in the SCP also allows all to be centralized in one place, instead of spread over paper-based information and web-based resources given to them by various HCPs. Despite patients' agreement on category 2, their opinions differed on other functions. For example, some accentuated the need for general information (category 1, information and education) and more specifically on treatment effectiveness, but others felt that knowing this might negatively influence their hope toward a positive outcome. A reason for this may be found in patients' motives for the composition of their ideal SCP, which were predominantly based on their personal, specific experiences with the disease and the knowledge they gained when they were affected and treated [29]. In addition, our findings align with previous research showing that providing certain information might be beneficial for some patients but can trigger fear among others [8,9], and thereby further underscore the importance of personalizing the SCP's content [30,31]. To adequately tailor the SCP to each patient's individual needs, future research should investigate ways for SCP tailoring, possibly through patient profile definition or integrating artificial intelligence (AI) methodologies. In addition, further quantitative research should follow to investigate the actual impact of our results on both SCPs' implementation and effectiveness in clinical practice, for example, through a classic randomized controlled trial or partially randomized patient preference trial, the latter incorporating patient preferences in the process of randomization [32].

HCPs and IT professionals, in contrast, mainly based their ideal SCP preferences on their professional expertise. HCPs primarily related their choices to their clinical knowledge and roles as medical professionals and indicated what is needed to provide patients with a complete treatment plan and in order to obtain informed consent. IT experts offered more practical reasons, predominantly concerning feasibility and privacy issues, thereby including perspectives the end users may have insufficient knowledge about [29,33]. Both HCPs and IT professionals regarded *oncological follow-up (category 3)* as most important. Within this category, information about (tools for) detecting

recurrences was added unanimously. This corresponds with previous literature that identified fear of recurrence as a prevalent psychological concern among melanoma survivors [6,17,34], and offering information and support concerning this has been shown to alleviate the intensity of such fears [35]. In the context of the ongoing digital health transformation and the workforce challenges in health care [36,37], AI offers promising solutions. AI could help in assessing skin abnormalities and thereby detecting recurrences in the future [38]. Although physicians have concerns about AI tools' accuracy and potential health inequality risks, it could lead to fewer unnecessary consultations, cost reductions, and improved care pathways [39]. As a result, improving SSC practices regarding *oncological follow-up* could also facilitate advancements in category 4 on *care coordination*, a topic that HCPs in our study also deemed important. This suggests that the HCPs viewed the SCP as a care coordination tool, which could potentially address areas for improvement that were stressed in previous literature [40].

The feasibility and privacy concerns highlighted by IT professionals were particularly related to functions deemed vital to end users, ie, patients and HCPs, such as an overview of where to go (category 2), a personal follow-up schedule (category 3), and tools to improve information transmission between HCPs (category 4). Although legitimate, particularly given the patients' unmet needs, it is important to investigate how to adequately address these issues. Suggestions put forth to mitigate some of these concerns included providing only the patients' specialist's name rather than their comprehensive contact information and linking the SCP to the electronic patient portal instead of incorporating a personal follow-up schedule directly within the SCP. However, since the use by HCPs in clinical practice and thereby the SCP's implementation could be facilitated by linking it directly to the electronic health record [6], which has been shown feasible before [41], it should be investigated how to overcome these privacy concerns.

The above discussion suggests that HCPs, patients, and IT professionals attribute different but complementary roles to SCPs. Whereas HCPs view SCPs primarily as a coordination tool, patients stress their informational and supportive roles, and IT professionals see them as a data-sharing tool that must function in a safe and reliable manner. These differing perspectives can be explained by the organizational structure of melanoma care, including SSC, which is organized in networks with centralization of administration of systemic treatments across the Netherlands. In the Rijnmond Region, it operates from a "shared-care model" [42], where HCPs lead the network and its coordination while patients are positioned at the receiving end of care, thus receiving care, information, and support. From this point of view, the role of IT professionals in relation to SCPs is to ensure that coordination, information sharing, and support are conducted in a reliable manner.

Recommendations for Future SCP Developers

The future SCP that we envision based on our results should address all the above roles, functioning both as a comprehensive information tool—facilitating the safe and reliable linking and sharing of information of multiple stakeholders through its digital aspect and connection with the patient's electronic health record—and as a means to improve the coordination of melanoma care. Being better informed and supported will likely enhance patient empowerment, allowing them to take a more active role in managing their disease and treatment coordination and thereby facilitating shared decision-making [43,44], rather than merely being passive recipients of care. Thus, future developers should create SCPs that contain functions and features with both personal and general information, information about/referral to reliable information on psychosocial and societal support for patients, as well as information on lifestyle and tools for detecting recurrences, and with functions and features that facilitate care coordination. This SCP can and should be further personalized, depending on patients' and HCPs preferences, by adding additional functions and features such as a personal follow-up schedule, information on where to go in case of questions [12], tips for informing relatives, and support for them.

Strengths and Limitations

The main strength of this study is its inclusive approach to cocreation. Unlike many other cocreation studies that often engage only 1 stakeholder group, typically patients [45,46], or more recent at most 2 (patients and HCPs [47,48]), our study uniquely incorporated IT professionals. Whilst they are not the primary users of SCPs [4], their involvement proved invaluable as they offered crucial insights into the development process, particularly highlighting feasibility and privacy concerns. Moreover, we actively engaged all stakeholders from the inception of the SCP development rather than limiting their participation to its evaluation, as seen in, and of which the importance was stressed in, previous literature [7,33]. Furthermore, this cocreation approach aligns with the current shift toward value-based health care, endorsing the principles of patient partnership and shared decision-making [49]. Such an approach empowers the target audience and other stakeholders to shape the outcome actively, ensuring that the final SCP aligns closely with their needs and preferences. Moreover, the composition of our sessions with mixed groups fostered mutual learning and encouraged interactive discussions [13,14,50]. These elements collectively enhance the likelihood of the tool being widely accepted, facilitating more effective implementation and practical effectiveness.

A limitation of our study was its regional sample, which raises questions about the transferability of our findings. However, since melanoma care is uniformly organized in networks throughout the Netherlands, we expect that our results will be applicable outside our region and possibly to other countries if melanoma care is similarly organized. However, to reach an optimal, inclusive SCP, perspectives and needs of patients with varying levels of (health) literacy, socio-economic status, and backgrounds should also be investigated and incorporated. Another aspect warranting attention is the inclusion of an IT professional being a relative of a patient with melanoma. Even though we believe their experiences did not greatly affect the composition of IT professionals' ideal SCP, it could be interesting to further investigate relatives' perspectives on ideal SCPs and their underlying motivations for it. Based on previous literature, we know that relatives of cancer survivors can encounter significant challenges and have unmet needs

throughout the patients' disease trajectory [17,51,52] and are therefore sometimes even included in the definition of a cancer survivor [3]. Lastly, our participant pool only included patients who had finished their treatment for some time. Since retrospective experiences might differ from those of patients currently undergoing treatment, it is important to focus on this latter group in future research, ensuring a complete understanding of patients' SSC needs throughout the whole disease trajectory.

Conclusions

This cocreation study provides insights into stakeholders' ideal melanoma SCP and the motivations behind them. Considering the diversity in both preferences and underlying motives regarding SCP composition between patients, HCPs, and IT specialists, it is crucial to develop a broad SCP that extends beyond traditional SCP content, emphasizing personalization. By understanding the motives and considerations of patients and HCPs in shaping their ideal SCPs, which we were able to elicit through the interaction and discussion between different stakeholders, thoughtful design can optimize patient care and support throughout the survivorship journey. At the same time, keeping the practical requirements of IT professionals in terms of feasibility and privacy in mind is important to ensure the ideal SCPs can be realized. In addition to continued stakeholder involvement, efforts should be focused on addressing the potential feasibility and privacy issues, particularly those related to personalization, to ensure the SCP meets the needs of both patients and HCPs.

Acknowledgments

This study is part of a larger project that is multi-sponsored by Bristol Myers Squibb (BMS), Merck Sharp & Dohme (MSD), Novartis, and Pierre Fabre. However, the sponsors were not involved in the design of the study, data collection, data analysis, manuscript preparation, and/or publication decisions. The authors thank all patients, HCPs, and IT professionals for sharing their experiences and particularly for their openness during the cocreation sessions. Furthermore, the authors thank Esther Bijl for helping with the data collection and analysis.

Authors' Contributions

NK, ML, JB, and KT designed the study. NK, JB, MW, and KT performed the patient selection. NK and JB performed data analysis under supervision of ML and KT. NK, ML, JB, and KT interpreted the data. NK and KT wrote the first version of the manuscript, and all authors revised the following versions of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of all functions and features per category of survivorship care.
[DOCX File , 17 KB - cancer_v11i1e55746_app1.docx]

References

1. Weiss SA, Kluger H. CheckMate-067: raising the bar for the next decade in oncology. J Clin Oncol 2022;40(2):111-113 [FREE Full text] [doi: 10.1200/JCO.21.02549] [Medline: 34855466]
2. Arnold M, Singh D, Laversanne M, Vignat J, Vaccarella S, Meheus F, et al. Global burden of cutaneous melanoma in 2020 and projections to 2040. JAMA Dermatol 2022;158(5):495-503 [FREE Full text] [doi: 10.1001/jamadermatol.2022.0160] [Medline: 35353115]
3. Reuben SH. Living beyond cancer: finding a new balance. In: President's Cancer Panel. Bethesda, MD: National Cancer Institute; 2004.
4. Hewitt M, Greenfield S, Stovall E. From cancer patient to cancer survivor. In: Lost in Transition. Washington, DC: The National Academic Press; 2005:1-506.
5. Health Council of The Netherlands. Follow-up in oncology-identify objectives, substantiate actions. The Hague. 2007. URL: https://www.healthcouncil.nl/documents/advisory-reports/2007/03/27/follow-up-in-oncology-identify-objectives-substantiate-actions [accessed 2024-11-15]
6. Kamminga NCW, Wakkee M, De Bruin RJ, van der Veldt AAM, Joosse A, Reeder S, et al. Oncological healthcare providers' perspectives on appropriate melanoma survivorship care: a qualitative focus group study. BMC Cancer 2023;23(1):278 [FREE Full text] [doi: 10.1186/s12885-023-10759-9] [Medline: 36973713]
7. Hill RE, Wakefield CE, Cohn RJ, Fardell JE, Brierley MEE, Kothe E, et al. Survivorship care plans in cancer: a meta-analysis and systematic review of care plan outcomes. Oncologist 2020;25(2):e351-e372 [FREE Full text] [doi: 10.1634/theoncologist.2019-0184] [Medline: 32043786]

8. Keesing S, McNamara B, Rosenwax L. Cancer survivors' experiences of using survivorship care plans: a systematic review of qualitative studies. J Cancer Surviv 2015;9(2):260-268 [FREE Full text] [doi: 10.1007/s11764-014-0407-x] [Medline: 25343971]
9. de Rooij BH, Ezendam NPM, Vos MC, Pijnenborg JMA, Boll D, Kruitwagen RFPM, et al. Patients' information coping styles influence the benefit of a survivorship care plan in the ROGY care trial: new insights for tailored delivery. Cancer 2019;125(5):788-797 [FREE Full text] [doi: 10.1002/cncr.31844] [Medline: 30500067]
10. Birken SA, Mayer DK. Survivorship care planning: why is it taking so long? J Natl Compr Canc Netw 2017;15(9):1165-1169. [doi: 10.6004/jnccn.2017.0148] [Medline: 28874600]
11. Klemanski DL, Browning KK, Kue J. Survivorship care plan preferences of cancer survivors and health care providers: a systematic review and quality appraisal of the evidence. J Cancer Surviv 2016;10(1):71-86. [doi: 10.1007/s11764-015-0452-0] [Medline: 25911150]
12. Kamminga NCW, Wakkee M, Swart I, Nijsten TEC, Lugtenberg M. Determining the core content of a digital survivorship care plan for melanoma survivors: a multistakeholder Delphi consensus study. Br J Dermatol 2024;190(3):440-441. [doi: 10.1093/bjd/ljad450] [Medline: 37979160]
13. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: using cooperative experimental methods in developing health services and computer technology. Qual Health Res 2007;17(1):122-130. [doi: 10.1177/1049732306293664] [Medline: 17170250]
14. Sanders EBN, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008;4(1):5-18 [FREE Full text] [doi: 10.1080/15710880701875068]
15. Sanders E, Stappers PJ. Convivial Toolbox: Generative Research for the Front End of Design. Amsterdam: Bis; 2012:272.
16. Vandekerckhove P, de Mul M, Bramer WM, de Bont AA. Generative participatory design methodology to develop electronic health interventions: systematic literature review. J Med Internet Res 2020;22(4):e13780 [FREE Full text] [doi: 10.2196/13780] [Medline: 32338617]
17. Kamminga NCW, van der Veldt AAM, Joosen MCW, de Joode K, Joosse A, Grünhagen DJ, et al. Experiences of resuming life after immunotherapy and associated survivorship care needs: a qualitative study among patients with metastatic melanoma. Br J Dermatol 2022;187(3):381-391 [FREE Full text] [doi: 10.1111/bjd.21670] [Medline: 35596713]
18. Kamminga NCW, van der Veldt AAM, Wakkee M, van den Berge FR, van der Beek LAA, Joosen MCW, et al. From decision to reflection: understanding the experiences and unmet care needs of patients treated with immunotherapy for melanoma in the adjuvant or metastatic setting. BMC Cancer 2024;24(1):662 [FREE Full text] [doi: 10.1186/s12885-024-12410-7] [Medline: 38816701]
19. Green J, Thorogood N. Qualitative methodology in health research. In: Qualitative Methods for Health Research. London: SAGE Publications Ltd; 2018.
20. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
21. Vandekerckhove P, Timmermans J, de Bont A, de Mul M. Diversity in stakeholder groups in generative co-design for digital health: assembly procedure and preliminary assessment. JMIR Hum Factors 2023;10:e38350 [FREE Full text] [doi: 10.2196/38350] [Medline: 36787170]
22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
23. Salz T, Baxi SS, Blinder VS, Elkin EB, Kemeny MM, McCabe MS, et al. Colorectal cancer survivors' needs and preferences for survivorship information. J Oncol Pract 2014;10(4):e277-e282 [FREE Full text] [doi: 10.1200/JOP.2013.001312] [Medline: 24893610]
24. de Rooij BH, Thomas TH, Post KE, Flanagan J, Ezendam NPM, Peppercorn J, et al. Survivorship care planning in gynecologic oncology-perspectives from patients, caregivers, and health care providers. J Cancer Surviv 2018;12(6):762-774 [FREE Full text] [doi: 10.1007/s11764-018-0713-9] [Medline: 30209681]
25. Mayer DK, Nasso SF, Earp JA. Defining cancer survivors, their needs, and perspectives on survivorship health care in the USA. Lancet Oncol 2017;18(1):e11-e18. [doi: 10.1016/S1470-2045(16)30573-3] [Medline: 28049573]
26. Wolner ZJ, Flowers NI, Yushak ML, Chen SC, Yeung H. Exploring the melanoma survivorship experience: a qualitative study. Br J Dermatol 2021;185(1):221-223 [FREE Full text] [doi: 10.1111/bjd.19868] [Medline: 33544884]
27. Treanor C, Donnelly M. Late effects of cancer and cancer treatment--the perspective of the patient. Support Care Cancer 2016;24(1):337-346. [doi: 10.1007/s00520-015-2796-4] [Medline: 26066051]
28. Levy D, Dhillon HM, Lomax A, Marthick M, McNeil C, Kao S, et al. Certainty within uncertainty: a qualitative study of the experience of metastatic melanoma patients undergoing pembrolizumab immunotherapy. Support Care Cancer 2019;27(5):1845-1852. [doi: 10.1007/s00520-018-4443-3] [Medline: 30178142]
29. Ball S, Harshfield A, Carpenter A, Bertscher A, Marjanovic S. Patient and public involvement in research. In: Enabling Meaningful Contributions. California, USA: RAND Corporation; 2019.
30. King-Dowling S, Psihogios AM, Hill-Kayser C, Szalda D, O'Hagan B, Darabos K, et al. Acceptability and feasibility of survivorship care plans and an accompanying mobile health intervention for adolescent and young adult survivors of childhood cancer. Pediatr Blood Cancer 2021;68(3):e28884 [FREE Full text] [doi: 10.1002/pbc.28884] [Medline: 33416214]

31. Lai-Kwon J, Kelly B, Lane S, Biviano R, Bartula I, Brennan F, et al. Feasibility, acceptability, and utility of a nurse-led survivorship program for people with metastatic melanoma (MELCARE). Support Care Cancer 2022;30(11):9587-9596 [FREE Full text] [doi: 10.1007/s00520-022-07360-4] [Medline: 36136246]
32. Wasmann KA, Wijsman P, van Dieren S, Bemelman W, Buskens C. Partially randomised patient preference trials as an alternative design to randomised controlled trials: systematic review and meta-analyses. BMJ Open 2019;9(10):e031151 [FREE Full text] [doi: 10.1136/bmjopen-2019-031151] [Medline: 31619428]
33. Dalko K, Kraft B, Jahn P, Schildmann J, Hofstetter S. Cocreation of assistive technologies for patients with long COVID: qualitative analysis of a literature review on the challenges of patient involvement in health and nursing sciences. J Med Internet Res 2023;25:e46297 [FREE Full text] [doi: 10.2196/46297] [Medline: 37581906]
34. Kasparian NA, McLoone JK, Butow PN. Psychological responses and coping strategies among patients with malignant melanoma: a systematic review of the literature. Arch Dermatol 2009;145(12):1415-1427. [doi: 10.1001/archdermatol.2009.308] [Medline: 20026852]
35. Dieng M, Butow PN, Costa DSJ, Morton RL, Menzies SW, Mireskandari S, et al. Psychoeducational intervention to reduce fear of cancer recurrence in people at high risk of developing another primary melanoma: results of a randomized controlled trial. J Clin Oncol 2016;34(36):4405-4414. [doi: 10.1200/JCO.2016.68.2278] [Medline: 27998215]
36. Brommeyer M, Liang Z. A systematic approach in developing management workforce readiness for digital health transformation in healthcare. Int J Environ Res Public Health 2022;19(21):13843 [FREE Full text] [doi: 10.3390/ijerph192113843] [Medline: 36360722]
37. Weber P, Peltonen LM, Junger A. The essence and role of nurses in the future of biomedical and health informatics. Stud Health Technol Inform 2022;300:164-176. [doi: 10.3233/SHTI220948] [Medline: 36300409]
38. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115-118 [FREE Full text] [doi: 10.1038/nature21056] [Medline: 28117445]
39. Sangers TE, Wakkee M, Moolenburgh FJ, Nijsten T, Lugtenberg M. Towards successful implementation of artificial intelligence in skin cancer care: a qualitative study exploring the views of dermatologists and general practitioners. Arch Dermatol Res 2023;315(5):1187-1195 [FREE Full text] [doi: 10.1007/s00403-022-02492-3] [Medline: 36477587]
40. Collett GK, Durcinoska I, Rankin NM, Blinman P, Barnes DJ, Anderiesz C, et al. Patients' experience of lung cancer care coordination: a quantitative exploration. Support Care Cancer 2019;27(2):485-493. [doi: 10.1007/s00520-018-4338-3] [Medline: 29980906]
41. Hill-Kayser CE, Jacobs LA, Gabriel P, Palmer SC, Hampshire MK, Vachani C, et al. Feasibility study of an electronic interface between internet-based survivorship care plans and electronic medical records. J Oncol Pract 2016;12(4):e380-e387. [doi: 10.1200/JOP.2015.006841] [Medline: 27006355]
42. Hall SJ, Samuel LM, Murchie P. Toward shared care for people with cancer: developing the model with patients and GPs. Fam Pract 2011;28(5):554-564. [doi: 10.1093/fampra/cmr012] [Medline: 21467132]
43. Bomhof-Roordink H, Gärtner FR, Stiggelbout AM, Pieterse AH. Key components of shared decision making models: a systematic review. BMJ Open 2019;9(12):e031763 [FREE Full text] [doi: 10.1136/bmjopen-2019-031763] [Medline: 31852700]
44. Skorstad M, de Rooij BH, Jeppesen MH, Bergholdt SH, Ezendam NPM, Bohlin T, et al. Self-management and adherence to recommended follow-up after gynaecological cancer: results from the international InCHARGE study. Int J Gynecol Cancer 2021;31(8):1106-1115. [doi: 10.1136/ijgc-2020-002377] [Medline: 33858949]
45. Elsbernd A, Hjerming M, Visler C, Hjalgrim LL, Niemann CU, Boisen KA, et al. Using cocreation in the process of designing a smartphone app for adolescents and young adults with cancer: prototype development study. JMIR Form Res 2018;2(2):e23 [FREE Full text] [doi: 10.2196/formative.9842] [Medline: 30684439]
46. Alvarez-Perez Y, Perestelo-Perez L, Rivero-Santana A, Wagner AM, Torres-Castaño A, Toledo-Chávarri A, IC-Health Project Consortium. Cocreation of massive open online courses to improve digital health literacy in diabetes: pilot mixed methods study. JMIR Diabetes 2021;6(4):e30603 [FREE Full text] [doi: 10.2196/30603] [Medline: 34898453]
47. Yau KW, Tang TS, Görges M, Pinkney S, Amed S. Using human-centered design and cocreation to create the Live 5-2-1-0 mobile app to promote healthy behaviors in children: app design and development. JMIR Pediatr Parent 2023;6:e44792 [FREE Full text] [doi: 10.2196/44792] [Medline: 37195754]
48. Geerling B, Kelders SM, Stevens AWMM, Kupka RW, Bohlmeijer ET. A web-based positive psychology app for patients with bipolar disorder: development study. JMIR Form Res 2022;6(9):e39476 [FREE Full text] [doi: 10.2196/39476] [Medline: 35946327]
49. Porter ME, Teisberg EO. Redefining Health Care: Creating Value-Based Competition on Results. Boston: Harvard Business School Publishing; 2006.
50. Green J, Thorogood N. Qualitative methods for health research. In: Group Interviews and Discussions. London: SAGE Publications Ltd; 2018.
51. Prusak A, van der Zwan JM, Aarts M, Arber A, Cornelissen R, Burgers S, et al. The psychosocial impact of living with mesothelioma: experiences and needs of patients and their carers regarding supportive care. Eur J Cancer Care (Engl) 2021;30(6):e13498. [doi: 10.1111/ecc.13498] [Medline: 34339095]

52. Aung SHH, White K, Bloomfield J. The experiences and the needs of caregivers of patients with head and neck cancer: an integrative review. Cancer Nurs 2021;44(6):E361-E373. [doi: 10.1097/NCC.0000000000000833] [Medline: 32541207]

Abbreviations

AI: artificial intelligence
COREQ: Consolidated Criteria for Reporting Qualitative Research
HCP: health care professional
SCP: survivorship care plan
SSC: survivorship care

Edited by N Cahill; submitted 22.12.23; peer-reviewed by J Elias, Y Ke; comments to author 05.06.24; revised version received 14.06.24; accepted 31.10.24; published 02.01.25.

Please cite as:
Kamminga NCW, Lugtenberg M, Van den Broek JA, Nijsten T, Wakkee M, Tabeau K
Exploring Motives Behind Ideal Melanoma Survivorship Care Plans With Multiple Stakeholders: A Cocreation Study
JMIR Cancer 2025;11:e55746
URL: https://cancer.jmir.org/2025/1/e55746
doi:10.2196/55746
PMID:

Original Paper

Co-Designing Priority Components of an mHealth Intervention to Enhance Follow-Up Care in Young Adult Survivors of Childhood Cancer and Health Care Providers: Qualitative Descriptive Study

Sharon H J Hou[1,2], PhD; Brianna Henry[3], BSc; Rachelle Drummond[3], MPH; Caitlin Forbes[3], BSc; Kyle Mendonça[3], MD; Holly Wright[3], BSc; Iqra Rahamatullah[3], BSc; Perri R Tutelman[3,4], PhD; Hailey Zwicker[3], BSc; Mehak Stokoe[3], BSc; Jenny Duong[3], BSc; Emily K Drake[5], PhD; Craig Erker[6], MD, MSc; Michael S Taccone[7], MD, PhD; Liam Sutherland[8], MSc; Paul Nathan[9,10], MD, MSc; Maria Spavor[11], MD; Karen Goddard[12], MD; Kathleen Reynolds[13,14], MD; Fiona S M Schulte[3], PhD

[1]Faculty of Education, Simon Fraser University, Burnaby, BC, Canada
[2]BC Children's Hospital, Vancouver, BC, Canada
[3]Department of Oncology, University of Calgary, Calgary, AB, Canada
[4]Department of Psychology, University of Calgary, Calgary, AB, Canada
[5]Faculty of Health, Dalhousie University, Halifax, NS, Canada
[6]Department of Pediatrics, Dalhousie University, Halifax, NS, Canada
[7]Childhood Cancer Survivor Canada, Toronto, ON, Canada
[8]Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[9]Department of Pediatrics, University of Toronto, Toronto, ON, Canada
[10]Institute for Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[11]Department of Pediatrics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada
[12]Department of Surgery, Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada
[13]Department of Family Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[14]Long Term Survivors Clinic, Alberta Children's Hospital, Calgary, AB, Canada

Corresponding Author:
Fiona S M Schulte, PhD
Department of Oncology
University of Calgary
3395 Hospital Drive N.W.
Calgary, AB, T2N5G2
Canada
Phone: 1 403 698 8103
Email: fsmschul@ucalgary.ca

Abstract

Background: Survivors of childhood cancer are at risk of medical, psychological, and social late effects. To screen for their risks, receipt of consistent, cancer-specific follow-up care is crucial. However, <50% of survivors attend their aftercare, and only 35% of them recognize that they could have a serious health problem. The use of mobile health (mHealth) is a promising form of intervention to educate, connect, and empower survivors of childhood cancer on the importance of follow-up care.

Objective: This study aimed to use co-design to identify the priority components to include in an mHealth intervention with young adult (aged between 18 and 39 years) survivors of childhood cancer and health care providers.

Methods: This study was conducted between January and November 2022 in Canada and used patient-oriented research methods. Participants were recruited through local or provincial long-term follow-up clinics, using convenience sampling from patient partners who assisted in recruiting survivors across geographical areas in western, central, and eastern Canada, and social media outreach (X, formally known as Twitter; Facebook; and Instagram). Qualitative descriptive data (focus group interviews) from survivors of childhood cancer and health care providers (individual interviews) were gathered. We analyzed the collected data

using reflexive thematic analysis and verified it through member checking techniques through an online community engagement event.

Results: We conducted with patient partners 5 online (Zoom) focus groups with 22 survivors of childhood cancer (mean age 29.19, SD 4.78 y). We conducted individual telephone interviews with 7 health care providers. Participants identified five priority areas to be included in an mHealth intervention: (1) connections, (2) education and information, (3) engagement, (4) personalization, and (5) resources. Results were shared with and validated by survivors of childhood cancer, their families, health care providers, and academic researchers as part of a community engagement event. Small and large group discussions were facilitated to allow participants to review and discuss the accuracy of the themes derived regarding the core components to be included in mHealth. A graphic recording artist visually captured key ideas from the event. A subset of the participants also completed a web-based satisfaction survey, and responses indicated that the community engagement event was generally well received.

Conclusions: Results from this study have provided the necessary foundation to progress in intervention development. The next step of this multiphased project is to build an innovative and accessible mHealth intervention prototype that is based on the identified core components and is grounded in an established conceptual framework for co-design of mHealth.

(JMIR Cancer 2025;11:e57834) doi:10.2196/57834

KEYWORDS

mobile health; mHealth; pediatric oncology; cancer survivorship; qualitative research; patient-oriented research; co-design; intervention development

Introduction

Follow-Up Care for Survivors of Childhood Cancer

Due to advances in treatment and supportive care, the survival rate for pediatric cancer (age 0-19 years) now exceeds 80% [1]. In turn, the number of long-term survivors of childhood cancer has grown exponentially, with more than 45,000 survivors living in Canada today [2]. While this is encouraging, survivors of childhood cancer are at a lifelong risk of developing medical, psychological, and social late effects due to their cancer treatment [3]. By the age of 50 years, nearly 100% of survivors develop one or more chronic health conditions, many of which are disabling or life-threatening [4,5]. This prevalence suggests an important need for early screening and intervention of late effects to improve longer-term outcomes for this group considered vulnerable.

Consistent, cancer-specific follow-up care is associated with better long-term outcomes, including earlier identification of late effects or secondary cancers and minimized risk of morbidity and mortality [6-8]. However, less than 50% of survivors of childhood cancer attend long-term follow-up clinics [9,10]. Known barriers to attending follow-up care in Canada include limited knowledge of late effects and recommendations for follow-up care, treatment factors (eg, diagnosis and type of treatment), distance from cancer-specific follow-up clinics, and sociodemographic factors (eg, minority status and male sex) [7,10]. Furthermore, only 35% of survivors recognize that they could develop a serious health problem [11]. Young adults (aged between 18 and 39 years [12]) survivors of childhood cancer are especially at risk of not receiving consistent long-term follow-up care due to their unique developmental stage. Specifically, this is a critical and dynamic time during which major transitions occur, including greater autonomy, independent living, and financial independence, as well as a period of increased mental health concerns and risk-taking behaviors [13]. The transition from pediatric to adult health services is also particularly challenging and may result in diminished engagement in follow-up care [14].

Mobile Health Intervention for Survivors of Childhood Cancer

Development of innovative interventions is needed to better educate and engage survivors of childhood cancer in their follow-up care. An ideal intervention would meet survivors where they are, be accessible outside of formal follow-up care programs, account for the sociocultural context of survivors [15], and ultimately address the distinct needs and challenges of this population. Mobile health (mHealth) refers to the use of wireless technology in medical care to deliver health education. mHealth has shown potential to educate patients about preventative health care [16]. Furthermore, mHealth may serve to break down some of the geographic barriers and issues related to accessibility faced by survivors of childhood cancer residing in more remote or rural regions [17]. Given the widespread use of smartphones among young people [18], mHealth has the potential to address the unmet psychosocial and health care needs of young adult survivors of childhood cancer. Therefore, the use of mHealth is a promising new form of intervention to educate, connect, and empower survivors of childhood cancer on the importance of cancer-specific follow-up care.

mHealth interventions targeting young adult survivors of childhood cancer are in their infancy [19,20]. A systematic review of eHealth and mHealth interventions in pediatric cancer lends support for the feasibility and acceptability of technology-based approaches to improve outcomes of children with cancer. However, evidence of the effectiveness of interventions targeting specific outcomes (eg, emotional distress and health behaviors) remains mixed [19]. Limitations of existing interventions include restrictions to single-site pretest-posttest designs, a failure to consider an iterative process to intervention development, and a lack of engagement with patients as partners in the co-design of these programs, and intervention components must be guided by survivors' priorities [21]. Engaging patients as partners is a feasible and efficient way to conduct clinical research, as they are considered experts through their own lived experiences [21]. Research shows that patient engagement in health research has many benefits,

including higher participation rates, design of study protocols with more relevant outcomes, and more meaningful and accessible means of disseminating research to study participants and community members [21]. However, survivors are seldom included in the development of interventions directed at improving knowledge of the importance of long-term follow-up care. Therefore, co-designing with survivors of childhood cancer in addressing their own barriers to attending follow-up care may be a promising means to enable survivors of childhood cancer to understand and engage in their follow-up care and maximize their outcomes. There is also limited research that incorporates health care provider perspectives [22,23]. Health care providers that deliver follow-up care to survivors of childhood cancer may offer complementary insights on the mHealth intervention development in the context of Canadian health care systems.

Current Research

The overarching goal of this research was to co-design an mHealth intervention with young adult survivors of childhood cancer, as well as health care providers that deliver follow-up care to this population. In this research, we refer to co-design as a meaningful engagement with end users in the entirety of the research process [24]. In addition, our research approach adhered to the best practices outlined by the Strategy for Patient-Oriented Research (SPOR) of the Canadian Institutes of Health Research (CIHR) [25].

As a foundational step toward mHealth intervention development, we conducted a qualitative study to understand and identify the priority components to be included in an mHealth intervention that can engage, educate, and empower survivors of childhood cancer on the importance of attending their follow-up care. Subsequently, we validated the results by seeking feedback from multiple informants, including survivors, caregivers, health care providers, and researchers, through a community engagement event.

Methods

Overview

This study was conducted as part of a larger, multiphased project regarding survivors of childhood cancer and their follow-up care experiences [26,27]. We used patient-oriented research and qualitative descriptive study design. Qualitative description is a qualitative research framework that provides detailed experiences directly from participant perspectives [28]. This study design is well suited for health sciences research because it captures broad insights and helps address key clinical and health care services questions [29].

Study data were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) electronic data capture tools hosted at the University of Calgary. REDCap is a secure, web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for data integration and interoperability with external sources [29,30].

Ethical Considerations

Ethics approval for this study was granted by the Health Research Ethics Board of Alberta—Cancer Committee (HREBA.CC-20-0248). As this project involved human participants, informed consent was obtained from all participants. Data collected were deidentified before data analysis. Participants were compensated for their time in the study (refer to the Procedure section for compensation details).

Patient and Public Involvement

In line with best practices of CIHR SPOR [25], we have reported the background, objectives, methods, and results of this project based on the checklist from the Guidance for Reporting Involvement of Patient and the Public Short Form [31]. Patient partners were involved as coresearchers throughout the project, including study design, recruitment, data collection, interpretation of results, and knowledge dissemination. Patient partners were compensated financially for their time.

Participants

Survivors were eligible to participate in the qualitative study if they were (1) currently aged between 18 and 39 years, (2) diagnosed with cancer before the age of 18 years, (3) at least 5 years after diagnosis and/or 2 years after treatment, and (4) currently living in Canada. Health care providers were eligible to participate in the study if they were (1) delivering clinical care to survivors of childhood cancer, (2) practicing for >5 years, and (3) currently living in Canada.

Participants were eligible to participate in the community engagement event if they were (1) survivors of childhood cancer as defined earlier, (2) parents or caregivers of survivors of childhood cancer, (3) health care providers that provide follow-up care to survivors of childhood cancer as defined earlier, or (4) researchers of childhood cancer survivorship.

Recruitment

Several strategies were used to recruit individuals from diverse sociodemographic backgrounds as part of the larger, multiphased project [26]. First, survivors and health care providers were informed about the study by their local or provincial long-term follow-up clinics, where they were given the option to complete a consent-to-contact form and subsequently invited to participate by the study team. Second, patient partners assisted in recruiting survivors across geographical areas in western, central, and eastern Canada. The practice of patients recruiting patients to integrate diverse patient perspectives is supported by the principle of inclusiveness according to the CIHR SPOR framework [25]. Third, social media outreach was used, including X (formerly known as Twitter; X Corp), Facebook (Meta Platforms), and Instagram (Meta Platforms). This strategy involved recruiting through patient and young adult cancer advocacy groups and communities by sharing study graphics (eg, Childhood Cancer Survivor Canada and #AYACSM).

Participants that completed other phases of this project were invited to participate in this qualitative study. Participants that participated in other phases of the project, as well as those who did not participated in the project, were all welcome to attend the community engagement event.

Procedure

We conducted 5 online focus groups with survivors of childhood cancer on Zoom (Zoom Communications Inc) between March and July 2022. Focus group discussions lasted approximately 90 minutes. Eligible participants were contacted and provided with a link to consent to the study on the web through REDCap. They were then provided with a Zoom link and date for the focus group. Participants were compensated with a CAD $25 (US $17.91) e-gift card for their participation. Similarly, we conducted telephone interviews with health care providers who followed the same consent process as survivor participants. Each interview lasted 15 to 20 minutes. Health care providers were also compensated with a CAD $25 (US $17.91) e-gift card for their participation.

We hosted a half-day online community engagement event via Zoom in October 2022. The event was cofacilitated by 2 academic researchers (FSMS and SHJH) and 2 patient partners (RD and IR). A collaborator from a digital technology company, Cambian, attended the event to provide input from a feasibility and usability standpoint. Cambian specializes in providing collaborative health care information services. A graphic recording artist from Fuselight Creative attended the event to visually capture central and representative ideas generated from the community event. Fuselight Creative is a visual facilitation company that specializes in the strategic use of graphic recording to enhance interactive engagement and learning in community events.

Eligible participants were asked to register for the event and provide their email address to receive the Zoom link. Registered participants provided verbal consent to partake on the day of the event on the web via Zoom. Participants were compensated with a CAD $25 (US $17.91) honorarium for their participation in the community engagement event. Likewise, patient partners (RD and IR) were compensated for their cofacilitation of the event.

Measures

Sociocultural Demographics

For the qualitative study, survivors completed an initial questionnaire regarding their sociodemographic background, including age, sex, gender, relationship status, family composition, and race and ethnicity. Survivors also completed questions regarding their clinical history, such as their diagnosis, age at diagnosis, type of treatments received, and number of years after treatment. Health care providers completed a similar questionnaire to survivors regarding their sociodemographic background, as well as questions regarding their professional experiences, including the number of years practicing and years of experience delivering care to survivors of childhood cancer.

Interview Guide

Survivors engaged in an online focus group discussion, guided by a semistructured interview guide exploring the development of an mHealth intervention for survivors of childhood cancer to optimize their engagement in follow-up care. Questions were codeveloped with patient partners and centered on the development of content (eg, “What are your thoughts on creating a platform for survivors?”), education (eg, “What type of information would you want included?”), communication (eg, “Would you want an opportunity to interact with other survivors and/or healthcare providers?”), and engagement (eg, “What are some features that could help you engage in your follow-up care?”) related to the mHealth intervention.

Health care providers engaged in an individual telephone interview. The same set of semistructured interview questions was administered, with wording modified to reflect the health care providers’ perspective on the same topic. The complete set of questions asked during the focus groups can be found in Multimedia Appendix 1.

Small and Large Group Discussions

As part of the community event, participants engaged in interactive small and large interactive group discussions. Facilitators posed semistructured questions on the accuracy, representativeness, and resonance of the themes conceptualized from the data. During small group discussions (3-4 participants per group), differences in perspectives were discussed in the presence of an observer who was a member of the research team. The observer’s role is to ensure equal participation among members and that no one individual or group’s views are dominant over others. The observer takes notes, facilitates the flow of conversation and turn-taking among participants, and shares a summary of key discussion points with members to ensure the accuracy of notes taken. Key discussion points from the small group sessions were subsequently integrated into the larger group discussion (all participants) to achieve consensus, facilitated by 2 main facilitators (SHJH and FSMS) and 2 patient partners (RD and IR). Large group discussions were facilitated using interactive media features, including a Google Jamboard, Mentimeter, and polling on Zoom.

Satisfaction Survey

Participants completed a web-based questionnaire within 1 week following the event. Questions were derived and adapted from the Public and Patient Engagement Evaluation Tool [32]. Participants were asked to rate the extent to which they agreed with statements regarding their experience at the event, using a 5-point Likert scale (1=strongly disagree and 5=strongly agree). Questions were centered on participants’ understanding of the nature of the workshop (eg, “I had a clear understanding of the purpose of this community workshop”), as well as whether the workshop met the stated objectives described herein. We also solicited feedback from participants using an open-text response format.

Analysis Plan

Sociodemographic information gathered from participants was summarized using descriptive statistics and analyzed with SPSS (version 28.0; IBM Corp). Qualitative descriptive data gathered from focus groups and interviews were audio-recorded, transcribed verbatim, and deidentified. Transcripts were subsequently analyzed using reflexive thematic analysis [33]. Reflexive thematic analysis is a theoretically flexible analytical approach used to identify patterns of meaning and conceptualize them into themes. This type of analysis is commonly used in qualitative description studies [34]. In total, 5 researchers

(SHJH, BH, RD, KM, and HW) conducted the analysis according to the steps outlined by Braun and Clarke [33]. First, researchers read the transcripts several times to familiarize themselves with the data. Next, researchers developed a preliminary coding framework and coding book. Each researcher independently coded the same transcript and met as a team to compare coding choices and ensure alignment in the interpretation of the codes. Following this, the remaining transcripts were each analyzed by 2 researchers. Theme abstraction took place iteratively among the researchers by meeting regularly to discuss and review each researcher's interpretations. Any differences in theme abstraction were reconciled through open dialogue until a consensus was reached. Data analysis was supported using NVivo (version 14; Lumivero) [35], a qualitative data management software. Multimedia Appendix 2 includes a reflexivity statement from all researchers involved in the thematic analysis.

Member checking, also known as response validation, is a validation technique used in qualitative health research and has been recognized as a method of rigor to enhance the trustworthiness of qualitative data [36]. Member checking was conducted with multiple informants, including survivors, caregivers, health care providers, and researchers, through an online community engagement event. Member checking was achieved through multiple methods, including small and large group discussions, satisfaction surveys, and graphic recording throughout the community engagement event.

Results

Participant Characteristics

In total, we conducted 5 online focus groups with 22 survivors of childhood cancer, with a range of 4 to 6 participants present per focus group. The mean age of survivors was 29.19 (SD 4.78) years. Over 95% (21/22) of survivors indicated that their assigned sex was female and likewise that their gender identity was female. Most survivors (19/22, 86%) identified as White. Survivors came from 6 provinces, most prevalently from Alberta (10/22, 46%), Ontario (5/22, 23%), and Nova Scotia (4/22, 18%). Most survivors (18/22, 82%) lived in an urban geographical region. Survivors reported a history of leukemias (11/22, 50%), lymphomas (6/22, 27%), and solid tumors (5/22, 23%) as the most common diagnoses. The average age at diagnosis was 10.59 (SD 5.45) years, and the mean time off treatment was 17.45 (SD 6.81) years.

We conducted individual telephone interviews with 7 health care providers. In total, 6 health care providers indicated that their sex assigned at birth was female and that their gender identity was female. All health care providers identified as White. Health care providers included 3 registered nurses and 4 allied health professionals. All health care providers had experience delivering care to survivors of childhood cancer; 29% (2/7) of the participants had between 1 and 5 years of experience, 43% (3/7) of the participants had between 5 and 10 years of experience, 14% (1/7) of the participants had between 10 and 15 years of experience, and 14% (1/7) of the participants had >15 years of experience. Health care providers were from Alberta (2/7, 29%), British Columbia (3/7, 43%), Manitoba (1/7, 14%), and Ontario (1/7, 14%). All health care providers reported residing in urban regions. A summary of participant demographic characteristics is provided in Table 1.

Table 1. Demographic and clinical characteristics of survivors of childhood cancer (n=22) and demographic and professional characteristics of the health care providers of survivors of childhood cancer (n=7).

Demographic and characteristics	Value
Survivors of childhood cancer: demographic and clinical characteristics	
Current age (y), mean (SD)	29.19 (4.78)
Sex, n (%)	
Male	1 (5)
Female	21 (96)
Gender, n (%)	
Male	1 (5)
Female	21 (96)
Ethnicity[a], n (%)	
Indigenous, First Nations, Inuit, or Métis	2 (9)
Black, African, or Caribbean	1 (5)
East Asian	2 (9)
White or European	19 (86)
Other	2 (9)
Province of residence, n (%)	
Alberta	10 (46)
New Brunswick	1 (5)
Nova Scotia	4 (18)
Ontario	5 (23)
Quebec	1 (5)
Saskatchewan	1 (5)
Geographic region, n (%)	
Rural	4 (18)
Urban	18 (82)
Distance from major urban center (km), n (%)	
0	11 (50)
1-20	4 (18)
21-40	2 (9)
>40	3 (14)
No response	2 (9)
Age at diagnosis (y), mean (SD)	10.59 (5.45)
Time after treatment (y), mean (SD)	17.45 (6.81)
Cancer diagnosis, n (%)	
Leukemia (eg, acute lymphoblastic leukemia and acute myeloid leukemia)	11 (50)
Lymphoma (eg, Hodgkin and non-Hodgkin)	6 (27)
Solid tumor (eg, Wilms tumor and osteosarcoma)	5 (23)
Health care provider: demographic and professional characteristics	
Sex, n (%)	
Male	1 (14)
Female	6 (86)
Gender, n (%)	

Demographic and characteristics	Value
Male	1 (14)
Female	6 (86)
Ethnicity: White or European, n (%)	7 (100)
Profession, n (%)	
Registered nurse	3 (43)
Allied health professional	4 (57)
Time in profession (y), mean (SD)	22.43 (8.00)
Years delivering care to survivors of childhood cancer, n (%)	
>1 but <5	2 (29)
>5 but <10	3 (43)
>10 but <15	1 (14)
>15	1 (14)
Province of residence, n (%)	
Alberta	2 (29)
British Columbia	3 (43)
Manitoba	1 (14)
Ontario	1 (14)
Geographic region: urban, n (%)	7 (100)
Distance from major urban center (km), n (%)	
0	2 (29)
1-15	1 (14)
>15	0 (0)
No response	4 (57)

[a]Participants were able to select all that apply. Total responses may exceed the total number of participants. Those that indicated "other" were invited to specify if they desired. In this sample, participants did not specify "other."

Priority Components to Be Included in an mHealth Intervention for Survivors of Childhood Cancer and Health Care Providers

Connections: Establishing Connections Is a Top Priority

Survivors' Perspective

Survivors spoke of the importance of being able to connect with other survivors because they feel validated by those who understand that the cancer journey does not end when treatment ends and that, in fact, the cancer experience is lifelong (quote 1). Survivors also felt that the ability to communicate with other survivors on mHealth would allow them to establish a sense of community through a shared experience and shared understanding of having been impacted by cancer (quotes 2 and 3). Survivors specified that such a connection could be a therapeutic process as well as an opportunity to provide peer support (quotes 4 and 5). Survivors also discussed the possibility of being able to connect with their health care provider through additional means outside of clinical care (quote 6). Finally, survivors emphasized that being able to establish these connections with other survivors of childhood cancer as well as health care providers was a means of self-empowerment and self-advocacy because they could take control of their own care and experiences (quotes 7-9).

Health Care Providers' Perspective

Health care providers also recognized the importance of connecting survivors with one another. Health care providers shared that these social connections are likely to mitigate feelings of isolation for survivors in their cancer experience and, in turn, empower them to engage in their follow-up care (quotes 10 and 11). Health care providers also expressed a desire to connect with other health care providers across Canada for the opportunity to consult and learn from one another about pertinent matters in delivering follow-up care for survivors of childhood cancer (quote 12). In addition, health care providers shared that being able to connect with survivors (eg, navigating appointments and/or scheduling), as well as with primary care providers, may be a useful feature to incorporate in mHealth to enhance the efficiency with which they deliver follow-up care (quote 13; Textbox 1).

Textbox 1. Theme, subthemes, and representative quotes from survivors (n=22) and health care providers (n=7) on the development of an mHealth intervention to enhance follow-up care. This textbox covers theme (1), "connections: establishing connections is the top priority."

Survivors' perspective

- Connecting with other survivors feels validating because not everyone understands the cancer journey and that it is lifelong.
  - Quote 1: "...I think in just hearing others communicate about their experiences to speak to the volume and importance of it's not done after, we're not done after treatment, it's a lifelong thing that we'll always re-engage in at various points in time in different ways." [S6F2]I
  - Quote 2: "...the idea of a platform is really great, especially when you think about people's readiness to engage in that kind of stuff...So having that available when people are ready to engage gives them access and gives them a sense of control about when they are actually feeling they need that support versus when you have that follow-up first talk..." [S6F2]
  - Quote 3: "And you can be surrounded with people that do care about how you're feeling and how you're doing but no one knows what it's like. So being able to connect to people that have gone through the same things as you, that can validate your feelings and all of that..." [S4F5]
- Connecting with other survivors can help create a sense of community.
  - Quote 4: "Almost kind of like meet-up where you could get a bunch of survivors together and like, all go for coffee or something and we'll just hang out and talk about our stories or like how we're moving forward with our lives or what's going on...I think it would be really cool to see something on a bigger scale that's perhaps all the way across the country or just something along that line, as well as having resources." [S4F3]
  - Quote 5: "And also, maybe an option to talk to people who have been through the same thing, but not talk about cancer at all...Build up a bigger community around it without feeling super nervous or weirded out by it." [S4F3]
- Connecting with health care providers may help ensure a shared understanding of what occurs at follow-up care and increase access to health care providers.
  - Quote 6: "When I said, 'I am anxious about ABC' or like, 'XYZ happened to me in my life,' did they hear that or did they just say, 'Worried about cancer but looks good. Goodbye,' like I want to know that they had the same thing. And it would be lovely to have some kind of ability to keep on top of that or keep track of that whether that be some variation of like there's a passport where it's just like, 'We had this conversation on this date' and then I know for future that it has been brought up or it has been talked about." [S1F4]
- Self-advocacy is integral to being a survivor of childhood cancer, and teaching people how to empower themselves through connections and knowledge is key to their health care.
  - Quote 7: "They put you kind of, you're more involved in the whole process if you have really easy access to all the information. Like [participant 3] was saying about her binder, and how she's sort of being involved in the process with keeping track. I think that just helps you feel like you're taking control of your own health. And it's like, that would help engagement." [S2F1]
  - Quote 8: "I think for some of us who are a little more like question shy when it comes to like speaking to doctors, like I know, I often find myself listening and not asking. I think maybe including information on how to like advocate for yourself as a patient would be important." [S3F3]
  - Quote 9: "I'm not 100 percent sure on how this would fit in with the platform yet, but I think that something in aftercare that is kind of missing that might be beneficial in kind of helping people to be able to advocate and engage in their care a little bit more is like—I don't know if it's the first appointment you ever have when you go to adult care or the last one or somewhere in-between." [S3F4]

Health care providers' perspective

- Health care providers recognize that connecting survivors with other survivors is a priority. These social connections can potentially help mitigate feelings of isolation and empower them to engage in their follow-up care.
  - Quote 10: "Yeah well, I agree with the idea of a platform, it helps people to move from kind of the more insular cancer experience to broadening and connecting in the community and identifying what they have in common with other people because I think that would help people and with the focus on mitigating of effects." [HP2]
  - Quote 11: "Really help them with like social opportunities because a lot of them are so isolated socially. Some of them have a really hard time making friends, because they don't like, it's not like a muscle memory they learned in high school. Like where they don't—they just have a hard time knowing how to make and maintain friendships. So that I think would be a useful sort of like, a place where people like me, across the country, could refer to get people connected socially." [HP5]
- Health care providers want to connect with other health care providers to discuss, ask questions, and share expertise.
  - Quote 12: "No. I think one interesting thing too is that we, as a long-term follow-up group, have a Canadian group that we talk about. Like we need every, say, quarterly online and discuss, you know, different issues and what are you doing here? What are you doing here? So, I think having that expertise, Canadian expertise of those groups would be good. So, you know, when this is up and running it would be great if that was presented at our group. Like, you could join our email Zoom, and you know, get everyone on the same page of engagement." [HP6]

- Health care providers indicated that navigating appointments and scheduling with survivors and communication with primary care providers may be useful features.
    - Quote 13: "So, I don't know if adding the primary care information and being able for them to amend it as it changes. I don't know if there's a functionality where when they see an update, they could forward that to their primary care provider. I mean I'm thinking really big here and I know it's nearly impossible to do but, you know, there must be a way to maybe get things together even if it means, you know, they have to physically show it to their primary care provider when they go for their next visit." [HP1]

Education and Information: Having a Cancer History Profile Can Be Useful

Survivors' Perspective

Features that provided information on and tracking survivors' cancer history were some of the most prevalent themes discussed by survivors. Specifically, survivors emphasized the importance of having their cancer treatment history summarized in the form of a cancer profile, along with information regarding the type of follow-up care they receive, a summary of their follow-up care visits, and a summary of past and current medications and treatments (quote 14). Participants highlighted several benefits of this feature in an mHealth intervention, including being able to recall their medical history during appointments with new health care providers, establishing a shared understanding between survivors and health care providers, addressing any gaps in communication, and making their follow-up care more accessible (quotes 15 and 16).

Health Care Providers' Perspective

Health care providers identified a need for a cancer profile so that information such as treatment history (eg, cumulative doses and exposures) can be summarized in 1 hub (quote 17). Furthermore, health care providers indicated that a cancer profile may serve as an effective tool to communicate with survivors about their follow-up care, including the importance of screening for late effects, provided that the cancer profile can include a summary of potential risks of late effects related to diagnosis and treatment (quote 18). Similarly, health care providers explored the possibility of linking updated survivorship guidelines to a cancer profile, which may be informative for survivors of childhood cancer (quote 19; Textbox 2).

Textbox 2. Theme, subthemes, and representative quotes from survivors (n=22) and health care providers (n=7) regarding the development of a mobile health (mHealth) platform to enhance follow-up care. This textbox covers theme (2), "education and information: creating a personalized cancer profile would be helpful."

Survivors' perspective

- A treatment summary can help with communicating with health care providers at follow-up care visits and remembering treatment history details.
    - Quote 14: "Yeah, I mean, I feel like, earlier like I was saying, I have that binder which I feel like is kind of like a one stop shop. But if that was in a digital form, I feel like that would be, that would be useful. Some sort of platform where, yeah it would have some sort of resource thing, but customized to what treatment you had, what drugs specifically you were on. Something that would have—even when you go to aftercare, a digital version of your test results as well. That way it's stored there and there's a history you can refer back to." [S3F1]
- Being able to track and view the history of follow-up care visits would make care more convenient, transparent, and accessible.
    - Quote 15: "...I guess a lot of things are getting more digital now so just being able to have access to records, maybe possible notes of previous visits you had. So, something just to refer to would be very convenient." [S1F1]
    - Quote 16: "...let's say if you want to refer back to a visit that you did a couple of years ago, have some notes that you can kind of refer to. Other than, kind of see how you've improved. And I think that would just help to refer back and then see how it's progressed, from any concerns from past follow-ups." [S1F1]

Health care providers' perspective

- A treatment summary can help health care providers consolidate information and inform treatment planning.
    - Quote 17: "I think that making sure that it has like a comprehensive treatment summary, and by that, I mean...like the significant cumulative doses, that they know what complications that they have, and then the things at risk. But I think the number one thing on the very first page after their treatment summary would be the very practical things you have to do, echoes, when's your next test due; and to also make sure that they have a primary care physician" [HP1]
- A cancer profile may help health care providers communicate the importance of understanding and screening for late effects for survivors of childhood cancer.
    - Quote 18: People have to be prepared kind of for the long haul without knowing exactly what the long haul is. But for some people we have a fairly good idea that they will like pretty significant cognitive or physical disabilities depending on what they've gone through. So, the challenge of figuring out how to help people with that process as it evolves." [HP2]
- Linking updated survivorship guidelines to a cancer profile may be informative for survivors of childhood cancer.
    - Quote 19: "I think it could be linked to the survivorship guidelines somehow but that the less we put and the most important stuff that we put in that will actually make it useful." [HP1]

Engagement: Issues Related to Accessibility and Health Equity Impact Survivors' Engagement in Their Follow-Up Care

Survivors' Perspective

Survivors discussed the challenges related to accessing their follow-up care, referring to the ease of reaching and using health care services, and how these barriers reflected issues related to health inequities that some experience. For example, survivors felt that all survivors requiring follow-up care should be able to access their follow-up care, and yet, disparities exist in this regard (quote 20). Survivors spoke about their difficulty in accessing follow-up care due to geographic barriers, especially for those that reside in more rural or remote communities (quotes 21-23). In addition, survivors shared that features such as appointment reminders, accommodation provided for transportation, and being able to ask questions to health care providers in an mHealth intervention may be ways that can help increase access to their follow-up care (quote 24).

Health Care Providers' Perspective

Health care providers recognized that health inequities exist in the current service delivery of follow-up care to survivors of childhood cancer and highlighted the need to make mHealth easily usable and accessible (quote 24). Health care providers also identified practical strategies that can help survivors engage in their follow-up care, such as displaying a list of the upcoming appointments on mHealth or outlining follow-up care in a stepwise fashion so that survivors that face cognitive effects of their treatment can be accommodated and supported in their engagement with follow-up care (quotes 25 and 26; Textbox 3).

Textbox 3. Theme, subthemes, and representative quotes from survivors (n=22) and health care providers (n=7) on the development of an mHealth intervention to enhance follow-up care. This textbox covers theme (3), "engagement: issues related to accessibility and health equity impact survivors' engagement in their follow-up care."

Survivors' perspective

- Appointment reminders, flexibility, and accommodations for transportation are ways to increase access and engagement to follow-up care.
  - Quote 20: "I would love to have an appointment at a certain time and for the appointment to be like within one hour of that time rather than that's the only thing on my day because I might be pushed four or five hours. And I understand like oncology is a field that does have emergencies and things happen and yes, like those precedent absolutely. But when I'm coming there just to get a rubber stamp that said I've done it and they've already looked at my blood work, where it definitely could've been a phone call, but I was spending my whole day in clinic instead, those are the things that would make it easier for me to engage more with my follow-up rather than treat it like a check box." [S3F1]
- Health inequities exist, and people from remote or rural communities have a harder time accessing their follow-up care. Telehealth mitigates some of these challenges but not all.
  - Quote 21: "But aside from that a lot of the more helpful things to engage with follow-up are things that unfortunately, unless you're building a city from scratch, are very hard to accomplish. I would like to be near my area, I would like more my follow-up is being happening, I would like for there to either be parking or public transport that is easy to access and get me there. All of those things." [S1F4]
  - Quote 22: "And if that's something that, like I said out here, it was shocking to live the reality that health equity is a dream, not a reality in Canada as much as we want it to be. Because like people in [city], the health care here is, it's just not accessible, it's brutal...I mean, I think Zoom and all of the telecommunication stuff will really help now, hopefully where people are more open to doing these sorts of things virtually, rather than having to drive four hours to [city], that sounds horrible." [S4F5]
  - Quote 23: "But just, if there was something that could address that for people who maybe can't access care where they live remotely, and they can actually be connected with practitioners or—yeah, I don't know. I think that's all I wanted to say about the app." [S3F5]
- Being able to communicate with the health care team and ask questions may help increase access to follow-up care.
  - Quote 24: "I also think a message feature would probably be efficient because I think of like, if I have to call the clinic sometimes that can be a little bit of a hassle with my phone tag. But if I knew that on the App, I could just send a message saying something like, 'appointment needs to be changed,' whatever. 'Can you give me a call?' Or even just being able to do it that way completely would be useful." [S3F1]

Health care providers' perspective

- Health inequities exist, and access to care can be increased by showing survivors how to navigate mHealth and making mHealth readily available.
  - Quote 24: "I think that, if I had to be very sort of realistic, I think that making sure that they have things to access is the most important. I still think that that probably would be like a big barrier to building this if we had to add that functionality [to the mHealth platform]." [HP1]
- There are practical strategies that may help survivors of childhood cancer engage in their follow-up care.
  - Quote 25: "But I think the number one thing on the very first page after their treatment summary would be the very practical things you have to do, echoes, when's your next test due." [HP1]
  - Quote 26: "Yeah, just that transition to adulthood thing of helping people see what are the normal kinds of tasks that people learn to become adults; so that's kind of more concrete for them. Perhaps break some of those tasks down into concrete steps so it's not so overwhelming." [HP2]

Resources: Providing Personalized Resources Can Help Enhance the Survivorship Experience, Psychosocial Well-Being, and Reproductive Health

Survivors' Perspective

Survivors highlighted the need for resources specific to the unique needs of cancer survivorship, highlighting that personalization helps normalize the experience of survivors of childhood cancer (quotes 27 and 28). Survivors also expressed a desire for more mental health resources. They readily endorsed the impact of diagnosis and treatment on their social, emotional, and mental well-being, especially wanting to know more about the notion of posttraumatic growth and posttraumatic stress in survivorship (quotes 29-30). Survivors also highlighted major gaps in resources dedicated to reproductive health, even though there is a negative impact of cancer treatment on fertility and family planning (quote 31). Survivors also noted that educational resources to support survivors' transition back to school are largely absent (quote 32). Finally, survivors discussed the importance of being able to access resources that are up-to-date and reflect recent advances in science and research (quote 33).

Health Care Providers' Perspective

Health care providers identified, as a priority, the need for resources dedicated to survivorship experience, financial assistance, and health care transitions (quote 34). They also identified a gap in psychosocial and educational supports (quote 35). Moreover, health care providers spoke of the need for better resourcing to help survivors understand and navigate their fertility and family planning (quote 36). Finally, health care providers discussed the importance of being able to share resources that are updated and reflect advances in research and science (quote 37; Textbox 4).

Textbox 4. Themes, subthemes, and representative quotes from survivors (n=22) and health care providers (n=7) on the development of an mHealth intervention to enhance follow-up care. This textbox covers theme (4), "resources: providing tailored resources can help enhance the survivorship experience, psychosocial well-being, and female and reproductive health."

Survivors' perspective

- Offering resources on the unique survivorship experience can help to normalize being a survivor of a childhood cancer.
  - Quote 27: "And like it would be nice to have a platform where you can exchange stories and have some of those anxieties assuaged of like this isn't a unique to you experience, it's a unique to the childhood cancer experience. Like there are other people who went through the same thing." [S1F4]
  - Quote 28: "But I think that maybe people would like to access, maybe there's testimonials from people. If it's about people feeling like it normalizes their situation." [S3F5]
- Mental health resources are lacking. Survivors are interested in better understanding the impact of diagnosis and treatment on their social, emotional, and mental well-being, including experiences of posttraumatic growth and posttraumatic stress.
  - Quote 29: "But what I was going to say was, I would say like social, emotional, mental health resources, just that larger I think important too. Or even contacts specifically connected to the aftercare clinic with relation to those. Whether it be social work, or child and youth, depending on the age, that type of thing." [S3F4]
  - Quote 30: "So, I think education maybe around mental health, what certain terms might mean like post-traumatic growths, what is that? Or post-traumatic stress disorder, or survivor's guilt, education content around what kind of complications could pop up for people that have had cancer treatment as a pediatric patient." [S7F2]
- Resources on female and reproductive health of survivors of childhood cancer is inadequate. Survivors would like to be better educated and equipped with the impact of diagnosis and treatment on their female and reproductive health.
  - Quote 31: "I was never told any—I had to ask if I was going to be able to have kids. It was never brought up to me. But I mean, I don't know, I only had chemotherapy. So, I don't know if radiation has something to do with that, but yeah, they've never brought it up." [S2F5]
- Knowing what accommodations and modifications are available to survivors of childhood cancer in their learning and education would be a helpful guide.
  - Quote 32: "I think maybe for deciding where this is for me, at home schooling. Because I know when I was going through treatment they had to—I skipped out like a year and a bit. So, yeah, I guess just resource things or places to refer to, to kind of continue that education if you're not able to be in class." [S1F1]
- Survivors would like resources that are updated and reflect advances in research and science.
  - Quote 33: "...just what's the most up to date evidence, because I think it's constantly evolving, so I think that's a huge piece of that information need and education need. I think it also helps you problem-solve, too, when things come up and you're 'is this normal?' And then you go to research and it's 'yeah, 20 percent or more end up with mental healthcare needs. OK, it's not abnormal for me to be feeling this way or for me to have this challenge.' So, I think there's a normalization element to having information that can support you in navigating follow-up care, for sure." [S6F2]

Health care providers' perspective

- Resources dedicated to the survivorship experience, financial assistance, and health care transitions are priorities.
  - Quote 34: "So, if we can figure out parking, just sort of if you think about another practical suggestion, how do you engage, if you need financial assistance with parking or travel, links to some resources that might be helpful. Not that there's not many, but just sort of how to eliminate barriers." [HP4]
- Health care providers identified a major gap in psychosocial and educational supports.
  - Quote 35: "Well, the big mental health ones I think is key, mental health, like the, I know [location] Health Services have their, has their cope, hard, coping in hard times kind of link or something, just those would be I think standard, would be helpful to have that. And even counselling specific, psychosocial counselling, counselors that are, identify, have experience in working in this area. Private ones would be helpful, although, yeah, and then just financial resources as well too, just links to that. I know, you could kind of there's a whole bunch of different ones I'm sure that you could link there, but the counselling, vocational supports." [HP4]
- Better resourcing is needed to help survivors of childhood cancer understand and navigate their fertility and family planning.
  - Quote 36: "You know some common big things that, not so much for our young patients for our adolescent and young adults, I do think that sort of resources about fertility and family planning needs to be somewhere in there because I think that that's something that some patients whether it be cultural or just personal can't openly ask, and I think that we'll agree that that's probably the most complex late effect that we deal with from a physical and mental-health perspective. I think that that—I don't know if it has to be something on its own, but I wish we could be better at telling a young person that, you know, family planning is different for everybody and that there are different ways to have families and the definition of family is changing every day. So, I don't know if that part is just—if I won $10 million I would put the money toward fertility salvage in survivors; that's what I would do." [HP1]

- Health care providers would like resources that are updated and reflect advances in research and science.
  - Quote 37: "Also, a place where they could see where there's a change in practice, like how often we do echoes or, you know, new recommendations for vaccinations, you know, like the very specific stuff that they, you know—flags or alerts I think would be helpful." [HP1]

Personalization: There Is a Need for Personalized Features on mHealth, Including Consideration of Survivors' Readiness; Emotional Impact of Accessing Follow-Up Care; Privacy; and Right to Accurate, Moderated Information

Survivors' Perspective

Survivors highlighted several important considerations to help personalize the mHealth intervention. For instance, survivors described the importance of providing options so that survivors can feel that they have a choice in when and how they wish to navigate mHealth in their cancer journey (quotes 38 and 39). Furthermore, survivors discussed the emotional impact of accessing their follow-up care, highlighting that the survivorship experience is diverse and that they want to be able to establish their own boundaries in accessing their follow-up care and cancer history through mHealth (quote 40). In addition, survivors discussed the importance of incorporating privacy features to protect their identity and health information, whether that is from their family members and/or other members of the mHealth community (quotes 41 and 42). Finally, survivors recognized that, while they would value being able to communicate with other survivors in a forum on mHealth, there is a need to modulate all communication to ensure that any information or resources shared are accurate and not to be mistaken with professional recommendations (quote 43).

Health Care Providers' Perspective

Health care providers expressed their concern for any information that survivors receive that is not communicated directly from a health care provider (quote 44). They expressed a desire to ensure that survivors receive accurate knowledge and resources to support their follow-up care (quote 44). Health care providers also shared their concern for the limited capacity of current health care providers in being able to support the growing population of survivors (quote 45; Textbox 5).

Textbox 5. Themes, subthemes, and representative quotes from survivors (n=22) and health care providers (n=7) on the development of an mHealth intervention to enhance follow-up care. This textbox covers theme (5), "personalization: there is a need to account for personalized features, including survivors' readiness; emotional impact of accessing follow-up care; privacy; and right to accurate, moderated information, in accessing mHealth."

Survivors' perspective

- Being able to choose when to access, and have control over access, mHealth is an important feature. Not all survivors feel ready or want to engage in their mHealth at all times.
  - Quote 38: "And I don't think that it should be alert based, or maybe if there's a way that you opt out of alerts, because I think that that would be probably a stressor, when you're at work, and you like, 'Ping. Have you thought about cancer today?'" [S3F5]
  - Quote 39: "...the idea of a platform is really great, especially when you think about people's readiness to engage in that kind of stuff. Oftentimes...I wasn't ready at first to engage with other youths at that time. It took some time, and it actually was, I made my closest friendships with other survivors five years post, maybe even more than that. So, it was—I was quite defended from that, but there was a readiness there, that I was ready to do that, but it came at an appropriate time. So having that available when people are ready to engage gives them access and gives them a sense of control about when they are actually feeling they need that support versus when you have that follow-up first talk, it's 'hey, here are the resources available to you, have at them when you need them,' and that can be a little intimidating." [S6F2]
- For survivors, there is an emotional impact to accessing their follow-up care. These experiences are diverse, and survivors want to be able to establish their own boundaries in accessing their follow-up care and cancer history.
  - Quote 40: "But my concern with peer support has always been that the people who tend to gravitate toward looking for support, are people who aren't doing well. And so, I don't want to go or be part of something that becomes a co-rumination." [S2F5]
- Establishing privacy features is important for survivors to protect their identity and health information from their family and others.
  - Quote 41: "And I think it's not just me that you're kind of targeting with a platform like this, it's also especially for people who might be younger and still living at home. You're now getting the whole family. Because, with iPads and phones and emails and all these things now, if folks are still living at home, their parents may also be seeing these notifications." [S4F5]
  - Quote 42: "I'd love to meet up with other individuals that have shared my experience one on one or even in a small group setting like this, whether it be like an optional pen pal thing through the platform where you could talk one on one. And, you know, if you wanted to talk to them through video that is your choice but make something so it's secure so you're not giving out personal information if you don't want to." [S4F5]
- It is important that any communication on mHealth is moderated and not to be mistaken with medical or professional advice.
  - Quote 43: "...I do kind of feel as though there might need to be some sort of like regulation on what is said. As much as like 100 percent lived experiences and what people have gone through is so important and can be relayed, but I just would also worry that things would be maybe not communicated clearly and I would just worry that like, 'Oh, someone said that high dose vitamin C treatments at this random clinic did something.' And just that is the only like little thing in my ear that's saying like just be careful is all." [S3F5]

Health care providers' perspective

- Any information shared with survivors of childhood cancer needs to be properly communicated by a health care provider.
  - Quote 44: "You know I think it's a good idea to have some sort of platform where survivors can [pause] I can't think of the right word, that they can look at and give them information. I do sometimes think I worry if they get information without back-up to discuss it and all, like you can't just tell somebody, 'You're going to be infertile' and not go through it all, you know, discuss your options and that sort of thing. That's what I'm thinking from that perspective. [HP3]
- There is a concern for the capacity of health care providers to be able to serve the growing population of survivors of childhood cancer.
  - Quote 45: "I mean, yeah, it's, like I said, the big piece is just increasing capacity and then, yeah, looking at, if we're going to continue on this trend in terms of survivorship. And if we're looking at having the seeing people indefinitely within [location], I don't know if that's the case, then what's the future hold...That's what my concern is, is that we're really, are we really providing the best service possible if we're, we have these volumes and we just don't have the staff to really provide that support, that's my main concern, I guess, where are things going and is there going to be capacity to support it." [HP4]

Validating Findings Facilitated Through Community Engagement

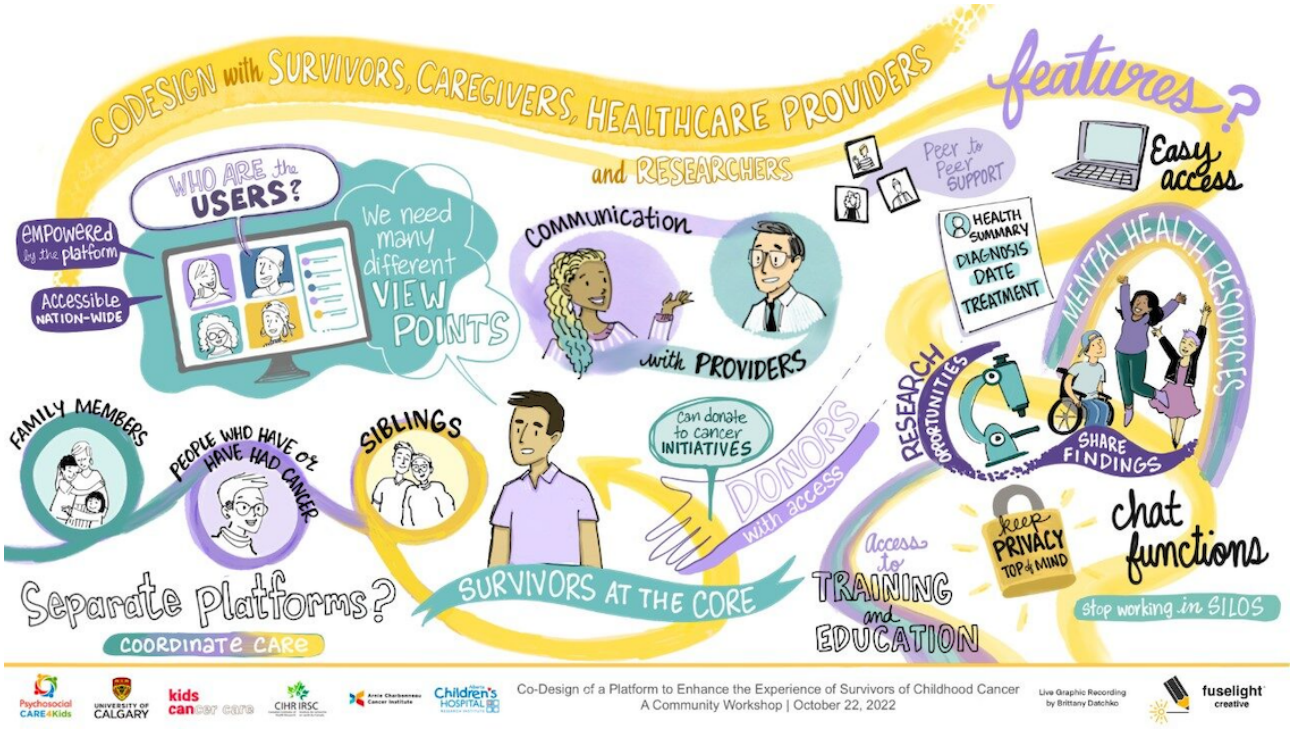
A total of 31 participants, including survivors of childhood cancer (n=10, 32%) and their caregivers (n=2, 6%), health care providers (n=6, 19%), researchers (n=11, 35%, including 2 patient partners), and collaborators (n=2, 6%) participated in the community engagement event to review and verify the qualitative results gathered.

Small and Large Group Discussions

Participants endorsed the accuracy of themes derived regarding the core components to be included in mHealth. Participants explored additional contexts in the development of an mHealth intervention, including considerations for family members (eg,

families and siblings) and the practicalities of preserving privacy and confidentiality on a technology-based intervention. Participants also reported feeling heard through the graphic recording process. The final graphic recording is shown in Figure 1.

Figure 1. Graphic recording of the community engagement event on October 22, 2022. Illustration completed by Fuselight Creative.



Satisfaction Survey

Some participants (8/22, 36%) completed a web-based questionnaire assessing participant engagement at the event. On average, participants *strongly agreed* that they felt that their views were heard (mean 5.00, SD 0.00) and that they were able to discuss their views freely (mean 4.71, SD 0.45). Participants also strongly agreed that they felt they had enough information to contribute to the topics discussed (mean 4.71, SD 0.35) and that a wide range of views on the topics were shared (mean 4.71, SD 0.45). Participants *mostly agreed* that they had a clear understanding of the purpose of the community workshop (mean 4.14, SD 0.35). In addition, participants were asked to indicate the extent to which they agreed that the community engagement event objectives (ie, to co-design mHealth, to engage with others, and to provide feedback on study results) were achieved, and their responses ranged from *mostly* to *strongly* agreed (mean scores ranging from 4.00 to 4.71, SD scores ranging from 0.45 to 0.83). Importantly, participants *strongly agreed* that they felt confident that the input provided through this event would be incorporated by the study team. A summary of participant responses is reported in Tables 2 and 3.

Table 2. Participant responses to community engagement event evaluation (n=8)[a] for section 1 (overview of the workshop)[b].

Overview of the workshop	Responses, mean (SD)
"I had a clear understanding of the purpose of this community workshop."	4.14 (0.35)
"I had enough information to contribute to the topics discussed."	4.86 (0.35)
"I was able to express my views freely."	4.71 (0.45)
"I felt that my views were heard."	5.00 (0.00)
"A wide range of views on the topics discussed were shared."	4.71 (0.45)

[a]A total of 8 attendees, including 4 survivors of childhood cancer, 1 caregiver, 2 family members, and 1 researcher, completed the evaluation.

[b]Attendees were asked to rate the extent to which they agreed with statements in this table, using a rating scale where 1=strongly disagree and 5=strongly agree.

Table 3. Participant responses to community engagement event evaluation (n=8)[a] for section 2 (workshop objectives)[b].

Workshop objectives	Responses, mean (SD)
"To describe the preliminary results from the study on improving follow-up care for survivors of childhood cancer."	4.14 (0.83)
"To co-design a platform for survivors of childhood cancer that is agreed upon within this community."	4.00 (0.76)
"To engage with other core community members (health care professionals, researchers, families, and caregivers)."	4.71 (0.45)
"I am confident that the input provided through this community workshop will be used by the CARE4Kids team at the University of Calgary."	4.71 (0.45)

[a]A total of 8 attendees, including 4 survivors of childhood cancer, 1 caregiver, 2 family members, and 1 researcher, completed the evaluation.

[b]Attendees were asked to rate the extent to which they agreed that the objectives listed in this table were achieved, using the rating scale where 1=strongly disagree and 5=strongly agree.

Discussion

Principal Findings

It is important that survivors of childhood cancer receive routine health care follow-up, yet many survivors do not understand the importance of their follow-up care or have limited knowledge on their need for follow-up care. Our study aimed to amplify the voices of young adult survivors of childhood cancer in identifying the priority components to be included in an mHealth intervention that can help educate and engage survivors in their long-term follow-up care.

Our approach was unique in using co-design and a qualitative research framework and adhering patient-oriented research principles to engage in mHealth intervention development [21,37]. We uniquely incorporated perspectives from survivors of childhood cancer and health care providers that deliver cancer-specific follow-up care. Patient partners collaborated with our team over the entirety of the research process, including study design, recruitment, data collection (ie, cofacilitation of focus groups and interviews), interpretation of results, and knowledge dissemination. Furthermore, a community engagement activity was cofacilitated with patient partners to enhance the validity of qualitative descriptive data gathered.

Limited mHealth interventions have been created *with* survivors of childhood cancer despite its rising potential to improve health [38]. Most efforts have focused on patients who are on active cancer treatment. For instance, a systematic review indicated a positive effect of mHealth interventions on improving the health-related quality of life of adult patients with cancer [38]. Other studies have explored delivering a survivorship care plan and an app for enhancing self-management for adolescents and young adults [39]. Few studies have effectively engaged in the co-design of mHealth, and even less work has been conducted with young adult survivors of childhood cancer in this regard, as well as with health care providers. Therefore, our work provides important and novel insights from the perspectives of individuals with lived experience of cancer and health care providers regarding their follow-up care experiences. Specifically, results shed light on the priority areas necessary to increase knowledge of and engagement in follow-up care for survivors of childhood cancer, establishing a critical foundation in mHealth intervention development.

In total, 5 major themes were conceptualized as the priority components of the mHealth intervention. Many of the themes are consistent with the existing literature on the unmet needs of survivors of childhood cancer. For example, survivors and health care providers from our study identified a lack of knowledge of diagnosis and treatment, as well as associated late effects, for survivors. These results are consistent with past work showing that most survivors of childhood cancer fail to recognize their risk for developing a serious health condition [11,40]. Furthermore, previous research indicates that survivors of childhood cancer need a better way to learn about and engage in their own health information [10,41]. Indeed, our findings indicate a need for advanced means to deliver follow-up care knowledge, such as the use of a mobile app or website, as well as the need for creative features to enhance the follow-up care experience, such as reminders for appointments and a cancer profile to consolidate health history information.

Another notable theme generated was the need for more education and support during health care transitions. This builds on an extensive body of literature on health care transitions for young adults impacted by cancer [42]. There are cancer-specific risks and health care needs of survivors of childhood cancer that are distinct from those with other chronic illnesses [43]. Therefore, our research lends further support for the importance of personalized care for young adult survivors during their health care transition from pediatric to adult health care, one that prioritizes educating survivors on their potential late effects related to their diagnosis and treatment, as well as the utility of engaging in consistent surveillance to promote longer-term optimal outcomes.

Survivors expressed some hesitation toward having regular access to their cancer history information and therefore discussed the importance of being able to have choice and control in how they navigate mHealth. These findings are consistent with existing work documenting experiences related to posttraumatic stress experienced by survivors [44] and the importance of prioritizing a trauma-informed approach in intervention development.

Several other themes were generated from this study that are novel and relevant to building an mHealth intervention for survivors of childhood cancer. Survivors and health care providers emphasized the importance of establishing connections with other survivors and health care providers. These results likely reflect a sense of disconnect that survivors experience and a desire to broaden opportunities for meaningful connections. Indeed, survivors spoke about the need to explore their identity beyond their cancer journey. Taken together, these

results are consistent with past work documenting that adolescents and young adults with cancer experience substantial psychosocial challenges, including peer and family relationships and personal growth stresses [45]. Likewise, health care providers echoed the importance of cultivating connections among survivors to address feelings of isolation on their cancer journey. Importantly, health care providers also highlighted a desire to connect with other health care providers that provide follow-up care across Canada, emphasizing the importance of connecting with and consulting one another to stay informed of current concerns and practices in survivorship. Few studies have addressed the unmet needs of health care providers in delivering follow-up care to survivors of childhood cancer [23]. Our study provides unique insight of both survivors and health care providers, offering a more contextualized understanding of how to improve follow-up care from multiple perspectives.

Issues related to accessibility of follow-up care and health equity were a prominent finding from our research. These outcomes contribute to our understanding of some of the geographic barriers faced by survivors from rural or remote regions, as reflected in our study sample. Developing an mHealth intervention will aim to address these barriers by connecting survivors to their follow-up care through technology. However, barriers to health care are complex and dynamic and require consideration of factors beyond the individual, including health care providers and health care systems factors, to alleviate health disparities [46]. Therefore, further work is needed to capture the complete and intersecting effects of accessibility factors, as well as other inequities faced by survivors of childhood cancer, on their receipt of high-quality health care. These are important considerations in intervention development to ensure that the burden of change is not placed solely on survivors of childhood cancer but rather recognized as a systemic problem that requires multilevel intervention.

Our research incorporated a response validation technique, member checking, to enhance the rigor of qualitative data gathered. Previous qualitative work using member checking lacked detail and discussion in the implementation of the technique. Absence of this reporting may be confounded by epistemological and methodological challenges [36]. In this research, we conducted a comprehensive assessment and report of member checking using a multiinformant, multimodal approach to strengthen the credibility and validity of data gathered. The strength of this approach is to demonstrate the true and iterative process that we took to achieve consensus among researchers and those with lived experiences in the research process. Our goal with this undertaking is to enhance the transparency, accessibility, and replicability of best practices in qualitative health research.

Limitations and Future Directions

We review several important limitations to be considered when interpreting the results of this study. A major strength of this research was leveraging technology for participation (ie, online focus groups and community engagement). This meant an increase in accessibility to those residing in remote or rural regions. However, reliance on technology also meant that participants from lower levels of income, or some individuals from geographically more remote or rural regions of Canada, may face greater barriers to participating because they are less likely to have access to technology. Future research incorporating community outreach, phone-based participation, or compensated travel to the local context can help to mitigate this challenge in research recruitment. Input on how we can account for potential technological limitations in reaching individuals from more remote or rural regions of Canada will be important as we build the intervention.

Research shows that a diversity of perspectives drives innovation [47]. Participants from our study offered important insights into accessibility issues that contribute to health inequities for survivors of childhood cancer. However, our sample nonetheless lacked representation and voice from individuals from diverse backgrounds (eg, diversity in ethnicity, gender identity, sex, language use, and geographic regions) and/or who are not engaged in their follow-up care. We recognize that a potential bias of our sample is that most of our sample reported attending their follow-up care. We are missing the voices of individuals who are not engaged in their care at this time, and future research leveraging purposive sampling of those not regularly attending their follow-up care would offer us important insights into the barriers they face in attending care. In addition, there is a notable disparity in gender representation of participants, with 86% (19/22) of participants identifying as female gender. Studies have found that female gender predicted attendance to follow-up care, such that female participants were more likely than male participants to attend follow-up care [48]. This bias in our sampling reinforces that our findings reflect the views of those who are attending their follow-up care. In addition, there are many factors that contribute to a lack of gender representation in cancer research, including researcher bias, gender stereotypes, and unequal social opportunities [49]. Importantly, a review of gender representation trends in psychosocial survivorship research showed that there is a trend toward a more balanced representation of men and women over a 15-year period (2007 and 1992) [49]. Taken together, our skewed sample indicates an important need to continue to engage with survivors from equity-deserving groups, particularly given the increasing diversity of the Canadian population.

Research shows that members of equity-deserving groups face significant barriers to accessing high-quality and accessible health care [50,51]. Without knowledge of the perspectives and experiences of these individuals, we are missing critical information that can help to address health disparities and, in turn, bring greater awareness to the importance of follow-up care surveillance and attendance for those from underrepresented and underserved communities. Implementation of safe, inclusive, and culturally responsive recruitment strategies is needed to increase representation in pediatric cancer research [52].

Finally, this study solicited feedback from participants and other important parties to enhance the credibility and validity of the results. We demonstrated rigor of data by using a multiinformant and multimodal approach to achieve consensus. However, the response rate for completing the satisfaction survey for our community event was 36%, so responses gathered on patient engagement may not be representative of feedback from most attendees. Another limitation of our community engagement

event is that we did not determine whether attendees included those who previously participated in the focus groups and interviews. Therefore, the feedback gathered at the community event may be influenced by participants' familiarity with the study. Future research establishing and implementing a comprehensive and systematic evaluation of patient and public engagement is necessary to strengthen the rigor of our evaluative framework and enhance patient engagement and research.

The next step of this multiphased project is to build an innovative and accessible mHealth intervention prototype based on the core components identified and grounded in an established conceptual framework for co-design of intervention development. Results from this study have provided the bedrock to progress in our development of an mHealth intervention for survivors of childhood cancer to enhance their knowledge of and engagement in their follow-up care.

Conclusions

In this study, we identified core components to be included in an mHealth intervention to increase the knowledge of and enhance follow-up care engagement for survivors of childhood cancer. We engaged in a rigorous and iterative co-design process with survivors of childhood cancer and health care providers. We incorporated a community engagement event to validate our findings with a broader audience of community members. Findings will inform the next phase of our multiphased, co-design project, ultimately aiming to improve follow-up care and long-term outcomes for survivors of childhood cancer.

Acknowledgments

The authors acknowledge Sharon Paulse and Annette Flanders for their support and insights on the recruitment process. This study is supported by funding from the Strategy for Patient-Oriented Research of the Canadian Institutes of Health Research Catalyst Grant, Alberta Children's Hospital Research Institute, and Kids Cancer Care. Postdoctoral fellow SHJH is funded by the Arnie Charbonneau Cancer Institute and Cumming School of Medicine at the University of Calgary Postdoctoral Fellowship Awards.

Data Availability

Data generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

FSMS secured funding acquisition. FSMS and SHJH conceptualized the study and methodology. SHJH led investigation, data curation, project administration, formal analysis, and supervision. FSMS oversaw investigation, data curation, project administration, formal analysis, and supervision. CF and BH supported project administration, data curation, and formal analysis. RD, KM, and HW contributed to portions of formal analysis. IR, PRT, HZ, MS and JD contributed to portions of data curation. EKD, CE, MST, LS, PN, MS, KG, and KR contributed to portions of investigation. SHJH prepared the original draft. FSMS, CF, BH, RD, KM, HW, IR, PRT, HZ, MS, EKD, and JD contributed to reviewing and editing the manuscript.

Conflicts of Interest

EKD owns and operates a consultancy (EmilyDrake.ca). CE is a DSMB member for CONNECT. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Focus group or interview prompts for survivors and health care providers.
[DOCX File , 14 KB - cancer_v11i1e57834_app1.docx]

Multimedia Appendix 2
Researcher's reflexivity statements.
[DOCX File , 13 KB - cancer_v11i1e57834_app2.docx]

References

1. New WHO report highlights scale of childhood cancer inequalities in the European region. World Health Organization. 2022 Feb 15. URL: https://www.who.int/europe/news/item/15-02-2022-new-who-report-highlights-scale-of-childhood-cancer-inequalities-in-the-european-region [accessed 2023-06-12]
2. Childhood Cancer Survivor Canada homepage. Childhood Cancer Survivor Canada. URL: https://childhoodcancersurvivor.org/ [accessed 2023-06-12]
3. Landier W, Bhatia S, Eshelman DA, Forte KJ, Sweeney T, Hester AL, et al. Development of risk-based guidelines for pediatric cancer survivors: the Children's Oncology Group Long-Term Follow-Up Guidelines from the Children's Oncology Group Late Effects Committee and Nursing Discipline. J Clin Oncol 2004 Dec 15;22(24):4979-4990. [doi: 10.1200/JCO.2004.11.032] [Medline: 15576413]

4. Hudson MM, Bhatia S, Casillas J, Landier W, Section on Hematology/Oncology, Children's Oncology Group, American Society of Pediatric Hematology/Oncology. Long-term follow-up care for childhood, adolescent, and young adult cancer survivors. Pediatrics 2021 Sep;148(3):e2021053127 [FREE Full text] [doi: 10.1542/peds.2021-053127] [Medline: 34462344]
5. Bhakta N, Liu Q, Ness KK, Baassiri M, Eissa H, Yeo F, et al. The cumulative burden of surviving childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE). Lancet 2017 Dec 09;390(10112):2569-2582 [FREE Full text] [doi: 10.1016/S0140-6736(17)31610-0] [Medline: 28890157]
6. Michel G, Mulder RL, van der Pal HJ, Skinner R, Bárdi E, Brown MC, et al. Evidence-based recommendations for the organization of long-term follow-up care for childhood and adolescent cancer survivors: a report from the PanCareSurFup Guidelines Working Group. J Cancer Surviv 2019 Oct 08;13(5):759-772. [doi: 10.1007/s11764-019-00795-5] [Medline: 31396878]
7. Reynolds K, Spavor M, Brandelli Y, Kwok C, Li Y, Disciglio M, et al. A comparison of two models of follow-up care for adult survivors of childhood cancer. J Cancer Surviv 2019 Aug 27;13(4):547-557. [doi: 10.1007/s11764-019-00774-w] [Medline: 31250352]
8. American Academy of Pediatrics Section on Hematology/Oncology Children's Oncology Group. Long-term follow-up care for pediatric cancer survivors. Pediatrics 2009 Mar;123(3):906-915 [FREE Full text] [doi: 10.1542/peds.2008-3688] [Medline: 19255020]
9. Edgar AB, Borthwick S, Duffin K, Marciniak-Stepak P, Wallace WH. Survivors of childhood cancer lost to follow-up can be re-engaged into active long-term follow-up by a postal health questionnaire intervention. Eur J Cancer 2012 May;48(7):1066-1073. [doi: 10.1016/j.ejca.2011.06.009] [Medline: 21737254]
10. Oeffinger KC, Mertens AC, Hudson MM, Gurney JG, Casillas J, Chen H, et al. Health care of young adult survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. Ann Fam Med 2004 Jan 01;2(1):61-70 [FREE Full text] [doi: 10.1370/afm.26] [Medline: 15053285]
11. Kadan-Lottick NS, Robison LL, Gurney JG, Neglia JP, Yasui Y, Hayashi R, et al. Childhood cancer survivors' knowledge about their past diagnosis and treatment: childhood cancer survivor study. JAMA 2002 Apr 10;287(14):1832-1839. [doi: 10.1001/jama.287.14.1832] [Medline: 11939869]
12. Adolescents and young adults with cancer. National Institutes of Health National Cancer Institute. URL: https://www.cancer.gov/types/aya [accessed 2024-02-25]
13. Docherty SL, Kayle M, Maslow GR, Santacroce SJ. The adolescent and young adult with cancer: a developmental life course perspective. Semin Oncol Nurs 2015 Aug;31(3):186-196. [doi: 10.1016/j.soncn.2015.05.006] [Medline: 26210197]
14. Szalda D, Pierce L, Hobbie W, Ginsberg JP, Brumley L, Wasik M, et al. Engagement and experience with cancer-related follow-up care among young adult survivors of childhood cancer after transfer to adult care. J Cancer Surviv 2016 Apr 25;10(2):342-350. [doi: 10.1007/s11764-015-0480-9] [Medline: 26303367]
15. Hou SH, Schulte FS. An investigation of cultural influences in survivors of paediatric cancer: a systematic review protocol. BMJ Open 2022 Nov 22;12(11):e062452 [FREE Full text] [doi: 10.1136/bmjopen-2022-062452] [Medline: 36414296]
16. Devine KA, Viola AS, Coups EJ, Wu YP. Digital health interventions for adolescent and young adult cancer survivors. JCO Clin Cancer Inform 2018 Dec;2:1-15 [FREE Full text] [doi: 10.1200/CCI.17.00138] [Medline: 30652583]
17. Gonzalez BD. Promise of mobile health technology to reduce disparities in patients with cancer and survivors. JCO Clin Cancer Inform 2018 Dec;2:1-9. [doi: 10.1200/cci.17.00141]
18. Wesley KM, Fizur PJ. A review of mobile applications to help adolescent and young adult cancer patients. Adolesc Health Med Ther 2015;6:141-148 [FREE Full text] [doi: 10.2147/AHMT.S69209] [Medline: 26316835]
19. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: a systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37 [FREE Full text] [doi: 10.1002/pon.5280] [Medline: 31692183]
20. McCann L, McMillan KA, Pugh G. Digital interventions to support adolescents and young adults with cancer: systematic review. JMIR Cancer 2019 Jul 31;5(2):e12071 [FREE Full text] [doi: 10.2196/12071] [Medline: 31368438]
21. Domecq JP, Prutsky G, Elraiyah T, Wang Z, Nabhan M, Shippee N, et al. Patient engagement in research: a systematic review. BMC Health Serv Res 2014 Feb 26;14:89 [FREE Full text] [doi: 10.1186/1472-6963-14-89] [Medline: 24568690]
22. Berg C, Stratton E, Esiashvili N, Mertens A, Vanderpool RC. Providers' perspectives of survivorship care for young adult survivors of childhood cancer. J Cancer Educ 2016 Mar 7;31(1):31-38 [FREE Full text] [doi: 10.1007/s13187-015-0848-6] [Medline: 25943901]
23. Howard AF, Kazanjian A, Pritchard S, Olson R, Hasan H, Newton K, et al. Healthcare system barriers to long-term follow-up for adult survivors of childhood cancer in British Columbia, Canada: a qualitative study. J Cancer Surviv 2018 Jun 8;12(3):277-290 [FREE Full text] [doi: 10.1007/s11764-017-0667-3] [Medline: 29222704]
24. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17 [FREE Full text] [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
25. Strategy for patient-oriented research. Canadian Institutes of Health Research. 2018. URL: https://cihr-irsc.gc.ca/e/41204.html [accessed 2023-03-15]

26. Hou S, Henry B, Schulte F. An assessment of young people living beyond cancer and factors that impact their attendance of their aftercare. In: Proceedings of the 54th Congress of the International Society of Pediatric Oncology. 2022 Presented at: SIOP 2022; September 28-October 1, 2022; Barcelona, Spain.
27. Henry B, Hou S, Drummon R, Canton K, Forbes C, Reynolds K, et al. Adherence to follow-up care among adolescent and young adult survivors of childhood cancer in Canada. In: Proceedings of the 38th Annual Canadian Association of Psychosocial Oncology Conference. 2023 Presented at: CAPO 2023; June 20-22, 2023; Montreal, QC.
28. Colorafi KJ, Evans B. Qualitative descriptive methods in health science research. HERD 2016 Jul;9(4):16-25 [FREE Full text] [doi: 10.1177/1937586715614171] [Medline: 26791375]
29. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
30. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
31. Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. BMJ 2017 Aug 02;358:j3453 [FREE Full text] [doi: 10.1136/bmj.j3453] [Medline: 28768629]
32. Abelson J, Tripp L, Kandasamy S, Burrows K, PPEET Implementation Study Team. Supporting the evaluation of public and patient engagement in health system organizations: results from an implementation research study. Health Expect 2019 Oct 02;22(5):1132-1143 [FREE Full text] [doi: 10.1111/hex.12949] [Medline: 31373754]
33. Braun V, Clarke V. Thematic analysis. In: APA Handbook of Research Methods in Psychology: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. Washington, DC: American Psychological Association; 2023:65-81.
34. Sandelowski M, Barroso J. Classifying the findings in qualitative studies. Qual Health Res 2003 Sep;13(7):905-923. [doi: 10.1177/1049732303253488] [Medline: 14502957]
35. NVivo 14. Lumivero. URL: https://lumivero.com/ [accessed 2025-03-26]
36. Birt L, Scott S, Cavers D, Campbell C, Walter F. Member checking: a tool to enhance trustworthiness or merely a nod to validation? Qual Health Res 2016 Nov;26(13):1802-1811. [doi: 10.1177/1049732316654870] [Medline: 27340178]
37. Frisch N, Atherton P, Doyle-Waters MM, MacLeod ML, Mallidou A, Sheane V, et al. Patient-oriented research competencies in health (PORCH) for researchers, patients, healthcare providers, and decision-makers: results of a scoping review. Res Involv Engagem 2020;6:4 [FREE Full text] [doi: 10.1186/s40900-020-0180-0] [Medline: 32055415]
38. Buneviciene I, Mekary RA, Smith TR, Onnela JP, Bunevicius A. Can mHealth interventions improve quality of life of cancer patients? A systematic review and meta-analysis. Crit Rev Oncol Hematol 2021 Jan;157:103123 [FREE Full text] [doi: 10.1016/j.critrevonc.2020.103123] [Medline: 33190065]
39. King-Dowling S, Psihogios AM, Hill-Kayser C, Szalda D, O'Hagan B, Darabos K, et al. Acceptability and feasibility of survivorship care plans and an accompanying mobile health intervention for adolescent and young adult survivors of childhood cancer. Pediatr Blood Cancer 2021 Mar 08;68(3):e28884 [FREE Full text] [doi: 10.1002/pbc.28884] [Medline: 33416214]
40. Shuldiner J, Shah N, Corrado AM, Hodgson D, Nathan PC, Ivers N. Determinants of surveillance for late effects in childhood cancer survivors: a qualitative study using the theoretical domains framework. J Cancer Surviv 2022 Jun 06;16(3):552-567. [doi: 10.1007/s11764-021-01050-6] [Medline: 33954882]
41. Lee JL, Gutierrez-Colina A, Williamson Lewis R, Wasilewski-Masker K, Meacham LR, Mertens AC, et al. Knowledge of late effects risks and healthcare responsibility in adolescents and young adults treated for childhood cancer. J Pediatr Psychol 2019 Jun 01;44(5):557-566. [doi: 10.1093/jpepsy/jsy102] [Medline: 30624691]
42. Casillas J, Kahn KL, Doose M, Landier W, Bhatia S, Hernandez J, et al. Transitioning childhood cancer survivors to adult-centered healthcare: insights from parents, adolescent, and young adult survivors. Psychooncology 2010 Sep;19(9):982-990 [FREE Full text] [doi: 10.1002/pon.1650] [Medline: 20017115]
43. Wilkins KL, D'Agostino N, Penney AM, Barr RD, Nathan PC. Supporting adolescents and young adults with cancer through transitions: position statement from the Canadian Task Force on Adolescents and Young Adults with cancer. J Pediatr Hematol Oncol 2014 Oct;36(7):545-551. [doi: 10.1097/MPH.0000000000000103] [Medline: 24390448]
44. Lown EA, Phillips F, Schwartz LA, Rosenberg AR, Jones B. Psychosocial follow-up in survivorship as a standard of care in pediatric oncology. Pediatr Blood Cancer 2015 Dec 23;62 Suppl 5(Suppl 5):S514-S584 [FREE Full text] [doi: 10.1002/pbc.25783] [Medline: 26700918]
45. Zebrack B, Isaacson S. Psychosocial care of adolescent and young adult patients with cancer and survivors. J Clin Oncol 2012 Apr 10;30(11):1221-1226. [doi: 10.1200/jco.2011.39.5467]
46. Kilbourne AM, Switzer G, Hyman K, Crowley-Matoka M, Fine MJ. Advancing health disparities research within the health care system: a conceptual framework. Am J Public Health 2006 Dec;96(12):2113-2121. [doi: 10.2105/ajph.2005.077628]
47. Díaz-García C, González-Moreno A, Jose Sáez-Martínez F. Gender diversity within R & D teams: its impact on radicalness of innovation. Innovation 2014 Dec 17;15(2):149-160. [doi: 10.5172/impp.2013.15.2.149]

48. Klosky JL, Cash DK, Buscemi J, Lensing S, Garces-Webb DM, Zhao W, et al. Factors influencing long-term follow-up clinic attendance among survivors of childhood cancer. J Cancer Surviv 2008 Dec 12;2(4):225-232 [FREE Full text] [doi: 10.1007/s11764-008-0063-0] [Medline: 18787958]
49. Hoyt MA, Rubin LR. Gender representation of cancer patients in medical treatment and psychosocial survivorship research: changes over three decades. Cancer 2012 Oct 01;118(19):4824-4832 [FREE Full text] [doi: 10.1002/cncr.27432] [Medline: 22294480]
50. Reeves TJ, Mathis TJ, Bauer HE, Hudson MM, Robison LL, Wang Z, et al. Racial and ethnic disparities in health outcomes among long-term survivors of childhood cancer: a scoping review. Front Public Health 2021;9:741334 [FREE Full text] [doi: 10.3389/fpubh.2021.741334] [Medline: 34778176]
51. Bhatia S. Disparities in cancer outcomes: lessons learned from children with cancer. Pediatr Blood Cancer 2011 Jun 15;56(6):994-1002 [FREE Full text] [doi: 10.1002/pbc.23078] [Medline: 21328525]
52. Duong J, Forbes C, Henry B, Bilash T, Rahamatullah I, Hou S, et al. Exploring lived experiences and inclusive engagement practices for under-represented adolescents and young adults in psychosocial oncology research: a qualitative study. In: Proceedings of the 38th Annual Canadian Association of Psychosocial Oncology. 2023 Presented at: CAPO 2023; June 20-22, 2023; Montreal, QC.

Abbreviations

CIHR: Canadian Institutes of Health Research
mHealth: mobile health
REDCap: Research Electronic Data Capture
SPOR: Strategy for Patient-Oriented Research

Edited by N Cahill; submitted 27.02.24; peer-reviewed by E Vashishtha, S King-Dowling, J Mistry, Z Ehtesham, SL Coven; comments to author 12.08.24; revised version received 08.10.24; accepted 21.01.25; published 25.04.25.

Please cite as:
Hou SHJ, Henry B, Drummond R, Forbes C, Mendonça K, Wright H, Rahamatullah I, Tutelman PR, Zwicker H, Stokoe M, Duong J, Drake EK, Erker C, Taccone MS, Sutherland L, Nathan P, Spavor M, Goddard K, Reynolds K, Schulte FSM
Co-Designing Priority Components of an mHealth Intervention to Enhance Follow-Up Care in Young Adult Survivors of Childhood Cancer and Health Care Providers: Qualitative Descriptive Study
JMIR Cancer 2025;11:e57834
URL: https://cancer.jmir.org/2025/1/e57834
doi:10.2196/57834
PMID:

Predicting Overall Survival in Patients with Male Breast Cancer: Nomogram Development and External Validation Study

Wen-Zhen Tang[1*], PhD; Shu-Tian Mo[1*], PhD; Yuan-Xi Xie[2], MSc; Tian-Fu Wei[3], BSc; Guo-Lian Chen[4], BSc; Yan-Juan Teng[1], MSc; Kui Jia[3], MSc

[1]Department of Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

[2]Department of Central Sterile Supply, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

[3]Department of Gastrointestinal Surgery, The First Affiliated Hospital of Guangxi Medical University, No 6 Shuangyong Road, Nanning, China

[4]Department of Medical Oncology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

[*]these authors contributed equally

Corresponding Author:
Kui Jia, MSc
Department of Gastrointestinal Surgery, The First Affiliated Hospital of Guangxi Medical University, No 6 Shuangyong Road, Nanning, China

Abstract

Background: Male breast cancer (MBC) is an uncommon disease. Few studies have discussed the prognosis of MBC due to its rarity.

Objective: This study aimed to develop a nomogram to predict the overall survival of patients with MBC and externally validate it using cases from China.

Methods: Based on the Surveillance, Epidemiology, and End Results (SEER) database, male patients who were diagnosed with breast cancer between January 2010, and December 2015, were enrolled. These patients were randomly assigned to either a training set (n=1610) or a validation set (n=713) in a 7:3 ratio. Additionally, 22 MBC cases diagnosed at the First Affiliated Hospital of Guangxi Medical University between January 2013 and June 2021 were used for external validation, with the follow-up endpoint being June 10, 2023. Cox regression analysis was performed to identify significant risk variables and construct a nomogram to predict the overall survival of patients with MBC. Information collected from the test set was applied to validate the model. The concordance index (C-index), receiver operating characteristic (ROC) curve, decision curve analysis (DCA), and a Kaplan-Meier survival curve were used to evaluate the accuracy and reliability of the model.

Results: A total of 2301 patients with MBC in the SEER database and 22 patients with MBC from the study hospital were included. The predictive model included 7 variables: age (hazard ratio [HR] 1.89, 95% CI 1.50 - 2.38), surgery (HR 0.38, 95% CI 0.29 - 0.51), marital status (HR 0.75, 95% CI 0.63 - 0.89), tumor stage (HR 1.17, 95% CI 1.05 - 1.29), clinical stage (HR 1.41, 95% CI 1.15 - 1.74), chemotherapy (HR 0.62, 95% CI 0.50 - 0.75), and HER2 status (HR 2.68, 95% CI 1.20 - 5.98). The C-index was 0.72, 0.747, and 0.981 in the training set, internal validation set, and external validation set, respectively. The nomogram showed accurate calibration, and the ROC curve confirmed the advantage of the model in clinical validity. The DCA analysis indicated that the model had good clinical applicability. Furthermore, the nomogram classification allowed for more accurate differentiation of risk subgroups, and patients with low-risk MBC demonstrated substantially improved survival outcomes compared with medium- and high-risk patients (P<.001).

Conclusions: A survival prognosis prediction nomogram with 7 variables for patients with MBC was constructed in this study. The model can predict the survival outcome of these patients and provide a scientific basis for clinical diagnosis and treatment.

(JMIR Cancer 2025;11:e54625) doi:10.2196/54625

KEYWORDS

male breast cancer; specific survival; prediction model; nomogram; Surveillance, Epidemiology, and End Results database; SEER database

Introduction

Male breast cancer (MBC) is an infrequent type of malignancy [1,2]. The incidence of MBC accounts for less than 1% of all breast cancer (BC) instances, and MBC accounts for 0.31% of all BC cases in China [3-5]. The incidence of MBC varies by region and ethnicity, with higher rates observed in Africa, North America, and Australia, and the lowest rates are observed in Asia [6]. In China, there are only 4 cases of MBC per million people, but this figure has been increasing gradually in recent

years [6]. Due to the low incidence of MBC, current research on BC primarily focuses on female patients [7]. Therefore, the current treatment for MBC is based on the guidelines for treating female BC [8]. However, MBC possesses unique tumor, molecular, and clinicopathological characteristics, and no consensus has been established regarding its diagnosis, treatment, and assessment of prognostic risk factors. A previous study revealed that the median age at diagnosis for BC in men is 67 years old, which is 5 - 10 years later than that in women [9]. Despite this, the overall survival of MBC is significantly lower than that of female BC, largely due to late diagnosis [10].

The TNM (tumor, extent of spread to the lymph nodes, and presence of metastasis) staging system is the most commonly used clinical instrument to evaluate the prognosis of individuals with cancer [11-13]. However, in MBC, the limited amount of breast tissue and the frequent involvement of the chest wall at early stages reduce the prognostic value of TNM staging [14]. Many studies have demonstrated that factors such as age, tumor type, and other factors significantly influence the prognosis of BC [15,16]. Compared to using the clinical stage alone, comprehensive multivariate models can provide numerical estimates of practice-specific risk and the accuracy of prognostic predictions for patients with cancer [17]. Therefore, various clinical medical records need to be combined to construct a prognostic model for MBC, thus enabling a more accurate judgment of the prognosis of patients and an accurate, individual evaluation of the prognosis of patients.

Current clinical approaches for constructing risk prediction models include the nomogram, a scoring system, and other methods, which can serve as a guide for clinical decision-making and individualized treatment [18-20]. The nomogram, as a straightforward and intuitive prediction tool with strong predictive ability, has the advantages of accurate predictive ability and calibration ability, and it has been widely used in prognosis research [21-23].

This study aimed to identify the prognostic indicators of patients with MBC by using the Surveillance, Epidemiology, and End Results (SEER) database; establish a predictive model on the basis of the independent predictors of overall survival; and internally and externally validate the model to guide clinical staff in evaluating the prognosis of patients more accurately and formulating more personalized diagnosis and treatment plans. We present the study in accordance with the TRIPOD reporting checklist.

Methods

Data Sources

The SEER database collects information on new cancer cases and survival rates from 18 population-based cancer registries, which currently cover approximately 30% of the US population [22]. Clinical data on male patients with pathologically confirmed BC from 2010 to 2015 were gathered using the SEER database to establish a training set and an internal validation set. Data from patients with MBC admitted to the First Affiliated Hospital of Guangxi Medical University between 2013 and 2021 were used for the external validation set of the model. Clinical data for MBC were retrospectively collected from the hospital database, and follow-up information was obtained through telephone interviews. Patients with missing follow-up data or other essential clinical information were excluded.

Patient Inclusion and Exclusion Criteria

The criteria for patient inclusion were as follows: (1) male patients; (2) an *International Classification of Diseases for Oncology, Third Revision* code; (3) breast as the primary site; and (4) complete survival data. The exclusion criteria were (1) missing clinical information, including TNM staging and tumor laterality; (2) unknown demographic characteristics, such as age at diagnosis and marital status; and (3) instances without records of follow-up (0-month survival time code). The enrolled patients were randomly assigned in a 7:3 ratio to two sets: a training set and an internal validation set. The training set was used to develop the prediction model, and the internal validation set was used for internal validation. The data obtained at the hospital were applied for external validation.

Variable Selection

The outcome variable in this study was overall survival. The selection of predictor variables was informed by previous reports in the literature. The variables collected included year of diagnosis, age, marital status, pathological grade, breast subtype, estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2 (HER2), American Joint Committee on Cancer stage, chemotherapy, surgery, radiotherapy, duration of follow-up, and death.

Follow-Up of Patients

Male patients with BC were followed up by telephone in the hospital, and the follow-up ended on June 10, 2023. The index used for follow-up was overall survival time, with the outcome event being mortality.

Statistical Analysis

Data analysis was conducted using the R software version 4.1.1 (IBM Corp). Percentages were used to represent categorical variables, and the χ^2 test or Fisher exact test was used to compare the baseline characteristics of the training set, internal validation set, and external validation set. The Kaplan-Meier model was applied to describe the overall survival curve, and the log-rank test was used to evaluate the disparities in survival among various subgroups of each variable. First, variables that had a significance value of $P<.05$ in the univariate analysis were chosen to be incorporated into the multivariate Cox proportional hazard model to obtain variables affecting the survival of patients with MBC. Second, stepwise regression was performed based on the Akaike information criterion. The nomogram prediction model was constructed using R software (via the *rms* and *survival* R packages) to assess the influence of risk factors on the overall survival of patients with MBC. Predictions were made for the 1-, 3-, and 5-year overall survival rates of patients with MBC by constructing the nomogram.

The performance of the nomogram was evaluated through internal and external validations. Bootstrapping was used to perform 1000 instances of resampling to internally validate the predictive performance of the nomogram to ensure the stability

and reliability of the model's performance. The discrimination of the nomogram was assessed using the Concordance index (C-index) and receiver operating characteristic (ROC) curve. A calibration curve was created to assess the degree of calibration of the nomogram to ensure its accuracy and reliability. Furthermore, decision curve analysis (DCA) was conducted using *ggDCA* in the R package, to evaluate the clinical utility and application value of the nomogram. Finally, X-tile software (version 3.6.1, Yale University School of Medicine) was used for risk stratification on the basis of the total score of the nomogram for each individual. An α level of .05 was used.

Ethical Considerations

The data used in this study were extracted from a publicly accessible SEER database. This study was reviewed and approved by the Medical Ethics Committee of the First Affiliated Hospital of Guangxi Medical University (2023-E320-01). During the follow-up, informed consent was obtained orally from individual participants included in the study, and the investigator explained the purpose of the study to the patient or caregiver. Participants were also made aware of their right to withdraw at any time without penalty or prejudice to their future care, a principle that was strictly adhered to throughout the study period. In addition, participants who completed the survey received a complimentary disease knowledge resource as a token of appreciation and compensation for their participation, All participants' information was confidential, and each patient was assigned an ID to keep the study data and results anonymous.

Results

Patients' Baseline Characteristics

Figure 1 depicts the screening procedure in the SEER database. In accordance with the inclusion and exclusion criteria, a cohort of 2301 eligible patients with MBC was selected from the SEER database and randomly divided into a training set (n=1595) and an internal validation set (n=706). A total of 22 patients with MBC were chosen from the institution to serve as an external validation set. Significant variations in age were observed among the 3 groups in relation to demographic characteristics (P=.01). The proportion of older men in the SEER database (training set: 1180/1595, 74%; internal validation set: 505/706, 71.5%) was substantially greater than that in the external validation set (9/22, 41%). Significant differences were found in chemotherapy, lung metastasis, breast subtype, and HER2 status among the 3 groups (all P<.05). The proportion of men with breast cancer who received chemotherapy was higher in the external validation set (17/22, 77%) than in the SEER database (training set: 601/1595, 37.7%; internal validation set: 257/706, 36.4%). The incidence of lung metastasis in patients with MBC in the external validation set (3/22, 14%) was higher than that in the SEER database (training set: 49/1595, 3.1%; internal validation set: 24/706, 3.4%). There was a high prevalence of luminal A among men, with rates of 86.7% (1379/1595) and 85% (607/706) in the training set and internal validation set, respectively, as well as 41% (9/22) in the external validation set. A notable detail is that a significant portion of the total population exhibited a HER2-negative status accounting for 88.6% (1413/1595) in the training set, 87.5% (618/706) in the internal validation set, and 59% (13/22) in the external validation set. Table 1 displays the demographic and clinicopathological characteristics.

Figure 1. Flow chart for inclusion and partition of patients.

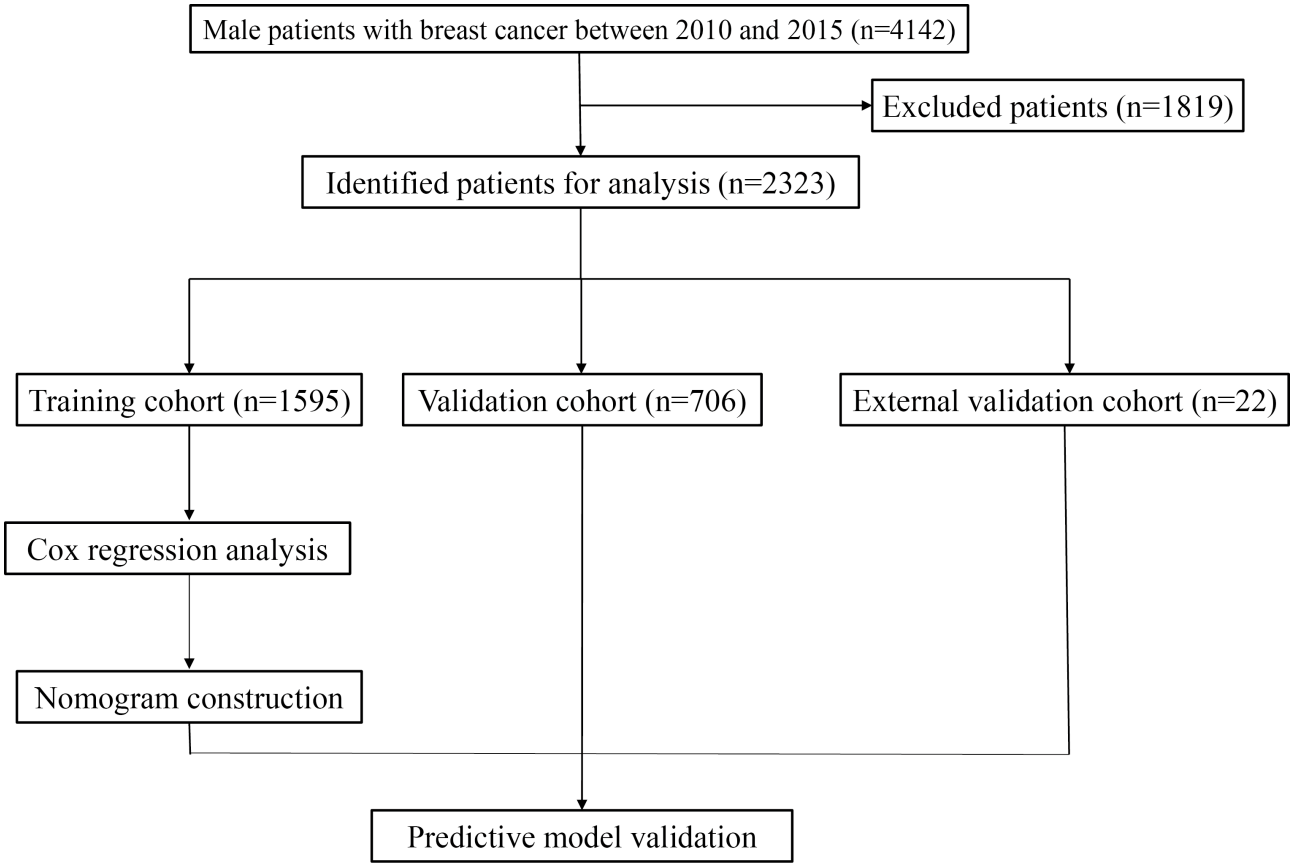


Table . Demographics and clinicopathologic characteristics of male breast cancer.

Variables	Total (n=2323)	Training set (n=1595)	Internal validation set (n=706)	External validation set (n=22)	*P* value
Marital status, n (%)					.10[a]
Unmarried	797 (34.3)	544 (34.1)	250 (35.4)	3 (14)	
Married	1526 (65.7)	1051 (65.9)	456 (64.6)	19 (86)	
Age, n (%)					.01[a]
≤60 years	629 (27.1)	415 (26)	201 (28.5)	13 (59)	
>60 years	1694 (72.9)	1180 (74)	505 (71.5)	9 (41)	
T[b] stage, n (%)					.07[c]
T_0	39 (1.7)	25 (1.6)	11 (1.6)	3 (14)	
T_1	1047 (45.1)	717 (45)	323 (45.8)	7 (32)	
T_2	971 (41.8)	675 (42.3)	285 (40.4)	11 (50)	
T_3	70 (3)	46 (2.9)	23 (3.3)	1 (5)	
T_4	196 (8.4)	132 (8.3)	64 (9.1)	0 (0)	
N stage, n (%)[d]					.30[c]
N_0	1314 (56.6)	896 (56.2)	409 (57.9)	9 (41)	
N_1	704 (30.3)	488 (30.6)	205 (29)	11 (50)	
N_2	188 (8.1)	132 (8.3)	56 (7.9)	0 (0)	
N_3	117 (5)	79 (5)	36 (5.1)	2 (9)	
M[e] stage, n (%)					.53[a]
M_0	2131 (91.7)	1466 (91.9)	646 (91.5)	19 (86)	
M_1	192 (8.3)	129 (8.1)	60 (8.5)	3 (14)	
Clinical stage, n (%)					.79[c]
0	1 (0)	1 (0.1)	0 (0)	0 (0)	
I	733 (31.6)	496 (31.1)	232 (32.9)	5 (23)	
II	1020 (43.9)	704 (44.1)	304 (43.1)	12 (55)	
III	377 (16.2)	265 (16.6)	110 (15.6)	2 (9)	
IV	192 (8.3)	129 (8.1)	60 (8.5)	3 (14)	
Laterality, n (%)					.93[c]
Right	1073 (46.2)	731 (45.8)	331 (46.9)	11 (50)	
Left	1245 (53.6)	860 (53.9)	374 (53)	11 (50)	
Bilateral	5 (0.2)	4 (0.2)	1 (0.1)	0 (0)	
Surgery, n (%)					.65[a]
Yes	2119 (91.2)	1449 (90.8)	650 (92.1)	20 (91)	
No	204 (8.8)	146 (9.2)	56 (7.9)	2 (9)	
Radiation, n (%)					.37[a]
Yes	747 (32.2)	515 (32.3)	228 (32.3)	4 (18)	
No	1576 (67.8)	1080 (67.7)	478 (67.7)	18 (82)	
Chemotherapy, n (%)					<.001[a]
Yes	875 (37.7)	601 (37.7)	257 (36.4)	17 (77)	

Variables	Total (n=2323)	Training set (n=1595)	Internal validation set (n=706)	External validation set (n=22)	*P* value
No	1448 (62.3)	994 (62.3)	449 (63.6)	5 (23)	
Bone metastasis, n (%)					.59[c]
Yes	133 (5.7)	89 (5.6)	42 (5.9)	2 (9)	
No	2187 (94.1)	1503 (94.2)	664 (94.1)	20 (91)	
Unknown	3 (0.1)	3 (0.2)	0 (0)	0 (0)	
Brain metastasis, n (%)					.49[c]
Yes	14 (0.6)	12 (0.8)	2 (0.3)	0 (0)	
No	2302 (99.1)	1577 (98.9)	703 (99.6)	22 (100)	
Unknown	7 (0.3)	6 (0.4)	1 (0.1)	0 (0)	
Liver metastasis, n (%)					.42[c]
Yes	24 (1)	15 (0.9)	9 (1.3)	0 (0)	
No	2293 (98.7)	1574 (98.7)	697 (98.7)	22 (100)	
Unknown	6 (0.3)	6 (0.4)	0 (0)	0 (0)	
Lung metastasis, n (%)					.04[a]
Yes	76 (3.3)	49 (3.1)	24 (3.4)	3 (14)	
No	2240 (96.4)	1539 (96.5)	682 (96.6)	19 (86)	
Unknown	7 (0.3)	7 (0.4)	0 (0)	0 (0)	
Breast subtype, n (%)					<.001[a]
Luminal A	1995 (85.9)	1379 (86.5)	607 (86)	9 (41)	
Luminal B	260 (11.2)	168 (10.5)	82 (11.6)	10 (45)	
HER2[f] -positive	22 (0.9)	14 (0.9)	6 (0.8)	2 (9)	
Triple-negative	46 (2)	34 (2.1)	11 (1.6)	1 (5)	
Estrogen receptor status, n (%)					.14[a]
Negative	72 (3.1)	52 (3.3)	18 (2.5)	2 (9)	
Positive	2251 (96.9)	1543 (96.7)	688 (97.5)	20 (91)	
Progesterone receptor status, n (%)					.06[a]
Negative	224 (9.6)	167 (10.5)	54 (7.6)	3 (14)	
Positive	2099 (90.4)	1428 (89.5)	652 (92.4)	19 (86)	
HER2 status, n (%)					.01[a]
Negative	2044 (88)	1413 (88.6)	618 (87.5)	13 (59)	
Positive	279 (12)	182 (11.4)	88 (12.5)	9 (41)	

[a]Chi-square test was performed.

[b]T: tumor.

[c]Fisher precision probability test was performed.

[d]N: lymph nodes.

[e]M: metastasis.

[f]HER2: human epidermal growth factor receptor 2.

The data of 22 male patients diagnosed with BC in the study hospital were collected. Statistically significant differences were observed between the two groups in terms of age, T stage, chemotherapy, breast subtype, and HER2 status (P<.05). In the unit set, ≤60 years old, T_2 stage, receiving chemotherapy, luminal B, and HER2-positive status accounted for a greater proportion of patients with MBC. The clinicopathological characteristics of the SEER set and the unit set are provided in Multimedia Appendix 1.

Univariate and Multivariate Cox Regression Analysis

Cox regression risk analysis was applied to conduct univariate and multivariate survival analysis for the patients with MBC in the training set. The findings indicated that age (hazard ratio [HR] 1.89, 95% CI 1.50 - 2.38), marital status (HR 0.75, 95% CI 0.63 - 0.89), T stage (HR 1.17, 95% CI 1.05 - 1.29), clinical grade (HR 1.41, 95% CI 1.15 - 1.74), surgery (HR 0.38, 95% CI 0.29 - 0.51), chemotherapy (HR 0.62, 95% CI 0.50 - 0.75), and HER2 status (HR 2.68, 95% CI 1.20 - 5.98) were risk variables for the survival of patients with MBC (all $P<.05$; Table 2).

Table . Univariate and multivariate analysis of male breast cancer risk factors in the training set.

Variable	Univariate		Multivariate	
	HR[a] (95% CI)	*P* value	HR (95% CI)	*P* value
Marital status	0.71 (0.60 - 0.84)	<.001	0.75 (0.63 - 0.89)	.01
Age	1.74 (1.39 - 2.16)	<.001	1.89 (1.50 - 2.38)	<.001
T[b] stage	1.45 (1.33 - 1.58)	<.001	1.17 (1.05 - 1.29)	.04
N[c] stage	1.23 (1.12 - 1.35)	<.001	1.03 (0.90 - 1.17)	.67
M[d] stage	5.19 (4.18 - 6.45)	<.001	1.08 (0.61 - 1.91)	.81
Clinical stage	1.86 (1.70 - 2.04)	<.001	1.41 (1.15 - 1.74)	.01
Laterality	1.00 (0.85 - 1.19)	.99	—[e]	—
Surgery	0.18 (0.15 - 0.22)	<.001	0.38 (0.29 - 0.51)	<.001
Radiation	0.98 (0.82 - 1.18)	.84	—	—
Chemotherapy	0.78 (0.65 - 0.93)	.006	0.62 (0.50 - 0.75)	<.001
Bone metastasis	0.18 (0.14 - 0.24)	<.001	0.71 (0.46 - 1.10)	.12
Brain metastasis	0.21 (0.09 - 0.49)	<.001	0.53 (0.24 - 1.17)	.12
Liver metastasis	0.20 (0.10 - 0.41)	<.001	0.91 (0.47 - 1.80)	.79
Lung metastasis	0.25 (0.17 - 0.35)	<.001	0.98 (0.69 - 1.40)	.93
Breast subtype	1.16 (0.93 - 1.44)	.19	2.03 (0.95 - 4.30)	.07
Estrogen receptor status	0.44 (0.29 - 0.65)	<.001	1.86 (0.45 - 7.76)	.39
Progesterone receptor status	0.64 (0.50 - 0.83)	<.001	0.91 (0.67 - 1.24)	.56
HER2[f] status	1.31 (1.02 - 1.68)	.03	2.68 (1.20 - 5.98)	.02

[a]HR: hazard ratio.

[b]T: tumor.

[c]N: lymph node.

[d]M: metastasis.

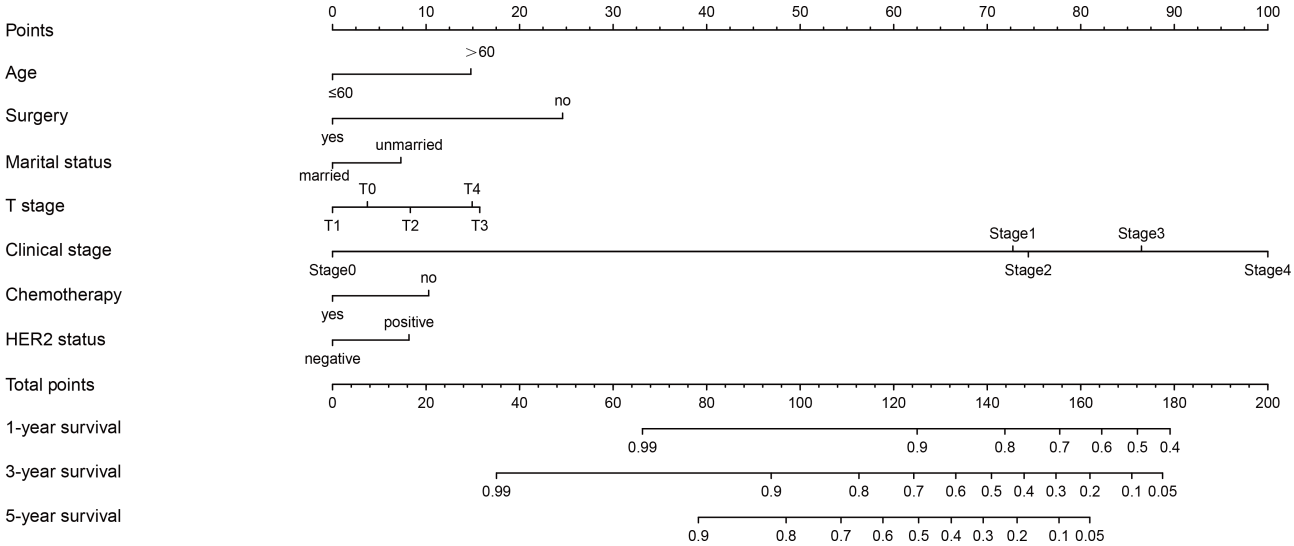
[e]Variables that were not significant in the univariate analysis do not have specific data in the multivariate analysis.

[f]HER2: human epidermal growth factor receptor 2.

Construction and Validation of Nomogram

The construction of a nomogram for the overall survival prognosis of MBC was based on the results of the Cox regression analysis conducted on the training set. This analysis identified 7 variables that were subsequently used in the development of the nomogram (Figure 2). The nomogram demonstrated that clinical stage and surgery were the primary vital risk variables that affect the survival outcomes of patients with MBC. The total score could predict the 1-, 3-, and 5-year survival rates of patients with MBC by summing the scores of each variable.

Figure 2. Nomogram for predicting 1-, 3- and 5-year overall survival in patients with male breast cancer. HER2: human epidermal growth factor receptor 2; T: tumor.



The discrimination ability of the nomogram was evaluated in this study by using the ROC curve and the C-index. The area under the curve values of the nomogram at 1-, 3-, and 5-year survival probabilities had excellent discrimination efficacy in the training set (Figure 3A-C). The area under the curve values in the internal validation set were 0.736, 0.773, and 0.765 (Figure 3D-F), and the external validation set values were 1, 0.947, and 0.825 (Figure 3G-I). The C-index of the training set was determined using the bootstrap method, and the C-index of the external validation set was 0.72, 0.747, and 0.981 for the at 1-, 3-, and 5-year survival, respectively, indicating that the nomogram exhibited a favorable discriminatory capability in the American and Chinese populations.

The calibration curves were used to evaluate the consistency of the nomogram. The findings indicated a high degree of uniformity between the predicted and observed probabilities of survival in the training set (Figure 4A-C) and internal validation set (Figure 4D-F).

The DCA curve demonstrated that the nomogram exhibited superior performance in terms of net clinical benefit and predictive accuracy for 3- and 5-year survival outcomes in the training set (Figure 5A) and internal validation set (Figure 5B).

Figure 3. Receiver operating characteristic curve of prediction of 1-, 3-, and 5-year survival in the training set (A-C), internal validation set (D-F), and external validation set (G–I). AUC: area under the curve.

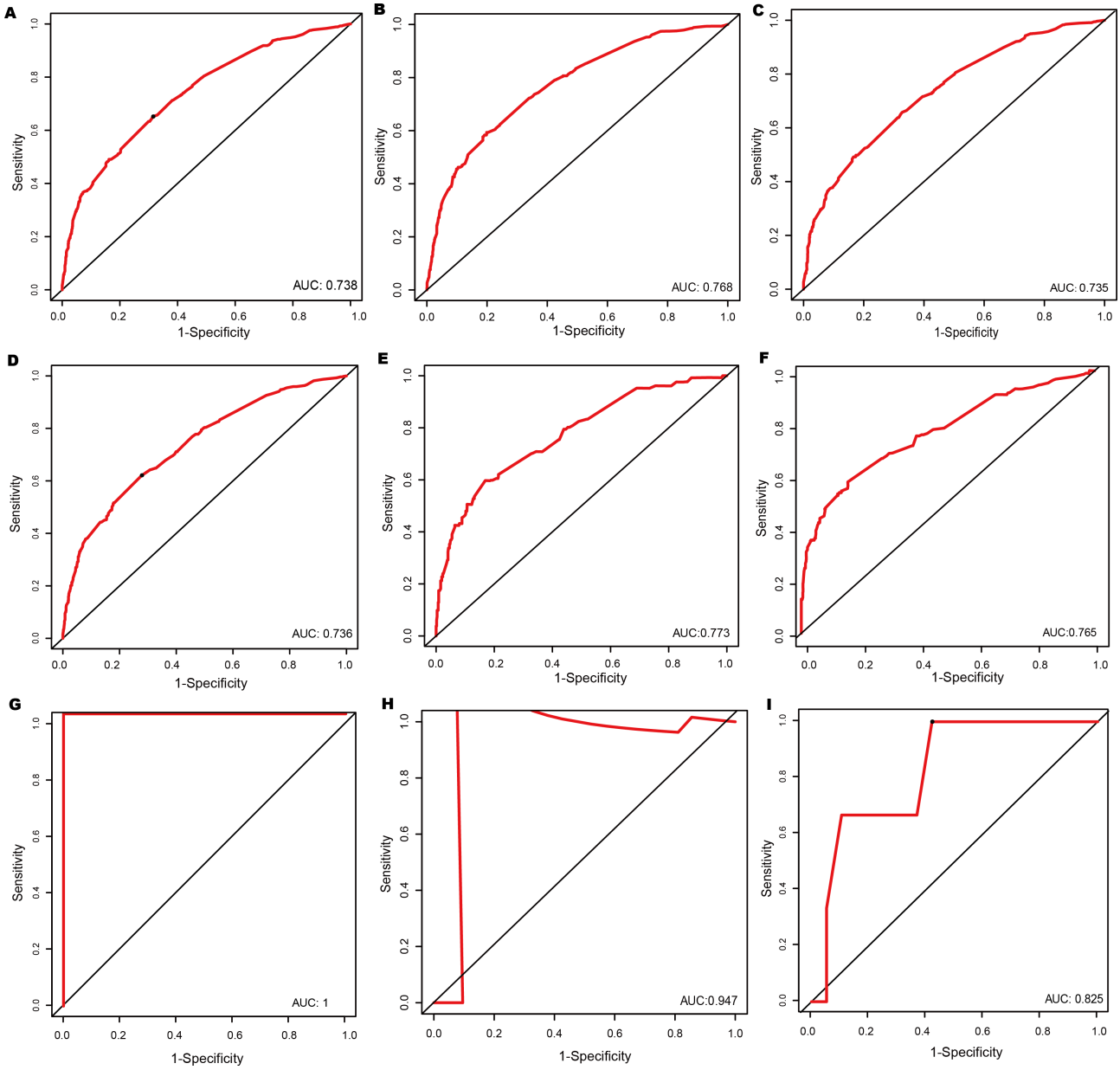


Figure 4. Calibration curve of 1-, 3-, and 5-year overall survival in the training set (A-C) and internal validation set (D-F). The errors bars represent the 95% CI of these estimates.

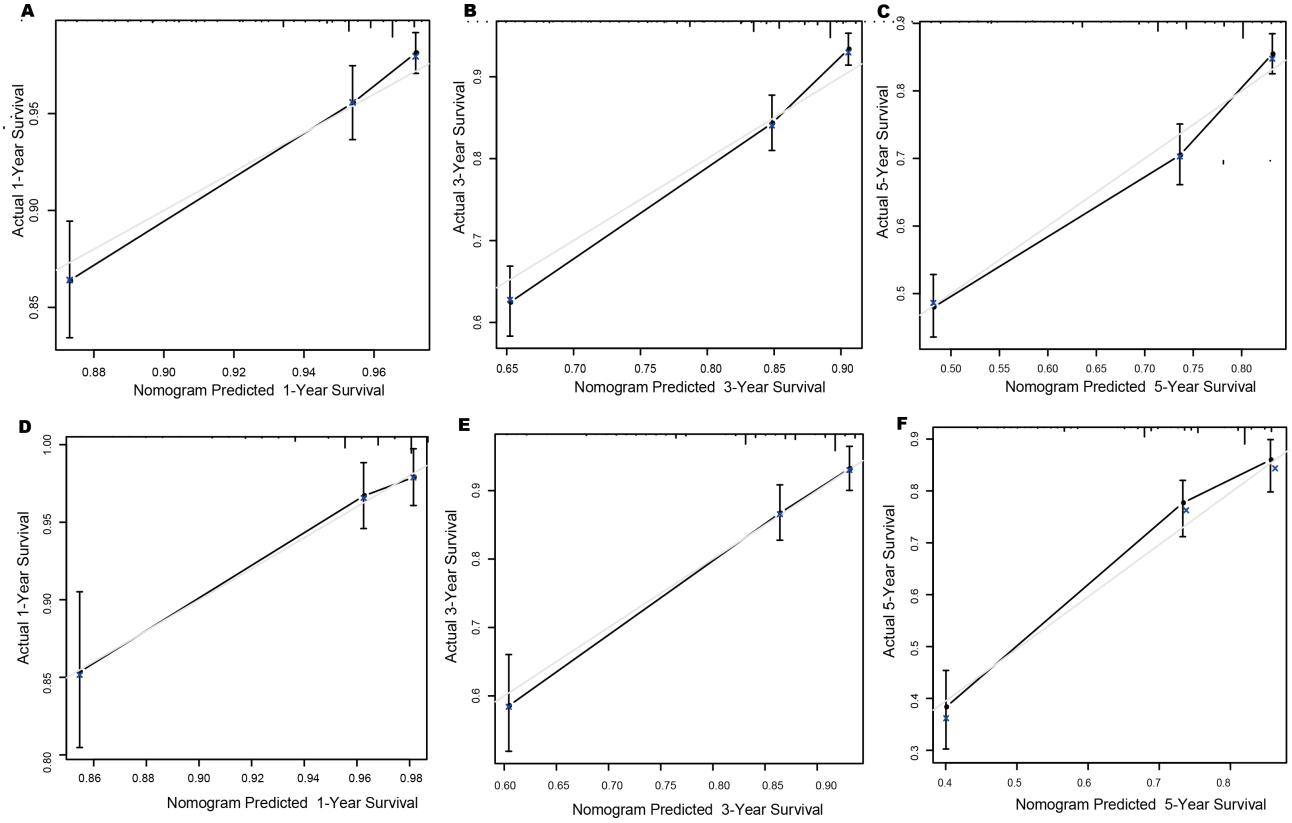
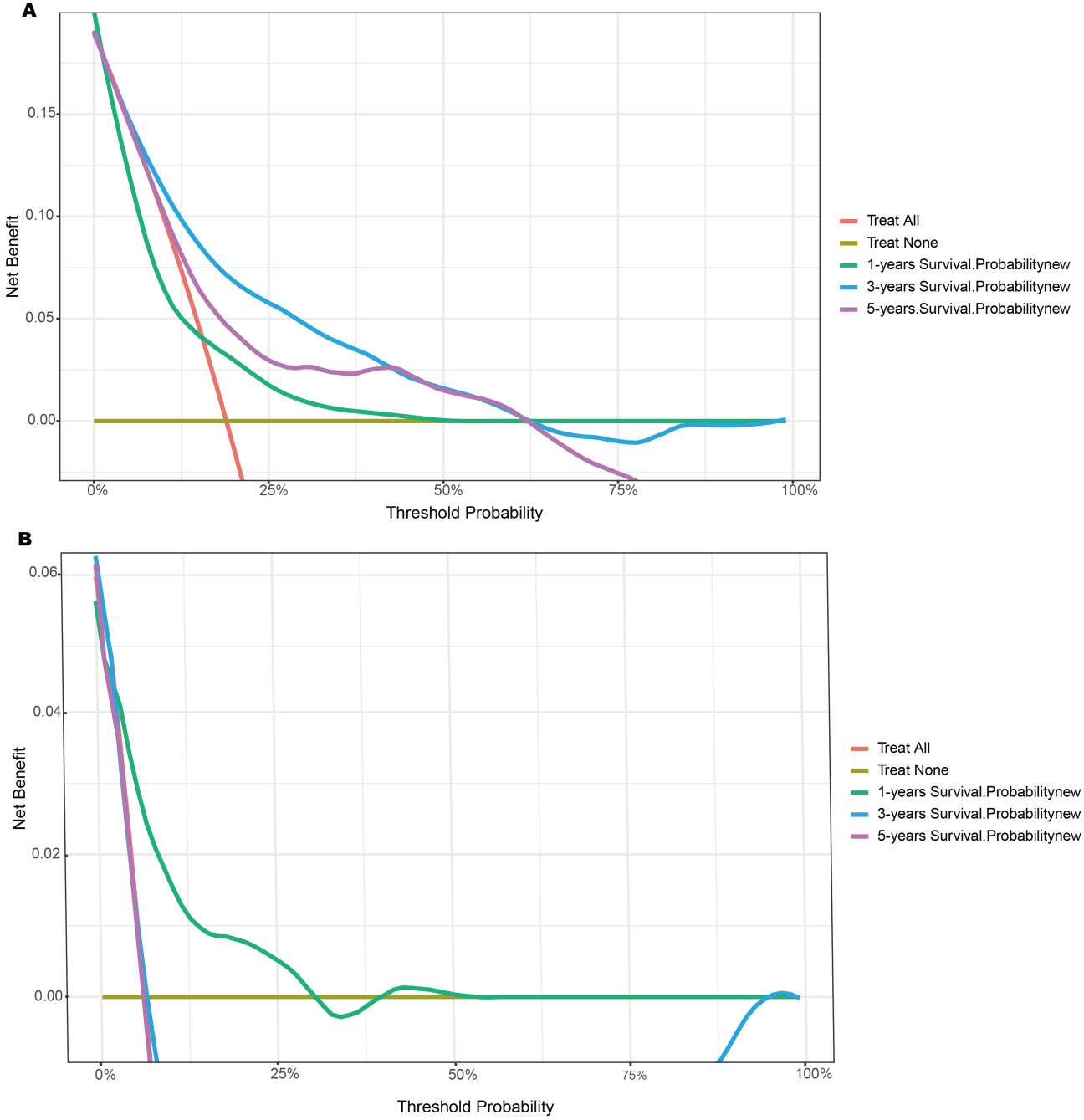


Figure 5. Decision curve analysis of 1-, 3- and 5-year survival in the training set (A) and the internal validation set (B).



Nomogram Prediction Score Risk Stratification

Finally, risk stratification was conducted by calculating the nomogram total score of each individual in the training set (Table 3). After the cut-off values were determined using X-tile software, all patients with MBC were divided into 3 groups: low-risk group (points≤93), medium-risk group (93<points≤117), and high-risk group (points>117). The survival curves of each risk group were depicted using the Kaplan-Meier model in the training set (Figure 6A) and internal validation set (Figure 6B). A log-rank test was used to compare the differences between the groups to assess the accuracy of risk stratification on the basis of the nomogram score.

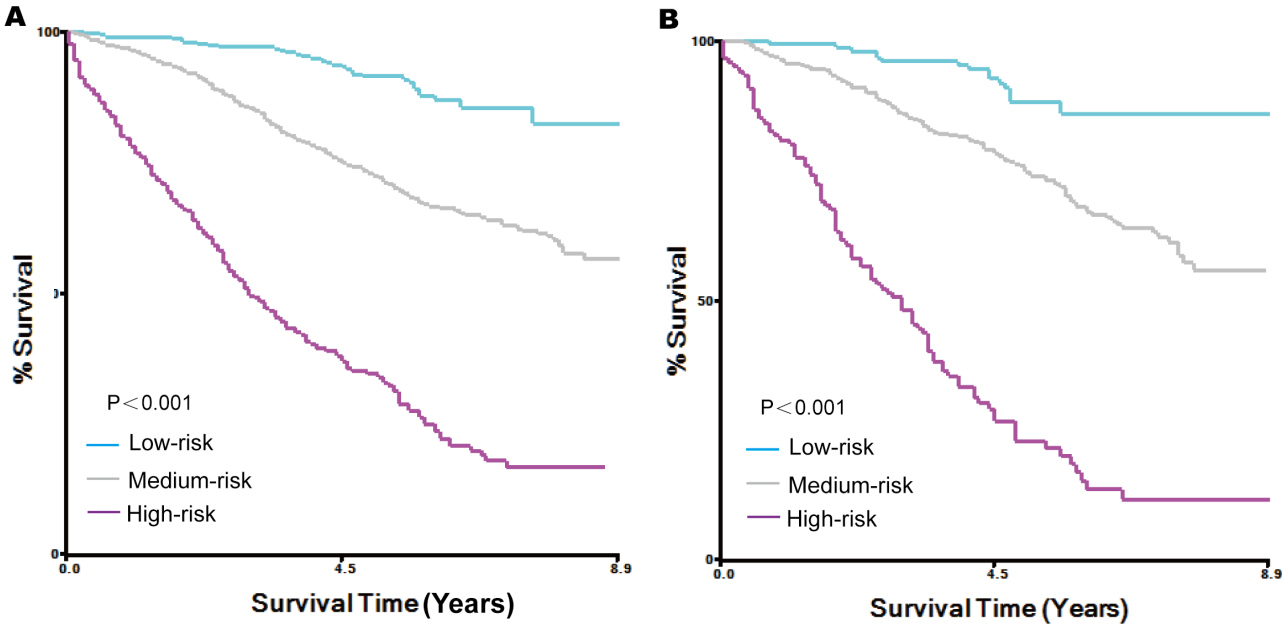
Table . Nomogram score of male breast cancer survival.

Variable	Points
Age	
≤60 years	0
>60 years	15
Surgery	
Yes	0
No	25
Marital status	
Unmarried	7
Married	0
T[a] stage	
T_0	4
T_1	0
T_2	8
T_3	16
T_4	15
Clinical stage	
0	0
I	73
II	74
III	86
IV	100
Chemotherapy	
Yes	0
No	10
HER2[b] status	
Positive	8
Negative	0

[a]T: tumor.

[b]HER2: human epidermal growth factor receptor 2.

Figure 6. Analysis of survival based on risk stratification. Kaplan-Meier curve for patients categorized as low-risk, medium-risk, or high-risk in the training set (A) and internal validation set (B).



Discussion

Principal Findings

In this study, we developed a nomogram to predict survival in MBC based on the SEER database and validated it using both an internal validation dataset from this database and an external validation cohort from a single center. We identified 7 independent risk factors and incorporated them into the nomogram to predict the survival of patients with MBC. The results of both internal and external validation demonstrated that the nomogram exhibited good accuracy and discriminative power, confirming the robustness of the prediction model.

MBC has a low incidence and is a rare malignancy [24]. However, MBC exhibits a delayed onset, presents as a more advanced disease, and has a less unfavorable prognosis than female BC [25]. Due to its rarity, MBC is often overlooked in clinical practice. The assessment of prognosis in MBC holds considerable importance for facilitating the implementation of comprehensive treatment strategies. This study used the clinical data of 1595 patients with MBC from the SEER database to establish a nomogram for prognosticating the survival in MBC. The bootstrap method was used for internal validation, and external validation was performed in the hospital cohort. The ROC curve, C-index, and calibration curve were used to assess the discrimination and reliability of the nomogram. Additionally, the clinical benefit and application value of the nomogram were evaluated using the DCA curve. The findings demonstrated that the nomogram can accurately and individually predict the survival outcomes of patients with MBC. This predictive tool holds the potential for informing clinical decision-making and guiding the development of appropriate diagnostic and therapeutic strategies.

Previous studies analyzed the influencing factors of MBC survival by using univariate and multivariate Cox proportional hazards regression models [26,27]. Compared with traditional multivariate regression, least absolute shrinkage and selection operator (LASSO) regression is widely regarded as a superior approach for variable selection owing to its ability to mitigate model complexity, minimizing overfitting by incorporating a loss function or a penalty term into the objective function. In this study, the LASSO regression algorithm identified 7 variables (age, surgery, marital status, T stage, clinical stage, chemotherapy, and HER2 status) as factors that are associated with the prognosis of MBC. This detail has also been recognized in other studies [9,28]. Based on the aforementioned variables, a nomogram prediction model that significantly enhanced the clinical applicability within various clinical scenarios was developed. The nomogram exhibited good discrimination, consistency, and clinical validity in the training set and validation set. It may guide clinical decision-making for these patients more effectively.

Age was identified as a significant risk factor for the survival of patients with MBC in the nomogram, and individuals aged over 60 years had higher mortality, consistent with prior studies [28,29]. This finding may be related to the presence of more comorbidities in older patients [28]. Surgery and chemotherapy are essential for determining the prognosis of patients with MBC who are undergoing treatment, a finding that is similar to that of previous studies [30-32]. A recent study conducted by Wang et al [33] indicated that patients diagnosed with MBC who received surgery or chemotherapy exhibited a more favorable prognosis than individuals who did not undergo these treatments. The prognostic significance of marital status was observed, with unmarried patients exhibiting a poorer prognosis [34-36]. The reason for this result may be that unmarried patients with MBC experience more significant psychological distress, including feelings of sadness and anxiety, compared to married patients [37], and they may demonstrate greater adherence to treatment regimens [38], which could improve cancer management. Additionally, this study provided evidence to support the notion that the T stage and clinical stage are prognostic indicators for MBC [39,40]. Among the 7 parameters in the nomogram, the clinical stage showed the most significant influence on overall

survival, and patients with stages III and IV MBC had the worst prognosis. One study in Serbia showed that low initial disease stage and low tumor grade are independent predictors of a good prognosis in patients with MBC [41]. In addition, having a HER2-positive tumor is widely acknowledged as a significant prognostic factor for MBC, and this observation has been corroborated by other investigations [42-44].

In this study, the average age of onset of Chinese patients with MBC may be younger than that of patients in the SEER database, which is similar to the onset characteristics of female patients with BC [45]. In addition, the hospital exhibited a higher proportion of patients in the early T stage than the SEER database. The proportion of patients undergoing chemotherapy was significantly greater than that in the SEER database, contributing to the favorable prognosis observed in Chinese patients with MBC.

Constructing a nomogram for the survival of patients with MBC can be beneficial for medical staff to intuitively analyze the weight of risk factors and the corresponding survival probabilities of patients. These survival probabilities can be used as a basis for stratification. The patients were classified into 3 groups: low-, medium-, and high-risk. For example, a patient with MBC that is over 60 years old, is married, has undergone surgery and chemotherapy, has a T grade of T_2, has a tumor stage of II, and has a HER2-positive tumor, would have a total score of approximately 105, belonging to the medium-risk group for survival. Therefore, the medical staff should take relevant measures to timely manage and improve the prognosis of this patient. In clinical practice, the proposed model can be used to determine and evaluate the survival rate and prognosis of patients with MBC. This approach aims to provide personalized and accurate survival rate and prognosis and then develop targeted clinical decisions for patients with MBC.

Strength and Limitations

Our study has the following strengths. First, the existing prognostic models for BC have a focus on female BC, and little focus has been given to MBC. The study aims to develop a prognostic model specifically for this group. Second, the SEER database included a large and diverse cohort, ensuring robust and representative results. In addition, external validation using datasets from our own hospital further confirmed the model's accuracy and generalizability.

However, this study has some limitations. First, as a retrospective study, it is subject to selection bias. Second, important variables, such as endocrine therapy, BMI, and the cellular proliferation marker Ki-67, are not included in the SEER database, which may limit the accuracy and effectiveness of the nomogram. Finally, the external validation sample size in this study was limited, including only retrospective data from a single health care institution, and the predictive ability of the model for the Chinese population needs to be further verified using a large sample of data.

Future Directions

Future studies should consider incorporating data from multicenter cohorts to increase the sample size, thereby enhancing the accuracy and generalizability of survival prediction models for MBC. By collecting data from diverse geographic locations, researchers can ensure that the model captures a broader range of clinical variables, improving its robustness and applicability. Additionally, prospective cohort studies should be conducted to externally validate the model in real-world clinical settings and assess its practical utility in daily clinical decision-making for MBC. Furthermore, integrating additional datasets that include critical variables, such as BMI and the cellular proliferation marker Ki-67, would further strengthen the model's predictive power.

Conclusion

In summary, a nomogram was developed using 7 variables to predict the prognosis of patients with MBC, and age, surgery, marital status, T stage, clinical stage, and HER2 status were identified as independent risk factors for predicting the survival of patients with MBC. Internal and external verifications proved that the model has good accuracy and reliability. Thus, it could serve as an accurate and individualized tool that clinicians could use for decision-making.

Acknowledgments

The authors thank all the patients and institutions for participating in the study, especially for the open access to the SEER database.

Data Availability

The original contributions of this study are detailed in the article and supplementary materials. For any further inquiries, please contact the corresponding author.

Authors' Contributions

Conceptualization: KJ, TFW, YJT, GLC
Data curation: STM, YXX
Formal analysis: WZT, STM
Writing – original draft: WZT
Writing – review & editing: KJ

Conflicts of Interest

None declared.

Multimedia Appendix 1
Clinicopathological characteristics of the Surveillance, Epidemiology, and End Results (SEER) dataset and our dataset.
[DOCX File, 22 KB - cancer_v11i1e54625_app1.docx]

Checklist 1
Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) checklist for prediction model development and validation.
[DOCX File, 94 KB - cancer_v11i1e54625_app2.docx]

References

1. Wang W, Xu X, Tian B, et al. Clinical features of patients with male breast cancer in Shanxi province of China from 2007 to 2016. J Investig Med 2019 Mar;67(3):699-705. [doi: 10.1136/jim-2018-000823] [Medline: 30368484]
2. Abdelwahab Yousef AJ. Male breast cancer: epidemiology and risk factors. Semin Oncol 2017 Aug;44(4):267-272. [doi: 10.1053/j.seminoncol.2017.11.002] [Medline: 29526255]
3. Gucalp A, Traina TA, Eisner JR, et al. Male breast cancer: a disease distinct from female breast cancer. Breast Cancer Res Treat 2019 Jan;173(1):37-48. [doi: 10.1007/s10549-018-4921-9] [Medline: 30267249]
4. Song AL, Ou JH. Clinical practice guideline for diagnosis and treatment of male breast cancer in China. Chin J Pract Surg 2023;43:139-143 [FREE Full text] [doi: 10.19538/j.cjps.issn1005-2208.2023.02.04]
5. Konduri S, Singh M, Bobustuc G, Rovin R, Kassam A. Epidemiology of male breast cancer. Breast 2020 Dec;54:8-14. [doi: 10.1016/j.breast.2020.08.010] [Medline: 32866903]
6. Skop M, Lorentz J, Jassi M, Vesprini D, Einstein G. “Guys don’t have breasts”: the lived experience of men who have BRCA gene mutations and are at risk for male breast cancer. Am J Mens Health 2018 Jul;12(4):961-972. [doi: 10.1177/1557988317753241] [Medline: 29400121]
7. Rojas K, Stuckey A. Breast cancer epidemiology and risk factors. Clin Obstet Gynecol 2016 Dec;59(4):651-672. [doi: 10.1097/GRF.0000000000000239] [Medline: 27681694]
8. Lin AP, Huang TW, Tam KW. Treatment of male breast cancer: meta-analysis of real-world evidence. Br J Surg 2021 Sep 27;108(9):1034-1042. [doi: 10.1093/bjs/znab279] [Medline: 34476472]
9. Leone JP, Zwenger AO, Iturbe J, et al. Prognostic factors in male breast cancer: a population-based study. Breast Cancer Res Treat 2016 Apr;156(3):539-548. [doi: 10.1007/s10549-016-3768-1] [Medline: 27039306]
10. Giordano SH. Breast cancer in men. N Engl J Med 2018 Oct 4;379(14):1385-1386. [doi: 10.1056/NEJMc1809194] [Medline: 30281998]
11. Rice TW, Ishwaran H, Ferguson MK, Blackstone EH, Goldstraw P. Cancer of the esophagus and esophagogastric junction: an eighth edition staging primer. J Thorac Oncol 2017 Jan;12(1):36-42. [doi: 10.1016/j.jtho.2016.10.016] [Medline: 27810391]
12. Wang J, Zhou J, Liu L, Wu SG. Stage-specific survival in breast cancer in chinese and white women: comparative data analysis. JMIR Public Health Surveill 2022 Nov 15;8(11):e40386. [doi: 10.2196/40386] [Medline: 36378507]
13. Yang SP, Su HL, Chen XB, et al. Long-term survival among histological subtypes in advanced epithelial ovarian cancer: population-based study using the surveillance, epidemiology, and end results database. JMIR Public Health Surveill 2021 Nov 17;7(11):e25976. [doi: 10.2196/25976] [Medline: 34787583]
14. Donegan WL, Redlich PN, Lang PJ, Gall MT. Carcinoma of the breast in males: a multiinstitutional survey. Cancer 1998 Aug 1;83(3):498-509. [doi: 10.1002/(sici)1097-0142(19980801)83:3<498::aid-cncr19>3.0.co;2-r] [Medline: 9690543]
15. Anderson WF, Jatoi I, Tse J, Rosenberg PS. Male breast cancer: a population-based comparison with female breast cancer. J Clin Oncol 2010 Jan 10;28(2):232-239. [doi: 10.1200/JCO.2009.23.8162] [Medline: 19996029]
16. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424. [doi: 10.3322/caac.21492] [Medline: 30207593]
17. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. Lancet Oncol 2015 Apr;16(4):e173-e180. [doi: 10.1016/S1470-2045(14)71116-7] [Medline: 25846097]
18. Palzer EF, Vempati S, Helgeson ES, Matas AJ. Long-term living kidney donor risk: a web-based calculator. J Am Soc Nephrol 2020 Dec;31(12):2968-2969. [doi: 10.1681/ASN.2020081238] [Medline: 33127859]
19. Gao Y, Li S, Jin Y, et al. An assessment of the predictive performance of current machine learning-based breast cancer risk prediction models: systematic review. JMIR Public Health Surveill 2022 Dec 29;8(12):e35750. [doi: 10.2196/35750] [Medline: 36426919]
20. Yang X, Qiu H, Wang L, Wang X. Predicting colorectal cancer survival using time-to-event machine learning: retrospective cohort study. J Med Internet Res 2023 Oct 26;25:e44417. [doi: 10.2196/44417] [Medline: 37883174]

21. Wu J, Zhang H, Li L, et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: a population-based analysis. Cancer Commun (Lond) 2020 Jul;40(7):301-312. [doi: 10.1002/cac2.12067] [Medline: 32558385]
22. Huang X, Luo Z, Liang W, et al. Survival nomogram for young breast cancer patients based on the SEER database and an external validation cohort. Ann Surg Oncol 2022 Sep;29(9):5772-5781. [doi: 10.1245/s10434-022-11911-8] [Medline: 35661275]
23. Wang Z, Gao L, Guo X, et al. Development and validation of a nomogram with an autophagy-related gene signature for predicting survival in patients with glioblastoma. Aging (Albany NY) 2019 Dec 17;11(24):12246-12269. [doi: 10.18632/aging.102566] [Medline: 31844032]
24. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66(2):115-132. [doi: 10.3322/caac.21338] [Medline: 26808342]
25. Malani AK. Male breast cancer: a different disease than female breast cancer? South Med J 2007 Feb;100(2):197. [doi: 10.1097/01.smj.0000209277.21066.c2] [Medline: 17330691]
26. Zhao J, Wang B, Zhao J, Mao Y, Liu J, Yang Y. Male breast cancer: a closer look at patient and tumor characteristics and factors associated with survival. Thorac Cancer 2020 Nov;11(11):3107-3116. [doi: 10.1111/1759-7714.13611] [Medline: 32930510]
27. Jylling AMB, Jensen V, Lelkaitis G, Christiansen P, Nielsen SS, Lautrup MD. Male breast cancer: clinicopathological characterization of a national Danish cohort 1980-2009. Breast Cancer (Auckl) 2020 Jul;27(4):683-695. [doi: 10.1007/s12282-020-01066-3] [Medline: 32108307]
28. Yadav S, Karam D, Bin Riaz I, et al. Male breast cancer in the United States: treatment patterns and prognostic factors in the 21st century. Cancer 2020 Jan 1;126(1):26-36. [doi: 10.1002/cncr.32472] [Medline: 31588557]
29. Khan NAJ, Tirona M. An updated review of epidemiology, risk factors, and management of male breast cancer. Med Oncol 2021 Mar 15;38(4):39. [doi: 10.1007/s12032-021-01486-x] [Medline: 33721121]
30. Singh R, Cao L, Sarode AL, Kharouta M, Shenk R, Miller ME. Trends in surgery and survival for T1-T2 male breast cancer: a study from the National Cancer Database. Am J Surg 2023 Jan;225(1):75-83. [doi: 10.1016/j.amjsurg.2022.09.043] [Medline: 36208958]
31. Zhang J, Jiang H, Zhang J, et al. Effectiveness and safety of pegylated liposomal doxorubicin versus epirubicin as neoadjuvant or adjuvant chemotherapy for breast cancer: a real-world study. BMC Cancer 2021 Dec 6;21(1):1301. [doi: 10.1186/s12885-021-09050-6] [Medline: 34872507]
32. Pan H, Zhang K, Wang M, Ling L, Wang S, Zhou W. The effect of chemotherapy on survival in patients with nonmetastatic male breast cancer: a population-based observational study. Cancer 2020 Aug 15;126 Suppl 16:3830-3836. [doi: 10.1002/cncr.32829] [Medline: 32710661]
33. Wang J, Sun Y, Qu J, et al. Survival analysis for male ductal and lobular breast cancer patients with different stages. Future Oncol 2019 Jan;15(2):167-180. [doi: 10.2217/fon-2018-0455] [Medline: 30088776]
34. Adekolujo OS, Tadisina S, Koduru U, Gernand J, Smith SJ, Kakarala RR. Impact of marital status on tumor stage at diagnosis and on survival in male breast cancer. Am J Mens Health 2017 Jul;11(4):1190-1199. [doi: 10.1177/1557988316669044] [Medline: 27694551]
35. Liu L, Chi YY, Wang AA, Luo Y. Marital status and survival of patients with hormone receptor-positive male breast cancer: a surveillance, epidemiology, and end results (SEER) population-based study. Med Sci Monit 2018 May 24;24:3425-3441. [doi: 10.12659/MSM.910811] [Medline: 29795054]
36. Krajc K, Miroševič Š, Sajovic J, et al. Marital status and survival in cancer patients: a systematic review and meta-analysis. Cancer Med 2023 Jan;12(2):1685-1708. [doi: 10.1002/cam4.5003] [Medline: 35789072]
37. Goldzweig G, Andritsch E, Hubert A, et al. Psychological distress among male patients and male spouses: what do oncologists need to know? Ann Oncol 2010 Apr;21(4):877-883. [doi: 10.1093/annonc/mdp398] [Medline: 19822532]
38. Tran BX, Fleming M, Do HP, Nguyen LH, Latkin CA. Quality of life improvement, social stigma and antiretroviral treatment adherence: implications for long-term HIV/AIDS care. AIDS Care 2018 Dec;30(12):1524-1531. [doi: 10.1080/09540121.2018.1510094] [Medline: 30099885]
39. Methamem M, Ghadhab I, Hidar S, Briki R. Breast cancer in men: a serie of 45 cases and literature review. Pan Afr Med J 2020;36:183. [doi: 10.11604/pamj.2020.36.183.22574] [Medline: 32952827]
40. André S, Pereira T, Silva F, et al. Male breast cancer: specific biological characteristics and survival in a Portuguese cohort. Mol Clin Oncol 2019 Jun;10(6):644-654. [doi: 10.3892/mco.2019.1841] [Medline: 31031981]
41. Sipetic-Grujicic SB, Murtezani ZH, Neskovic-Konstatinovic ZB, et al. Multivariate analysis of prognostic factors in male breast cancer in Serbia. Asian Pac J Cancer Prev 2014;15(7):3233-3238. [doi: 10.7314/apjcp.2014.15.7.3233] [Medline: 24815476]
42. Sokmen FC. Prognostic factors in male breast cancer: a single centre experience. J Coll Physicians Surg Pak 2020 Dec;30(12):1285-1290. [doi: 10.29271/jcpsp.2020.12.1285] [Medline: 33397054]
43. Abreu MH, Afonso N, Abreu PH, et al. Male breast cancer: looking for better prognostic subgroups. Breast 2016 Apr;26:18-24. [doi: 10.1016/j.breast.2015.12.001] [Medline: 27017238]

44. Leone J, Freedman RA, Lin NU, et al. Tumor subtypes and survival in male breast cancer. Breast Cancer Res Treat 2021 Aug;188(3):695-702. [doi: 10.1007/s10549-021-06182-y] [Medline: 33770314]
45. Fan L, Strasser-Weippl K, Li JJ, et al. Breast cancer in China. Lancet Oncol 2014 Jun;15(7):e279-e289. [doi: 10.1016/S1470-2045(13)70567-9] [Medline: 24872111]

Abbreviations

BC: breast cancer
C-index: concordance index
DCA: decision curve analysis
HER2: human epidermal growth factor receptor 2
LASSO: least absolute shrinkage and selection operator
MBC: male breast cancer
ROC: receiver operating characteristic
SEER: Surveillance, Epidemiology, and End Results
TNM: tumor, extent of spread to the lymph nodes, and presence of metastasis

Edited by N Cahill; submitted 16.11.23; peer-reviewed by K Matthew, K Zhang, S Matsuda; revised version received 12.12.24; accepted 18.12.24; published 04.03.25.

Please cite as:
Tang WZ, Mo ST, Xie YX, Wei TF, Chen GL, Teng YJ, Jia K
Predicting Overall Survival in Patients with Male Breast Cancer: Nomogram Development and External Validation Study
JMIR Cancer 2025;11:e54625
URL: https://cancer.jmir.org/2025/1/e54625
doi:10.2196/54625

Challenges of Cross-Sectoral Video Consultation in Cancer Care on Patients' Perceived Coordination: Randomized Controlled Trial

Fereshteh Baygi[1], PhD; Theis Bitz Trabjerg[1], PhD; Lars Henrik Jensen[2,3], PhD; Maria Munch Storsveen[1], MSc; Sonja Wehberg[1], PhD; Jeffrey James Sisler[4], PhD; Jens Søndergaard[1], PhD; Dorte Gilså Hansen[1], PhD

[1]Department of Public Health, Research Unit of General Practice, University of Southern Denmark, Campusvej 55, Odense, Denmark

[2]Department of Oncology, Lillebælt University Hospital, Vejle, Denmark

[3]The Department of Regional Health Research, Danish Colorectal Cancer Center South, Center of Clinical Excellence, Vejle Hospital, University of Southern Denmark, Vejle, Denmark

[4]Department of Family Medicine, Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, CanadaCanada

Corresponding Author:
Fereshteh Baygi, PhD
Department of Public Health, Research Unit of General Practice, University of Southern Denmark, Campusvej 55, Odense, Denmark

Abstract

Background: Patients with cancer need coordinated care for both treatment and concurrent health conditions. This requires collaboration among specialists when using telemedicine services, emphasizing the importance of care continuity.

Objective: This study aimed to explore the effects of cross-sectorial video consultation involving oncologists, general practitioners, and patients with cancer on patients' perceived coordination of care, compared with usual care.

Methods: This study describes the primary outcomes from a 7-month follow-up of patients in the Partnership Project, a randomized clinical trial. Patients in the intervention group were randomized to receive a "partnership consultation," a shared video consultation with an oncologist, general practitioners, and the patient, in addition to their usual care. Questionnaires were completed for both groups at baseline and 7 months to assess the primary outcome, "global assessment of inter-sectorial cooperation," from the Danish questionnaire "Patients' attitude to the health care service." The questionnaire also included 2 single items and 5 index scales, examining patients' attitude toward cooperation in the health care system. Change in perceived global coordination from baseline to 7 months was compared between intention-to-treat groups using generalized estimating equations in a linear regression model.

Results: A total of 278 participants were randomized with 1:1 allocation, with 80 patients receiving the intervention. Further, 210 patients completed the questionnaire at baseline, while 118 responded at 7-month follow-up. The estimated difference in the primary outcome between usual care (−0.13, 95% CI −0.38 to 0.12) and intervention (0.11, 95% CI −0.11 to 0.34) was 0.24 (95% CI −0.09 to 0.58) and not statistically significant (*P*=.15).

Conclusions: Low rates of intervention completion and high levels of missing data compromised the interpretability of our study. While we observed a high level of global assessment of coordination, the estimated intervention effect was smaller than anticipated, with no significant difference in perceived coordination between control and intervention groups. Future studies should explore strategies like patient incentives to increase response rate and improve the evaluation of this innovative health care model.

Trial Registration: ClinicalTrials.gov NCT02716168; https://clinicaltrials.gov/study/NCT02716168

International Registered Report Identifier (IRRID): RR2-10.1186/s12875-019-0978-8

(JMIR Cancer 2025;11:e60158) doi:10.2196/60158

KEYWORDS

randomized controlled trials; video consultations; outcome assessment; patients' satisfaction; patients' care coordination; interprofessional relations; cancer

Introduction

Health care systems increasingly use digital technology to improve quality of health care services across a spectrum of medical issues including critical conditions like cancer [1]. Notably, over the past 2 decades, there has been a growing deployment of telemedicine technologies, making a transformative shift in how health care is delivered and experienced [2].

Patients with cancer have distinctive medical requirements, including both cancer therapy and treatment of concurrent health conditions [3]. Addressing these needs involves engaging various specialties and health professionals, using specialized telemedicine care services, and ensuring continuity of care throughout and after cancer treatment. This necessitates a higher level of care coordination [4]. However, both patients and health care providers face challenges in coordinating care and communication patterns, as evidence by logistic issues such as technological problems which hinder effective telemedicine practices [5,6]. To mitigate such challenges, shared care models have been proposed as a promising approach [7]. These models allow patients to benefit from the expertise of specialists, while maintaining the care through the primary care providers. By bringing together the patients, general practitioners (GPs), and oncologist in a shared video consultation, telemedicine offers a powerful solution to improve care coordination. A recent study further supports the benefits of telemedicine as its ability to improve care coordination, and better management patients' health needs through enhanced communication [8]. Therefore, efforts to assist patients with cancer have shifted toward patient-centered communication approaches [6], so that over time, such approaches for these patients are rapidly expanding and diversifying [9]. Based on a previous study, these approaches may have varying impact on patients' outcomes and perceptions [9]. Furthermore, application of such approaches in combination with virtual consultations may have diverse effects on outcomes, as they may interact differently with each patient's unique health needs. This aligns with the health care providers' perspective, who advocate patient-centered approaches in cancer care [9]. Despite such widespread advocacy, there is limited consensus on definition and methods to achieve patient-centered care [9].

A previous meta-analysis on cancer care coordination suggests that implementation of cancer-care coordination approaches resulted in positive changes in majority of measured outcomes (eg, overall patients' experience on cancer care, quality of end-of-life care, etc). The study recommended the development of new intervention models and care coordination strategies to enhance patients' self-management [10]. Notably, none of the studies included in this meta-analysis applied a virtual intervention mode [10].

We hypothesized that virtual shared models involving specialists, primary care providers, and patients could more effectively address optimal outcomes for patients with cancer [3]. Hence, this study aims to investigate the effects of a shared video consultation including oncologists, GPs, and patients with cancer on the patients' perceived coordination of care.

Methods

Study Design

This study is a randomized controlled trial entitled "The Partnership Project' (PSP)" [11]. The protocol and details of the study have been published previously [11-13]. This paper now presents the primary outcome based on a 7-month follow-up survey on patients' participation in a shared video-based consultation.

Participants and Setting

All newly diagnosed patients with any type of cancer receiving treatment with chemotherapy for the first time at the Department of Oncology, Lillebælt Hospital, University Hospital of Southern Denmark were invited for the study. The eligibility criteria were age above 18 years, proficiency in speaking and reading Danish, and having an estimation from an oncologist indicating a survival time of more than 7 months.

Multimedia Appendix 1 provides an explanation of the initial sample size that was previously published [14]. Since patient inclusion matched the predetermined sample size, the trial was ended.

Usual Care

The control group was randomized to receive "usual care" in terms of standard information exchanging between the department of oncology and primary care. This involved sending an electronic summary letter to the GP after each visit to the department of oncology. GPs and the hospital can communicate by phone if needed. In addition, patients may reach out to their GP or a designated coordinator at the department of oncology.

Intervention

Patients in the intervention group were randomized to receive a "partnership consultation," which was a shared video consultation involving an oncologist, GP, and patients with cancer, alongside their "usual care." GPs were contacted only after obtaining patients' consent, and the GP had the option to refuse participation. Three to 6 weeks in advance, the sessions were scheduled during regular clinic hours. Patients were given the option to choose either the GPs' office or the oncologist's office for their consultations. In case the patient preferred to sit by the GP, the video consultation took place in that way, with the oncologist alone in his or her office at the hospital. GPs or oncologists may have had more than 1 patient in the intervention group. However, we do not have specific information about the individual oncologist for each patient in our database. The consultations were chaired by an oncologist within 12 weeks from the time of inclusion. Before the consultation, oncologists and GPs received information including a consultation guide with themes that may be relevant (Multimedia Appendix 2). Typically, the oncologist was assisted by an oncology nurse. A summary of the consultation was recorded in the hospital electronic patient record, shared with the GP, and accessible for the patient at an online portal (sundhed) for medical reports in Denmark.

Three to 6 weeks in advance, the sessions were scheduled during regular clinic hours. Patients were given the option to choose either the GPs' office or the oncologist's office for their consultations. In case the patient preferred to sit by the GP, the video consultation took place in that way, with the oncologist alone in his or her office at the hospital.

GPs or oncologists may have had more than 1 patient in the intervention group. However, we do not have specific information about the individual oncologist for each patient in our database.

The consultations were chaired by an oncologist within 12 weeks from the time of inclusion.

Prior to the consultation, oncologists and GPs received information including a consultation guide with themes that may be relevant (Multimedia Appendix 2). Typically, the oncologist was assisted by an oncology nurse.

A summary of the consultation was recorded in the hospital electronic patient record, shared with the GP, and accessible for the patient at an online portal (sundhed.dk) for medical reports in Denmark.

Randomization and Blinding

Following informed consent, patients were assigned in a 1:1 ratio through block randomization, with block sizes and sequences known only to the REDCap (Research Electronic Data Capture; Vanderbilt University) [15] data manager from our collaboration partners. The allocation was transparent for the patients, GPs, and oncologist. However, during baseline data collection, patients in the intervention group and enrolling nurse were kept blind of the randomization.

A project nurse at the research unit in the department of oncology conducted the randomization process and enrolled patients in the study following the patients' consent. Neither patients nor their GPs and oncologists in the intervention group were blinded to the patients' allocation status. Data analysts remained blinded to the allocation. GPs of patients in the control arm were not formally informed until they received the survey.

Primary Outcomes and Instruments

Patients were asked to complete questionnaires at baseline, and after 4 and 7 months. Upon arrival at the department of oncology, patients received information, a consent form, and a paper-based baseline questionnaire, which outpatient nurse collected after enrollment. Follow-up questionnaires were sent electronically using REDCap [15], which securely managed distribution and response collection. However, patients could request paper-based follow-up with prepaid return envelope.

An overview of primary and secondary outcomes can be found in Multimedia Appendix 3. The primary outcome included the single item "global assessment of inter-sectorial cooperation," which was part of the Danish questionnaire "Patients' attitude to the health care service" [14]. The English questionnaire "The patient cancer diary" [16] served as the basis for the 26-item Danish questionnaire. The adaptation was done based on interviews with Danish cancer patients and caregivers [17] and the English questionnaire template [16]. The questionnaire was chosen because it measures the study's aim, patients' perceptions of cross-sector cooperation, and has previously been used in a Danish cancer study [14]. Single items were scored on a 5-point Likert scale from strongly agree (1) to strongly disagree (5). The "not relevant" category was coded as missing.

This questionnaire comprises also 2 other single items and 5 index scales (secondary outcome; Multimedia Appendix 3), examining patients' attitude toward the cooperation in the health care system. There is no manual available for the questionnaire; however, 2 papers have been published that describe the validation and usage of the questionnaire [14,17]. For the 5 subscales, at most 1 missing was replaced by the mean of the other items in the corresponding subscale. A subscale was coded as missing if more than 1 single item in the scale were missing. The direction of the answer scale varied depending on the item. For instance, in the case of the primary outcome, a low score indicated a positive attitude toward the question, while for secondary outcomes (eg, global feeling of left in limbo), a high score indicated a positive attitude. However, for analysis purpose, all items were aligned so that a higher value indicated a positive attitude toward the questions. Primary and secondary outcomes were measured at the following time points: baseline, 4, and 7 months after baseline. Coding was done separately for each time point.

Other Parameters

Demographic data for patients including age, gender, education, marital status, having child, work status, comorbidity, diagnosis or cancer type were assessed through questionnaire which was completed by patients at baseline.

Deviation From the Protocol in Statistical Analyses

As outlined in the published protocol [11], the original statistical analyses plan aimed to conduct a simple group comparison at 7 months using *t* tests or Wilcoxon rank-sum tests. However, a deviation from the initial analysis strategy was decided due to the large amount of missing data associated with the primary variable at 7 months.

Definition of Intervention and Control Groups for Analysis

The main analysis strategy followed the intention-to-treat (ITT) approach, defining groups by random allocation (control and intervention). As a second approach, we defined 2 as-treated grouping approaches. First, we split the intervention group by degree of intervention fidelity: 1 group had the intervention as defined by protocol, the second did not receive the intervention due to technical issues, while the third did not receive the intervention for other reasons. Then, as-treated $group_1$ (AT1) comprised patients completing the video consultations; this group was compared with patients who did not receive the intervention (randomized to control or $subgroup_1$ or $subgroup_2$). As-treated $group_2$ (AT2) comprised patients with planned video consultations, regardless of completion due to technical reasons (AT1 and $subgroup_1$); this group was compared with patients randomized to controls and $subgroup_2$. As-treated groups were defined post hoc. Unless stated explicitly otherwise, we based the group definition on the ITT approach.

Revised Statistical Analysis

As a first step, we compared the change from baseline to 7 months between the 2 groups, as defined by allocation (ITT), through a linear regression model. This model was applied to measurements at both baseline and 7 months, using generalized estimating equations (GEE) to account for within-patient clusters. Robust variance estimation was used, and the group difference was modeled as time-by-group interaction. Similar estimates were presented for both post hoc and defined as-treated approaches. No additional covariates were considered in this primary analysis. The GEE approach was chosen to ensure

robustness of the statistical methods in the presence of missing data. Analyses specified in the protocol were also reported, restricted to complete cases. A corresponding analysis strategy was followed for secondary outcomes. Data analyses was done using Stata version 17 [18], and the significance level was set at 5%.

Ethical Considerations

Statement Regarding Human Subject Research Ethics Review

The Regional Ethics Committee on Biomedical Research in Denmark (S-20142000 - 138) and the Danish Data Protection Agency (2014-41-3534) peer reviewed and approved the study.

Informed Consent Descriptions

At the Oncological Department, Vejle Hospital, Denmark, outpatient clinic nurses obtained informed consent from patients for the PSP. Patients provided consent to participate in the randomized controlled trial, the video recordings, and the user perspective assessments on the same consent form. The consent forms were securely stored at the Clinical Research Unit, Department of Oncology, Vejle Hospital. The unit of randomization was the patient. Therefore, according to Danish law and the instructions of the Regional Ethics Committee on Biomedical Research, consent from GPs was not required. However, out of courtesy and to show consideration for their workload, oral consent was obtained from GPs when their patients were allocated to the intervention group. Before the study's start, written information about the trial was sent to all GPs in the Region of Southern Denmark. If a GP declined to participate, their patients were not invited to join the study.

GPs for patients in the control group were not contacted and were therefore unaware of their patients' participation in the study until they were asked to complete a questionnaire 4 months after the patients' inclusion.

Privacy and Confidentiality Protection Description

Data security in video consultations is essential. Patients demonstrate a high level of trust regarding data security, as they trust the health care staff using the technology. To ensure this trust, all video consultations were conducted on the Region of Southern Denmark's secure videoconference servers using virtual meeting rooms. These servers provide a highly secure connection with no third-party data processing, and meeting rooms could only be accessed by the participating parties. Before a video consultation, patients may have discussed confidential matters, such as alcohol consumption or smoking, which they had not shared with all health care providers. This could place patients in a dilemma. To address this, the intervention guide for oncologists and GPs includes a note to handle such situations sensitively.

Compensation Details

No compensation was provided to patients or oncologists. GPs were reimbursed through the standard payment system of the Region of Southern Denmark for participating in video consultations with a specialist at the hospital. The agreement used existing provisions for cross-sector cooperation and discharge follow-up. GPs received a fee for video consultations based on the time spent: €48 (at the time of the study, the exchange rate was approximately €1=US $1.10; fee number 4670) for up to 30 minutes and €97 (fee number 4669) for consultations exceeding 30 minutes. Therefore, they did not receive any additional payment for participation, nor were they paid by the study for their involvement in the video consultations. Furthermore, GPs were compensated for completing the questionnaires. Payment was provided by the Region of Southern Denmark and corresponded to one module (€18 at the time for the trial), equivalent to the payment for a standard patient consultation in their clinic.

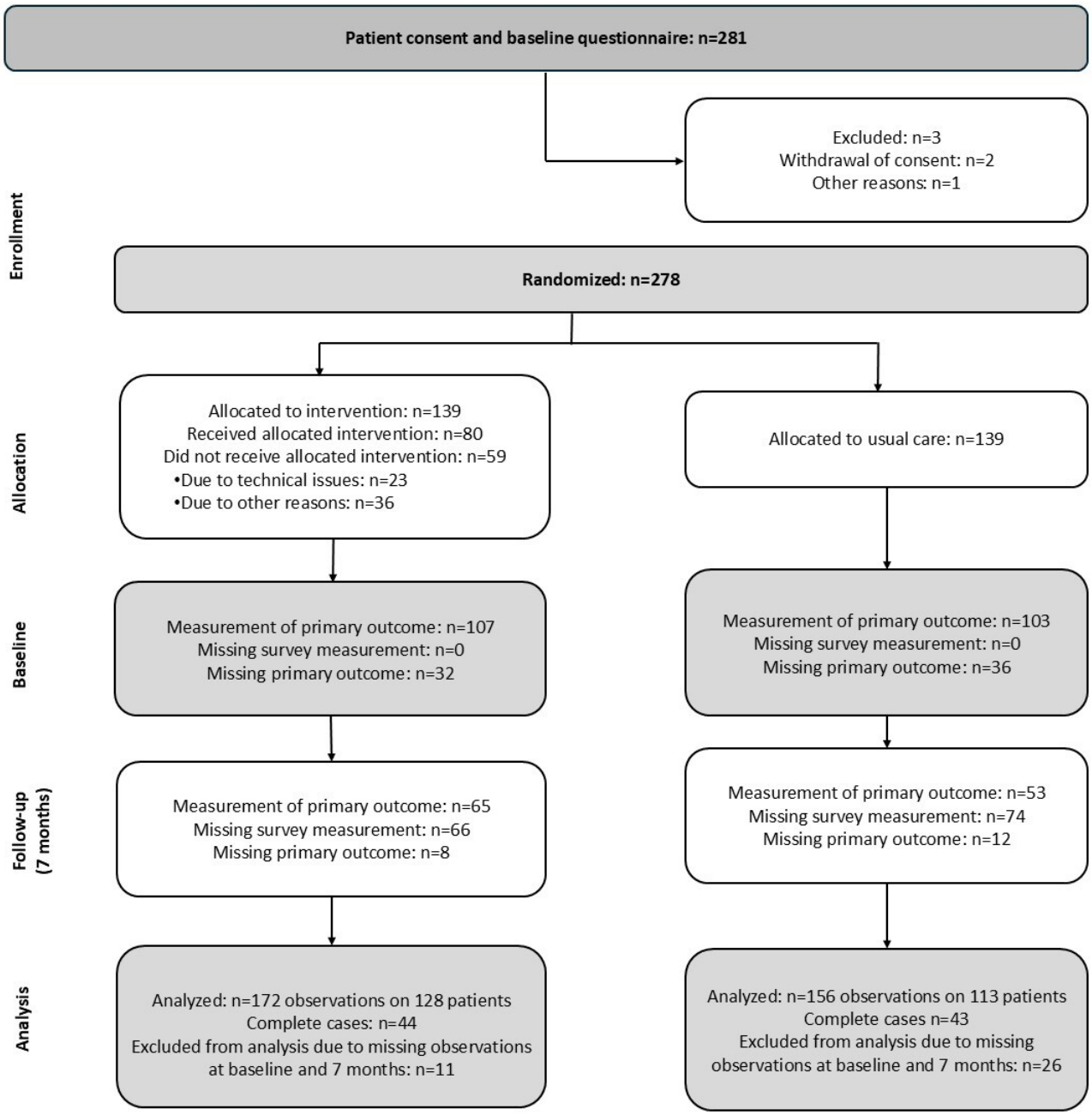
Results

Recruitment and Participant Flow

The patients were included between June 2016 and November 2019. In this study, 281 patients initially agreed to participate. Three patients were excluded due to withdrawal of the consent or other reasons. In total, 278 patients were randomized; 139 patients were allocated to the intervention group and another 139 to the control group. However, due to the following reasons, only 80 patients received the intervention as intended: GP refused participation for 22 patients; in 15 cases, IT failed; and for 8 patients, there were administrative (scheduling) issues. A total of 8 patients died before intervention, 3 patients were too ill, and 3 did not wish to participate in the intervention. See Figure 1 for the CONSORT (Consolidated Standards of Reporting Trials) flow diagram [19] (Checklist 1).

An overview of the GP-patient relationship can be found in Multimedia Appendix 4 to provide additional context.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



Baseline Data

As shown in Table 1, patients in the intervention and control groups had similar baseline characteristics. However, comorbidity was more frequent in the control group (58.3% vs 46.8%) than in the intervention group.

Table . Baseline characteristics of patients in the intervention and control groups.

Characteristics		Total (N=278), n (%)	Control group (n=139), n (%)	Intervention group (n=139), n (%)
Age (years), mean (SD)		65.2 (10.6)	63.8 (11)	66.6 (10)
Sex				
	Male	155 (55.6)	77 (55.4)	78 (56.1)
	Female	123 (44.4)	62 (44.6)	61 (43.9)
Education				
	Primary and upper secondary school	176 (63.3)	85 (61.1)	88 (63.3)
	Further education	76 (27.3)	41 (29.5)	35 (25.2)
	Higher education	16 (5.8)	7 (5)	9 (6.5)
Marital status				
	Single or missing[a]	81 (29.1)	48 (34.5)	33 (23.7)
	Married or residing with a companion	197 (70.9)	91 (65.5)	106 (76.3)
Children living at home				
	No children at home or missing[a]	244 (87.8)	120 (86.3)	124 (89.2)
	Children at home	34 (12.2)	19 (13.7)	15 (10.8)
Work status				
	Employed	89 (32)	46 (33.1)	43 (30.9)
	Public benefits	15 (5.4)	9 (6.5)	6 (4.3)
	Retired or missing[a]	174 (62.6)	84 (60.4)	90 (64.7)
Comorbidity				
	No	132 (47.5)	58 (41.7)	74 (53.2)
	Yes	146 (52.5)	81 (58.3)	65 (46.8)
Diagnosis or cancer type				
	Breast	33 (11.9)	17 (12.2)	16 (11.5)
	Gynecological	13 (4.7)	4 (2.9)	9 (6.5)
	Lung	106 (38.1)	53 (38.1)	53 (38.1)
	Gastrointestinal	110 (39.6)	56 (40.3)	54 (38.8)
	Other	16 (5.8)	9 (6.5)	7 (5)
	Incident cancer (yes or missing[a])	255 (91.7)	126 (90.6)	129 (92.8)

[a]There were less than 3 patients with missing information on marital status, number of children at home or work status, and 6 patients with missing information on incident cancer. These patients were grouped with the indicated categories.

Numbers Analyzed

In Table 2, an overview of the missing data for the primary outcome at different time points in both control and intervention groups is presented. Over time, there was a decline in participation in both the control group (38% at 7 months vs 74% at baseline) and the intervention group (47% at 7 months vs 77% at baseline), based on 278 randomized patients. In the ITT analyses, 172 observations on 128 patients from the intervention group and 156 observations on 113 patients from the control group were included. A total of 11 patients (intervention) and 26 patients (control) were excluded from the analysis due to missing observations at both baseline and 7 months.

A total of 59 participants failed to have the intervention as intended, due to technical problems or other reasons. Based on the subgroup analyses, the subgroup$_1$ had a higher percentage (78.3%) of nonmissing values, which gradually dropped to 39.1%% and 34.8% at the subsequent time points (Table 2). The subgroup$_2$ displayed a comparable pattern of missing data at various time points (83.8%, 35.1%, and 24.3%, respectively).

Table . Overview of the missing data for the primary outcome in both control and intervention groups at baseline and follow-up.

Group	Baseline					4 months				7 months			
	Total, n	Data available, n (%)	Data[a] missing, n (%)	Not relevant[b], n	Missing[c] survey, n (%)	Data available, n (%)	Data missing, n (%)	Not relevant, n	Missing survey, n (%)	Data available, n (%)	Data missing, n (%)	Not relevant, n	Missing survey, n (%)
Total	278	210 (75.5)	68 (24.5)	51	0	138 (49.6)	27 (97)	14	113 (40.6)	118 (42.4)	20 (7.2)	0	140 (50.4)
$Group_1$[d]	139	103 (74.1)	36 (25.9)	27	0	61 (43.9)	15 (10.8)	7	63 (45.3)	53 (38.1)	12 (8.6)	—[e]	74 (53.2)
$Group_2$[f]	139	107 (77)	32 (23)	24	0	77 (55.4)	12 (8.6)	7	50 (36)	65 (46.8)	8 (5.8)	—	66 (47.5)
$Subgroup_1$[g]	23	18 (78.3)	5 (21.7)	—	0	9 (39.1)	3 (13)	—	11 (47.8)	8 (34.8)	<3 (8.7)	—	13 (56.5)
$Subgroup_2$[h]	36	31 (83.8)	6 (16.2)	—	0	13 (35.1)	2 (5.4)	—	22 (59.5)	9 (24.3)	<3	—	27 (73)

[a]Data missing: the primary variable was absent, while other sections might have been answered.

[b]Not relevant: the primary variable was answered as “not relevant”; this is part of the data missing category.

[c]Missing survey: the entire questionnaire was missing for a specific time point.

[d]$Group_1$: randomized as control.

[e]Not applicable.

[f]$Group_2$: randomized as intervention.

[g]Intervention group was divided into the following subgroups based on reasons for not receiving the intervention. $Subgroup_1$: not received intervention due to technical problems.

[h]$Subgroup_2$: not received intervention due to other problems.

Outcomes and Estimations

Table 3 shows patients’ attitudes toward the cooperation between the primary sector and the department of oncology. The estimated within-group changes in the primary outcome between baseline and follow-up were −0.13 (95% CI –0.38 to 0.12) in the control group and 0.11 (95% CI −0.11 to 0.34) in the intervention group. The between-group difference was estimated as 0.24 (95% CI −0.09 to 0.58; *P*=.15). This suggest that, based on perceived global coordination, there was no noticeable differences between the ITT groups from the beginning to 7-month follow-up.

Table . Patients' attitudes toward the cooperation between the primary sector and the department of oncology.

Outcomes		Group	Baseline		7 months		Estimated change (95% CI)	Group-time interac-tion,(95% CI)	*P* value
			n	Mean (SD)	n	Mean (SD)			
Primary									
	ITT[a]	C[b]	103	3.73 (0.98)	53	3.62 (1.04)	−0.13 (−0.38 to 0.12)	—[c]	—
	ITT	I[d]	107	3.79 (0.96)	65	3.91 (0.98)	0.11 (−0.11 to 0.34)	0.24 (−0.09 to 0.58)	.15
	AT1[e]	C	151	3.75 (0.96)	70	3.67 (1.09)	−0.10 (−0.33 to 0.12)	—	—
	AT1	I	59	3.78 (1)	48	3.94 (0.89)	0.17 (−0.08 to 0.42)	0.27 (−0.07 to 0.61)	.11
	AT2	C	133	3.71 (0.95)	62	3.65 (1.06)	−0.09 (-0.33 to 0.14)	—	—
	AT2	I	77	3.86 (1)	56	3.93 (0.95)	0.10 (−0.15 to 0.34)	0.19 (−0.15 to 0.53)	.27
Secondary (subscores)									
	LIMBO[f]								
	ITT	C	111	3.77 (1.18)	55	3.73 (1.18)	−0.05 (−0.39 to 0.29)	—	—
	ITT	I	112	3.96 (1.03)	64	3.81 (1.15)	−0.13 (−0.44 to 0.18)	−0.08 (−0.55 to 0.38)	.73
	FAM-Global[g]								
	ITT	C	117	12.79 (27.91)	42	6.12 (14.73)	−6.68 (-13.41 to 0.06)	—	—
	ITT	I	108	19.69 (35.65)	4	8.06 (19.41)	−11.65 (−20.44 to −2.87)	−4.97 (−16.04 to 6.09)	.38
	FAM-Information[h]								
	ITT	C	105	14.30 (2.02)	42	14.15 (2.22)	−0.29 (−0.99 to 0.42)	—	—
	ITT	I	94	14.16 (2.07)	45	14.38 (2.19)	0.20 (−0.48 to 0.87)	0.48 (−0.49 to 1.46)	.33
	FAM-Care[i]								
	ITT	C	108	20.6 (3.92)	42	20.19 (3.98)	−0.14 (−1.10 to 0.82)	—	—
	ITT	I	99	20.42 (3.92)	45	20.57 (4.57)	0.35 (−0.92 to 1.62)	0.48 (−1.11 to 2.08)	.55
	FAM-knowledge[j]								
	ITT	C	105	10.7 (2.83)	41	10.66 (2.74)	−0.19 (−0.83 to 0.45)	—	—

Outcomes		Group	Baseline		7 months		Estimated change (95% CI)	Group-time interac-tion,(95% CI)	*P* value
			n	Mean (SD)	n	Mean (SD)			
	ITT	I	95	11.28 (2.44)	46	11.89 (2.44)	0.71 (−0.04 to 1.46)	0.90 (−0.09 to 1.88)	.07
LIMBO-Total									
	ITT	C	105	27.98 (5.59)	54	28.17 (5.34)	0.03 (−1.37 to 1.42)	—	—
	ITT	I	108	29.32 (4.81)	64	27.93 (5.63)	−1.58 (−2.95 to −0.21)	−1.61 (−3.56 to 0.35)	.11
Coordination-Total									
	ITT	C	97	13.89 (3.50)	53	13.75 (3.45)	−0.15 (−1.12 to 0.81)	—	—
	ITT	I	106	14.29 (3.46)	61	15.19 (3.12)	0.90 (0.04 to 1.75)	1.05 (−0.23 to 2.34)	.11

[a]ITT: intention-to-treat approach.

[b]C: control.

[c]Not applicable.

[d]I: intervention.

[e]AT: as-treated groups (AT1 and AT2).

[f]LIMBO: global feeling of left in limbo.

[g]FAM-Global: global support from general practitioner.

[h]FAM-Information: information from general practitioner subscale.

[i]FAM-Care: support from general practitioner subscale.

[j]FAM-knowledge: general practitioners' knowledge regarding treatment subscale.

In the context of the AT1 approach (Table 3), comparing patients who received the intervention with those who did not, the estimated within-group change in the primary outcome between baseline and follow-up was −0.10 (95% CI −0.33 to 0.12) in the control group and 0.17 (95% CI −0.08 to 0.42) in the AT1 group. The between-group difference was estimated as 0.27 (95% CI −0.07 to 0.61; *P*=.12).

For the AT2 approach (Table 3), the estimated within-group change in the primary outcome between baseline and follow-up was −0.09 (95% CI −0.33 to 0.14) in the control group and 0.10 (95% CI −0.15 to 0.34) in the AT2 group. The estimated between-group difference was 0.19 (95% CI −0.15 to 0.53) with the corresponding *P*=.27.

The estimated within-group and between-group changes in all secondary outcomes including 2 single items and 5 subscales showed no significant differences between the ITT groups from the beginning to 7-month follow-up (Table 3).

We also conducted the originally specified analyses on the primary outcomes at 7 months only, comparing intervention and control group in the ITT, AT1, and AT2 approach. The findings are presented in Table 4, showing similar *P* values.

Table . Additional analyses on primary outcomes at 7 months.

	Difference of means at 7 months (IV[a] minus control) (95% CI)	*P* value for *t* test	*P* value for Wilcoxon rank-sum test (exact)
ITT[b]	0.285 (−0.085 to 0.655)	.13	.11
AT1[c]	0.266 (−0.109 to 0.641)	.21	.21
AT2[d]	0.283 (−0.085 to 0.652)	.12	.12

[a]IV: intervention group.

[b]ITT: intention-to-treat approach.

[c]AT1: as-treated group$_1$.

[d]AT2: as-treated group$_2$.

Discussion

Principal Findings

This study found that the addition of a cross-sectoral video consultation to usual care did not significantly impact patients' perceived coordination of care. Both intervention and control groups showed high levels of perceived coordination at both time points, with no statistically significant differences over time or between the groups.

Comparison With Previous Work

Based on our knowledge, video consultation has been used for patients with cancer for many years [20]. A recently published systematic review showed virtual consultation over time has been developed and improved in many ways (eg, delivery platforms and stakeholder engagement) [21]. Despite this improvement, comparing our findings was challenging due to the lack of studies on patients' attitudes toward care coordination in multidisciplinary video consultations, which involve patients with cancer, oncologists, and GPs simultaneously.

A newly released scoping review revealed that specialist collaborations with GPs and patients can increase the effective quality of care in the follow-up phase for patients with cancer [22]. This comprehensive review did not report findings on patients' attitudes toward care coordination. Therefore, we believe that in this area, more studies should be initiated to capture a better picture of care coordination from patients' perspective and subsequently enhance the quality-of-care coordination for patients with cancer.

Limitations of the Study

Several limitations should be considered. First, a large percentage of participants in the intervention group did not proceed with the intervention, mostly due to GP refusal to participate, administrative or technical issues. This could affect the generalizability of our findings to the broader population or specific subgroups due to potential selection bias. However, it should be noted that the trial was carried out before the introduction of a standard, clinically available video setup during the COVID-19 pandemic. Since then, the technical aspects of video-based communication in everyday life and health care consultations have improved dramatically [23]. However, challenges related to establishing online meetings, achieving relevant views of all participants, and ensuring efficient sound quality persist. These issues can occasionally make scheduled consultations impracticable [24]. These facts highlight the relevance of the results of "the Partnership Study" and underscore the importance of our learnings for future health care, particularly in adapting the evolving landscape of telemedicine. Second, the low completion rate for video consultation (58%) and high rate of missing data in our study affected the quality of our data, consequently limited our ability to draw definitive conclusion on effectiveness of the intervention. As a result, we focused on addressing challenges encountered. Third, a considerable amount of missing data for the primary outcome at the 7-month time point might impact the statistical power and consequently lead to a lack of significance in our findings. Several factors can be contributed to this issue (eg, the lengthy follow-up period, focus of the process evaluation on the intervention rather than barriers to participants retention, inadequate assessment of the follow-up duration during the pilot test, reliance on survey distribution alone, particularly during COVID-19 pandemic). In addition, it is possible that a ceiling effect influenced the intervention's lack of superiority, as coordination scores were substantially higher, compared with findings from another department [14]. Fourth, the choice of single primary outcome variable that included "not relevant" response option may have constrained the depth of data obtained, because patients who had not experienced collaboration might have selected "not relevant." Fifth, in the as-treated analysis, some patients from the intervention group were combined with patients from the control group, which assumes that the nonreceipt of the intervention was unrelated to individual patients' characteristics. Therefore, our findings should be interpreted with caution. Sixth, the study does not provide insight into why some patients marked the primary outcome as "not relevant." It is possible that these patients had no experience with collaboration at the time of the survey, and this uncertainty limits our understanding of the factors contributing to missing data and how they may influence the study's conclusions. The handling of responses marked as "not relevant" as missing data may not actually capture patients' experiences, that raise concerns about the interpretation of the results (eg, generalizability of findings related to collaboration experiences).

Future Direction

Despite the lack of significant differences on primary and secondary outcomes of care coordination, limitations of our study may have implications for the research community in their future studies. For instance, our findings stress the need for further exploration into structural and engagement factors to strengthen future interventions.

The low completion rate for video consultation may indicate logistical and engagement challenges. Therefore, we encourage researchers in their future similar intervention to implement strategies that enhance patients' engagement and improve data completion rate.

We encountered challenges such as potential power problem in reaching statistical significance. Despite this, we believe it is crucial to delve deeper into these findings and explore underling factors. This could provide valuable insights for development of more effective interventions in future. Findings may also highlight more involvement of patients to address their concerns related to care coordination and consequently enhance their experiences with the health care system. Furthermore, the findings highlight the importance of continually evaluation of health interventions to understand the impacts over time and make timely and necessary improvement, particularly where we have clearly identified specific problems or challenges, like decreased patients' satisfaction.

Conclusion

Our study, conducted before the COVID-19 pandemic, found that the shared video consultation model, involving patients with cancer, oncologists, and GPs, did not result in a statistically

significant difference in patients' perceived coordination of care between the control and intervention groups. We suggest that technical issues impacting the intervention implementation and the potential ceiling effect may have contributed to these results. Therefore, we emphasize the necessity for additional evaluation of the conceptual notion of uniting patients with cancer, oncologists, and GPs, particularly considering the advancements in techniques, the adoption of virtual communication, and the expanding role of GPs in cancer care. Further exploration of specific aspects of care coordination may provide additional insights into areas for improvement in this innovative health care model. This study highlights the complexity of implementing collaborative health care interventions and emphasizes the importance of ongoing evaluation of the intervention to optimize patients' care coordination in cancer management. In addition, future research should focus on evolving trends in virtual communication among professionals and the public, as we think that leveraging post-COVID virtual communications could improve future health care interventions.

Acknowledgments

Financial support was provided by the Danish Cancer Society (R156-A10018-16-S49), the Region of Southern Denmark (14/39009), ML Jørgensen and GunnarHansen Foundation, and the University of Southern Denmark. The sponsors played no role in the planning, design, or implementation of the study. ChatGPT (version 3.5, OpenAI) was used to edit the first draft of the manuscript. However, the next versions were edited and revised by the authors.

Authors' Contributions

TBT and LHJ conceived the study, participated in study design and data collection, and revised the manuscript. JS, JJS, and DGH conceived the study, participated in study design, and revised the manuscript critically. MMS and SW analyzed the data, participated in the interpretation of the results, and revised the manuscript. FB made substantial contributions to the interpretation of the results and wrote the manuscript. All authors commented on and approved the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original sample size.
[DOCX File, 19 KB - cancer_v11i1e60158_app1.docx]

Multimedia Appendix 2
The consultation guide for general practitioners and oncologists, including themes potentially relevant for the consultation.
[DOCX File, 21 KB - cancer_v11i1e60158_app2.docx]

Multimedia Appendix 3
Overview of primary and secondary outcomes.
[DOCX File, 20 KB - cancer_v11i1e60158_app3.docx]

Multimedia Appendix 4
Overview of general practitioner–patient relation.
[DOCX File, 19 KB - cancer_v11i1e60158_app4.docx]

Checklist 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File, 1100 KB - cancer_v11i1e60158_app5.pdf]

References

1. Stoumpos AI, Kitsios F, Talias MA. Digital transformation in healthcare: technology acceptance and its applications. Int J Environ Res Public Health 2023 Feb 15;20(4):3407. [doi: 10.3390/ijerph20043407] [Medline: 36834105]
2. Chan RJ, Crichton M, Crawford-Williams F, et al. The efficacy, challenges, and facilitators of telemedicine in post-treatment cancer survivorship care: an overview of systematic reviews. Ann Oncol 2021 Dec;32(12):1552-1570. [doi: 10.1016/j.annonc.2021.09.001] [Medline: 34509615]
3. Duthie K, Strohschein FJ, Loiselle CG. Living with cancer and other chronic conditions: patients' perceptions of their healthcare experience. Can Oncol Nurs J 2017;27(1):43-48. [doi: 10.5737/2368807627143348] [Medline: 31148776]

4. Halpern MT, Cohen J, Lines LM, Mollica MA, Kent EE. Associations between shared care and patient experiences among older cancer survivors. J Cancer Surviv 2021 Apr;15(2):333-343. [doi: 10.1007/s11764-020-00934-3] [Medline: 32948992]
5. Badr H, Carmack CL, Diefenbach MA. Psychosocial interventions for patients and caregivers in the age of new communication technologies: opportunities and challenges in cancer care. J Health Commun 2015;20(3):328-342. [doi: 10.1080/10810730.2014.965369] [Medline: 25629218]
6. Hadian M, Jelodar ZK, Khanbebin MJ, Atafimanesh P, Asiabar AS, Dehagani SMH. Challenges of implementing telemedicine technology: a systematized review. Int J Prev Med 2024;15:8. [doi: 10.4103/ijpvm.ijpvm_48_23] [Medline: 38563036]
7. Trabjerg TB, Jensen LH, Sondergaard J, Wehberg S, Sisler JJ, Hansen DG. Cross-sectoral video consultation in cancer care: GPs' evaluation of a randomised controlled trial. BJGP Open 2021 Apr;5(2):BJGPO.2020.0114. [doi: 10.3399/BJGPO.2020.0114] [Medline: 33293407]
8. Larrabee Sonderlund A, Quirina Bang Van Sas T, Wehberg S, et al. Development of a video consultation patient-satisfaction questionnaire (vCare-PSQ): a cross-sectional explorative study. JMIR Form Res 2024 Aug 2;8:e58928. [doi: 10.2196/58928] [Medline: 39094110]
9. Elkefi S, Asan O. The impact of patient-centered care on cancer patients' QOC, self-efficacy, and trust towards doctors: analysis of a national survey. J Patient Exp 2023;10:23743735231151533. [doi: 10.1177/23743735231151533] [Medline: 36698621]
10. Gorin SS, Haggstrom D, Han PKJ, Fairfield KM, Krebs P, Clauser SB. Cancer care coordination: a systematic review and meta-analysis of over 30 years of empirical studies. Ann Behav Med 2017 Aug;51(4):532-546. [doi: 10.1007/s12160-017-9876-2] [Medline: 28685390]
11. Trabjerg TB, Jensen LH, Søndergaard J, Sisler JJ, Hansen DG. Improving continuity by bringing the cancer patient, general practitioner and oncologist together in a shared video-based consultation - protocol for a randomised controlled trial. BMC Fam Pract 2019 Jun 25;20(1):86. [doi: 10.1186/s12875-019-0978-8] [Medline: 31238886]
12. Trabjerg TB, Jensen LH, Søndergaard J, Sisler JJ, Hansen DG. Cross-sectoral video consultations in cancer care: perspectives of cancer patients, oncologists and general practitioners. Support Care Cancer 2021 Jan;29(1):107-116. [doi: 10.1007/s00520-020-05467-0] [Medline: 32318872]
13. Trabjerg TB, Jensen LH, Sondergaard J, Trabjerg ND, Sisler JJ, Hansen DG. Investigating whether shared video-based consultations with patients, oncologists, and GPs can benefit patient-centred cancer care: a qualitative study. BJGP Open 2020 Jun;4(2):bjgpopen20X101023. [doi: 10.3399/bjgpopen20X101023] [Medline: 32238390]
14. Nielsen JD, Palshof T, Mainz J, Jensen AB, Olesen F. Randomised controlled trial of a shared care programme for newly referred cancer patients: bridging the gap between general practice and hospital. Qual Saf Health Care 2003 Aug;12(4):263-272. [doi: 10.1136/qhc.12.4.263] [Medline: 12897359]
15. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
16. Baker R, Preston C, Cheater F, Hearnshaw H. Measuring patients' attitudes to care across the primary/secondary interface: the development of the patient career diary. Qual Health Care 1999 Sep;8(3):154-160. [doi: 10.1136/qshc.8.3.154] [Medline: 10847871]
17. Nielsen JD, Palshof T, Olesen F. Cross-sectorial cooperation regarding cancer patients in a recently started care program. Ideas and themes based on focus group interviews with general practitioners and oncologists. Ugeskr Laeger 1999 Apr 5;161(14):2074-2078. [Medline: 10354793]
18. What's new in STATA 17. Stata. 2021. URL: https://www.stata.com/stata17/ [accessed 2025-02-03]
19. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332. [doi: 10.1136/bmj.c332] [Medline: 20332509]
20. Fleissig A, Jenkins V, Catt S, Fallowfield L. Multidisciplinary teams in cancer care: are they effective in the UK? Lancet Oncol 2006 Nov;7(11):935-943. [doi: 10.1016/S1470-2045(06)70940-8] [Medline: 17081919]
21. Singh S, Fletcher GG, Yao X, Sussman J. Virtual care in patients with cancer: a systematic review. Curr Oncol 2021 Sep 8;28(5):3488-3506. [doi: 10.3390/curroncol28050301] [Medline: 34590602]
22. Nicolaisen A, Lauridsen GB, Haastrup P, Hansen DG, Jarbøl DE. Healthcare practices that increase the quality of care in cancer trajectories from a general practice perspective: a scoping review. Scand J Prim Health Care 2022 Mar;40(1):11-28. [doi: 10.1080/02813432.2022.2036421] [Medline: 35254205]
23. Wherton J, Shaw S, Papoutsi C, Seuren L, Greenhalgh T. Guidance on the introduction and use of video consultations during COVID-19: important lessons from qualitative research. BMJ Lead 2020 Sep;4(3):120-123. [doi: 10.1136/leader-2020-000262]
24. Warmoth K, Lynch J, Darlington N, Bunn F, Goodman C. Using video consultation technology between care homes and health and social care professionals: a scoping review and interview study during COVID-19 pandemic. Age Ageing 2022 Feb 2;51(2):afab279. [doi: 10.1093/ageing/afab279]

Abbreviations:

AT1: as-treated group$_1$
AT2: as-treated group$_2$
CONSORT: Consolidated Standards of Reporting Trials
GEE: generalized estimating equations
GP: general practitioner
ITT: intention-to-treat
PSP: Partnership Project
REDCap: Research Electronic Data Capture

Edited by N Cahill; submitted 03.05.24; peer-reviewed by A Poudel, GL Mazza; revised version received 17.12.24; accepted 16.01.25; published 11.02.25.

Please cite as:
Baygi F, Bitz Trabjerg T, Jensen LH, Munch Storsveen M, Wehberg S, Sisler JJ, Søndergaard J, Gilså Hansen D
Challenges of Cross-Sectoral Video Consultation in Cancer Care on Patients' Perceived Coordination: Randomized Controlled Trial
JMIR Cancer 2025;11:e60158
URL: https://cancer.jmir.org/2025/1/e60158
doi:10.2196/60158

Assessing Health Information Seeking Behaviors Among Targeted Social Media Users Using an Infotainment Video About a Cancer Clinical Trial: Population-Based Descriptive Study

Jonathan Sommers[1*], BA; Don S Dizon[2*], MD; Mark A Lewis[3*], MD; Erik Stone[1*], BA; Richard Andreoli[1*], BA; Vida Henderson[4*], PhD, PharmD, MPH, MFA

[1]Digital Health Networks, Los Angeles, CA, United States

[2]Brown University, Providence, RI, United States

[3]Intermountain Health, Murray, UT, United States

[4]Fred Hutchinson Cancer Center, University of Washington, 1100 Fairview Ave N, Seattle, WA, United States

[*]all authors contributed equally

Corresponding Author:
Vida Henderson, PhD, PharmD, MPH, MFA
Fred Hutchinson Cancer Center, University of Washington, 1100 Fairview Ave N, Seattle, WA, United States

Abstract

Background: The lack of information and awareness about clinical trials, as well as misconceptions about them, are major barriers to cancer clinical trial participation. Digital and social media are dominant sources of health information and offer optimal opportunities to improve public medical awareness and education by providing accurate and trustworthy health information from reliable sources. Infotainment, material intended to both entertain and inform, is an effective strategy for engaging and educating audiences that can be easily disseminated using social media and may be a novel way to improve awareness of and recruitment in clinical trials.

Objective: The purpose of this study was to evaluate whether an infotainment video promoting a clinical trial, disseminated using social media, could drive health information seeking behaviors.

Methods: As part of a video series, we created an infotainment video focused on the promotion of a specific cancer clinical trial. We instituted a dissemination and marketing process on Facebook to measure video engagement and health information seeking behaviors among targeted audiences who expressed interest in breast cancer research and organizations. To evaluate video engagement, we measured reach, retention, outbound clicks, and outbound click-through rate. Frequencies and descriptive statistics were used to summarize each measure.

Results: The video substantially increased health information seeking behavior by increasing viewership from 1 visitor one month prior to launch to 414 outbound clicks from the video to the clinical trial web page during the 21-day social media campaign period.

Conclusions: Our study shows that digital and social media tools can be tailored for specific target audiences, are scalable, and can be disseminated at low cost, making it an accessible educational, recruitment, and retention strategy focused on improving the awareness of clinical trials.

Trial Registration: ClinicalTrials.gov NCT03418961; https://clinicaltrials.gov/study/NCT03418961

(JMIR Cancer 2025;11:e56098) doi:10.2196/56098

KEYWORDS

cancer clinical trials; digital media; social media; infotainment; recruitment; education and awareness; edutainment; public engagement; cancer; lack of information; social media; health information; medical awareness; video series; public audience; low cost; research participants

Introduction

A total of 90% of Americans use social media [1] and over 40% of Americans watch web-based videos daily [2]. Digital media is a dominant source of health information, with 50% to 80% of internet users searching for health information on the web [3-5]. Unfortunately, misinformation and disinformation threaten the quality of health information available [6,7]. Given public interest in accessing web-based health information and the potential public reach, social media offers an optimal opportunity for public health practitioners and health care providers to improve public medical awareness and education

by providing accurate and trustworthy sources of health information. However, the attention span of social media audiences is typically short. Thus, digital content has to be engaging, meaningful, and attention grabbing [8,9].

One specific area of health care that can benefit from such strategies is cancer clinical trials. The lack of information and awareness about clinical trials, coupled with significant misconceptions about them, persist as major barriers to cancer clinical trial participation [10]. Only 3% to 5% of patients with cancer participate in clinical trials and over 50% of Americans report that the lack of awareness and information are major reasons for low participation rates [11-13]. This is the case despite one study reporting that 56% of respondents preferred the internet as a source of information about clinical trials [11].

Infotainment is defined as the delivery of broadcast material intended to entertain, engage, and inform [14]. Examples of this form of media in our culture vary but can include news talk shows, podcasts, and social media influencers. Its use may be a novel way to improve awareness and alleviate misconceptions that act as barriers to clinical trials. Moreover, since many trials close due to the lack of accrual [15], infotainment delivered on social media may be a viable clinical trial recruitment strategy.

The purpose of this study was to evaluate whether a clinical trial promotional video, disseminated on social media, could effectively drive health information seeking behaviors and, as a result, increase awareness about the clinical trial. To accomplish this goal, we instituted a dissemination and marketing process to measure health information seeking behaviors among public audiences on social media.

Methods

Video Development and Content

A clinical trial promotional video represents an underexplored method for engaging with targeted audiences. The trial we chose to feature was Southwest Oncology Group (SWOG) S1501, “Prospective Evaluation of Carvedilol in Prevention of Cardiac Toxicity in Patients with Metastatic HER-2+ Breast Cancer, Phase III.” [16] Collaborating with a clinical trial recruitment and retention specialist from the SWOG, we selected this trial based on its relative ease of understanding and to improve accrual performance. Our research team collaborated with the SWOG Cancer Research Network, a global cancer research community that designs and conducts federally funded clinical trials; The Hope Foundation for Cancer Research, a public charity with the mission of raising and contributing funds for the treatment and prevention of cancer; and Digital Health Networks (DHN), a media production studio and streaming service that specializes in producing, marketing, and distributing health care content and collecting viewership data.

The creation of the video was informed by the World Health Organization (WHO) Strategic Communications Framework [17] and the social cognitive theory [18]. The WHO Strategic Communications Framework asserts that effective health communications should be accessible, actionable, credible and trusted, relevant, timely, and understandable [17]. The social cognitive theory asserts that learning occurs in social contexts where individuals can influence and be influenced by others and their environment [18]. Using the social cognitive theory framework in the context of the health information seeking behavior of social media users, Zhang et al [19] found that (1) information quality, social media platform quality, and user experience have a significant positive effect on emotional arousal; (2) user experience, social support, and emotional arousal have a significant positive effect on self-efficacy; and (3) emotional arousal and self-efficacy have a significant positive effect on social media users’ health information seeking behavior. To this end, we sought to develop high-quality, emotionally engaging digital media that delivers accurate information about a clinical trial, with an embedded URL hyperlink to learn more information from a credible source, on a highly used social media platform to increase awareness about the trial and motivate users to seek additional information.

To ensure the information, language, and imagery provided in the video was medically accurate, a project steering committee, who chose the clinical trial of focus, was formed, consisting of SWOG Digital Engagement committee members, study investigators, and a representative from The Hope Foundation for Cancer Research. The video concept, script, and storyline were initially developed by a research team member who is also a filmmaker; cancer survivor; and long-standing member of the SWOG’s Digital Engagement, Adolescent and Young Adult, and Patient Advocates committees. The outline, script, and questions were submitted to the project steering committee for feedback, and regular communication was maintained throughout the process in multiple feedback loops. Updates were provided during committee meetings and were presented at biannual SWOG group meetings with storyboards, a rough and final cut of the video, and initial project results. This inclusive process encouraged committee members and meeting attendees to provide feedback, ask questions, and be part of the creative filmmaking process. The steering committee viewed the video and gave the final sign off prior to launch.

To engage audiences, the video attempted to parody cliché tropes often seen in pharmaceutical commercials. In our fictional video, Beth, a more relatable portrayal of a woman who is visually worn down from cancer treatment, is watching television and flipping through channels as a highly stylized, cheery commercial for the S1501 study comes on. It is akin to pharmaceutical advertisements commonly broadcast on television. Beth, feeling the side effects of treatment, dozes off but is awakened when she, in disbelief, is interacting with the commercial’s characters and asking the physicians questions about clinical trials and specifically the S1501 study. In this narrative structure, Beth’s role is to serve as a vehicle to convey the pertinent information the study team hoped the audience would learn. In the end, Beth is brought back to her reality. Appearing cautiously interested in the trial, she presumably seeks out further information about it. Video screenshots showing the main character (Beth), a paradoxical patient with cancer, and a scene with clinicians explaining the clinical trial to Beth is provided in Figure 1. The full video is available in Multimedia Appendix 1.

Figure 1. Video screenshots showing (left) the main character (Beth), (middle) a paradoxical patient with cancer, and (right) a scene with clinicians explaining the clinical trial to Beth.

Ethical Considerations

The National Cancer Institute Cancer Prevention and Control Central Institutional Review Board reviewed the recruitment/patient education video (protocol version dated June 6, 2019; study ID: S1501) and granted approval on July 24, 2019. The expedited review was conducted in accordance with the federally defined categories of expedited review stated in 45 CFR 46.110(b)(1)(ii) and 21 CFR 56.110(b)(2). The dissemination strategy was not determined to be human subjects research.

The study protocol and study materials for video 2 were reviewed and approved by the National Cancer Institute Institutional Review Board. We are publishing the results from a marketing strategy to see how it could inform research. The data presented herein represent information collected for the purposes of social media marketing, and we did not collect individual-level data. Rather, we collected population-level data to inform the effectiveness of marketing research. Institution review board approval and informed consent were not obtained as no human subject data were collected on an individual level, no data were obtained through interaction with any individuals, and we cannot ascertain the identities of individuals to whom data the belonged. The data collection was conducted by a medical entertainment company, and instead of keeping the data internal, we are sharing this information with the wider medical audience.

Identifiable individuals have granted consent for the use of their image in this publication.

Dissemination and Marketing Methods

For this work, we disseminated the video on Facebook (Meta) because of its number of users and daily interactions, campaign optimizations, low cost of advertising, analytic capabilities, and sophisticated audience targeting that identifies specific audiences based on users' interests and previous interactions [20]. DHN ran and coordinated the marketing campaign using their Facebook advertising account, and video performance data were captured using Facebook's advertising management platform.

The video ran daily from September 7, 2021, to September 27, 2021. The video length was 5 minutes and 10 seconds long, and the production cost was US $4000. Projects with a similar method would cost significantly more to produce; however, costs were kept at a minimum by filming at three free locations and DHN's in-kind production services, which included directing, editing, color correction, music and graphic licenses, and voice-over narration. The video consisted of two different campaigns. The purpose of the first campaign was to identify the optimal target audience who might be more engaged with the video content. For the first campaign, which was conducted from September 7, 2021, to September 10, 2021, we used Facebook targeting algorithms to disseminate the video to audiences who expressed interests in breast cancer research and to audiences who expressed interest in breast cancer organizations. We compared the engagement (number of views and outbound clicks to the S1501 clinical trial web page) between each group. The audience interested in breast cancer research had greater engagement and was chosen as our target audience for the second campaign, which ran from September 10, 2021, to September 27, 2021. Our second campaign disseminated the video to individuals who met the look-alike audience profile, which is an audience who shared similar indicated interests, characteristics, and behaviors to individuals who were interested in breast cancer research. This technique helped us identify potentially more engaged users who would more likely be interested in our video; this is where we spent the majority of our US $1000 advertising campaign budget.

Measures and Analysis

To evaluate video engagement, we measured four outcomes:

- *Reach* was defined as the number of people exposed to the video during the advertisement campaign.
- *Retention* was defined as the length of time that individuals watched the video. We measured the number of individuals who watched for at least 3 seconds and at intervals of 25%, 50%, 75%, and 100%.
- *Outbound clicks* were defined as the number of clicks on a "learn more button" under the video that was linked to the SWOG Cancer Research Network's S1501 patient information web page.
- *Outbound click-through rate* was defined as the percentage of times viewers saw a video and performed an outbound click.

We calculated frequencies and descriptive statistics to summarize each outcome.

Results

Reach, retention, outbound clicks, and outbound click-through rate are summarized in Table 1.

Table . Educational video engagement outcomes for the S1501 study.

Outcomes	Value
Length	5 min 10 s
Video reach, n	61,456
Length of video watched, n (% retained)	
At least 3 seconds	47,566 (77.4)[a]
25%	226 (0.5)[b]
50%	84 (37.2)[c]
75%	44 (52.4)[d]
100%	34 (77.3)[e]
Outbound clicks, n	414
Percentage increase in visitors	41,300
Outbound click-through rate, n/N (%)	414/61,456 (0.67)

[a]N=61,456.
[b]n=47,566.
[c]n=266.
[d]n=84.
[e]n=44.

In our first campaign, we found that the group interested in breast cancer research performed better with video views and outbound clicks than the group interested in breast cancer organizations (9764 views and 54 outbound clicks vs 2513 views and 25 outbound clicks). This informed our target look-alike audience for the second campaign. The video had a total reach of 61,456 individuals. A total of 77.4% (47,566/61,456) watched at least 3 seconds and among those, 0.5% (226/47,566) watched 25% of the video. The number of viewers dropped at each consecutive retention interval; however, the retention rate increased at each consecutive retention interval past 25%. A month prior to launch, the S1501 patient information clinical trial web page had only 1 visitor. During the 21-day dissemination campaign period, the video received 414 outbound clicks from Facebook to the patient-facing clinical trial page. The outbound click-through rate was 0.67% (414/61,456).

Discussion

Principal Findings

We found that with an active digital marketing dissemination strategy and a very modest marketing budget (US $1000), the video substantially increased health information seeking behavior by increasing visitors to the SWOG's S1501 patient information page from only 1 web page visit the month prior to campaign initiation to 414 web page visits during the campaign. Although research organizations have passively disseminated clinical trial–related content on social media platforms, this very active and intentional campaign to engage with the public is promising as it relates to clinical trial recruitment and illustrates the utility of using engaging digital content coupled with social media marketing as an effective strategy for clinical trial recruitment. Additionally, our study shows that digital and social media tools can be disseminated to specific target audiences at low cost, making it an accessible educational, recruitment, and retention strategy for clinical trials.

Our study shows that the health information seeking behavior of social media users may be impacted by immediate attrition. The average social media video watch time benchmark is 10 seconds, and Facebook's best practices suggest that video advertisements be 15 seconds or less [9]. Considering that the average Facebook video watch time is 4.6 seconds [21], our video (5 min 10 s), which only retained 0.5% of 3-second viewers, illustrates the challenge of delivering important information in a succinct way. In addition to video length and the lack of engagement, another factor that may have contributed to this decline is that although we used Facebook algorithms to target individuals who were interested in breast cancer research and breast cancer organizations, it is plausible that only a subset of these individuals were interested in learning about clinical trials. Interestingly, viewership retention across other intervals was drastically higher (37.2% retention at the 50% interval, 52.4% retention at the 75% interval, and 77.3% retention at the 100% interval). Viewers who watched 25% of the video (1 min 17 s) tended to stay for the duration of the video, suggesting an engaged audience. Although a drop-off in retention is expected, our data suggest that the first 45-60 seconds may be the most instrumental in capturing audiences and maximizing engagement among viewers who are interested in the video topic area. It is noteworthy that the number of outbound clicks (n=414) was substantially higher than the number of people who watched the full video (n=34). This indicates that certain individuals were compelled to seek further health information before they finished watching the video, although we do not know when during the duration of the video that they performed an outbound click. These data show that *completely watching* the video was not necessarily correlated with seeking additional information

(ie, outbound clicks) and further illustrates the importance of delivering compelling and engaging content within the first 60 seconds of video content. Finally, the outbound click-through rate (0.67%) was below the average outbound click-through rates of 0.89% across industries and 0.83% in health care, but above the 0.45% average in science [22,23], which may also provide insight into the level of engagement in the video.

In this burgeoning and ever-evolving social media landscape, digital marketing and its advanced analytical techniques allow researchers to engage with patients and potential trial participants where they are already consuming health information. It also allows for real-time data on what worked well and what did not work well, so that researchers can pivot their tact and capitalize on the most effective strategies to engage with an audience that is more likely to be interested in or eligible for their trials. Although this medium is novel in oncology, social media analytics have proven to be successful in informing engagement strategies for other industries, and future oncology research could benefit from these digital tools.

Comparison to Prior Work

Our research is aligned with limited previous studies that found web-based infotainment videos to be an effective approach in increasing public understanding about science and health care among web-based health information seekers [14,24,25]. One study found that narrative infotainment videos compared to expository videos resulted in more likes by viewers without a university education and better information recall among viewers [14]. Study findings that assess study recruitment with and cost effectiveness of social media advertisement for study enrollment are mixed [26-29]. A recent scoping review reported that 9 (27%) out of 33 studies that used both social media and traditional methods for recruitment to clinical trials achieved or exceeded their enrollment target, and one study reported that social media outperformed other recruitment methods. The review did not specify how many of these studies used video as a recruitment method, highlighting the need for more research to understand if infotainment videos disseminated on social media to target audiences may be a significant strategy for facilitating clinical trial recruitment [24,30].

Limitations

Our study is limited in that we were not the administrators of the S1501 study's public facing web page and did not retain detailed information about visitor activities or characteristics (eg, location, frequency, and time of visit). This limited our ability to ascertain if each outbound click was from a unique or repeat visitor and our ability to understand additional characteristics that may have impacted health information seeking behaviors. We recommend that research teams be the administrators of both outgoing (where the advertisement played) and incoming (where visitors were sent) visitor analytics to maximize data utility. Additionally, our study did not test whether outbound clicks resulted in actual recruitment; therefore, we cannot make assertions that the increase in web page visitors contributed to study recruitment [31]. In addition, our video did not include subtitles in English nor any other language. Since 85% of Facebook users interact with the platform with the sound off [32], future digital media engagement studies should consider the use of language subtitles as this feature might facilitate engagement and accessibility with a larger audience. Another limitation is that our target audience indicators may not have been the most appropriate indicators to capture audiences interested in a specific clinical trial. We cannot ascertain that outbound clicks reached potential participants, as opposed to other individuals who may have been interested in clinical trials (eg, researchers and providers). More specialized or specific targeting (eg, those meeting study criteria) should be used to capture audiences who would most benefit from and be engaged by video media for specific clinical trials. Finally, we tested engagement using only one social media platform. Social media platforms vary in the average length of videos and average video watch times. It is plausible that social media platforms that are more conducive to longer viewing times (eg, YouTube) may have yielded different viewership and engagement outcomes.

Future Directions

Emerging internet technologies and social media are widely used sources for health information. Our study found that infotainment disseminated using social media is a useful and effective approach in relaying complex health information, motivating interested viewers to seek additional health information, and driving public audiences to credible and reliable sources of information. It has promising utility in facilitating recruitment and retention strategies for cancer clinical trials and generating increased awareness about clinical trials among patients and the general public.

Acknowledgments

We would like to acknowledge the Southwest Oncology Group (SWOG) Cancer Research Network, The Hope Foundation for Cancer Research, Digital Health Networks, and Dr Justin Floyd and Dr Monika Leja for supporting this work. This research was supported by funding from The Hope Foundation for Cancer Research and SWOG Cancer Research Network National Institutes of Health/National Cancer Institute (NCI) award U10CA180888 and was also supported by Digital Health Networks. VH is supported by the NCI (K01CA248852, P30CA015704, UGCA286954, and U54CA132381), the American Cancer Society, and the Andy Hill Cancer Research Endowment Fund. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders. JS is a filmmaker, cancer survivor, and patient advocate who works at Digital Health Networks. He is a long-standing member of the SWOG's Digital Engagement, Adolescent and Young Adult, and Patient Advocates committees.

Data Availability

Data generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

JS and VH wrote the main manuscript text and prepared the table and figure. JS, VH, RA, DSD, and MAL contributed to the interpretation of data and the writing and editing of the manuscript. JS, DSD, ES, and RA contributed to study development and study design. JS, RA, ES, and DSD contributed to the acquisition of data, data analysis, video development, and editing of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Southwest Oncology Group S1501 video.
[MP4 File, 151291 KB - cancer_v11i1e56098_app1.mp4]

References

1. Ruby D. Social media users 2023 — (global demographics). DemandSage. 2023 Jun 7. URL: https://www.demandsage.com/social-media-users/#:~:text=USA%2DSpecific%20Social%20Media%20Statistics&text=The%20USA%20has%20302.35%20million,74.2%25%20of%20adults%20using%20it [accessed 2023-06-07]
2. Online video usage in the United States - statistics & facts. Statista. 2023. URL: https://www.statista.com/topics/1137/online-video/#topicOverview [accessed 2023-09-08]
3. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520. [Medline: 25083128]
4. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. Online health information seeking among us adults: measuring progress toward a Healthy People 2020 objective. Public Health Rep 2019;134(6):617-625. [doi: 10.1177/0033354919874074] [Medline: 31513756]
5. Jia X, Pang Y, Liu LS. Online health information seeking behavior: a systematic review. Healthcare (Basel) 2021 Dec 16;9(12):1740. [doi: 10.3390/healthcare9121740] [Medline: 34946466]
6. Haslam K, Doucette H, Hachey S, et al. YouTube videos as health decision aids for the public: an integrative review. Can J Dent Hyg 2019 Feb 1;53(1):53-66. [Medline: 33240342]
7. Stimpson JP, Park S, Pruitt SL, Ortega AN. Variation in trust in cancer information sources by perceptions of social media health mis- and disinformation and by race and ethnicity among adults in the United States: cross-sectional study. JMIR Cancer 2024 May 8;10:e54162. [doi: 10.2196/54162] [Medline: 38717800]
8. Adam M, Gates J. The rise of “health entertainment” to convey lifesaving messages in the COVID-19 pandemic. Scientific American. 2020 May 1. URL: https://blogs.scientificamerican.com/observations/the-rise-of-health-entertainment-to-convey-lifesaving-messages-in-the-covid-19-pandemic [accessed 2023-09-08]
9. Best practices for mobile video ads. Meta. 2023. URL: https://www.facebook.com/business/help/144240239372256?id=603833089963720 [accessed 2023-09-06]
10. Hillyer GC, Beauchemin M, Garcia P, et al. Readability of cancer clinical trials websites. Cancer Control 2020;27(1):1073274819901125. [doi: 10.1177/1073274819901125] [Medline: 31973569]
11. America Speaks! volume 18. ResearchAmerica!. 2018. URL: https://www.researchamerica.org/wp-content/uploads/2022/09/RSAM_1801_PollDataSummary_FINALWEB_0.pdf [accessed 2023-09-08]
12. Unger JM, Vaidya R, Hershman DL, Minasian LM, Fleury ME. Systematic review and meta-analysis of the magnitude of structural, clinical, and physician and patient barriers to cancer clinical trial participation. J Natl Cancer Inst 2019 Mar 1;111(3):245-255. [doi: 10.1093/jnci/djy221] [Medline: 30856272]
13. Unger JM, Cook E, Tai E, Bleyer A. The role of clinical trial participation in cancer research: barriers, evidence, and strategies. Am Soc Clin Oncol Educ Book 2016;35:185-198. [doi: 10.1200/EDBK_156686] [Medline: 27249699]
14. Davis LS, León B, Bourk MJ, Finkler W. Transformation of the media landscape: infotainment versus expository narrations for communicating science in online videos. Public Underst Sci 2020 Oct;29(7):688-701. [doi: 10.1177/0963662520945136] [Medline: 32729396]
15. Barriers to patient enrollment in therapeutic clinical trials for cancer. American Cancer Society Cancer Action Network. 2018 Apr 11. URL: https://www.fightcancer.org/policy-resources/barriers-patient-enrollment-therapeutic-clinical-trials-cancer [accessed 2025-02-21]
16. S1501 carvedilol in preventing cardiac toxicity in patients with metastatic HER-2-positive breast cancer. ClinicalTrials.gov. 2024. URL: https://clinicaltrials.gov/study/NCT03418961 [accessed 2025-02-25]

17. WHO strategic communications framework for effective communications. World Health Organization. 2017. URL: https://www.who.int/docs/default-source/documents/communicating-for-health/communication-framework.pdf [accessed 2023-08-16]
18. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
19. Zhang X, Chen B, Li G, Dong Y. Exploring the health information seeking behavior of social media users under the background of COVID-19 pandemic: an empirical study based on social cognitive theory. Front Psychol 2022 Nov 8;13:1045914. [doi: 10.3389/fpsyg.2022.1045914] [Medline: 36425825]
20. Social media & clinical trial recruitment. PatientCentra. 2017. URL: https://www.patientcentra.com/patient-recruitment-insights/clinical-trials-awareness-problem
21. Social media trends report. Socialbakers. 2019. URL: https://emplifi.io/ [accessed 2023-09-08]
22. Irvine M. Facebook ad benchmarks for your industry. WordStream Blog. 2019. URL: https://www.wordstream.com/blog/ws/2019/11/12/facebook-ad-benchmarks [accessed 2019-08-19]
23. Irvine M. Facebook ad benchmarks for your industry [data]. WordStream Blog. 2017. URL: https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks
24. Nolan TS, Bell AM, Chan YN, Leak Bryant A, Bissram JS, Hirschey R. Use of video education interventions to increase racial and ethnic diversity in cancer clinical trials: a systematic review. Worldviews Evid Based Nurs 2021 Oct;18(5):302-309. [doi: 10.1111/wvn.12539] [Medline: 34561957]
25. Riley AH, Sangalang A, Critchlow E, Brown N, Mitra R, Campos Nesme B. Entertainment-education campaigns and COVID-19: how three global organizations adapted the health communication strategy for pandemic response and takeaways for the future. Health Commun 2021 Jan;36(1):42-49. [doi: 10.1080/10410236.2020.1847451] [Medline: 33225758]
26. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Nov 7;18(11):e286. [doi: 10.2196/jmir.5698] [Medline: 27821383]
27. Guillory J, Wiant KF, Farrelly M, et al. Recruiting hard-to-reach populations for survey research: using Facebook and Instagram advertisements and in-person intercept in LGBT bars and nightclubs to recruit LGBT young adults. J Med Internet Res 2018 Jun 18;20(6):e197. [doi: 10.2196/jmir.9461] [Medline: 29914861]
28. Admon L, Haefner JK, Kolenic GE, Chang T, Davis MM, Moniz MH. Recruiting pregnant patients for survey research: a head to head comparison of social media-based versus clinic-based approaches. J Med Internet Res 2016 Dec 21;18(12):e326. [doi: 10.2196/jmir.6593] [Medline: 28003174]
29. Kayrouz R, Dear BF, Karin E, Titov N. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. Internet Interv 2016 Jan 6;4:1-10. [doi: 10.1016/j.invent.2016.01.001] [Medline: 30135786]
30. Darmawan I, Bakker C, Brockman TA, Patten CA, Eder M. The role of social media in enhancing clinical trial recruitment: scoping review. J Med Internet Res 2020 Oct 26;22(10):e22810. [doi: 10.2196/22810] [Medline: 33104015]
31. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. JMIR Res Protoc 2016 Aug 10;5(3):e161. [doi: 10.2196/resprot.5747] [Medline: 27511829]
32. Patel S. 85 percent of facebook video is watched without sound. Digiday. 2016 May 17. URL: https://digiday.com/media/silent-world-facebook-video [accessed 2023-08-18]

Abbreviations

DHN: Digital Health Networks
SWOG: Southwest Oncology Group
WHO: World Health Organization

Edited by N Cahill; submitted 05.01.24; peer-reviewed by G Pearson, R Ma; revised version received 31.01.25; accepted 13.02.25; published 03.03.25.

Please cite as:
Sommers J, Dizon DS, Lewis MA, Stone E, Andreoli R, Henderson V
Assessing Health Information Seeking Behaviors Among Targeted Social Media Users Using an Infotainment Video About a Cancer Clinical Trial: Population-Based Descriptive Study
JMIR Cancer 2025;11:e56098
URL: https://cancer.jmir.org/2025/1/e56098
doi:10.2196/56098

Evaluation of Douyin Short Videos on Mammography in China: Quality and Reliability Analysis

Hongwu Yang*, MM; Chuangying Zhu*, BS; Chunyan Zhou*, BS; Ruibin Huang, MD; Lipeng Huang, MM; Peifen Chen, MM; Shanshan Zhu, BS; Huanpeng Wang, MM; Chunmin Zhu, MM

Department of Radiology of First Affiliated Hospital, Shantou University Medical College, Shantou, China

*these authors contributed equally

Corresponding Author:
Chunmin Zhu, MM
Department of Radiology of First Affiliated Hospital, Shantou University Medical College, Shantou, China

Abstract

Background: Breast cancer is the most common malignant tumor and the fifth leading cause of cancer death worldwide, imposing a significant disease burden in China. Mammography is a key method for breast cancer screening, particularly for early diagnosis. Douyin, a popular social media platform, is increasingly used for sharing health information, but the quality and reliability of mammography-related videos remain unexamined.

Objective: This study aimed to evaluate the information quality and reliability of mammography videos on Douyin.

Methods: In October 2023, a search using the Chinese keywords for “mammography” and “mammography screening” was conducted on Douyin. From 200 retrieved videos, 136 mammography-related videos were selected for analysis. Basic video information, content, and sources were extracted. Video content was assessed for comprehensiveness across 7 categories: conception, examination process, applicable objects, precautions, combined examinations, advantages, and report. Completeness was evaluated using a researcher-developed checklist, while reliability and quality were measured using 2 modified DISCERN (mDISCERN) tool and the Global Quality Score (GQS). Correlations between video quality and characteristics were also examined.

Results: Among the video sources, 82.4% (112/136) were attributed to health professionals, and 17.6% (24/136) were attributed to nonprofessionals. Among health professionals, only 1 was a radiologist. Overall, 77.2% (105/136) of the videos had useful information about mammography. Among the useful videos, the advantages of mammography were the most frequently covered topic (53/105, 50.5%). Median values for the mDISCERN and GQS evaluations across all videos stood at 2.5 (IQR 1.63 - 3) and 2 (IQR 1 - 2), respectively. Within the subgroup assessment, the median mDISCERN score among the useful and professional groups stood at 2 (IQR 2 - 3) and 3 (IQR 2 - 3), respectively, surpassing the corresponding score for the unhelpful and nonprofessional groups at 0 (IQR 0 - 0) and 0 (IQR 0 - 0.75; $P<.001$). Likewise, the median GQS among the useful and professional groups was evaluated at 2 (IQR 1.5 - 2) and 2 (IQR 1 - 2), respectively, eclipsing that of the unhelpful and nonprofessional groups at 1 (IQR 1 - 1) and 1 (IQR 1 - 1.37; $P<.001$). The GQS was weak and negatively correlated with the number of likes ($r=-0.24$; $P=.004$), comments ($r=-0.29$; $P<.001$), and saves ($r=-0.20$; $P=.02$). The mDISCERN score was weak and negatively correlated with the number of likes ($r=-0.26$; $P=.002$), comments ($r=-0.36$; $P<.001$), saves ($r=-0.22$; $P=.009$), and shares ($r=-0.18$; $P=.03$).

Conclusions: The overall quality of mammography videos on Douyin is suboptimal, with most content uploaded by clinicians rather than radiologists. Radiologists should be encouraged to create accurate and informative videos to better educate patients. As Douyin grows as a health information platform, stricter publishing standards are needed to enhance the quality of medical content.

(JMIR Cancer 2025;11:e59483) doi:10.2196/59483

KEYWORDS

breast cancer; mammography; Douyin; information quality; social media; video; DISCERN; Global Quality Score; web-based education; cancer screening; health information; medical content

Introduction

Breast cancer is the second most common cancer and the fourth leading cause of cancer death worldwide. In 2022, an estimated 2.3 million new cases (11.6% of all cancer cases) were diagnosed, and 666,000 deaths (6.9% of all cancer deaths) occurred, and the number of new breast cancer cases is projected to reach 4.4 million by 2070 [1]. Among women, breast cancer is the most commonly diagnosed cancer, and it is the leading

cause of cancer deaths globally. In 2022, breast cancer accounted for approximately 15.4% of all deaths in global female patients and 6.9% of all cancer deaths [2]. As the second most common cancer in Chinese women, an estimated number of 357,200 new cases of breast cancer occurred in 2022, accounting for approximately 7.4% of total new cancer cases in China and 15.5% of global breast cancer cases [3]. The World Health Organization recently launched the Global Breast Cancer Initiative with the aim of reducing breast cancer mortality by fostering timely diagnosis and adequate treatment and patient management [4]. The 5-year survival rate in patients with early breast cancer is very high; thus, early screening, detection, and treatment are important [5]. Measures used for breast cancer screening in the "Guidelines for breast cancer diagnosis and treatment by China Anti-Cancer Association (2024 edition)" include mammography, ultrasonography, clinical breast examination, breast self-examination, and magnetic resonance imaging. Guidelines recommend that the starting age for breast cancer screening in the general risk population is 40 years. However, for people at high risk of breast cancer, the start of screening may be earlier than the age of 40 years. For those older than 70 years of age may consider opportunistic screening [6]. Mammography is one of the most effective methods for breast cancer screening, especially for early breast cancer diagnosis, and it has been a major contributor to the decline in breast cancer mortality rates [7-9]. At present, there is no nationwide screening program for breast cancer in China. A cross-sectional survey conducted with a convenience sample of 494 Chinese women indicated that participation in screening practices ranged from 27.5% for BSE, 36.4% for clinical breast examination, 23.5% for mammography, and 40% for ultrasonography [10].

With the widespread adoption of internet technology, web-based platforms have become a primary channel for accessing public information. As of June 2023, China's internet user base has expanded to 1.079 billion, with short video users reaching 1.026 billion, representing 95.2% of the total internet population [11]. The short video format has emerged as a dominant force in the new media landscape, due to its low barrier to entry, concise format, and rapid dissemination capabilities, making it one of the most preferred mediums for health information acquisition. While TikTok stands as a global social media giant, operating in over 160 countries with more than 1 billion monthly active users [12], its services are unavailable in China due to internet regulations. Instead, Douyin (the Chinese equivalent of TikTok, literally meaning "shaking sound") has established itself as a national phenomenon, boasting over 750 million daily active users and ranking among the country's most popular applications [13]. The platform's influence on health communication is particularly noteworthy. The Douyin Health Science Data Report indicates that daily health science content reaches more than 200 million users as of March 2023 [14]. This trend is further supported by data from the 2023 Douyin Health Lifestyle New Paradigm White Paper, which reveals that during the first half of 2023, the platform hosted more than 10 million creators specializing in health care knowledge content [15]. Notably, industry reports highlight that 42% of Douyin's user base comprises individuals aged 40 years and older [16], suggesting a significant engagement of mature audiences with health-related content on the platform.

Mammography screening often evokes feelings of anxiety and discomfort among patients, prompting many to seek preparatory information and clarification through social media platforms. High-quality educational videos can serve as valuable resources in this context, potentially contributing to improved health outcomes. Research evidence underscores the effectiveness of video interventions in promoting mammography screening. A study focusing on Chinese immigrant women demonstrated that culturally adapted videos, developed based on the health belief model, significantly enhanced screening intentions, breast cancer knowledge, risk perception, and understanding of mammography benefits [17]. Furthermore, empirical evidence indicates that brief preprocedure video interventions can substantially increase both physician referrals for screening mammography and patient compliance with screening completion [18]. These findings highlight the potential of video-based educational tools in addressing patient concerns and facilitating informed decision-making regarding breast cancer screening.

Despite the growing reliance on social media for health information, significant challenges persist regarding the reliability and accuracy of such content. The diverse backgrounds of content creators and viewers, coupled with the absence of robust verification mechanisms, make it difficult to assess the quality and credibility of health-related information on these platforms [19]. A comprehensive systematic review of reviews revealed that the prevalence of health misinformation on social media ranges from 0.2% to 28.8% [20], posing substantial risks to users. Exposure to inaccurate health information through videos may lead to severe consequences, including delays in seeking appropriate care or even life-threatening situations [21,22]. Previous research has extensively evaluated the quality of health-related content on traditional video-sharing platforms like YouTube and TikTok, covering various medical topics such as cervical spondylosis, gastroesophageal reflux disease, and broken heart syndrome [23-25]. However, the examination of mammography-related video content remains limited. To date, only 2 studies have assessed the quality of mammography videos on YouTube [26,27], both of which identified inconsistencies in the quality of information presented. Notably, no studies have yet evaluated mammography-related content on Douyin, the Chinese counterpart of TikTok. This research gap underscores the need for systematic evaluation of mammography-related short videos on Douyin, particularly considering its massive user base in China. Therefore, this study aims to comprehensively assess the quality and reliability of mammography-related short videos on Douyin by analyzing their characteristics, sources, and content.

Methods

Search Strategy

To minimize the bias introduced by personalized recommendation algorithms, we used 3 tactics: creating a new Douyin account specifically for evaluation, disabling Douyin's personalized recommendations to eliminate differential content

recommendations caused by user habits, and banning access to mobile location services. All videos were viewed without any actions such as downloading, liking, commenting, collecting, or sharing. Evaluation tasks were carried out by 2 qualified radiologists (Chuangying Zhu and HY) from the division of radiology in a tertiary teaching hospital.

The keywords "钼靶" ("mammography" in Chinese) and "钼靶检查" ("mammography screening" in Chinese) were searched in the Douyin app on October 22, 2023, with no limits placed on the release time. Douyin offers 3 ways to filter videos: overall ranking, most recent, and most likes. We used the overall ranking mode to retrieve the top 100 videos because most consumers use this default sorting option. We chose the threshold number of 100 for 2 reasons. First, Douyin's search function takes topic relevance into account; the most relevant mammography videos tend to appear at the top of the results list, and it is difficult to observe any pertinent videos when the results exceed 100. Second, most general health users apply the "least effort" principle when searching for information on the web; they tend to concentrate on the top search results. In this study, we included videos directly related to mammography. The exclusion criteria were videos not in Chinese, videos not related to mammography, duplicate videos, videos shorter than 10 seconds in length, and videos that were unavailable.

Data Collection

A Microsoft Excel spreadsheet was created by a researcher for data collection. Video information analyzed in this study was the identity of the uploader; the duration in seconds; the number of "likes" as indicated by the heart icon; the number of comments, saves, and shares the video received; and the number of days since the video was uploaded.

We divided the videos into 2 main groups according to whether the uploaders were professional or nonprofessional. Professional videos consisted of videos uploaded directly by board-certified physicians, health channels, and hospital channels. Most health channels and hospital videos were narrated by doctors. Nonprofessional videos included those uploaded by patients and other individuals.

Quality and Reliability Assessment

The quality and reliability of the video were evaluated based on the following criteria: the accuracy and comprehensiveness of the content, the clarity and fluency of information delivery, and the overall usefulness of the video to its intended audience.

No validated tools for assessing mammography video content are available in the literature. According to the American Cancer Society recommendations for the early detection of breast cancer [28] and the China Anti-Cancer Association Breast Cancer Diagnosis and Treatment Guide and Standard (2024 Edition) [6], 2 qualified radiologists (Chunmin Zhu and RH) from the division of radiology in a tertiary teaching hospital, with more than 10 years of experience in the radiological profession, developed a completeness checklist to assess the quality of mammography video content (Table 1). The 7 categories cover most aspects of mammography: conception, examination process, applicable objects, precautions, combined other examinations, advantages, and report. A video was awarded 1 point in each domain if it mentioned the content listed in Table 1, resulting in a final score ranging from 0 to 11. A score of 0 indicated that there was no accurate content in any of the 7 earlier-mentioned areas of mammography, whereas a score of 11 indicated that a video contained accurate information in all areas. Videos were then further categorized as useful or unhelpful according to the final score. Videos with a score of 0 were considered unhelpful if they only dealt with personal experiences or testimonies without providing any scientific content, whereas useful videos received a score of ≥1. We used the modified DISCERN (mDISCERN) tool and the Global Quality Score (GQS), previously used in many studies of Douyin, as instruments to assess the quality of information in each video.

Table . Completeness checklist.

Content	Description
Conception	1. Basic principles: mention that a mammogram is done with a machine designed to look only at breast tissue, with low-dose x-rays.
	2. Radiation: mention that mammogram exposes the breasts to small amounts of radiation.
Examination process	3. Remove upper body clothing: mention that the patient must remove clothing above the waist to have a mammogram.
	4. Pain: mention that it might feel some discomfort when the breasts are compressed, and for some women, it can be painful.
	5. Two positions for unilateral breast: mention that x-ray pictures of each breast are taken, typically from 2 different angles.
Applicable objects	6. Age: mention when to start a mammogram and how often.
	7. High risk: mention that women who are at high risk for breast cancer based on certain factors should get a mammogram every year, typically starting before 40 years of age.
Precautions	8. Special period: mention that women who are in a special period, such as preparing for pregnancy, pregnant, or breastfeeding, and those who have undergone breast augmentation surgery need to inform doctor in advance. Mammograms are generally not recommended for pregnant women. It is best to schedule the examination about a week after her period.
Combined other examinations	9. Mention that breast ultrasound and magnetic resonance imaging can help find some breast cancers that cannot be seen on mammograms.
Advantage	10. Mention that mammograms have a great advantage in detecting calcifications.
Report	11. Mention that what is the breast imaging reporting and data system.

The DISCERN criteria are a validated scoring system developed by an Oxford University research team to assess the information quality and reliability of content related to consumer health information on treatment options [29]. The mDISCERN tool was modified by Singh et al [30] and is based on a 5-point Likert scale that examines goals, reliability of information sources, bias, areas of uncertainty, and additional sources. According to the mDISCERN score, the reliability of video content is considered good for a DISCERN score of >3 points, moderate for a DISCERN score of 3 points, and poor for a DISCERN score of <3 points.

The GQS, which was developed by Bernard et al [31], is a 5-point Likert scale used to assess the quality of a video based on the flow of information, completeness of the information presented on a particular topic, and usefulness of information to patients. A GQS of 1 is considered very poor, 2 is considered poor, 3 is considered fair, 4 is considered good, and 5 is considered excellent. The detailed information for mDISCERN and GQS is available on the web as in Multimedia Appendices 1 and 2.

Before starting to score the videos, radiologists first reviewed the official DISCERN and GQS instructions and referred to a simplified Chinese version [32], the latter is more adapted to the Chinese language and culture. To ensure consistency, prescoring discussions were mandatory. After reaching a consensus on the first 20 videos, the evaluators independently reviewed the subsequent entries. The original scores of the 2 radiologists (C Zhou and HY) were independently recorded. The scores of mDISCERN and GQS given by the 2 researchers (C Zhou and HY) were averaged to obtain an overall score, which was then used in the analysis. Any disagreements about the completeness checklist were resolved by consensus.

Statistical Analysis

SPSS (version 27.0; IBM Corp) was used for data entry and analysis. Data are summarized as frequency (n) and percentage (%) for categorical variables and median (IQR) for ordinal variables. The normality of the data was analyzed using the Shapiro-Wilk test. Because the data were not normally distributed, the Mann-Whitney *U* test was used to compare the continuous variables between the 2 groups. Cronbach α coefficients were used to calculate the agreement between the 2 researchers. Spearman correlation tests were used to assess relationships between parameters. The correlations were interpreted based on the magnitude of the Spearman correlation coefficient (r), with the following thresholds used as a guide to describe the strength of the relationships: $r<0.1$ is considered a negligible correlation, $0.1\leq r<0.4$ is a weak correlation, $0.4\leq r<0.7$ is a moderate correlation, $0.7\leq r< 0.9$ is a strong correlation, and $r\geq 0.9$ is a very strong correlation. These thresholds were adapted from conventional guidelines for interpreting correlation coefficients, as discussed in the literature [33]. Differences were considered statistically significant at a P value of <.05.

Ethical Considerations

No clinical data, human specimens, or laboratory animals were involved in this study. All information used in this study was obtained from publicly released Douyin videos, and none of the data involved personal privacy. In addition, the study did not involve any interaction with users; therefore, no ethics review

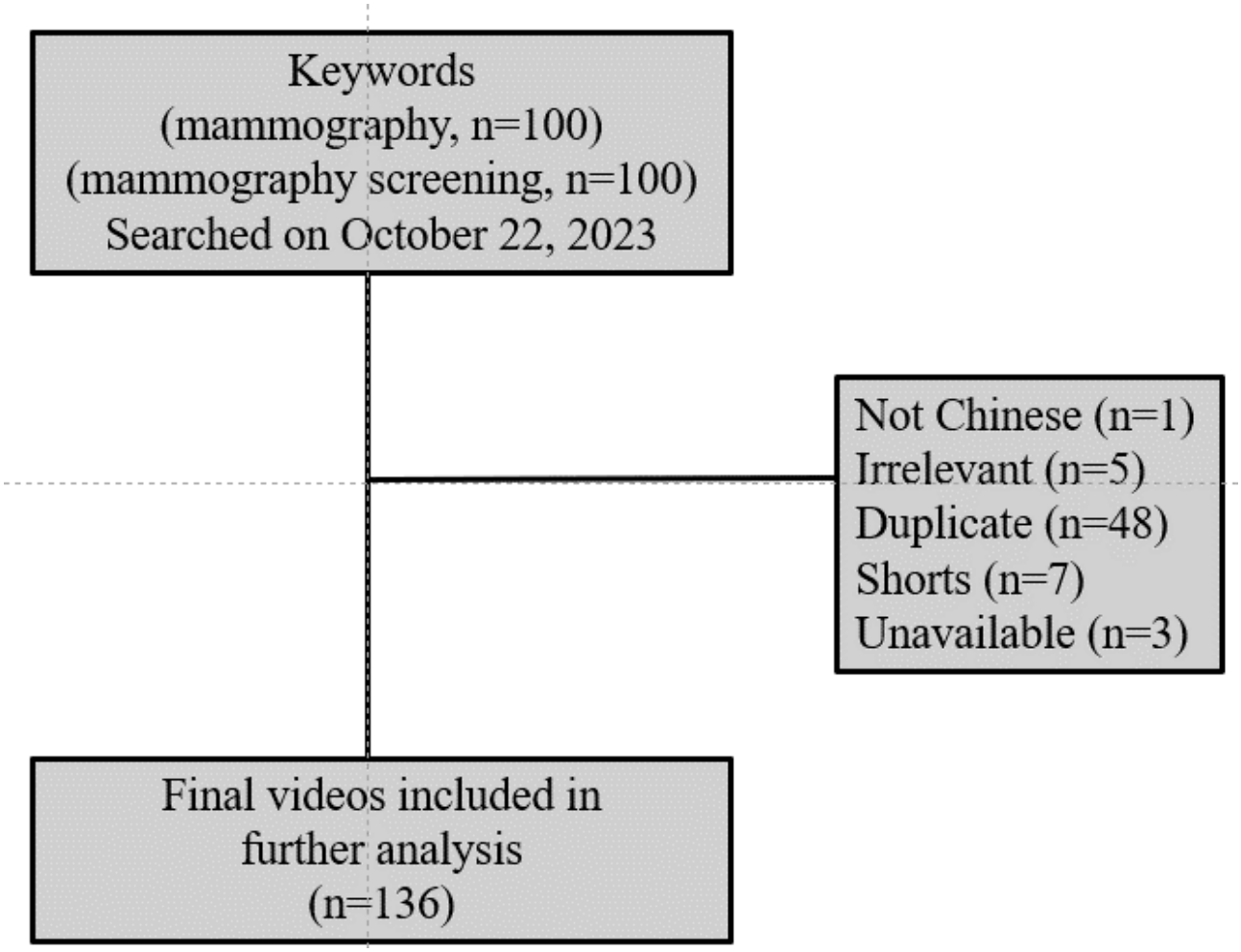
was required. All data were deidentified, and no individual users, videos, or screenshots are identifiable in this manuscript or its supplementary materials.

Results

Video Selection Process

In total, 200 videos were screened, and 136 were included in the study. The 64 excluded videos were 1 video in a non-Chinese language, 5 irrelevant videos, 48 duplicate videos, 7 short videos, and 3 unavailable videos (Figure 1).

Figure 1. Flowchart of videos included in the study.



Video Characteristics

The statistical analysis showed that the mammography videos ranged from 11 to 876 seconds. At the time of this study, the 136 short mammography videos had received 1,788,786 likes, 288,802 comments, 110,224 saves, and 598,393 shares. Each short video received 2 to 256,000 likes, 0 to 81,000 comments, 0 to 21,000 saves, and 0 to 145,000 shares. The most recent video was posted 21 days before the data collection, while the oldest had been on Douyin for more than 3 years. The median duration of the videos was 49.5 (IQR 32.5 - 76.75) seconds; the median indicators of engagement comprised 414 (IQR 155.75 - 1887.25) likes, 50.5 (IQR 20 - 286.25) comments, 56 (IQR 19-201.75) saves, and 61.5 (IQR 12 - 275.75) shares; and the median time since upload was 382.5 (IQR 116.25 - 635.75) days. The characteristics of the included videos are shown in Table 2.

Table . Characteristics of videos about mammography on Douyin.

Characteristics	Median (IQR)	Range
Duration (seconds)	49.50 (32.50 - 76.75)	11 - 876
Number of likes	414 (155.75 - 1887.25)	2 - 256,000
Number of comments	50.50 (20 - 286.25)	0 - 81,000
Number of saves	56 (19 - 201.75)	0 - 21,000
Number of shares	61.50 (12 - 275.75)	0 - 145,000
Days since upload	382.50 (116.25 - 635.75)	21 - 1208
Global Quality Score	2 (1 - 2)	0 - 4
DISCERN score	2.5 (1.63 - 3)	0 - 3.5

Uploader Douyin Account Characteristics

Most of the videos in our sample were contributed by professional users (112/136, 82.4%), while a relatively small proportion were contributed by nonprofessional users (24/136, 17.6%). Among professional users, most videos were created by board-certified physicians, followed by hospital channels and health channels (Table 3). Only 1 imaging physician was involved in the posting of 3 videos. The median video duration was significantly longer (P<.001) in the nonprofessional group and received significantly more comments (P=.004; Table 4).

Table . Proportion of videos by different types of uploaders.

Source	Description	Videos (n=136), n (%)
Professionals	Individuals or mechanisms who describe themselves as health professionals with certification	112 (82.4)
Board-certified physicians	Medical specialist who diagnoses, treats, and manages diseases and conditions related to breast cancer	106 (77.9)
Health channels	Organizations providing health knowledge	2 (1.6)
Hospital channels	Hospital platforms share health care information	4 (2.9)
Nonprofessionals	Individuals who share mammography experiences or medical personnel without certification	24 (17.6)

Table . Analysis of video characteristics by source.

Characteristics	Professional (n=112), median (IQR)	Nonprofessional (n=24), median (IQR)	*P* value
Duration (seconds)	46 (31.25 - 69.50)	96.50 (53 - 132.25)	<.001
Number of likes	382.50 (154 - 1658.50)	641.50 (155.75 - 15,223)	.33
Number of comments	44 (20 - 192.75)	165.50 (42.25 - 2900)	.004[a]
Number of saves	55.50 (19.5 - 182.5)	75.50 (10.25 - 2122.50)	.58
Number of shares	66 (12 - 198.75)	59 (11.5 - 4561.25)	.57
Days since upload	397 (137 - 648.25)	310.5 (78.25 - 611.50)	.66
Global Quality Score	2 (1 - 2)	1 (1 - 1.37)	<.001
DISCERN score	3 (2-3)	0 (0 - 0.75)	<.001

[a]P<.01.

As mentioned earlier, the selected videos were divided into useful and unhelpful groups based on scores of the completeness checklist. Of the 136 selected videos, the number of videos containing useful and unhelpful information was 105 (77.2%) and 31 (22.8%), respectively. Notably, despite uniformity in video days since upload between groups, uploads by unhelpful groups garnered more engagement metrics such as likes (median 6892, IQR 585 - 104,000), comments (median 1305, IQR 130 - 4103), saves (median 748, IQR 53-4381), and shares (median 1056, IQR 50 - 6071) relative to useful group, and this differential attains statistical significance (P<.01 for all; Table 5). Because the number of nonprofessional uploaders in the useful group was small (7/105), we could not compare this group.

Table . Analysis of video characteristics by usefulness.

Characteristics	Useful group (n=105), median (IQR)	Unhelpful group (n=31), median (IQR)	*P* value
Duration (seconds)	46 (31-69)	76 (46-114)	<.001
Number of likes	276 (131.50 - 815.50)	6892 (585 - 104,000)	<.001
Number of comments	35 (17 - 113.50)	1305 (130 - 4103)	<.001
Number of saves	47 (17-107)	748 (53 - 4381)	<.001
Number of shares	48 (10.50 - 158)	1056 (50 - 6071)	<.001
Days since upload	365 (104.50 - 675)	410 (151-538)	.77
Global Quality Score	2 (1.50 - 2)	1 (1 - 1)	<.001
DISCERN score	2 (2-3)	0 (0 - 0)	<.001

Information Content Comprehensiveness

Useful videos were analyzed based on the information they contained. Among all the categories, the advantages of mammography were the most frequently covered topic (53/105, 50.5%), followed in descending order by applicable objects (50/105, 47.6%), conception (47/105, 44.8%), examination process (44/105, 41.9%), combined other examinations (42/105, 40%), report (26/105, 24.8%), and precautions (11/105, 10.5%; Multimedia Appendix 3). Most of these videos (97/105, 92.4%) scored <5 points, and only 1 video received a maximum score of 7.

Video Reliability and Quality

The median (IQR) mDISCERN score and GQS of all videos were 2 (1 - 2) and 2.5 (1.63 - 3), respectively. The Cronbach α coefficients for reliability between the raters were 0.94 and 0.97 for the GQS and mDISCERN, respectively. The mDISCERN score and GQS of the videos in the useful and professional groups were significantly higher than those in the unhelpful and nonprofessional groups (all $P<.001$).

Correlation Analysis

Spearman correlation analysis revealed certain correlations among the characteristics of the videos. The video duration was positively correlated with the number of comments ($r=0.23$; $P=.008$), saves ($r=0.20$; $P=.02$), and shares ($r=0.19$; $P=.02$). Across all videos, Spearman correlation analysis revealed positive and significant correlations among the number of likes, comments, saves, shares, and days since upload ($P<.05$ for each pair).

The GQS was negatively or positively correlated with the number of likes ($r=-0.24$; $P=.004$), comments ($r=-0.29$; $P<.001$), and saves ($r=-0.20$; $P=.02$) as well as with the mDISCERN score ($r=0.65$; $P<.001$). The mDISCERN score was found to be negatively correlated with the number of likes ($r=-0.26$; $P=.002$), comments ($r=-0.36$; $P<.001$), saves ($r=-0.22$; $P=.009$), and shares ($r=-0.18$; $P=.03$). The correlation coefficients (r) reported in this study are generally below 0.39, indicating weak associations. In cases where the correlation coefficients are below 0.1, we consider these to be negligible. We acknowledge that the statistical significance of these correlations may be influenced by the sample size; therefore, we place greater emphasis on the magnitude of the correlation coefficients to better reflect the strength of the relationships. More detailed analytical results are shown in Table 6.

Table . Correlation analysis (Pearson *r* and 2-tailed *P* value) among the research variables.

Variable	Duration	Likes	Comments	Saves	Shares	Days since upload	GQS[a]	mDISCERN[b]
Duration								
r value	1	0.168	0.227[c]	0.201[d]	0.194[d]	–0.088	0.003	–0.144
P value	—[e]	.05	.008	.02	.02	.31	.98	.09
Likes								
r value	0.168	1	0.909[c]	0.91[c]	0.865[c]	0.284[c]	–0.245[c]	–0.262[c]
P value	.05	—[e]	<.001	<.001	<.001	<.001	.004	.002
Comments								
r value	0.227[c]	0.909[c]	1	0.851[c]	0.815[c]	0.252[c]	–0.289[c]	–0.361[c]
P value	.008	<.001	—[e]	<.001	<.001	.003	<.001	<.001
Saves								
r value	0.201[d]	0.91[c]	0.851[c]	1	0.915[d]	0.194[d]	–0.204[d]	–0.222[c]
P value	.02	<.001	<.001	—[e]	<.001	.02	.02	.009
Shares								
r value	0.194[d]	0.865[c]	0.815[c]	0.915[c]	1	0.353[c]	–0.111[c]	–0.181[d]
P value	.02	<.001	<.001	<.001	—[e]	<.001	.20	.03
Days since upload								
r value	–0.088	0.284[c]	0.252[c]	0.194[d]	0.353[c]	1	0.087	0.16
P value	.31	<.001	.003	.02	<.001	—[e]	.31	.06
GQS								
r value	0.003	–0.245[c]	–0.289[c]	–0.204[d]	–0.111	0.087	1	0.651[c]
P value	.98	.004	<.001	.02	.20	.31	—[e]	<.001
mDISCERN								
r value	–0.144	–0.262[c]	–0.361[c]	–0.222[c]	–0.181[d]	0.16	0.651[c]	1
P value	.09	.002	<.001	.009	.03	.06	<.001	—[e]

[a]GQS: Global Quality Score.

[b]mDISCERN: modified DISCERN.

[c]The correlation is significant at a significance level of .05 (2-tailed).

[d] The correlation is significant at a significance level of .01 (2-tailed).

[e]Not applicable.

Discussion

Principal Findings

This is the first study in the literature to evaluate Douyin content on mammography videos. According to the findings of 2 independent reviewers (C Zhou and HY), more than three-quarters of the videos were uploaded by professional individuals or institutions, and videos containing content primarily concerned with disease knowledge were of higher quality and more reliable. Nevertheless, the overall quality of the mammography videos was poor according to the completeness checklist, GQS, and mDISCERN score. Additionally, the fact that seekers gave higher ratings to the lower-quality videos than the higher-quality videos suggests that most health viewers are not able to identify poor-quality medical information in videos.

The rapid development of digital technology and the widespread application of mobile intelligent terminals have caused various new media to become important platforms for sharing and exchanging scientific knowledge. This has further expanded the channels through which the public can understand and obtain information, broadening the breadth and depth of knowledge. There was an unprecedented reliance on social media platforms to seek information during the COVID-19 pandemic [34]. Douyin is a representative national short video platform, and

watching videos every day has become a part of many people's lives.

Currently, uploaders who share health information on the Douyin app are required to obtain certification materials that verify their affiliation with tertiary A hospitals as doctors. In our study, approximately 80% of mammography-related Douyin contents were uploaded by professional users. Most of them were clinicians; only 1 was an imaging specialist. These findings show that clinicians in tertiary A hospitals with a high level of expertise are enthusiastic about participating in the popularization of mammography-related information. A previous study also showed that radiology-related content on the increasingly popular social media platform TikTok is mainly posted by nonphysician radiology personnel [35]. In addition, our results suggest that the videos cannot cover all aspects of mammography, which may be due to the limited short length of Douyin videos. Furthermore, the most prevalent content of the videos was the advantages of mammography in detecting calcifications; few videos fully addressed other types of content during the examination. This finding may indicate that most publishers believe that the unique advantage of mammography is to help detect breast cancer at an earlier stage. As radiologists, they may be more likely to focus on pain and positioning or precautions during the examination and have a more accurate understanding of diagnostic reports [36]. Pain and discomfort during mammography may influence participation in screening programs and be detrimental to cancer prevention efforts [37]. More senior radiologists should be encouraged to become involved in mammography popularization. Specialized training and publicity should be provided to meet the public's need for knowledge about mammography.

The current results indicate that the reliability and educational quality of mammography-related videos on Douyin are unsatisfactory, with median mDISCERN and GQS evaluations across all videos stood at 2.5 (IQR 1.63-3), and 2 (IQR 1-2), respectively. This finding is in accord with previous studies that have examined low-quality videos on various health topics and found that this information may not be reliable on Douyin [38,39]. Studies on other video platforms, such as YouTube, also showed that the overall quality of videos providing disease information was poor [40,41]. Because the content of most videos lacks peer or institutional quality review, many may not be subject to quality control and may not be evidence-based; it is therefore not surprising that much of this content is inadequate [42]. Thus, patients should access certified organizations and sites such as those certified with the Health on the Net Foundation Code of Conduct certificate to obtain professional information and avoid being misled by social media. The Health on the Net Foundation Code of Conduct was created as a practical solution to help internet users recognize reliable health-related information on the internet while distinguishing it from potentially erroneous or hazardous content [43]. However, contrary to all these findings, in previous studies of the quality of Douyin videos on children with humeral supracondylar fractures, chronic obstructive pulmonary disease, and cosmetic surgery, the overall information quality and reliability of these short videos were satisfactory in China [44-46]. This might be explained by the assessment instrument's lack of comparability between different disease categories and the bias introduced by the use of different scoring criteria among different researchers.

The results also showed that videos posted by professionals had significantly higher reliability and GQS than those posted by individuals. This finding indicates that ownership is an important element that can be used to assess the reliability of videos. Video content may be considered trustworthy when produced by professionals such as doctors, medical organizations, and health information websites [47]. Unfortunately, our regression analysis revealed that the number of likes, comments, and saves had a weak negative correlation with both the mDISCERN score and GQS. The results showed that lay users had difficulty distinguishing useful information from a large number of videos. A common misconception is that digital information accuracy is directly related to the number of hits or views [48]. There are thousands of health-related videos promoting misleading information that get millions of views, such as videos that disparage vaccinations [49,50]. These results also indicate that effective regulatory measures are needed to control scientifically accredited information. In the future, it would be beneficial to develop an algorithm that ranks videos preferentially uploaded by a trusted medical center or professional. If the public had less access to unhelpful videos, the damage could be less.

Limitations

This study has some limitations. First, this was a cross-sectional study that examined a very small portion of a very large amount of data. The number of views, likes, and dislikes of health-related videos on the internet changes over time. The "snapshot" approach to data collection seems to be the main limitation of this study because the results may vary with the use of different search terms and according to the date and time of the search. Second, because of the limitations of the search criteria, it was not possible to include all video resources that fit the topic of this study. Although we included a relatively small percentage of videos, we considered it to be sufficiently representative, as videos beyond the top 100 have no significant impact on the analysis. Third, we only included videos uploaded on Douyin, which is a Chinese video-sharing platform; thus, the findings may not be generalizable to other social media platforms (eg, YouTube) or to other countries. Subsequent cross-linguistic research is required to fill this gap. Finally, the GQS and DISCERN are subjective assessment tools. Although 2 independent experts (C Zhou and HY) determined the ratings iteratively and used Cronbach α coefficients to quantify the agreement between the 2 raters, subjective differences still cannot be ignored. Looking ahead, future research should include broader cross-linguistic comparative studies, using more appropriate assessment instruments to validate our findings.

Conclusions

According to the findings of our study, a majority of the Douyin videos concerning mammography were uploaded by clinicians and exhibited poor quality and reliability. Patients should not use these videos as the only source of information about mammography because they may lead to misdirected or inappropriate interventions. Douyin is often used to obtain health-related information, and radiologists should be

encouraged to provide useful and accurate videos and to instruct patients appropriately. From the standpoint of preventing and curing breast cancer, there is a need for stricter standards and procedures for video publishing to improve the quality of medical content.

Acknowledgments

The authors thank Angela Morben, DVM, ELS, from Liwen Bianji (Edanz) for editing the English text of a draft of this manuscript.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

The study was designed and planned by RH and Chunmin Zhu. Data collection, results interpretation, and manuscript writing were all carried out by HY and C Zhou. Chuangying Zhu reviewed the work. LH, HW, PC, and SZ contributed to the reagents, materials, and analysis tools. HY and C Zhou drafted the manuscript. Chunmin Zhu and RH reviewed and edited the manuscript. All authors have reviewed and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Modified DISCERN criteria.
[DOCX File, 17 KB - cancer_v11i1e59483_app1.docx]

Multimedia Appendix 2
Global Quality Score criteria.
[DOCX File, 16 KB - cancer_v11i1e59483_app2.docx]

Multimedia Appendix 3
Characteristics of useful videos related to each topic.
[PNG File, 1545 KB - cancer_v11i1e59483_app3.png]

References

1. Soerjomataram I, Bray F. Planning for tomorrow: global cancer incidence and the role of prevention 2020-2070. Nat Rev Clin Oncol 2021 Oct;18(10):663-672. [doi: 10.1038/s41571-021-00514-z] [Medline: 34079102]
2. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263. [doi: 10.3322/caac.21834] [Medline: 38572751]
3. Han B, Zheng R, Zeng H, et al. Cancer incidence and mortality in China, 2022. J Natl Cancer Cent 2024 Mar;4(1):47-53. [doi: 10.1016/j.jncc.2024.01.006] [Medline: 39036382]
4. Anderson BO, Ilbawi AM, Fidarova E, et al. The Global Breast Cancer Initiative: a strategic collaboration to strengthen health care for non-communicable diseases. Lancet Oncol 2021 May;22(5):578-581. [doi: 10.1016/S1470-2045(21)00071-1] [Medline: 33691141]
5. Survival rates for breast cancer. American Cancer Society. 2024. URL: https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html [accessed 2025-02-10]
6. Shao Z. Guidelines for breast cancer diagnosis and treatment by China Anti-cancer Association (2024 edition). China Oncol 2023;33(12):1092-1186. [doi: 10.19401/j.cnki.1007-3639.2023.12.004]
7. Duffy SW, et al. Beneficial effect of consecutive screening mammography examinations on mortality from breast cancer: a prospective study. Radiology 2021;299(3):541-547. [doi: 10.1148/radiol.2021203935] [Medline: 33650900]
8. Tabar L, et al. The incidence of fatal breast cancer measures the increased effectiveness of therapy in women participating in mammography screening. Cancer 2019;125(4):515-523. [doi: 10.1002/cncr.31840] [Medline: 30411328]
9. Duffy SW, et al. Mammography screening reduces rates of advanced and fatal breast cancers: results in 549,091 women. Cancer 2020;126(13):2971-2979. [doi: 10.1002/cncr.32859] [Medline: 32390151]
10. Bao Y, Kwok C, Lee CF. Breast cancer screening behaviors among Chinese women in Mainland China. Nurs Health Sci 2018;20(4):445-451. [doi: 10.1111/nhs.12533] [Medline: 29920900]

11. Statistical report on internet development in China. China Internet Network Information Center (CNNIC). 2023. URL: https://www.cnnic.cn/NMediaFile/2023/0908/MAIN1694151810549M3LV0UWOAV.pdf [accessed 2025-02-10]
12. Doyle B. Tiktok statistics—updated May 2024. Wallaroo Media. 2024. URL: https://wallaroomedia.com/blog/social-media/tiktok-statistics/ [accessed 2025-02-14]
13. Iqbal M. TikTok revenue and usage statistics (2025). Business of Apps. 2025. URL: https://www.businessofapps.com/data/tik-tok-statistics/ [accessed 2025-02-14]
14. TikTok health science data report. Douyin. 2023. URL: https://www.douyin.com/note/7219193337868340535 [accessed 2025-02-10]
15. TikTok health lifestyle new paradigm white paper. Ocean Engine. 2023. URL: https://trendinsight.oceanengine.com/arithmetic-report/detail/989 [accessed 2025-02-10]
16. QuestMobile 2024 new media ecosystem insights. Quest Mobile. 2024. URL: https://www.questmobile.com.cn/research/report/1871401635281997826 [accessed 2025-02-14]
17. Wang JH, Liang W, Schwartz MD, Lee MM, Kreling B, Mandelblatt JS. Development and evaluation of a culturally tailored educational video: changing breast cancer-related behaviors in Chinese women. Health Educ Behav 2008 Dec;35(6):806-820. [doi: 10.1177/1090198106296768] [Medline: 17602099]
18. Goel MS, O'Conor R. Increasing screening mammography among predominantly Spanish speakers at a federally qualified health center using a brief previsit video. Patient Educ Couns 2016 Mar;99(3):408-413. [doi: 10.1016/j.pec.2015.09.007] [Medline: 26456634]
19. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85. [doi: 10.2196/jmir.1933] [Medline: 23615206]
20. Borges do Nascimento IJ, Beatriz Pizarro A, Almeida J, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022 Sep 1;100(9):544-561. [doi: 10.2471/BLT.21.287654]
21. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552. [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
22. Muhammed T S, Mathew SK. The disaster of misinformation: a review of research in social media. Int J Data Sci Anal 2022;13(4):271-285. [doi: 10.1007/s41060-022-00311-6] [Medline: 35194559]
23. Wang H, Yan C, Wu T, et al. YouTube online videos as a source for patient education of cervical spondylosis-a reliability and quality analysis. BMC Public Health 2023;23(1):1831. [doi: 10.1186/s12889-023-16495-w] [Medline: 37730621]
24. Aydin MF, Aydin MA. Quality and reliability of information available on YouTube and Google pertaining gastroesophageal reflux disease. Int J Med Inform 2020 May;137:104107. [doi: 10.1016/j.ijmedinf.2020.104107] [Medline: 32146372]
25. Liang J, Wang L, Song S, et al. Quality and audience engagement of takotsubo syndrome-related videos on TikTok: content analysis. J Med Internet Res 2022 Sep 26;24(9):e39360. [doi: 10.2196/39360] [Medline: 36155486]
26. Basch CH, Hillyer GC, MacDonald ZL, Reeves R, Basch CE. Characteristics of YouTube(TM) videos related to mammography. J Cancer Educ 2015 Dec;30(4):699-703. [doi: 10.1007/s13187-014-0769-9]
27. Lee SC, Monga AK, Kawashita T, Dunnsiri T, Woo BKP. Analyzing the quality of mammography-related YouTube videos. Breast J 2020 Nov;26(11):2327-2328. [doi: 10.1111/tbj.14039] [Medline: 32924242]
28. American Cancer Society recommendations for the early detection of breast cancer. American Cancer Society. 2023. URL: https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/american-cancer-society-recommendations-for-the-early-detection-of-breast-cancer.html [accessed 2025-02-10]
29. Charnock D, Shepperd S, Needham G, Gann R. Discern: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111. [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
30. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis—a wakeup call? J Rheumatol 2012 May;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
31. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007 Sep;102(9):2070-2077. [doi: 10.1111/j.1572-0241.2007.01325.x] [Medline: 17511753]
32. Shan Y, Xing Z, Dong Z, Ji M, Wang D, Cao X. Translating and adapting the DISCERN instrument into a simplified Chinese version and validating its reliability: development and usability study. J Med Internet Res 2023 Feb 2;25:e40733. [doi: 10.2196/40733] [Medline: 36729573]
33. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018 May;126(5):1763-1768. [doi: 10.1213/ANE.0000000000002864] [Medline: 29481436]
34. Adebesin F, Smuts H, Mawela T, Maramba G, Hattingh M. The role of social media in health misinformation and disinformation during the COVID-19 pandemic: bibliometric analysis. JMIR Infodemiology 2023 Sep 20;3:e48620. [doi: 10.2196/48620] [Medline: 37728981]
35. Lovett JT, Munawar K, Mohammed S, Prabhu V. Radiology content on TikTok: current use of a novel video-based social media platform and opportunities for radiology. Curr Probl Diagn Radiol 2021;50(2):126-131. [doi: 10.1067/j.cpradiol.2020.10.004] [Medline: 33250298]

36. Ding S, Fontaine T, Serex M, Sá Dos Reis C. Strategies enhancing the patient experience in mammography: a scoping review. Radiography (Lond) 2024 Jan;30(1):340-352. [doi: 10.1016/j.radi.2023.11.016] [Medline: 38141428]
37. Montoro CI, Alcaraz MDC, Galvez-Sánchez CM. Experience of pain and unpleasantness during mammography screening: a cross-sectional study on the roles of emotional, cognitive, and personality factors. Behav Sci (Basel) 2023 May 4;13(5):377. [doi: 10.3390/bs13050377] [Medline: 37232614]
38. Sun F, Zheng S, Wu J. Quality of information in gallstone disease videos on TikTok: cross-sectional study. J Med Internet Res 2023 Feb 8;25:e39162. [doi: 10.2196/39162] [Medline: 36753307]
39. Yao L, Li Y, Lian Q, Sun J, Zhao S, Wang P. Health information sharing on social media: quality assessment of short videos about chronic kidney disease. BMC Nephrol 2022 Nov 28;23(1):378. [doi: 10.1186/s12882-022-03013-0] [Medline: 36443741]
40. Kılınç DD. Is the information about orthodontics on YouTube and TikTok reliable for the oral health of the public? A cross sectional comparative study. J Stomatol Oral Maxillofac Surg 2022 Oct;123(5):e349-e354. [doi: 10.1016/j.jormas.2022.04.009] [Medline: 35470122]
41. Aurlene N, Shaik SS, Dickson-Swift V, Tadakamadla SK. Assessment of usefulness and reliability of YouTube videos on denture care. Int J Dent Hyg 2024 Feb;22(1):106-115. [doi: 10.1111/idh.12771] [Medline: 37793081]
42. Comp G, Dyer S, Gottlieb M. Is TikTok the next social media frontier for medicine? AEM Educ Train 2021 Jul;5(3). [doi: 10.1002/aet2.10532] [Medline: 34095694]
43. Boyer C, Selby M, Scherrer JR, Appel RD. The Health On the Net Code of Conduct for medical and health websites. Comput Biol Med 1998 Sep;28(5):603-610. [doi: 10.1016/s0010-4825(98)00037-7] [Medline: 9861515]
44. Zhu Z, Zheng Y, Zhu D. Douyin as a source of information and education on humeral supracondylar fracture of children during the COVID-19 pandemic in Chinese Mainland: an observational study. Medicine (Baltimore) 2023 Jun 23;102(25):e34104. [doi: 10.1097/MD.0000000000034104] [Medline: 37352075]
45. Song S, Xue X, Zhao YC, Li J, Zhu Q, Zhao M. Short-video apps as a health information source for chronic obstructive pulmonary disease: information quality assessment of TikTok videos. J Med Internet Res 2021 Dec 20;23(12):e28318. [doi: 10.2196/28318] [Medline: 34931996]
46. Zhang J, Han P, Tang Y, Xi W, Xiao X, Yang F. Popular science and education of cosmetic surgery in China: quality and reliability evaluation of Douyin short videos. Health Expect 2023 Jun;26(3):1221-1226. [doi: 10.1111/hex.13737] [Medline: 36807969]
47. Anderer S. Patients are turning to TikTok for health information—here's what clinicians need to know. JAMA 2024 Apr 16;331(15):1262. [doi: 10.1001/jama.2024.1280]
48. Syed-Abdul S, Fernandez-Luque L, Jian WS, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013 Feb 13;15(2):e30. [doi: 10.2196/jmir.2237] [Medline: 23406655]
49. Fieselmann J, Annac K, Erdsiek F, Yilmaz-Aslan Y, Brzoska P. What are the reasons for refusing a COVID-19 vaccine? A qualitative analysis of social media in Germany. BMC Public Health 2022 Apr 27;22(1):846. [doi: 10.1186/s12889-022-13265-y] [Medline: 35484619]
50. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]

Abbreviations

GQS: Global Quality Score
mDISCERN: modified DISCERN

Edited by N Cahill; submitted 13.04.24; peer-reviewed by D Filos, I Moreira; revised version received 24.01.25; accepted 27.01.25; published 19.02.25.

Please cite as:
Yang H, Zhu C, Zhou C, Huang R, Huang L, Chen P, Zhu S, Wang H, Zhu C
Evaluation of Douyin Short Videos on Mammography in China: Quality and Reliability Analysis
JMIR Cancer 2025;11:e59483
URL: https://cancer.jmir.org/2025/1/e59483
doi:10.2196/59483

Comparison of Electronic Surveillance With Routine Monitoring for Patients With Lymphoma at High Risk of Relapse: Prospective Randomized Controlled Phase 3 Trial (Sentinel Lymphoma)

Katell Le Dû[1*], MD; Adrien Chauchet[2*], MD; Sophie Sadot-Lebouvier[1*], MD; Olivier Fitoussi[3*], MD; Bijou Fontanet[4*], MD; Arnaud Saint-Lezer[5*], MD; Frédéric Maloisel[6*], MD, PhD; Cédric Rossi[7*], MD, PhD; Sylvain Carras[8*], MD; Anne Parcelier[9*], MD; Magali Balavoine[10*]; Anne-Lise Septans[10*], PhD

[1]Department of Hematology, Confluent Private Hospital, 2 rue Eric Tabarly, Nantes, France

[10]Department of Biostatistics, Institut inter-régional de cancérologie, Le Mans, France

[2]Department of Hematology, University Hospital, Besançon, France

[3]Department of Hematology, Polyclinique Bordeaux Nord Aquitaine, Bordeaux, France

[4]Department of Hematology, Bergonié Institute, Bordeaux, France

[5]Department of Hematology, Mont de Marsan Hospital, Mont de Marsan, France

[6]Department of Hematology, Clinique Saint-Anne, Strasbourg, France

[7]Department of Hematology, University Hospital, Dijon, France

[8]Department of Hematology, University Hospital, Grenoble, France

[9]Department of Hematology, Centre hospitalier Bretagne Atlantique, Vannes, France

[*]all authors contributed equally

Corresponding Author:
Katell Le Dû, MD
Department of Hematology, Confluent Private Hospital, 2 rue Eric Tabarly, Nantes, France

Abstract

Background: Relapse is a major event in patients with lymphoma. Therefore, early detection may have an impact on quality of life and overall survival. Patient-reported outcome measures have demonstrated clinical benefits for patients with lung cancer; however, evidence is lacking in patients with lymphoma. We evaluated the effect of a web-mediated follow-up application for patients with lymphoma at high risk of relapse.

Objective: This study aims to demonstrate that monitoring patients via a web application enables the detection of at least 30% more significant events occurring between 2 systematic follow-up consultations with the specialist using an electronic questionnaire.

Methods: We conducted a prospective, randomized phase 3 trial comparing the impact of web-based follow-up (experimental arm) with a standard follow-up (control arm). The trial was based on a 2-step triangular test and was designed to have a power of 90% to detect a 30% improvement in the detection of significant events. A significant event was defined as a relapse, progression, or a serious adverse event. The study covered the follow-up period after completion of first-line treatment or relapse (24 months). Eligible patients were aged 18 years and older and had lymphoma at a high risk of relapse. In the experimental arm, patients received a 16-symptom questionnaire by email every 2 weeks. An email alert was sent to the medical team based on a predefined algorithm. The primary objective was assessed after the inclusion of the 40th patient. The study was continued for the duration of the analysis.

Results: A total of 52 patients were included between July 12, 2017, and April 7, 2020, at 11 centers in France, with 27 in the experimental arm and 25 in the control arm. The median follow-up was 21.3 (range 1.3 - 25.6) months, and 121 events were reported during the study period. Most events occurred in the experimental arm (83/119, 69.7%) compared with 30.2% (36/119) in the control arm. A median number of 3.5 (range 1-8) events per patient occurred in the experimental arm, and 1.8 (range 1-6) occurred in the control arm (P=.01). Progression and infection were the most frequently reported events. Further, 19 patients relapsed during follow-up: 6 in the experimental arm and 13 in the control arm (P<.001), with a median follow-up of 7.7 (range 2.8 - 20.6) months and 6.7 (range 1.9 - 16.4) months (P=.94), respectively. Statistical analysis was conducted after including the 40th patient, which showed no superiority of the experimental arm over the control arm. The study was therefore stopped after the 52nd patient was enrolled.

Conclusions: The primary objective was not reached; however, patient-reported outcome measures remain essential for detecting adverse events in patients with cancer, and the electronic monitoring method needs to demonstrate its effectiveness and comply with international safety guidelines.

Trial Registration: ClinicalTrials.gov NCT03154710; https://clinicaltrials.gov/ct2/show/NCT03154710

(JMIR Cancer 2025;11:e65960) doi:10.2196/65960

KEYWORDS

patient-reported outcome measures; lymphoma; risk of relapse; relapse; randomized trial; web-based; quality of life; survival; detection; progression; T-cell lymphoma; Hodgkin lymphoma

Introduction

Relapse or progression is a major event in the management of lymphoma. Predictive factors for relapse include histological subtype, extranodal involvement, high metabolic volume, and elevated serum lactate dehydrogenase (LDH) levels [1]. Early detection of relapse correlates with survival. In most cases, relapse is detected by the appearance of symptoms, clinical signs, or biological abnormalities [2-4]. Repeated surveillance computed tomography (CT) detects asymptomatic recurrence in only 1.7% of patients and increases the risk of secondary cancers because of radiation overexposure [5-8]. Circulating tumor DNA monitoring may be used to detect early recurrence before the onset of symptoms; however, this method has not been validated [9]. Electronic patient-reported outcome measures (ePROMs) based on the Common Terminology Criteria for Adverse Events have emerged as a method of early detection. This has increased survival rates in some cases (locally advanced lung cancer) [10]. ePROMs affect early event detection and overall survival in patients with cancer [11-15]; however, such evidence is lacking for patients with lymphoma. In this study, we compare the effect of web-based follow-up with that of standard follow-up.

Methods

Overview

We conducted an open-label, longitudinal, prospective study between July 12, 2017, and April 7, 2020, at 14 centers in France.

Ethical Considerations

This study was conducted according to the 1975 Declaration of Helsinki, revised in 2008, and the guidelines of the International Conference on Harmonization of Good Clinical Practice in Biomedical Research. The Ouest II national ethics committee in Angers approved the study on November 8, 2016, and the Agence Nationale de Sécurité du Médicament approved it on November 22, 2016 (approval: 2021–A01670–41). All of the patients provided written informed consent, which included the points of analysis, the method of data collection, and the primary and potential secondary statistical analyses. All patient data were anonymized and no financial compensation was provided.

Study Population, Inclusion Criteria, and Exclusion Criteria

Patients with lymphoma who were aged 18 years or older and had a high risk of relapse were considered eligible for this trial. They could have T-cell lymphoma in the first partial or complete response, Hodgkin lymphoma in the second partial or complete response, or diffuse large B-cell lymphoma in the first partial or complete response with a revised high International Prognostic Index score (≥3) or in the second partial or complete response. Patients who had undergone autologous stem cell transplantation were not excluded. Eastern Cooperative Oncology Group performance status between 0 and 2, an internet connection, and affiliation to the French social security system were required. Patients were recruited during follow-up consultations by the referring physician at each center.

The exclusion criteria were an initial symptom score <7, progression within 3 months of the last treatment, brain or meningeal involvement, history of another cancer treated within 3 years—with the exception of skin cancer (except melanoma) and in situ cervical cancer—pregnancy, breastfeeding, and any psychiatric pathology that may prevent compliance with the protocol.

An initial symptom score was established in the previous Sentinel study. The e-request algorithm was more sensitive for patients who were not very symptomatic at inclusion and had an initial score of less than 7 (by summing scores from 0 to 3 for symptoms concerning cough, dyspnea, pain, anorexia, and asthenia: 0=no problem, 1=mild problem, 2=moderate problem, and 3=severe problem) [12].

Randomization

Randomization was planned through minimization once patients were enrolled in the study and programmed using ENNOV Clinical data management software. Patients were randomly assigned 1:1 to a routine follow-up (control arm) or web-mediated follow-up (experimental arm). Stratification was conducted at inclusion according to the center, performance status, autologous stem cell transplantation history, relapse, and lymphoma subtype.

Follow-Up

Patients were included no later than 3 months at the end of their last treatment. Follow-up was 24 months after enrollment. A medical consultation and a biological assessment were performed every 3 months. In the control arm, CT scans were performed every 6 months. In the experimental arm, scans were performed when medically necessary. Quality of life (QoL) was assessed by 2 questionnaires every 3 months (European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire of Cancer Patients [QLQ-C30] and Patient Health Questionnaire-9 [PHQ-9]) [16,17]. Patient satisfaction with the application was evaluated by an internal questionnaire for patients in the experimental group during the 6-month visit from inclusion.

Web Application

The Moovcare patient-reported outcome (PRO) system is a class 1 medical device registered by Sivan Innovation, Ltd.,

with Conformité Européenne marking obtained in July 2017. Versions 1.7 (from July 2017 to October 2019) and 1.8 (from November 2019 to April 2020) were used in this study. “The reimbursed indication of the MOOVCARE Lung device is the early detection of recurrences or complications for patients over the age of 16 with nonprogressive lung cancer after the last medical treatment, regardless of the histological type of the tumor”—an excerpt from the user guide [18]. The indication was validated by the data of the Sentinel Lung study (published in 2019), which demonstrated a survival benefit of 9 months for patients monitored by the application compared to standard monitoring (P=.005) [13].

A scientific committee adapted the questionnaire for patients who were being monitored for lymphoma, whereas the technical monitoring and the decision algorithm remained the same.

PRO data were collected using questionnaires sent by the application directly to the patient through a clickable link to their email address. Patients were asked to complete a 16-question self-assessment every 14 days for 24 months after being randomized to the experimental arm (smartphone or email). They were also able to report an event in the web application between the 2 questionnaires. The study coordinators provided them with a short training session on the application. If the patient failed to complete the questionnaire, a reminder was sent after 24 hours, and the health care team contacted the patient as necessary.

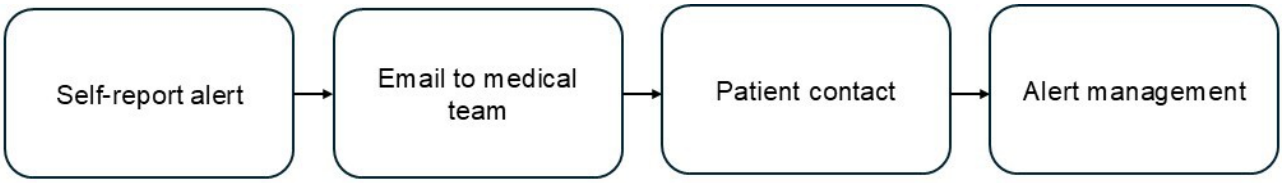
The questionnaire included 4 items, namely weight, LDH level (optional), hemoglobin level (optional), and a free comment (for other symptoms or remarks), and 12 following questions:

- Are you tired?
- Have you lost your appetite?
- Are you in pain?
- Are you short of breath?
- Do you feel depressed?
- Do you have a fever (temperature >38.1 °C, checked once at 1-hour intervals)?
- Do you have chills?
- Do you have pimples?
- Are you sweating profusely?
- Are you itching all over your body?
- Have you detected a lump under the skin or a lymph node?
- Have you noticed any abnormal swelling of the face or legs?

Patients were assigned a score based on their symptoms as follows: 0=no problem, 1=mild problem, 2=moderate problem, and 3=severe problem. An alert was triggered in the event of weight loss greater than 2 kg over 1 month, in the event of symptoms rated 3, the presence of fever or night sweats on 2 consecutive occasions, or elevation of serum LDH above 2-fold the normal level, or anemia indicated by hemoglobin levels of <10 g/dL.

In the event of an alert triggered by the application, an email was sent to the care team, with a reminder every 24 hours if there was no response (Figure 1). Patients could also report an event by writing a free text.

Figure 1. The decision algorithm used in this study.



Outcome Measures

The primary outcome was to demonstrate that follow-up via a web application could detect more significant events (including relapses) occurring between 2 routine follow-up consultations with the specialist in patients with lymphoma who were at high risk of relapse compared with standard follow-up. Secondary outcomes were overall and progression-free survival at 2 years, relapse rate at 2 years, QoL or patients in both arms, and compliance and satisfaction for the experimental arm.

Adverse and Significant Events

An adverse event was defined as any symptom reported by the patient either during the protocol consultation in the control arm or, through the application in the experimental arm. An event was considered significant if, the grade was greater (≥2) based on the Common Terminology Criteria for Adverse Events v4.02 or, if it prompted an imaging examination, treatment (of any kind), supportive care, unscheduled consultation, or emergency hospitalization.

QoL, Adherence, and Satisfaction

QoL was evaluated using the QLQ-C30 and PHQ-9 questionnaires at inclusion and, follow-up visits at 3, 6, 9, and 12 months. Patient adherence to the use of the web application was assessed according to the number of electronic questionnaires completed. A questionnaire had to be completed every 14 days. Patients completing less than 1 electronic questionnaire every 42 days (6 weeks) were considered noncompliant. Patient satisfaction with web monitoring and the use of the web application was assessed using a self-questionnaire at their 6-month follow-up visit.

Data Management and Statistical Analysis

Data Management

One electronic case report form (e-CRF; ENNOV Clinical) was created for each patient. The information required by the protocol was collated into the e-CRF, which included the data necessary to confirm compliance with the protocol and detect any major deviations, as well as the data necessary for statistical analysis. The information was collected without mentioning the

surname and first name in the e-CRF, with an identification number for the center and a patient number. Only the first letters of the patient's surname and first were visible. This code was the only patient identifier that appeared in the e-CRF, which made it possible to link e-CRFs to the corrresponding patients.

Statistical Analysis

Determination of the Size of the Study Population

The trial was based on a 2-step triangular test and was designed to have a power of 90% to detect a 30% improvement in the detection of significant events outside of routine consultations during the 6 months of follow-up with the web application. This is compared with a 60% rate of detection for significant events outside routine consultations among patients randomly assigned to conventional follow-up, with a significance of 5%. This sequential method made it possible to evaluate the application's effectiveness while controlling the power and type I error (the risk of falsely rejecting our null hypothesis) [19,20]. 40 evaluable patients were to be included per arm, and an interim analysis was to be performed when 20 evaluable patients per arm had 6 months of follow-up. Inclusion was not suspended before the 6-month follow-up.

Analysis of Variables, Progression-Free Survival, Overall Survival, and QoL Questionnaires

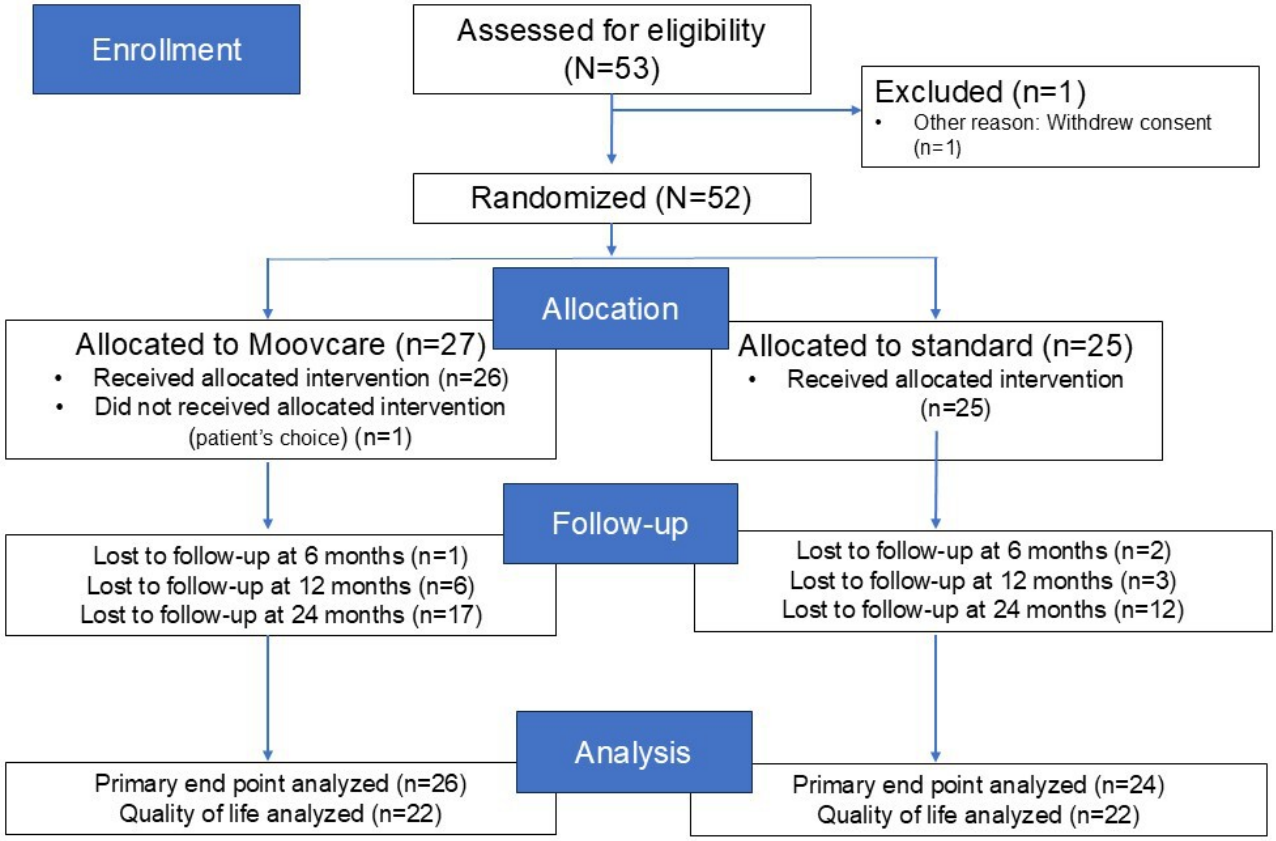
The analysis of the qualitative variables is presented in terms of numbers and percentages. The analysis of quantitative variables is presented as median or mean (SD) depending on the normality of the variable, whereas the minimum and maximum values are also indicated. The events are described in terms of frequency by etiological type (according to the Medical Dictionary for Egilatory Activities classification) and severity according to the NCI Common Terminology Criteria for Adverse Events version 4.02. For the analysis of censored data (overall survival and other event times), survival curves are plotted based on Kaplan-Meier estimates, and the median survival times and their 95% CIs are presented. For multivariate survival analyses, the Cox semiparametric model was used to calculate the odds ratios, which are presented with 95% CIs. The sensitivity of the application to detect a relapse and significant complications was calculated. The QoL scores were calculated according to the European Organisation for Research and Treatment of Cancer recommendations for the QLQ-C30 [16]. QoL is described for each measurement time, compared at inclusion, and then studied longitudinally using mixed analysis of variance models for repeated measures. PHQ-9 scores were calculated based on the recommendations described at each measurement time and compared at inclusion. Classes proposed in the literature (≤4; 5 - 14;>14) were used to describe the patients' state of depression [17]. Analyses were performed using SAS 9.3 software (SAS Institute, Inc).

Results

Study Design

A total of 53 patients were included between July 12, 2017, and April 7, 2020, from 14 centers in France: Le Mans (16/52, 31%), Besançon (9/52, 17%), Nantes (6/52, 11%), Bordeaux Bergonié (6/52,11%), Bordeaux Nord (6/52, 11%), Mont de Marson (4/52, 8%), Dijon (1/52, 2%), Grenoble (1/52, 2%), Paris (1/52, 2%), Strasbourg (1/52, 2%), and Vannes (1/52, 2%). One patient withdrew consent before randomization, and 27 patients were randomized to the experimental arm and 25 to the control arm. The median follow-up time was 21.3 (range 1.3 - 25.6) months. In total, 26 patients were evaluated at the primary end point in the experimental arm and 24 in the control arm (Figure 2). The Consolidated Standards of Reporting Trials (CONSORT) guideline (Checklist 1) was used to present the results.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flowchart diagram.



Patient Characteristics

The median age of the entire population was 65.6 (range 20.3 - 87.8) years; the median age was 64.3 (range 20.3 - 87.8) years in the experimental arm and 69.2 (range 23.9 - 84.4) years in the control arm (*P*=.99). Further, 80% of patients (40/50) had large diffuse B-cell lymphoma, 10% (5/50) had T-cell lymphoma, and 10% (5/50) had Hodgkin lymphoma. The 2 arms were well balanced with respect to age and histological subtype (Table 1).

The median time from initial diagnosis to randomization in the study was 7.7 (range 5.0 - 170.2) months. The median was 7.5 (range 5.3 - 58.2) months for the experimental arm and 7.8 (range 5.0 - 170.2) months for the control arm (*P*=.43).

Rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone were the primary first-line chemotherapeutic regimens. In the experimental arm, 17/26 (65%) patients received a single line of chemotherapy, and 9/26 (35%) received 2 lines. In the control arm, 16/24 (67%) patients received a line of chemotherapy, and 8/24 (33%) received a second line.

Table . Patient characteristics.

	Total	Web application	Control	*P* value[a]
Sex, n (%)				.80
Male	28 (56)	15 (57.7)	13 (54.2)	
Female	22 (44)	11 (42.3)	11 (45.8)	
ECOG[b], n (%)				.61
0	19 (38)	9 (34.6)	10 (41.7)	
1	31 (62)	17 (65.4)	14 (58.3)	
Histology, n (%)				.71
Lymphocyte-rich Hodgkin lymphoma	1 (2.0)	1 (3.8)	0 (0)	
Nodular sclerosis and Hodgkin lymphoma	4 (8)	2 (7.7)	2 (8.3)	
ALK-positive anaplastic large T-cell lymphoma	1 (2)	1 (3.8)	0 (0)	
Angio-immunoblastic T-cell lymphoma	2 (4)	1 (3.8)	1 (4.2)	
Peripheral T-cell lymphoma, NOS[c]	1 (2)	0 (0)	1 (4.2)	
Nasal NK[d] T-cell lymphoma	1 (2)	1 (3.8)	0 (0)	
Centroblastic B-cell lymphoma	5 (10)	4 (15.5)	1 (4.2)	
Diffuse large B-cell, NOS	29 (58)	14 (54)	15 (62.4)	
Primary cutaneous diffuse large B-cell lymphoma, leg-type	1 (2.0)	0 (0)	1 (4)	
Primary mediastinal B-cell lymphoma	1 (2.0)	1 (3.8)	0 (0)	
T-cell-rich large B-cell lymphoma	2 (4.0)	1 (3.8)	1 (4.2)	
Burkitt-like lymphoma	2 (4.0)	0 (0)	2 (8.3)	
Ann-Arbor classification, n (%)				.73
I	2 (4.0)	2 (7.7)	0 (0)	
II	7 (14.0)	3 (11.5)	4 (16.7)	
III	6 (12.0)	3 (11.5)	3 (12.5)	
IV	35 (70.0)	18 (69.3)	17 (70.8)	
Treatment, n				.61
First line				
ABVD[e]	4	2	2	
BEACOPP[f]	1	0	1	
CHOEP[g]	5	3	2	
R-ACVBP[h]	1	0	1	
R-CHOP[i]	27	13	14	
R-DA-EPOCH[j]	1	1	0	
R-MIV[k]	1	1	0	

	Total	Web application	Control	*P* value[a]
Radiotherapy	4	4	0	
Other	31	14	17	
Second line				N/A[l]
BEACOPP	1	0	1	
Brentuximab-bendamustine	1	1	0	
Brentuximab-ICE[m]	1	0	1	
DHAP[n]	1	0	1	
MIV[o]	1	0	1	
R-CHOP	1	0	1	
R-DA-EPOCH	1	1	0	
R-DHAP	1	1	0	
R-ESHAP[p]	1	1	0	
Radiotherapy	2	1	1	
Other	14	9	5	

[a]The *P* value was calculated using chi-sqaure test for qualitative variables, the Wilcoxon test for quantitative variables, and the Fisher test for the lower variables.

[b]ECOG: Eastern Cooperative Oncology Group.

[c]NOS: not otherwise specified.

[d]NK: natural killer.

[e]ABVD: adriamycin, bleomycin, vinblastine, dacarbazine.

[f]BEACOPP: bleomycin, etoposide, adriamycin, cyclophosphamide, vincristine, procarbazine, prednisone.

[g]CHOEP: cyclophosphamide, doxorubicin, vincristine, etoposide, prednisone.

[h]R-ACVBP: rituximab, doxorubicin, cyclophosphamide, vindesine, bleomycin, prednisone.

[i]R-CHOP: rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone.

[j]R-DA-EPOCH: rituximab and dose-adjusted etoposide, prednisone, vincristine, cyclophosphamide, doxorubicin.

[k]R-MIV: rituximab, mitoxantrone, ifosfamide, etoposide.

[l]N/A: not assessed (low variables).

[m]ICE: ifosfamide, carboplatin, etoposide.

[n]DHAP: dexamethasone, cytarabine, cisplatinum.

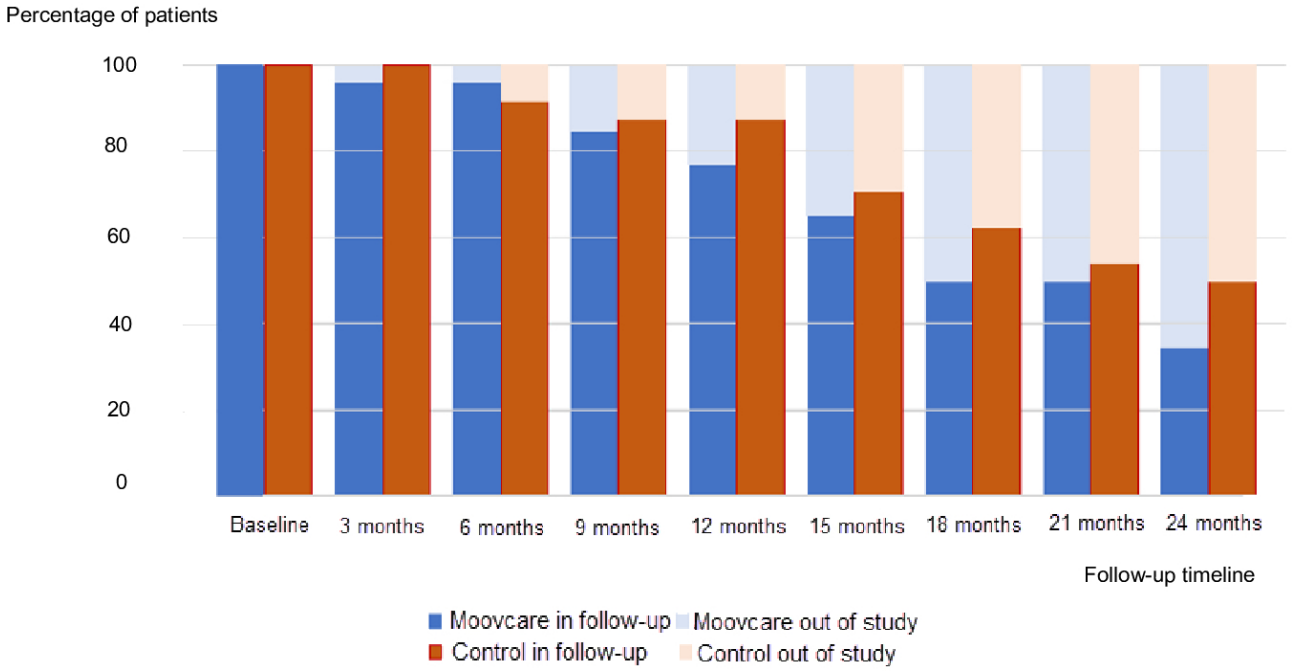
[o]MIV: mitoxantrone, ifosfamide, etoposide.

[p]R-ESHAP: rituximab, etoposide, cytarabine, cisplatin, methylprednisolone.

Follow-Up

In the experimental arm, 25/26 patients (96.1%) were still included in the study at the 6-month follow-up, 20/26 (76.9%) at 12 months, and 9/26 (34.6%) at 24 months. In the control arm, 22/24 (91.6%) were still being followed at 6 months, 21/24 (87.5%) at 12 months, and 12/24 (50%) at 24 months (Figure 3).

Figure 3. Patient follow-up.



The primary reasons for the loss of follow-up were the planned end of the protocol for 20 patients and the premature termination of the study by the sponsor (19 patients; Table 2).

Table . Reasons for discontinuing the study.

	Total, n	Web application, n (%)	Control, n (%)
Death	4	3 (75)	1 (25)
Investigator decision	2	0 (0)	2 (100)
Patient decision	1	1 (100)	0 (0)
End of follow-up	20	8 (40)	12 (60)
Missing patient	2	1 (50)	1 (50)
Premature termination	19	11 (58)	8 (42)
Missing data	2	2 (100)	0 (0)

Events

During the study period, 119 events were reported (Table 3). Most occurred in the experimental arm (83/119, 69.7%) versus 36/119 (30.2%) in the control arm, with a median number of events per patient of 3.5 (range 1-8) in the experimental arm and 1.8 (range 1-6) in the control arm (P=.004). In the experimental arm, 47/83 (56.6%) events were reported directly by the medical team after a scheduled consultation, whereas 36/83 (43.3%) were reported through the web application.

In the control arm, 19/36 (52.7%) events were detected during a scheduled consultation, 2/36 (5.6%) during an unscheduled consultation, 9/36 (25%) during a consultation with another specialist, 3/36 (8.3%) during hospitalization, and 3/36 (8.3%) during a patient call.

Progression and infection were the most frequently reported events. 19 patients relapsed during follow-up, with a median follow-up time of 7 (range 1.9 - 20.6) months; 6 in the web experimental arm and 13 in the control arm (P<.001), with a median follow-up of 7.7 (range 2.8 - 20.6) months and 6.7 (range 1.9 - 16.4) months (P=.94), respectively. 13 patients were treated for relapse; 11 by chemotherapy and 2 by radiotherapy.

30 patients were infected (23 in the experimental arm and 7 in the control arm; P=.59): 5 patients had an influenza infection, and the infectious agent was not reported for the others. 14 patients received treatment; 13 with antibiotics, 1 with antivirals, and 1 with antibiotics and antivirals. The grade of adverse events was not available in 60 out of 119 (50.4%) cases, which limited data interpretation.

Table . Description and classification of events.

	Total, n	Web application, n	Control, n	*P* value[a]
Events	119	83	36	N/A[b]
Number of events per patient	—[c]	3.5	1.8	.004
Grade (progression excluded)	98	75	23	
Grade 1/2	27	19	8	
Grade 3/4	10	7	3	
Unknown	61	49	12	
Progression	21	8	13	—
Suspected by the imaging data	2	2	0	—
Confirmed by biopsy	19	6	13	<.001
Infection	30	23	7	.60
Severity				
Unknown grade	20	15	5	
Grade 1 - 2	6	6	0	
Grade 3 - 4	4	2	2	
Subtype				
Pharyngitis	4	2	2	
Pneumopathy	4	2	2	
Bronchitis	2	1	1	
Influenzae infection	5	5	0	
Anal collection	1	0	1	
Urinary infection	2	2	0	
Gastroenteritis	3	3	0	
Not specified	9	8	1	
Pain	22	18	4	.13
Abdominal	7	7	0	
Thoracic	3	3	0	
Bone and muscle	5	3	2	
Not specified	7	5	2	
Secondary neoplasia	3	1	2	.41
Melanoma	1	1	0	
Colonic adenocarcinoma	1	0	1	
Uterine neoplasm	1	0	1	
Neurological events	5	4	1	.83
Peripheral neuropathy	2	2	0	
Dizziness	3	2	1	
Thrombosis	3	3	0	N/A[d]
Arterial	1	1	0	
Venous	2	2	0	
Bleeding events	1	1	0	N/A[d]

	Total, n	Web application, n	Control, n	*P* value[a]
Epistaxis	1	1	0	
Skin Rash	5	5	0	N/A[d]
Kidney-related events	3	2	1	.99
Increase in creatinine levels	2	1	1	
Kidney lithiasis	1	1	0	
Biological events	3	1	2	.41
Iron deficiency	1	0	1	
Hypercalcemia	1	0	1	
Elevated LDH[e] levels	1	1	0	
Other	23	17	6	.85
Fatigue	4	3	1	
Dyspnea or cough	3	3	0	
Itching	2	2	0	
Edema	4	4	0	
Gynecomastia	1	1	0	
Jugal cyst	1	0	1	
Colonic polyposis	2	0	2	
Hypertension	1	1	0	
Not specified radiological abnormalities	5	3	2	

[a]The *P* value was calculated using the chi-square test for qualitative variables, the Wilcoxon test for quantitative variables, and the Fisher test for the lower variables.

[b]N/A: not assessed (noncomparable values).

[c]Not applicable.

[d]N/A: not assessed (low values).

[e]LDH: lactate dehydrogenase.

Event Management

Overview

Events led to 19 additional medical consultations with the referring hematologist in the experimental arm (alert management resulted in 8 additional medical consultations and 11 without an alert) versus 15 in the control arm (*P*=.99; Table 4). 30 consultations were conducted with other specialists (15 in the experimental arm and 15 in the control arm). For some events, several specialists or referring hematologists were required to manage the patient.

Table . Management of events.

	Total, n	Web application, n	Control, n	*P*-value[a]
Consultation with the oncologist	31	16	15	N/A[b]
Referral to another specialist	30	15	15	N/A[b]
Imaging (scan)	33	22	11	.91
Hospitalization	8	5	3	.78
Progression	4	1	3	
Thoracic pain	1	1	0	
Myocardial infarction	1	1	0	
Balance disorders	1	1	0	
Respiratory distress	1	1	0	
Medical treatment	39	28	11	.94

[a]The *P* value was calculated using the chi-square test for qualitative variables, the Wilcoxon test for quantitative variables, and the Fisher test for the lower variables.

[b]N/A: not assessed.

Emergency hospitalization was required for 8 patients (4 for progression, 1 for myocardial infarction, 1 for respiratory distress, 1 for thoracic pain, and 1 for balance disorder), with no difference between the 2 arms (*P*=.78). 22 scans were performed in the web experimental arm versus 72 in the control arm (61 scanners scheduled for follow-up and 11 not scheduled for events*; P*<.001).

39 events required treatment, with a total of 24 patients receiving medical treatment (16 in the experimental arm and 8 in the control arm; *P*=.11; Table 5). 56 prescriptions were filled (37 in the web experimental arm and 19 in the control arm*; P*=.64).

Table . Pharmacological treatment of events.

	Total, n	Web application, n	Control, n	*P* value[a]
Total	56	37	19	.64
Antiinfection drugs	20	10	10	N/A[b]
Analgesics	5	3	2	N/A[b]
Neurological treatment	2	2	0	N/A[b]
Gastroenterological treatment	7	7	0	N/A[b]
Cardiological treatment	6	4	2	N/A[b]
Anticoagulant treatment	2	1	1	N/A[b]
Systemic corticoids	1	0	1	N/A[b]
Other not specified	13	10	3	N/A[b]

[a]The *P* value was calculated using the chi-square test for qualitative variables, the Wilcoxon test for quantitative variables, and the Fisher test for the lower variables.

[b]N/A: not assessed.

QoL and Depression

The patients were asked to complete the QLCQ-30 questionnaire every 3 months for 1 year. 44 patients completed at least 2 QoL questionnaires during the study; 22 patients per arm (ie, 22/26, 84.6% in the experimental arm and 22/24, 91.6% in the control arm). The higher the score, the poorer the QoL (maximum score: 114). The median scores did not differ between the 2 groups at 12 months with it being; 45 in the experimental arm (range 39-61) and 44 in the control arm (range 30-69*; P*=.94).

Regarding depression, 42 patients completed the questionnaire (21 per arm; ie, 21/26, 80.7% in the experimental arm and 21/24, 87.5% in the control arm]. The score was 1.0 (range 0 - 15) in the experimental arm versus 1.5 (range 0 - 13) in the control arm (*P*=.73).

Satisfaction (Experimental Arm)

In total, 20 patients in the experimental arm completed the satisfaction questionnaire (20/25, 80%). Further, 95% (19/20) of patients who responded to the satisfaction questionnaire were

satisfied and reassured by the application, whereas 90% (18/20) felt better informed.

Survival

Four patients died during the trial; 3 in the experimental arm and 1 in the control arm (*P*=.34). Overall survival at 12 months was 87.1% in the experimental arm (95% CI 65%-95.7%) and 95.2% in the control arm (95% CI 70.7-99.3%; *P*=.32; Figure 4). Progression-free survival at 12 months was 83.2% in the experimental arm (95% CI 61%-93.3%) and 68.5% in the control arm (95% CI 44.9%-83.6%; *P*=.27; Figure 5).

Figure 4. Overall survival. Exp.: experimental.

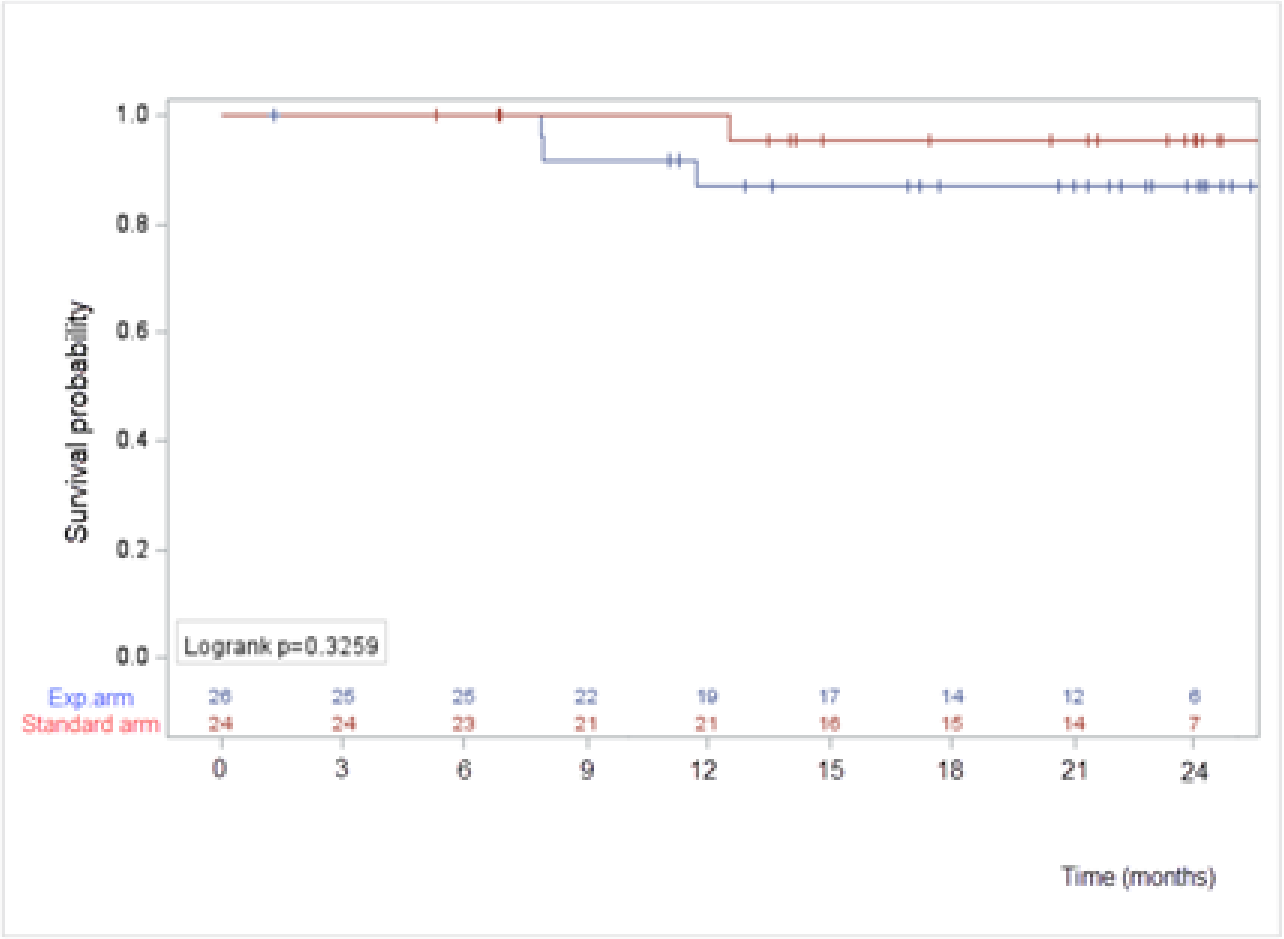
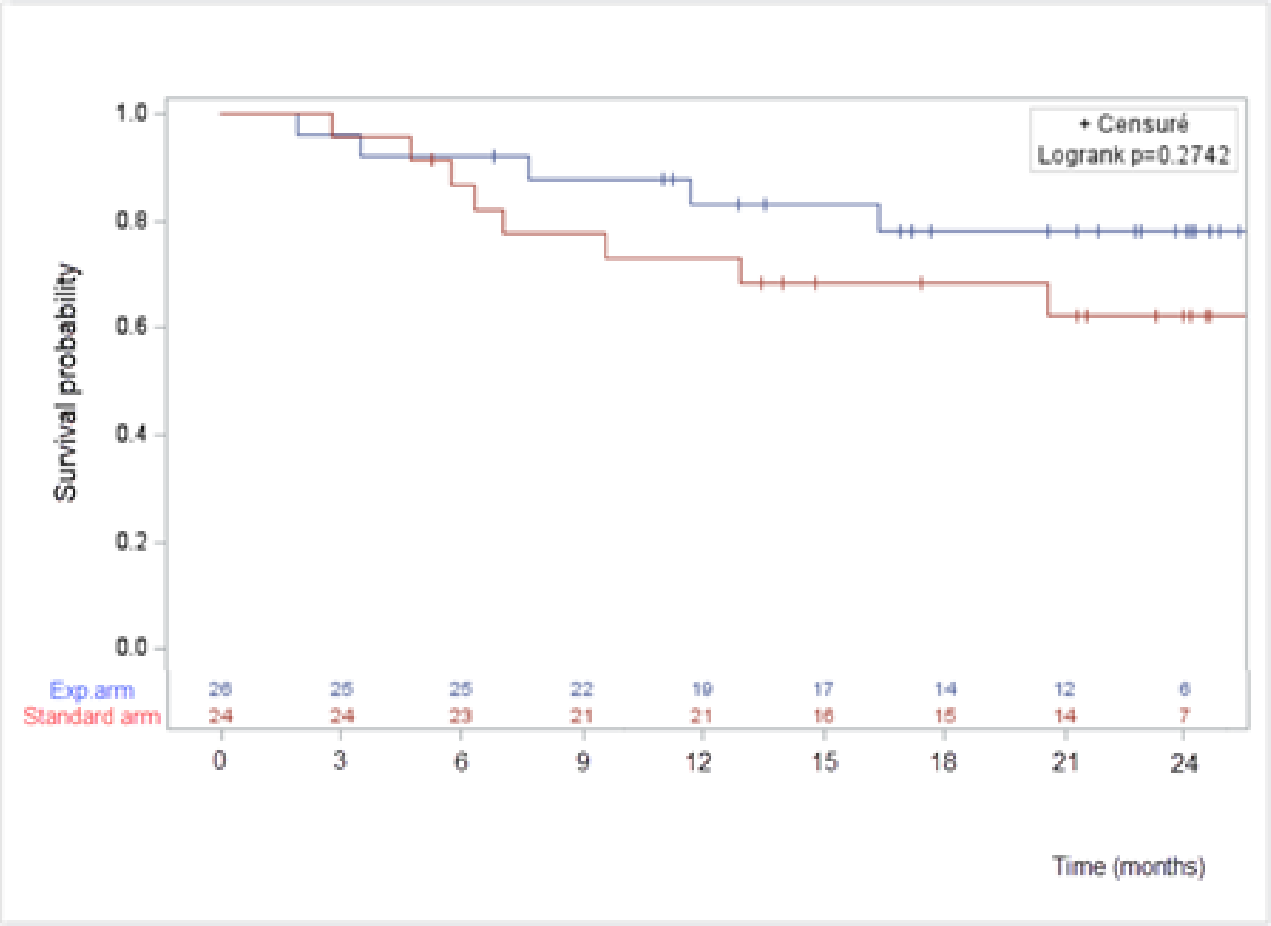


Figure 5. Progression-free survival. Exp.: experimental.



Protocol Deviations and Technical Failures in the Experimental Arm

Three alerts were not handled by the care team within the required time frame and were subsequently classified as minor (grade 1).

The automatic sending of questionnaires was stopped for 9/26 (35%) patients. For 3 patients, the questionnaires were sent in paper form. For the other 6, no solution could be found despite repeated interventions by the electronic application's technical department. One of these patients suffered 2 major events: myocardial infarction and relapse. These 2 events were not reported in the electronic application. Because of these technical problems, compliance could not be assessed.

Outcomes

An interim analysis was performed based on the protocol when the first 40 patients reached 6 months of follow-up. The results of this analysis did not reject the null hypothesis, which stated that there was no difference in the diagnosis of events between the 2 arms. The Sentinel Lymphoma Study Committee met on March 11, 2021, to oversee the analysis of the primary outcome, which indicated no difference in the diagnosis of significant events. A decision was made at the end of the meeting to discontinue the trial early. Based on the protocol, the study was terminated on March 15, 2021, following the sponsor's decision.

Discussion

Principal Findings

In this study, there is no difference in the occurrence of significant events between the 2 arms (median number per patient of 3.5 in the experimental arm and 1.8 in the control arm; P=.004). Progression, infection, and pain were the most frequently reported events. Patient satisfaction was very high and the patients felt reassured to have electronic monitoring. The patients included in the experimental arm underwent fewer scans compared with those in the control arm, without impacting overall survival, despite a short follow-up (P<.001).

Strengths and Limitations

First, the primary outcome of a 30% superiority of reporting significant events in the experimental arm has been overly optimistic. Thus, reducing the end point would have led to a substantial increase in the number of included patients. The number of events was probably not the best criterion for evaluating the effectiveness of remote monitoring. An improvement in QoL or a reduction in the risk of relapse would likely have been more relevant [13,15].

Second, technical problems with the web application occurred (electronic questionnaires not received, with major biases in event reporting). The incidents were not expected because of the experience of the software developer (Moovcare, Sivan Innovation, Ltd); however, there was a change in the technical

team between this publication on lung cancer and the start of our study [11-13]. The blocking of automatic questionnaires required 42 direct interventions by clinical study investigators with calls to the patient (firewalls and spam). IT support did not correct these recurring anomalies, despite the changes to the application in November 2019 (5 patients were included in the experimental arm after this date). These operational problems resulted in 15 meetings without resolution of the problems, with an average response time of 4.6 months from technical support (frequent changes to contact persons). As a result, the events in the experimental arm were not reported correctly, leading to study bias. The final report has been sent to the Agence Nationale de Sécurité du Médicament on June 29, 2021.

Finally, only 43% of the events were declared by the application in the experimental arm, which would indicate a problem with patient training.

Comparison With Prior Work

PROs are underestimated in clinical practice and trials for patients with lymphoma and have most often consisted of paper-based QoL questionnaires [21]. The measurement of PROs via electronic questionnaires has subsequently been evaluated in randomized trials with a low representation of patients with lymphoid malignancies [22]. However, Maguire et al [23] demonstrated that real-time electronic monitoring of symptoms was feasible during an initial chemotherapy cycle in patients with solid tumors and lymphoma, with a reduction in the intensity of side effects and anxiety, compared with a control group.

Future Directions

Proposals for the future include improving the study design by limiting the patient population to a single type of lymphoma, defining the objective to demonstrate an improvement in morbidity and possibly reduce cost, and guaranteeing the reliability of the electronic application. PROMs require standardization of analysis for comparative purposes. Therefore, it is necessary to regulate the use of health care applications to avoid malfunction and abuse [24]. Denis and Krakowski [25] defined 20 criteria of effectiveness, safety, and functionality that should govern the development of ePROMs. Telemonitoring applications should strive to improve patient compliance and prevent patients from dropping out due to a lack of understanding or receiving excessive notifications [26]. Telemonitoring applications must evolve with therapeutic innovations and be regularly reevaluated to demonstrate their long-term benefits on a larger scale [27]. Finally, guidelines are recommended for the design of clinical trials to evaluate the effectiveness of electronic solutions [28-30].

Conclusions

Sentinel Lymphoma is the first randomized phase 3 trial to evaluate the effect of remote monitoring on the detection of significant events in patients with hematological malignancies. Progression, infection, and pain were the most frequently reported events. Despite a high number of events (83 in the experimental arm against 36 in the control arm), the difference was not significant. A more targeted population, a more precise objective, and better security for remote surveillance solutions are recommended for subsequent projects.

Acknowledgments

The authors would like to thank the patients and their families, as well as Antoine Giraud for his advice. The authors would like to thank Enago for the English language review. This study was funded by Takeda.

Data Availability

All data from the study are freely available upon request to the corresponding author.

Authors' Contributions

KLD, MB, and ALS designed the study. ALS collected, analyzed, and interpreted the data. KLD drafted the first version of the manuscript. All authors participated in the preparation and revisions of the manuscript and approved the final version of the manuscript for submission.

Conflicts of Interest

KLD has received consulting fees from AstraZeneca, Abbvie, Incyte Janssen-Cilag, and Gilead.

Checklist 1
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File, 1143 KB - cancer_v11i1e65960_app1.pdf]

References

1. Vercellino L, Di Blasi R, Kanoun S, et al. Predictive factors of early progression after CAR T-cell therapy in relapsed/refractory diffuse large B-cell lymphoma. Blood Adv 2020 Nov 24;4(22):5607-5615. [doi: 10.1182/bloodadvances.2020003001] [Medline: 33180899]

2. Edelman MJ, Meyers FJ, Siegel D. The utility of follow-up testing after curative cancer therapy. A critical review and economic analysis. J Gen Intern Med 1997 May;12(5):318-331. [doi: 10.1046/j.1525-1497.1997.012005318.x] [Medline: 9159703]
3. Voss SD, Chen L, Constine LS, et al. Surveillance computed tomography imaging and detection of relapse in intermediate- and advanced-stage pediatric Hodgkin's lymphoma: a report from the Children's Oncology Group. J Clin Oncol 2012 Jul 20;30(21):2635-2640. [doi: 10.1200/JCO.2011.40.7841]
4. Xavier MF, Schuster SJ, Andreadis C, Downs L, Diccion B, Nasta SD. Detection of relapse in diffuse large B-cell lymphoma (DLBCL) and Hodgkin's lymphoma (HL): observations and implications for post-remission radiologic surveillance. Blood 2006 Nov 1;108(11):2428-2428. [doi: 10.1182/blood.V108.11.2428.2428]
5. Thompson CA, Charlson ME, Schenkein E, et al. Surveillance CT scans are a source of anxiety and fear of recurrence in long-term lymphoma survivors. Ann Oncol 2010 Nov;21(11):2262-2266. [doi: 10.1093/annonc/mdq215] [Medline: 20423914]
6. Armitage JO. Who benefits from surveillance imaging? J Clin Oncol 2012 Jul 20;30(21):2579-2580. [doi: 10.1200/JCO.2012.42.6189] [Medline: 22689799]
7. Brenner DJ, Hall EJ. Computed tomography—an increasing source of radiation exposure. N Engl J Med 2007 Nov 29;357(22):2277-2284. [doi: 10.1056/NEJMra072149] [Medline: 18046031]
8. Fazel R, Krumholz HM, Wang Y, et al. Exposure to low-dose ionizing radiation from medical imaging procedures. N Engl J Med 2009 Aug 27;361(9):849-857. [doi: 10.1056/NEJMoa0901249] [Medline: 19710483]
9. Frank MJ, Hossain NM, Bukhari A, et al. Monitoring of circulating tumor DNA improves early relapse detection after axicabtagene ciloleucel infusion in large B-cell lymphoma: results of a prospective multi-institutional trial. J Clin Oncol 2021 Sep 20;39(27):3034-3043. [doi: 10.1200/JCO.21.00377] [Medline: 34133196]
10. NCI common terminology criteria for adverse events (CTCAE) version 403. National Cancer Institute. URL: https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03 [accessed 2025-04-30]
11. Denis F, Viger L, Charron A, Voog E, Letellier C. Detecting lung cancer relapse using self-evaluation forms weekly filled at home: The Sentinel follow-up. Support Care Cancer 2014 Jan;22(1):79-85. [doi: 10.1007/s00520-013-1954-9]
12. Denis F, Viger L, Charron A, et al. Detection of lung cancer relapse using self-reported symptoms transmitted via an internet web-application: pilot study of the Sentinel follow-up. Support Care Cancer 2014 Jun;22(6):1467-1473. [doi: 10.1007/s00520-013-2111-1]
13. Denis F, Basch E, Septans AL, et al. Two-year survival comparing web-based symptom monitoring vs routine surveillance following treatment for lung cancer. JAMA 2019 Jan 22;321(3):306. [doi: 10.1001/jama.2018.18085]
14. Basch E, Deal AM, Dueck AC, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017 Jul 11;318(2):197-198. [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
15. Mody GN, Stover AM, Wang M, et al. Electronic patient-reported outcomes monitoring during lung cancer chemotherapy: a nested cohort within the PRO-TECT pragmatic trial (AFT-39). Lung Cancer 2021 Dec;162:1-8. [doi: 10.1016/j.lungcan.2021.09.020]
16. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 3;85(5):365-376. [doi: 10.1093/jnci/85.5.365]
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
18. Guide d'utilisation de l'application moovcare à destination des professionnels de santé [article in french]. : Moovcare; 2024 URL: https://www.moovcare.com/fr/clinical/
19. Ranque S, Badiaga S, Delmont J, Brouqui P. Triangular test applied to the clinical trial of azithromycin against relapses in Plasmodium vivax infections. Malar J 2002 Nov 12;1:13. [doi: 10.1186/1475-2875-1-13] [Medline: 12473182]
20. Bellissant E, Benichou J, Chastang C. A comparison of methods for phase II cancer clinical trials: advantages of the triangular test, a group sequential method. Lung Cancer 1994 Mar;10 Suppl 1:S105-S115. [doi: 10.1016/0169-5002(94)91673-x] [Medline: 8087499]
21. Oliva EN, Ionova T, Laane E, et al. Patient-reported outcomes in Hodgkin lymphoma trials: a systematic review. Front Oncol 2024;14:1353101. [doi: 10.3389/fonc.2024.1353101] [Medline: 38544841]
22. Basch E, Schrag D, Henson S, et al. Effect of electronic symptom monitoring on patient-reported outcomes among patients with metastatic cancer: a randomized clinical trial. JAMA 2022 Jun 28;327(24):2413-2422. [doi: 10.1001/jama.2022.9265] [Medline: 35661856]
23. Maguire R, McCann L, Kotronoulas G, et al. Real time remote symptom monitoring during chemotherapy for cancer: European multicentre randomised controlled trial (eSMART). BMJ 2021 Jul 21;374:n1647. [doi: 10.1136/bmj.n1647] [Medline: 34289996]
24. Pe M, Alanya A, Falk RS, et al. Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials-Innovative Medicines Initiative (SISAQOL-IMI): stakeholder views, objectives, and procedures. Lancet Oncol 2023 Jun;24(6):e270-e283. [doi: 10.1016/S1470-2045(23)00157-2]

25. Denis F, Krakowski I. How should oncologists choose an electronic patient-reported outcome system for remote monitoring of patients with cancer? J Med Internet Res 2021 Sep 9;23(9):e30549. [doi: 10.2196/30549] [Medline: 34499046]
26. Sacchi L, Veggiotti N, Girani E, Tibollo V, Quaglini S. Personalising symptoms reporting in telemonitoring applications for cancer patients. Stud Health Technol Inform 2022 May 25;294:900-904. [doi: 10.3233/SHTI220621] [Medline: 35612239]
27. Wagneur N, Callier P, Zeitoun JD, Silber D, Sabatier R, Denis F. Assessing a new prescreening score for the simplified evaluation of the clinical quality and relevance of eHealth apps: instrument validation study. J Med Internet Res 2022 Jul 5;24(7):e39590. [doi: 10.2196/39590] [Medline: 35788102]
28. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: The SPIRIT-PRO Extension. JAMA 2018 Feb 6;319(5):483-494. [doi: 10.1001/jama.2017.21903] [Medline: 29411037]
29. Bolinger E, Tyl B. Key considerations for designing clinical studies to evaluate digital health solutions. J Med Internet Res 2024 Jun 17;26:e54518. [doi: 10.2196/54518] [Medline: 38885020]
30. Di Maio M, Basch E, Denis F, et al. The role of patient-reported outcome measures in the continuum of cancer clinical care: ESMO Clinical Practice Guideline. Ann Oncol 2022 Sep;33(9):878-892. [doi: 10.1016/j.annonc.2022.04.007] [Medline: 35462007]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CT: computed tomography
e-CRF: electronic case report form
ePROM: electronic patient-reported outcome measure
LDH: lactate dehydrogenase
PHQ-9: Patient Health Questionnaire-9
PRO: patient-reported outcome
PROM: patient-reported outcome measure
QLQ-C30: Quality of Life Questionnaire of Cancer Patients
QoL: quality of life

Edited by N Cahill; submitted 31.08.24; peer-reviewed by R Mu, S Oerlemans; revised version received 15.03.25; accepted 31.03.25; published 06.05.25.

Please cite as:
Le Dû K, Chauchet A, Sadot-Lebouvier S, Fitoussi O, Fontanet B, Saint-Lezer A, Maloisel F, Rossi C, Carras S, Parcelier A, Balavoine M, Septans AL
Comparison of Electronic Surveillance With Routine Monitoring for Patients With Lymphoma at High Risk of Relapse: Prospective Randomized Controlled Phase 3 Trial (Sentinel Lymphoma)
JMIR Cancer 2025;11:e65960
URL: https://cancer.jmir.org/2025/1/e65960
doi:10.2196/65960

Preliminary Effectiveness of a Telehealth-Delivered Exercise Program in Older Adults Living With and Beyond Cancer: Retrospective Study

Emily R Dunston[1], MS, PhD; Sonal Oza[2], MD; Yang Bai[1], PhD; Maria Newton[1], PhD; Leslie Podlog[3,4], PhD; Kish Larson[5], BS; Darren Walker[5], BS; Rebecca W Zingg[5,6], DO; Pamela A Hansen[5,6], MD; Adriana M Coletta[1,7], MS, PhD, RD

[1]Department of Health and Kinesiology, University of Utah, Salt Lake City, UT, United States

[2]Department of Rehabilitation Medicine, Emory University School of Medicine, Atlanta, GA, United States

[3]Faculty of Medicine, School of Kinesiology and Physical Activity Sciences, Université de Montréal, Montréal, QC, Canada

[4]CHU Sainte-Justine Research Center, Montréal, QC, Canada

[5]Huntsman Cancer Institute, Salt Lake City, UT, United States

[6]Division of Physical Medicine and Rehabilitation, University of Utah, Salt Lake City, UT, United States

[7]Cancer Control and Population Sciences Program, Huntsman Cancer Institute, University of Utah, 2000 Circle of Hope Drive, Research South Building Rm 4747, Salt Lake City, UT, United States

Corresponding Author:
Adriana M Coletta, MS, PhD, RD
Department of Health and Kinesiology, University of Utah, Salt Lake City, UT, United States

Abstract

Background: Exercise can attenuate the deleterious combined effects of cancer treatment and aging among older adults with cancer, yet exercise participation is low. Telehealth exercise may improve exercise engagement by decreasing time and transportation barriers; however, the utility of telehealth exercise among older adults with cancer is not well established.

Objective: We aimed to evaluate the preliminary effectiveness of a one-on-one, supervised telehealth exercise program on physical function, muscular endurance, balance, and flexibility among older adults with cancer.

Methods: In this retrospective study, we analyzed electronic health record data collected from the Personal Optimism With Exercise Recovery clinical exercise program delivered via telehealth among older adults with cancer (≥65 y) who completed a virtual initial program telehealth assessment between March 2020 and December 2021. The virtual initial assessment included the following measures: 30-second chair stand test, 30-second maximum push-up test, 2-minute standing march, single leg stance, plank, chair sit and reach, shoulder range of motion, and the clock test. All baseline measures were repeated after 12-weeks of telehealth exercise. Change scores were calculated for all assessments and compared to minimal clinically important difference (MCID) values for assessments with published MCIDs. Paired samples *t* tests (2-tailed) were conducted to determine change in assessment outcomes.

Results: Older adults with cancer who chose to participate in the telehealth exercise program (N=68) were 71.8 (SD 5.3) years of age on average (range 65 - 92 y). The 3 most common cancer types in this sample were breast (n=13), prostate (n=13), and multiple myeloma (n=8). All cancer stages were represented in this sample with stage II (n=16, 23.5%) and III (n=18, 26.5%) being the most common. A follow-up telehealth assessment was completed by 29.4% (n=20) of older adults with cancer. Among those who completed a follow-up telehealth assessment, there were significant increases in the 30-second chair stand (n=19; mean change +2.00 repetitions, 95% CI 0.12 to 3.88) and 30-second maximum push-up scores (n=20; mean change +2.85 repetitions, 95% CI 1.60 to 4.11). There were no significant differences for the 2-minute standing march, plank, single leg stance, sit and reach, shoulder mobility, or clock test (P>.05). Nine (47.3%) older adults with cancer had a change in 30-second chair stand scores greater than the MCID of 2 repetitions.

Conclusions: Our findings suggest a one-on-one, supervised telehealth exercise program may positively influence measures of physical function, muscular endurance, balance, and flexibility among older adults with cancer, but more adequately powered trials are needed to confirm these findings.

(JMIR Cancer 2025;11:e56718) doi:10.2196/56718

KEYWORDS

physical activity; physical function; telerehabilitation; remote exercise; digital health; cancer survivors; older adults; smartphone

Introduction

Adults aged 65 years and older currently account for 67% of cancer survivors (eg, individuals living with and beyond a cancer diagnosis) in the United States [1]. By 2040 it is projected that 73% of cancer survivors in the United States will be aged 65 years and older [2]. Cancer treatment compounds the normal effects of aging resulting in an accelerated aging effect [3]. A hallmark characteristic of accelerated aging is poor physical functioning [3]. Older adults with cancer experience worse physical function than their younger counterparts [4] and older adults without cancer [5,6]. Physical function plays a critical role in the health of older adults with cancer and poor function is associated with decreased cancer survival [7], increased all-cause mortality [8], and increased symptom severity [9].

Regular participation in exercise is one strategy to help mitigate declines in physical function among cancer survivors of all ages. Among older adults with cancer, individualized, in-person supervised exercise programming, including combined aerobic and resistance training, for at least 12-weeks significantly improves physical function [10,11], quality of life [12], muscular strength [10,11], aerobic endurance [10,11], and symptoms of anxiety and depression [11]. Despite the numerous benefits of exercise for older adults with cancer, participation in exercise in this population is low with only 12% of older cancer survivors meeting both the aerobic and strength training guidelines [13]. Reasons for low engagement among this population include: lack of available exercise programming in convenient locations [14], transportation concerns [14,15], lack of time [14,16], physical symptoms (eg, fatigue) [14,16,17], and comorbidities [14,17]. Strategies to reduce barriers to participating in exercise among older adults with cancer are needed to improve exercise engagement and physical function in this population.

Delivering exercise programs using telehealth is a useful strategy in attenuating these barriers. Telehealth delivery of exercise detaches the exercise program from a physical location, resulting in exercise engagement in a more convenient location, such as the home, eliminating the need for travel, and reducing overall time commitment [18]. Telehealth delivered exercise can also lower the cost of participation as participants do not need to pay for transportation or parking [18]. After transitioning two trials from in-person to telehealth exercise, due to the impact of the COVID-19 pandemic on in-person research, Winters-Stone et al [19] observed better adherence and retention for telehealth exercise compared to in-person exercise among adult cancer survivors of all ages. In addition to addressing these barriers, supervised, telehealth-delivered exercise programs among adult cancer survivors of all ages have demonstrated improvements in physical symptoms and comorbidities such as: physical function [20,21], aerobic endurance [20,22], muscular endurance [20,22], and fatigue [22]. Specific to older adults living with cancer, telehealth delivery of exercise programming is considered acceptable [23], feasible [24], and safe [25]. Moreover, older cancer survivors view telehealth delivery of exercise positively and report limited technology related barriers to telehealth exercise participation [26]. Barriers and facilitators to participating in telehealth exercise reported by older cancer survivors are similar to those reported by their younger counterparts [26]. To our knowledge, only two studies to date have evaluated effectiveness of supervised telehealth exercise programming (ie, delivered via telehealth in real time) exclusively in older adults with cancer [24,27]. Both trials delivered group resistance training programs and observed significant improvements in markers of physical function after participating in the program [24,27]. However, little is known regarding the effectiveness of one-on-one telehealth supervised exercise in older adults with cancer. Given the dearth of research, we sought to address the issue in this investigation.

The purpose of this investigation was to explore the preliminary effectiveness of a one-on-one, supervised telehealth clinical exercise oncology program among older adults with cancer on physical function, muscular endurance, balance, and flexibility. We hypothesized that participation in telehealth exercise would result in a statistically significant improvement in physical function, muscular endurance, aerobic endurance, balance, and flexibility among older adults with cancer.

Methods

Study Design and Sample

This was a retrospective analysis of electronic health record data collected between March 2020 and December 2021 from the Huntsman Cancer Institute (HCI) at the University of Utah's clinical exercise oncology program, the Personal Optimism With Exercise Recovery (POWER) program. This study was approved by the University of Utah Institutional Review Board (IRB_00072431). To be included in this analysis participants must have met the following inclusion criteria: (1) ≥65 years of age, (2) diagnosis of invasive cancer, and (3) completion of an initial POWER program assessment via telehealth. Demographic and clinical data including age, sex, race, ethnicity, cancer site, cancer stage, and cancer treatment history, were pulled from the medical record. Initial and follow-up assessment data were abstracted from the POWER program clinical database by a trained researcher (ERD) with support from certified exercise physiologists within our hospital-based exercise oncology program using a study specific spreadsheet developed in partnership with this study's team. Data were cleaned to ensure all measures were within a physiologically reasonable range and units were consistent within measures (eg, all plank assessments were reported in seconds). Cancer treatment history from manual data abstraction was verified with the electronic health record.

Exercise Program

The POWER program is a hospital-based exercise oncology program embedded into clinical practice at the HCI. Details of this clinical program have been previously published [28]; therefore, only pertinent details will be discussed here. While the program has traditionally been offered both in-person and via telehealth, the POWER program shifted to exclusive telehealth delivery due to the COVID-19 pandemic in March 2020 and continued to operate primarily via telehealth through December 2021. Anyone seeking care at the HCI is eligible to participate in the POWER program and patients can enroll in the program through self-referral or physician referral.

POWER provides personalized exercise prescriptions, including both aerobic and resistance training, to program participants based on an initial assessment conducted by a physiatrist and certified exercise physiologist with expertise in cancer via telehealth. The typical length of the program was 12-weeks, but varied based on participant preference. After about 12-weeks, participants were encouraged to complete a telehealth follow-up assessment to evaluate their progress (ie, reassess all baseline measures) and revise the exercise prescription to promote continued progress. Ultimately, the POWER program aims to help survivors become comfortable and capable of safely engaging in exercise independently.

The exercise prescription was individualized to each participant's needs and was informed by the initial telehealth assessment which included a review of medical and cancer treatment history, physical examination, review of current exercise behavior, and assessment of physical function, muscular endurance, and flexibility. Following the initial assessment each participant met with a certified exercise physiologist twice weekly, via telehealth, for the duration of their program, for a supervised, 60-minute resistance training session. Body weight training and resistance bands were the primary mode of resistance training delivery; however, the resistance training program may have also included resistance machines or free weights per participant access and preference. No equipment was provided to participants by the exercise program. Prescribed aerobic exercise was completed unsupervised by each participant. The goal of each participant's program was to work toward meeting the physical activity guidelines for cancer survivors [29].

Participants accessed the telehealth exercise visits directly through their online patient portal using any electronic device that was capable of video calls (eg, smartphone, tablet, laptop, etc). Telehealth visits were conducted directly through electronic health records (Epic Systems Corporation) which allow certified exercise physiologists easy access to the participants address, contact information, and emergency contacts. Participants' location for each telehealth exercise session and contact information, in case the telehealth session was disconnected, was verified by the certified exercise physiologist at the start of each session. While survivors had an out-of-pocket cost of approximately US $8 per telehealth exercise training visit, the baseline and follow-up assessments were covered by medical insurance reimbursement.

Measures

Overview

The following measures were included in the telehealth initial and follow-up assessments in the POWER program. When developing the telehealth assessment procedures, decisions about which measures to include were based on the feasibility of carrying out measures in a telehealth format and alignment with the in-person POWER program assessment [28]. When administering the telehealth assessments, the video camera angle was adjusted for each assessment so that the certified exercise physiologist could observe the full range of motion and ensure proper form was being used.

30-Second Chair Stand Test

Lower extremity function was evaluated with the 30-second chair stand test. Participants stood from a seated position, with arms crossed across their chest, and were instructed to stand up and sit down as many times as they could in 30-seconds [30]. The number of repetitions (ie, return to seated) completed in 30-seconds were recorded. Repetitions were counted using consistent methods across assessments and assessors to optimize the reliability of this assessment. The 30-second chair stand test has been shown to be a good predictor of lower extremity function in older adults [31] and safe to conduct using telehealth in adults with cancer [32]. Moreover, the 30-second chair stand test has good test-retest reliability in older adults with cancer (intraclass correlation [ICC]=0.89) [33,34]. A minimal clinically important difference (MCID) of 2.0 has been established for the 30-second chair stand test [35].

30-Second Maximum Push-Up Test

Muscular endurance was assessed using the 30-second push-up test. The starting position for push-ups was with the hands on the floor approximately shoulder width apart and arms straight. Participants were instructed to lower themselves down toward the floor until their chest was one fist width above the floor and then return to the starting position; this is one repetition. Participants were asked to complete as many push-ups as possible in 30-seconds. If the participant was unable to perform a standard push-up (on toes), they were able to modify by starting on their knees or performing wall push-ups depending on ability [36]. Any modifications made at baseline were replicated at follow-up.

2-Minute Standing March

Aerobic endurance was assessed using the 2-minute standing march test. Participants stepped in place with a step height no lower than the midpoint between the patella and iliac crest. The number of steps (right and left equals one) completed in 2-minutes were recorded. If necessary, participants could use one hand on a counter-top or a chair to assist with balance. The 2-minute standing march has been shown to be a good alternative to the 6-minute walk test [37,38] with strong test-retest reliability (ICC=0.99) when assessed among older adults via telehealth [39].

Single Leg Stance

Balance was assessed using a single leg stance. Participants were instructed to lift one foot off of the ground and balance on one leg without holding onto anything for support for as long as possible with their eyes open. The single leg stance was performed once on each leg. Time balancing without assistance (from hands or the other foot) was recorded for each leg. No maximum time cap was imposed for the single leg stance. The single leg stance test has demonstrated good reliability (ICC=0.86) among older adults [40].

Plank

Participants were asked to hold a forearm plank on either their toes or knees, self-selected based on their ability, for as long as they were able to assess torso muscular endurance. Each participant was instructed to keep their elbows directly under

their shoulders with forearms extended forward and a neutral spine and neck. The variation (ie, knees or toes) participants selected and total time participants were able to hold the plank were recorded. The plank assessment was not performed in cases where contraindications, such as cardiovascular concerns or upper extremity injuries, were present. Telehealth plank assessment has demonstrated good reliability (ICC=0.97) among adults [41].

Chair Sit and Reach

Hamstring flexibility was assessed using the chair sit and reach test. Participants sat on a chair near the front edge of the seat with one leg extended (ie, heel on the floor and foot dorsiflexed at approximately 90 degrees) and the other leg bent with the sole of the foot flat on the floor. Then they were asked to place one hand on top of the other with palms facing down. Participants were then instructed to slowly bend forward at the hips, keeping their back flat, as they reached down the extended leg as far as they could. A score was assigned based on how far participants were able to reach: a 2 for the toes, 1.5 for the ankle, 1.0 for the shin, 0 for anything above the shin. The chair sit and reach test has demonstrated good reliability (ICC=0.95) and validity among older adults [35,42].

Shoulder Range of Motion

Range of motion in the shoulder joint was assessed by measuring shoulder flexion, shoulder extension, and shoulder abduction. Range of motion for each movement was observed and visually estimated to the nearest 10 degrees during the telehealth initial assessment. Visual estimation of shoulder range of motion has demonstrated acceptable reliability (ICC=0.57 - 0.70) among adults [43].

Clock Test

The clock test is a modified back scratch test used to assess shoulder internal rotation. Participants were instructed to reach behind their back with their palm facing out with the goal of reaching their hand as far up their back as possible. The test was conducted on both the right and left sides. The test is scored by visually estimating the position of the arm in correspondence to a position on the face of a clock during the telehealth initial assessment. Scores range from six to eleven on the right and six to one on the left with eleven and one indicating the highest levels of shoulder flexibility, respectively.

Statistical Analysis

Descriptive statistics were reported as means and SDs or medians and IQRs for continuous variables and frequencies and percentages for categorical variables. Differences in age, BMI, and continuous initial assessment variables between older adults with cancer who did and did not complete a follow-up assessment were determined using independent samples *t* tests. Differences in categorical demographic, clinical, and initial assessment variables were assessed using chi-square tests. Among the older adults with cancer that completed a follow-up assessment, mean change variables were computed as the difference between the follow-up and baseline values. Missing assessment data were excluded case-wise to maximize the sample size for each variable. Change scores were compared to values considered to be the MCID. The 30-second chair stand test was the only assessment with a published MCID value [35]. Paired samples *t* tests were conducted to determine if there were significant differences in assessment outcomes following the exercise intervention. Cohen *d* effect sizes are reported as an indicator of effect size. A Cohen *d* of 0.2 was considered a small effect, 0.5 was considered medium, and 0.8 was considered large. For categorical outcomes mean change scores and 95% CIs were calculated to determine change across the intervention. All data were analyzed in SPSS (version 29.0; IBM Corp).

Ethical Considerations

The protocol and waiver of informed consent was approved by the University of Utah Institutional Review Board (IRB_00072431) in accordance with the Declaration of Helsinki. All data presented were deidentified using study identification numbers prior to analysis. Compensation was not included for this study.

Results

Participants

A total of 68 older adults with cancer completed an initial assessment via telehealth and participated in the POWER program between March 2020 and December 2021. Older adults with cancer who participated in POWER via telehealth were 71.8 (SD 5.3) years of age on average (range 65 - 92 y) and had a median BMI of 26.7 kg/m^2 (IQR 7.3; Table 1). Most older adults with cancer were female (n=45, 66.2%) and were not actively receiving treatment during their participation in POWER (n=40, 58.8%). The most common cancer types among older adults were breast (n=18, 26.5%), prostate (n=13, 19.1%), and multiple myeloma (n=8, 11.8%).

Table . Participant demographic and clinical characteristics.

Variable			Total sample (N=68)	Follow-up assessment completed (n=20)	Follow-up assessment not completed (n=48)	Baseline differences between groups, *P* value
Age (years), mean (SD)			71.8 (5.3)	72.8 (4.6)	71.4 (5.5)	.33
BMI (kg/m^2), median (IQR)			26.7 (7.3)	26.7 (7.3)	26.2 (7.1)	.62
Sex, n (%)						.90
	Male		23 (33.8)	7 (35)	16 (33.3)	
	Female		45 (66.2)	13 (65)	32 (66.7)	
Race, n (%)						—[a]
	White		68 (100)	20 (100)	48 (100)	
Ethnicity, n (%)						.25
	Non-Hispanic		65 (95.6)	20 (100)	45 (93.8)	
	Hispanic		3 (4.4)	0 (0)	3 (6.3)	
Cancer stage, n (%)						.72
	I		12 (17.6)	3 (15)	9 (18.8)	
	II		16 (23.5)	3 (15)	13 (27.1)	
	III		18 (26.5)	6 (30)	12 (25)	
	IV		10 (14.7)	3 (15)	7 (14.6)	
	Unstaged		11 (16.2)	5 (25)	6 (12.5)	
	Unknown		1 (1.5)	0 (0)	1 (2.1)	
Active treatment[b], n (%)						.04[c]
	Yes		28 (41.2)	12 (60)	16 (33.3)	
	No		40 (58.8)	8 (40)	32 (66.7)	
Treatment history[d], n (%)						
	Chemotherapy					.04[c]
		Yes	38 (55.9)	15 (75)	23 (47.9)	
		No	30 (44.1)	5 (25)	25 (52.1)	
	Hormone therapy					.78
		Yes	29 (42.6)	8 (40)	21 (43.8)	
		No	39 (57.4)	12 (60)	27 (56.3)	
	Immunotherapy					.67
		Yes	18 (26.5)	6 (30)	12 (25)	
		No	50 (73.5)	14 (70)	36 (75)	
	Surgery					.81
		Yes	49 (72.1)	14 (70)	35 (72.9)	
		No	19 (27.9)	6 (30)	13 (27.1)	
	Radiation					.20
		Yes	26 (38.2)	10 (50)	16 (33.3)	
		No	42 (61.8)	10 (50)	32 (66.7)	
Number of treatment types, n (%)						
	None		3 (4.4)	0 (0)	3 (6.2)	—
	Unimodal[e]		12 (17.7)	3 (15)	9 (18.8)	.71

Variable		Total sample (N=68)	Follow-up assessment completed (n=20)	Follow-up assessment not completed (n=48)	Baseline differences between groups, *P* value
	Bimodal[f]	19 (27.9)	6 (30)	13 (27.1)	.81
	Multimodal[g]	34 (50)	11 (55)	23 (47.9)	.60
Cancer type, n (%)					.62
	Bladder	2 (2.9)	0 (0)	2 (4.2)	
	Brain	1 (1.5)	1 (5)	0 (0)	
	Breast	18 (26.5)	5 (25)	13 (27.1)	
	Colon	2 (2.9)	0 (0)	2 (4.2)	
	Endometrial	3 (4.4)	1 (5)	2 (4.2)	
	Fallopian tube	2 (2.9)	0 (0)	2 (4.2)	
	Gallbladder	1 (1.5)	0 (0)	1 (2.1)	
	Kidney	2 (2.9)	0 (0)	2 (4.2)	
	Leukemia	1 (1.5)	1 (5)	0 (0)	
	Lung	1 (1.5)	0 (0)	1 (2.1)	
	Lymphoma	2 (2.9)	1 (5)	1 (2.1)	
	Melanoma	1 (1.5)	0 (0)	1 (2.1)	
	Multiple myeloma	8 (11.8)	3 (15)	5 (10.4)	
	Multiple cancer types	2 (2.9)	1 (5)	1 (2.1)	
	Ovarian	4 (5.9)	3 (15)	1 (2.1)	
	Peritoneal	2 (2.9)	0 (0)	2 (4.2)	
	Prostate	13 (19.1)	4 (20)	9 (18.8)	
	Rectal	1 (1.5)	0 (0)	1 (2.1)	
	Squamous cell carcinoma	1 (1.5)	0 (0)	1 (2.1)	
	Uterine	1 (1.5)	0 (0)	1 (2.1)	

[a]Not able to detect differences between groups.

[b]Active treatment: receiving any curative treatment during participation in the Personal Optimism With Exercise Recovery program.

[c]Statistical significance (*P*<.05).

[d]Treatment history: receiving the treatment type at any point in their care.

[e]Unimodal: 1 treatment type.

[f]Bimodal: 2 treatment types.

[g]Multimodal: 3 or more treatment types.

Of the 68 older adults who completed an initial assessment and participated in POWER, 29.4% (n=20) completed a telehealth follow-up assessment. The median time elapsed between initial and follow-up assessments was 16.5 weeks (IQR 5.75). The majority of older adults with cancer who completed a follow-up were on active treatment (n=12, 60%). Statistically significant differences were not observed among the following clinical and demographic variables among older adults with cancer who did and did not complete a follow-up assessment: age, BMI, sex, race, ethnicity, cancer stage, history of hormone therapy, immunotherapy, surgery, and radiation, number of treatment types, or cancer type. A statistically significant difference was observed for the proportion of older adults with cancer who reported being on active treatment (*P*=.04) and having received chemotherapy (*P*=.04) between those who did and did not complete a follow-up assessment.

Change in Measured Outcomes

Values for each measured outcome from the initial telehealth assessment are reported in Table 2. There were no significant differences in initial assessment outcomes between the follow-up and no follow-up groups (*P*>.05).

Table . Initial assessment data.

Variable	Follow-up assessment completed		Follow-up assessment not completed		Between group difference	Total sample	
	n	Mean (SD)	n	Mean (SD)	*P* value	n	Mean (SD)
Standing march	19	77 (35.4)	42	77.2 (33.3)	.98	61	77.1 (4.3)
30-s maximum push-up	20	12.9 (4.3)	38	13.2 (3.5)	.77	58	13.1 (3.8)
30-s chair stand	19	12.3 (6.7)	42	12.7 (5.8)	.84	61	12.6 (6)
Plank (s)	14	79.1 (43.4)	31	63.8 (57.2)	.38	55	68.6 (53.3)
Single leg stance (s)							
Right	19	29.2 (26.9)	33	20.8 (23.8)	.25	52	23.8 (25.1)
Left	19	25.7 (26.5)	33	26.7 (24.5)	.89	52	26.4 (25)
Shoulder flexion (degrees)							
Left	20	168 (8.8)	44	164.4 (16.5)	.37	64	165.6 (14.6)
Right	20	165.8 (20)	44	165.9 (16)	.97	64	165.9 (17.2)
Shoulder extension (degrees)							
Left	20	58.5 (9.6)	43	59.4 (12.2)	.78	63	59.1 (11.3)
Right	20	58 (9.1)	43	59.9 (11)	.50	63	59.3 (10.4)
Shoulder abduction (degrees)							
Left	20	171.3 (12.1)	44	168.5 (17.1)	.51	64	169.3 (15.6)
Right	20	169.3 (21.5)	44	168.9 (18.2)	.94	64	169 (19.1)
Clock test							
Left	20	3.5 (3)	39	4.5 (3.4)	.27	59	4.2 (3.3)
Right	20	9.4 (1.1)	40	8.5 (2.6)	.52	60	8.8 (2.2)
Seated sit and reach							
Left	20	1.4 (0.6)	42	1.6 (0.4)	.16	62	1.6 (0.5)
Right	20	1.4 (0.6)	42	1.6 (0.5)	.40	62	1.5 (0.5)

Change in measured outcomes are reported in Table 3. Statistically significant changes were observed for the 30-second chair stand test (mean change +2.00 repetitions, 95% CI 0.12 to 3.88, Cohen *d*=0.51) and 30-second maximum push-up test (mean change +2.85 repetitions, 95% CI 1.60 to 4.11, Cohen *d*=1.06). Nine (47.3%) older adults with cancer had a change in 30-second chair stand scores that exceeded the MCID of 2.0 repetitions [35], and 14 (73.7%) older adults maintained their 30-second chair stand scores across the intervention. Although not statistically significant, positive changes were observed for the 2-minute standing march (mean change +12.79 repetitions, 95% CI −0.64 to 26.22, Cohen *d*=0.46), single leg stance on the left (mean change +4.80 s, 95% CI −0.67 to 10.27, Cohen *d*=0.44) and right (mean change +1.0 s, 95% CI −8.04 to 10.05, Cohen *d*=0.06), and shoulder abduction on the left (mean change +2.25 degrees, 95% CI −3.75 to 8.25, Cohen *d*=0.18) and right (mean change +0.25 degrees, 95% CI −4.34 to 4.84, Cohen *d*=0.03). The results from univariate analysis of covariance paralleled results from paired samples *t* tests.

Table . Change in assessment variables across the exercise intervention.

Variable	Quantity, n	Mean change	95% CI	*P* value	Cohen *d*[a]
Standing march	19	12.79	−0.64 to 26.22	.06	0.46
30-s maximum push-up	20	2.85	1.60 to 4.11	<.001[b]	1.06
30-s chair stand	19	2	0.12 to 3.88	.04[b]	0.51
Plank (s)	13	−5	−22.51 to 12.51	.55	0.17
Single leg stance (s)					
Left	18	4.8	−0.67 to 10.27	.08	0.44
Right	18	1	−8.04 to 10.05	.82	0.06
Shoulder flexion (degrees)					
Left	20	−1.25	−6.56 to 4.06	.63	0.11
Right	20	−3.5	−8.54 to 1.54	.16	0.33
Shoulder extension (degrees)					
Left	20	−0.75	−4.64 to 3.14	.69	0.09
Right	20	−0.5	−3.97 to 2.97	.76	0.07
Shoulder abduction (degrees)					
Left	20	2.25	−3.75 to 8.25	.44	0.18
Right	20	0.25	−4.34 to 4.84	.91	0.03
Clock test					
Left	20	1.7	−0.96 to 4.33	—[c]	—
Right	20	−0.53	−1.73 to 0.68	—	—
Seated sit and reach					
Left	20	0.13	−0.11 to 0.36	—	—
Right	20	0.15	−0.09 to 0.39	—	—

[a]Cohen *d* interpretation: small=0.2, medium=0.5, and large=0.8.

[b]Statistical significance (*P*<.05).

[c]Not applicable.

Discussion

Principal Findings

This study aimed to evaluate the preliminary effectiveness of a hospital-based telehealth exercise oncology program on physical function, muscular endurance, balance, and flexibility among older adults with cancer. Our findings demonstrate that supervised, one-on-one telehealth exercise may positively influence physical function among older adults with cancer. Additionally, nearly half (n=9) of individuals who completed a follow-up assessment exceeded an MCID change in the 30-second chair-stand test, a marker of lower extremity function.

Comparison to Prior Work

The majority of research surrounding telehealth supervised exercise programs for older adults without [44,45] and with cancer [24,27] has focused on group exercise. Less is known about one-on-one telehealth exercise. Among older adults without cancer, participating in at least 12-weeks of supervised group telehealth exercise training prevents declines in physical function [44,45]. Among older adults with cancer, a feasibility study by Sattar et al [24] evaluated an 8-week group telehealth strength and balance training program and observed significant improvements in five time chair stand test scores. Additionally, Gell et al [27] carried out a pilot trial examining a 16-week group telehealth aerobic and resistance training program and observed significant improvements in 30-second chair stand test scores. Collectively, group telehealth exercise programming among older adults with and without cancer is effective at improving physical function.

Findings from this study contribute to the literature by addressing an important gap in our understanding regarding the effectiveness of one-on-one telehealth exercise among older cancer survivors. Among cancer survivors of all ages (range 14 - 83 y), effectiveness of one-on-one supervised telehealth exercise has been evaluated [22]. Following 12-weeks of one-on-one training with a cancer exercise trainer once per week, cancer survivors significantly improved cardiovascular endurance, muscular endurance, and flexibility [22]. Findings from our study support previous research suggesting one-on-one telehealth exercise programs may positively influence physical function among older adults with cancer. Without exercise

intervention, we would expect to see little to no change in physical function parameters over short durations in older adults living with cancer. Over 13-weeks Mikklesen et al [11] found a mean change of +0.4 repetitions in the 30-second chair stand test and −1.0 points in self-reported physical function among older cancer survivors receiving standard of care and no exercise intervention. Over longer durations (eg, ≥1 y) functional declines are greater and can persist for years following diagnosis [6]. Preventing declines in physical function is important in this population because physical function has been shown to have a protective effect against all-cause mortality in older adults with cancer [8].

In addition to examining the effect of one-on-one, supervised telehealth exercise on physical function in older adults with cancer, we also characterized older adults with cancer who chose to participate in a telehealth exercise program. This information adds to our body of knowledge by demonstrating that older adults with cancer can engage in one-on-one telehealth delivered exercise programs. Additionally, we observed that a significantly greater proportion of older adults with cancer who were on active treatment or had received chemotherapy completed a follow-up assessment. This finding suggests that older cancer survivors on active cancer treatment are willing to engage in telehealth exercise which is important as recent recommendations from the American Society of Clinical Oncology encourage cancer survivors on active treatment to participate in aerobic and resistance exercise [46]. Understanding who we are reaching with telehealth exercise programs, who may be missing, and who completes a telehealth follow-up assessment can inform the development of interventions to improve the engagement and reach of telehealth delivered exercise programming among cancer survivors of all ages.

Strengths and Limitations

To our knowledge, this is one of the first studies that has evaluated effectiveness of one-on-one telehealth exercise programming exclusively among older adults with cancer and characterized the older adults who used this programming. We consider this a strength of our work. Additionally, the use of an established hospital-based exercise oncology program with over 15 years of experience offering telehealth exercise ensured high-quality exercise programming in this study. However, our study is not without limitations. First, the low follow-up assessment completion rate resulted in a small sample size and an underpowered analysis to demonstrate statistically significant changes in all outcomes measured. Therefore, the findings from this study are preliminary and additional research with a larger sample is needed. Second, the retrospective design of this study may have resulted in selection bias of participants who were more motivated to exercise and follow-up. Highly motivated individuals may be more likely to complete a follow-up assessment in the program potentially confounding the effects of the telehealth exercise program. Third, baseline and follow-up assessments were scheduled based on the participants' availability which resulted in different staff conducting baseline and follow-up assessments for some participants. However, a small team of certified exercise physiologists administered all virtual assessments, adhering to standardized program procedures to minimize interrater variability. Fourth, the lack of a nonexercise control group limits the conclusions that can be made regarding the ability of telehealth exercise to prevent declines in physical function. Future work should consider a prospective study design and inclusion of potential confounding variables as covariates to determine the effectiveness of one-on-one telehealth exercise on markers of physical function.

Conclusions

In summary, older adults living with and beyond cancer are able to participate in an exercise oncology program delivered via telehealth. Our findings provide preliminary evidence that telehealth may be a beneficial tool to facilitate exercise program delivery among older adults following a cancer diagnosis. However, telehealth exercise should not be considered a one-size-fits-all all approach as in-person, telehealth, or a combination of the two may be a better fit for some older adults with cancer, based on their needs and preferences. Further research is needed to understand the magnitude of the effects of one-on-one, supervised telehealth exercise on physical function among older adults with cancer.

Acknowledgments

Research reported in this publication was supported by the University of Utah Graduate Research Fellowship and Alice Oakes Bronson Research Fellowship.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ERD and AMC were responsible for study conceptualization, data acquisition, quality control of data, data analysis, and paper preparation. ERD was responsible for statistical analysis. SO, PAH, RWZ, DW, and KL participated in data collection. ERD, AMC, LP, MN, YB, and SO were responsible for study design and paper editing. All authors reviewed this paper and approved the final version.

Conflicts of Interest

None declared.

References

1. Miller KD, Nogueira L, Devasia T, et al. Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin 2022 Sep;72(5):409-436. [doi: 10.3322/caac.21731] [Medline: 35736631]
2. Bluethmann SM, Mariotto AB, Rowland JH. Anticipating the "silver tsunami": prevalence trajectories and comorbidity burden among older cancer survivors in the United States. Cancer Epidemiol Biomarkers Prev 2016 Jul;25(7):1029-1036. [doi: 10.1158/1055-9965.EPI-16-0133] [Medline: 27371756]
3. Hurria A, Jones L, Muss HB. Cancer treatment as an accelerated aging process: assessment, biomarkers, and interventions. Am Soc Clin Oncol Educ Book 2016;35:e516-e522. [doi: 10.1200/EDBK_156160] [Medline: 27249761]
4. Derks MGM, de Glas NA, Bastiaannet E, et al. Physical functioning in older patients with breast cancer: a prospective cohort study in the TEAM trial. Oncol 2016 Aug;21(8):946-953. [doi: 10.1634/theoncologist.2016-0033] [Medline: 27368882]
5. Leach CR, Bellizzi KM, Hurria A, Reeve BB. Is it my cancer or am I just getting older?: impact of cancer on age-related health conditions of older cancer survivors. Cancer 2016 Jun 15;122(12):1946-1953. [doi: 10.1002/cncr.29914] [Medline: 27159822]
6. Feliciano EMC, Vasan S, Luo J, et al. Long-term trajectories of physical function decline in women with and without cancer. JAMA Oncol 2023 Mar 1;9(3):395-403. [doi: 10.1001/jamaoncol.2022.6881] [Medline: 36656572]
7. Sehl M, Lu X, Silliman R, Ganz PA. Decline in physical functioning in first 2 years after breast cancer diagnosis predicts 10-year survival in older women. J Cancer Surviv 2013 Mar;7(1):20-31. [doi: 10.1007/s11764-012-0239-5] [Medline: 23232922]
8. Ezzatvar Y, Ramírez-Vélez R, de Asteasu MLS, et al. Physical function and all-cause mortality in older adults diagnosed with cancer: a systematic review and meta-analysis. J Gerontol A Biol Sci Med Sci 2021 Jul 13;76(8):1447-1453. [doi: 10.1093/gerona/glaa305] [Medline: 33421059]
9. Kenzik KM, Morey MC, Cohen HJ, Sloane R, Demark-Wahnefried W. Symptoms, weight loss, and physical function in a lifestyle intervention study of older cancer survivors. J Geriatr Oncol 2015 Nov;6(6):424-432. [doi: 10.1016/j.jgo.2015.08.004] [Medline: 26362355]
10. Olsen AP, Vinther A, Ragle AM, et al. Physical exercise program to older frail patients receiving chemotherapy for colorectal cancer - feasibility and experiences from the GERICO study. J Geriatr Oncol 2023 Jan;14(1):101382. [doi: 10.1016/j.jgo.2022.09.011] [Medline: 36328878]
11. Mikkelsen MK, Lund CM, Vinther A, et al. Effects of a 12-week multimodal exercise intervention among older patients with advanced cancer: results from a randomized controlled trial. Oncol 2022 Feb 3;27(1):67-78. [doi: 10.1002/onco.13970] [Medline: 34498352]
12. Paulo TRS, Rossi FE, Viezel J, et al. The impact of an exercise program on quality of life in older breast cancer survivors undergoing aromatase inhibitor therapy: a randomized controlled trial. Health Qual Life Outcomes 2019 Jan 18;17(1):17. [doi: 10.1186/s12955-019-1090-4] [Medline: 30658629]
13. Zhang J, Su M, Cheng J, Zhou S, Liu L, Yao NA. A nationally representative study of aerobic activity and strength training in older cancer survivors and their psychological distress and sleep difficulties. Support Care Cancer 2022 Nov;30(11):9597-9605. [doi: 10.1007/s00520-022-07370-2] [Medline: 36171399]
14. Perry CK, Ali W, Solanki E, Winters-Stone K. Attitudes and beliefs of older female breast cancer survivors and providers about exercise in cancer care. Oncol Nurs Forum 2020 Jan 1;47(1):56-69. [doi: 10.1188/20.ONF.56-69] [Medline: 31845920]
15. Mikkelsen MK, Michelsen H, Nielsen DL, Vinther A, Lund CM, Jarden M. "Doing What Only I Can Do": experiences from participating in a multimodal exercise-based intervention in older patients with advanced cancer-a qualitative explorative study. Cancer Nurs 2022;45(2):E514-E523. [doi: 10.1097/NCC.0000000000000987] [Medline: 34294645]
16. Cheung DST, Takemura N, Chau PH, Ng AYM, Xu X, Lin CC. Exercise levels and preferences on exercise counselling and programming among older cancer survivors: a mixed-methods study. J Geriatr Oncol 2021 Nov;12(8):1173-1180. [doi: 10.1016/j.jgo.2021.05.002] [Medline: 33985930]
17. Mikkelsen MK, Nielsen DL, Vinther A, Lund CM, Jarden M. Attitudes towards physical activity and exercise in older patients with advanced cancer during oncological treatment - a qualitative interview study. Eur J Oncol Nurs 2019 Aug;41:16-23. [doi: 10.1016/j.ejon.2019.04.005] [Medline: 31358249]
18. Dennett A, Harding KE, Reimert J, Morris R, Parente P, Taylor NF. Telerehabilitation's safety, feasibility, and exercise uptake in cancer survivors: process evaluation. JMIR Cancer 2021 Dec 21;7(4):e33130. [doi: 10.2196/33130] [Medline: 34854817]
19. Winters-Stone KM, Boisvert C, Li F, et al. Delivering exercise medicine to cancer survivors: has COVID-19 shifted the landscape for how and who can be reached with supervised group exercise? Support Care Cancer 2022 Mar;30(3):1903-1906. [doi: 10.1007/s00520-021-06669-w] [Medline: 34741653]
20. Crosby BJ, Newton RU, Galvão DA, et al. Feasibility of supervised telehealth exercise for patients with advanced melanoma receiving checkpoint inhibitor therapy. Cancer Med 2023 Jul;12(13):14694-14706. [doi: 10.1002/cam4.6091] [Medline: 37184115]
21. Myers SA, Weller S, Schwartz S, et al. Feasibility of a supervised virtual exercise program for women on hormone therapy for breast cancer. Transl J ACSM 2022;7(4):e000213. [doi: 10.1249/TJX.0000000000000213]

22. Wonders KY, Gnau K, Schmitz KH. Measuring the feasibility and effectiveness of an individualized exercise program delivered virtually to cancer survivors. Curr Sports Med Rep 2021 May 1;20(5):271-276. [doi: 10.1249/JSR.0000000000000846] [Medline: 33908915]
23. Lafaro KJ, Raz DJ, Kim JY, et al. Pilot study of a telehealth perioperative physical activity intervention for older adults with cancer and their caregivers. Support Care Cancer 2020 Aug;28(8):3867-3876. [doi: 10.1007/s00520-019-05230-0] [Medline: 31845007]
24. Sattar S, Haase K, Penz K, et al. Feasibility of a remotely delivered strength and balance training program for older adults with cancer. Curr Oncol 2021 Nov 2;28(6):4408-4419. [doi: 10.3390/curroncol28060374] [Medline: 34898562]
25. Sattar S, Papadopoulos E, Smith GVH, et al. State of research, feasibility, safety, acceptability, and outcomes examined on remotely delivered exercises using technology for older adult with cancer: a scoping review. J Cancer Surviv 2024 Dec;18(6):1861-1888. [doi: 10.1007/s11764-023-01427-9] [Medline: 37418170]
26. Dunston ER, Malouf A, Podlog LW, et al. Experiences participating in a telehealth exercise program among older adults with cancer: a qualitative study. J Cancer Surviv 2024 Mar 16. [doi: 10.1007/s11764-024-01564-9] [Medline: 38492170]
27. Gell NM, Dittus K, Caefer J, Martin A, Bae M, Patel KV. Remotely delivered exercise to older rural cancer survivors: a randomized controlled pilot trial. J Cancer Surviv 2024 Apr;18(2):596-605. [doi: 10.1007/s11764-022-01292-y] [Medline: 36374436]
28. Coletta AM, Rose NB, Johnson AF, et al. The impact of a hospital-based exercise oncology program on cancer treatment-related side effects among rural cancer survivors. Support Care Cancer 2021 Aug;29(8):4663-4672. [doi: 10.1007/s00520-021-06010-5] [Medline: 33502590]
29. Campbell KL, Winters-Stone KM, Wiskemann J, et al. Exercise guidelines for cancer survivors: consensus statement from international multidisciplinary roundtable. Med Sci Sports Exerc 2019 Nov;51(11):2375-2390. [doi: 10.1249/MSS.0000000000002116] [Medline: 31626055]
30. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Res Q Exerc Sport 1999 Jun;70(2):113-119. [doi: 10.1080/02701367.1999.10608028] [Medline: 10380242]
31. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. Gerontologist 2013 Apr;53(2):255-267. [doi: 10.1093/geront/gns071] [Medline: 22613940]
32. Bowman A, Denehy L, Benjemaa A, et al. Feasibility and safety of the 30‐second sit‐to‐stand test delivered via telehealth: an observational study. PM R 2023 Jan;15(1):31-40. [doi: 10.1002/pmrj.12783] [Medline: 35138036]
33. Blackwood J, Rybicki K. Physical function measurement in older long-term cancer survivors. J Frailty Sarcopenia Falls 2021 Sep;6(3):139-146. [doi: 10.22540/JFSF-06-139] [Medline: 34557613]
34. Eden MM, Tompkins J, Verheijde JL. Reliability and a correlational analysis of the 6MWT, ten-meter walk test, thirty second sit to stand, and the linear analog scale of function in patients with head and neck cancer. Physiother Theory Pract 2018 Mar;34(3):202-211. [doi: 10.1080/09593985.2017.1390803] [Medline: 29068767]
35. Wright AA, Cook CE, Baxter GD, Dockerty JD, Abbott JH. A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis. J Orthop Sports Phys Ther 2011 May;41(5):319-327. [doi: 10.2519/jospt.2011.3515] [Medline: 21335930]
36. Whitehurst MA, Johnson BL, Parker CM, Brown LE, Ford AM. The benefits of a functional exercise circuit for older adults. J Strength Cond Res 2005 Aug;19(3):647-651. [doi: 10.1519/R-14964.1] [Medline: 16095420]
37. Węgrzynowska-Teodorczyk K, Mozdzanowska D, Josiak K, et al. Could the two-minute step test be an alternative to the six-minute walk test for patients with systolic heart failure? Eur J Prev Cardiol 2016 Aug;23(12):1307-1313. [doi: 10.1177/2047487315625235] [Medline: 26743588]
38. Bohannon RW, Crouch RH. Two-minute step test of exercise capacity: systematic review of procedures, performance, and clinimetric properties. J Geriatr Phys Ther 2019;42(2):105-112. [doi: 10.1519/JPT.0000000000000164] [Medline: 29210933]
39. Ogawa EF, Harris R, Dufour AB, Morey MC, Bean J. Reliability of virtual physical performance assessments in veterans during the COVID-19 pandemic. Arch Rehabil Res Clin Transl 2021 Sep;3(3):100146. [doi: 10.1016/j.arrct.2021.100146] [Medline: 34589696]
40. Goldberg A, Casby A, Wasielewski M. Minimum detectable change for single-leg-stance-time in older adults. Gait Posture 2011 Apr;33(4):737-739. [doi: 10.1016/j.gaitpost.2011.02.020] [Medline: 21444208]
41. Güngör F, Ovacık U, Harputlu Ö, et al. Tele-assessment of core performance and functional capacity: reliability, validity, and feasibility in healthy individuals. J Telemed Telecare 2024 Jul;30(6):1017-1025. [doi: 10.1177/1357633X221117335] [Medline: 35916001]
42. Jones CJ, Rikli RE, Max J, Noffal G. The reliability and validity of a chair sit-and-reach test as a measure of hamstring flexibility in older adults. Res Q Exerc Sport 1998 Dec;69(4):338-343. [doi: 10.1080/02701367.1998.10607708] [Medline: 9864752]
43. Hayes K, Walton JR, Szomor ZR, Murrell GA. Reliability of five methods for assessing shoulder range of motion. Aust J Physiother 2001;47(4):289-294. [doi: 10.1016/s0004-9514(14)60274-9] [Medline: 11722295]

44. Buckinx F, Aubertin-Leheudre M, Daoust R, et al. Impacts of remote physical exercises on functional status and mobility among community-dwelling pre-disabled seniors during the COVID-19 lockdown. J Nutr Health Aging 2023;27(5):354-361. [doi: 10.1007/s12603-023-1914-1] [Medline: 37248759]
45. Carrapatoso SM, Abdalla PP, Cadete CM, Carvalho JM, Santos MPM, Bohn L. Does real-time online physical exercise improve physical fitness in seniors? Act Adapt Aging 2022 Oct 2;46(4):286-302. [doi: 10.1080/01924788.2022.2100686]
46. Ligibel JA, Bohlke K, May AM, et al. Exercise, diet, and weight management during cancer treatment: ASCO guideline. J Clin Oncol 2022 Aug 1;40(22):2491-2507. [doi: 10.1200/JCO.22.00687] [Medline: 35576506]

Abbreviations:

HCI: Huntsman Cancer Institute
ICC: intraclass correlation
MCID: minimal clinically important difference
POWER: Personal Optimism With Exercise Recovery

Edited by N Cahill; submitted 24.01.24; peer-reviewed by DS Ting Cheung, RJ Marker; revised version received 31.10.24; accepted 31.10.24; published 13.01.25.

Please cite as:
Dunston ER, Oza S, Bai Y, Newton M, Podlog L, Larson K, Walker D, Zingg RW, Hansen PA, Coletta AM
Preliminary Effectiveness of a Telehealth-Delivered Exercise Program in Older Adults Living With and Beyond Cancer: Retrospective Study
JMIR Cancer 2025;11:e56718
URL: https://cancer.jmir.org/2025/1/e56718
doi:10.2196/56718

Original Paper

A Novel Telehealth Exercise Program Designed for Rural Survivors of Cancer With Cancer-Related Fatigue: Single-Arm Feasibility Trial

Ryan J Marker[1], DPT, PhD; Andrew J Kittelson[2,3], DPT, PhD; Jared J Scorsone[4], BS; Ian A Moran[4], BS; John C Quindry[3], PhD; Heather J Leach[5], PhD

[1]Department of Physical Medicine and Rehabilitation, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[2]School of Physical Therapy and Rehabilitation Science, University of Montana, Missoula, MT, United States
[3]School of Integrative Physiology and Athletic Training, University of Montana, Missoula, MT, United States
[4]Anschutz Health and Wellness Center, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[5]Department of Health and Exercise Science, Colorado State University, Fort Collins, CO, United States

Corresponding Author:
Ryan J Marker, DPT, PhD
Department of Physical Medicine and Rehabilitation
University of Colorado Anschutz Medical Campus
12348 E Montview Boulevard
Aurora, CO, 80045
United States
Phone: 1 13037240819
Email: ryan.marker@cuanschutz.edu

Abstract

Background: Exercise interventions are among the best-known interventions for cancer-related fatigue (CRF). Rural survivors of cancer, however, report specific barriers to engaging in exercise programs and lack overall access to effective programs.

Objective: The purpose of this investigation was to assess the feasibility of a novel telehealth exercise program designed specifically for rural survivors of cancer with CRF.

Methods: A single-arm clinical trial of the BfitBwell Telehealth Program was performed. Based on an established clinical program, this adapted 12-week program addressed barriers previously reported by rural survivors by providing synchronous videoconference exercise sessions (2 per program), asynchronous exercise sessions using a personal training smartphone or internet app (3-5 per week), and regular symptom (CRF) monitoring using automated emailed surveys (every 2 weeks). Personalized exercise prescriptions containing aerobic and resistance activities were implemented by cancer exercise specialists. Symptom-triggered synchronous sessions were initiated for participants failing to improve in CRF, as identified by a reference chart of CRF improvements observed during a supervised exercise program. Eligible participants were adult survivors of any cancer diagnosis who had completed treatment with curative intent in the past 12 months or had no planned changes in treatment for the duration of the study, lived in a rural area, and were currently experiencing CRF. Feasibility was assessed by objective measures of recruitment, data collection, intervention acceptability and suitability, and preliminary evaluations of participant responses. CRF was the primary clinical outcome (assessed using the Functional Assessment of Chronic Illness Therapy—Fatigue Scale [FACIT-Fatigue]) and was measured before, after, and 6 months after program completion.

Results: In total, 19 participants enrolled in the study, 16 initiated the exercise program, and 15 completed the program. A total of 14 participants were recruited through internet advertisements, and the total recruitment rate peaked at 5 participants per month. Participants completed 100% of initial and final assessments (30 assessments across all participants) and 93% (70/75 possible surveys across all participants) of emailed surveys and attended 97% (29/30 possible sessions across all participants) of synchronous exercise sessions. In total, 6 participants initiated symptom-triggered sessions, with 6 of 7 initiated sessions attended. The mean FACIT-Fatigue scores significantly improved (P=.001) by 11.2 (SD 6.8) points following the completion of the program. A total of 13 participants demonstrated at least a minimal clinically important difference in FACIT-Fatigue scores (≥ +3 points) at this time. FACIT-Fatigue scores did not significantly change from program completion to 6-month follow-up (n=13; mean change –1.1, SD 3.4 points; P=.29).

Conclusions: Results from this investigation support the feasibility of the BfitBwell Telehealth Program and a subsequent efficacy trial. Novel program components also provide potential models for improving exercise program efficacy and efficiency through asynchronous exercise prescription and symptom monitoring.

Trial Registration: ClinicalTrials.gov NCT04533165; https://clinicaltrials.gov/study/NCT04533165

(JMIR Cancer 2025;11:e59478) doi:10.2196/59478

KEYWORDS

cancer-related fatigue; telehealth; physical activity; survivorship; digital health; lifestyle intervention; videoconference; symptom burden; symptom monitoring; geographic disparities; mHealth

Introduction

Cancer-related fatigue (CRF) is one of the most common and functionally limiting symptoms reported by survivors of cancer, with an estimated prevalence of 49% to 62% [1-4]. It is present in over a quarter of survivors up to 10 years after the completion of treatment [5], and survivors regularly report CRF as the symptom preventing them from living a "normal" life as well as the cause of major life events such as employment changes [6]. Participation in exercise interventions is an established intervention for the improvement of CRF in survivors of cancer [7,8], and a multidisciplinary roundtable pronouncement by the American College of Sports Medicine (ACSM) indicated that there is strong evidence that exercise can significantly reduce CRF. The current ACSM recommendation for an effective exercise prescription to remediate CRF is moderate-intensity aerobic training at least 3 times per week for at least 30 minutes and moderate-intensity resistance training at least 2 times per week [9].

Survivors of cancer living in rural locations [10], however, commonly lack access to many supportive services compared to nonrural survivors, including clinical exercise programs [11]. Rural survivors commonly report many specific barriers to engaging in exercise programs including lack of knowledge of available programs, distance and transportation to programs, lack of access to a knowledgeable exercise provider, and lack of flexibility in programming (in terms of time and location) [12,13]. An overall lack of accessible exercise programs for rural survivors has been identified in a recent review [14]. A survey of rural survivors reported that only 38% and 10% were currently meeting aerobic and resistance training guidelines for survivors, respectively [12], demonstrating a clear need to increase accessible exercise programs for this population.

While current initiatives are designed to increase the accessibility of exercise programs for rural survivors of cancer [15,16], additional opportunities exist in the development of novel rural-focused program designs. The Exercise for Cancer to Enhance Living Well study in Canada [15] provides an exemplary model of using and improving clinical networks to increase awareness of and access to exercise oncology programs for rural survivors. The intervention itself, however, mirrors the available supervised programs by replacing in-person services with telehealth services, similar to other telehealth adaptations in the United States [17], which may not address all barriers experienced by survivors in rural areas. Beyond these examples, most other currently accessible exercise programs for rural survivors are phone-based walking programs [14]. Given that supervised exercise programs consistently demonstrate greater improvements in CRF compared to unsupervised programs [8], the efficacy of these current programs may be limited by the lack of consistent supervision and the recommended resistance exercise. Continued innovation in program designs specifically for the rural population is required to truly reduce geographic disparities targeting both improved accessibility and efficacy.

The purpose of this single-arm feasibility study was to assess a novel telehealth exercise program designed specifically for rural survivors of cancer with CRF, with an emphasis on replicating the CRF improvements seen in clinically supervised exercise programs. The program addresses known participation barriers for rural survivors and uses a validated reference chart of CRF improvement in a supervised program combined with regular symptom monitoring and symptom-triggered intervention. Data collection and outcome selection were designed to assess recommended objectives for feasibility studies [18] including recruitment, data collection, intervention acceptability and suitability, and a preliminary evaluation of participant responses. It was hypothesized that the program would demonstrate overall feasibility based on these objectives, providing support for a larger efficacy study.

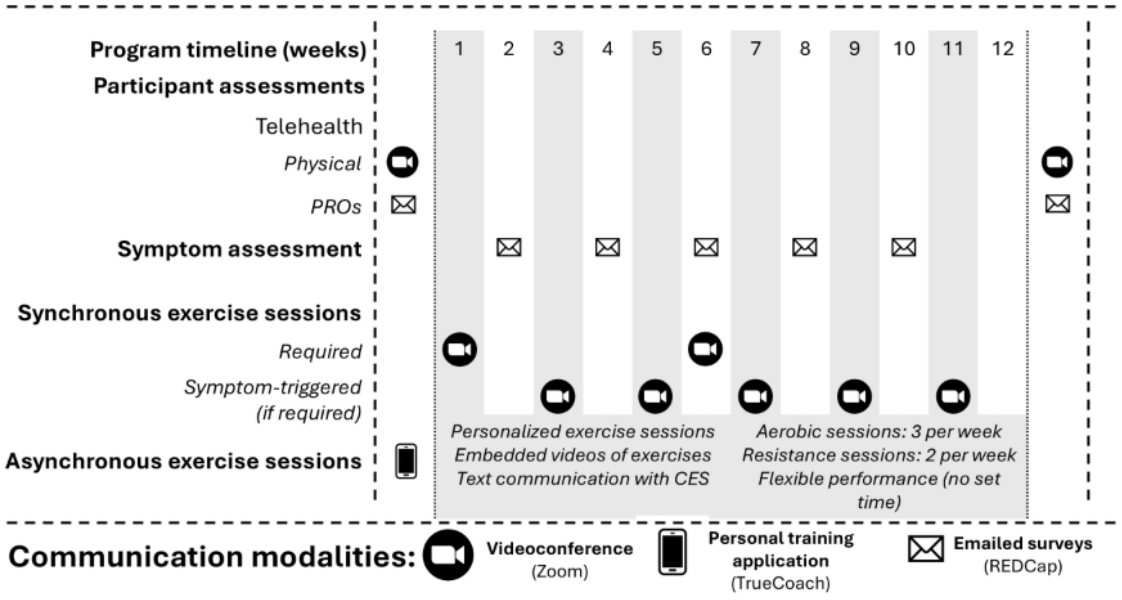
Methods

Study and Program Design

This was a single-arm clinical trial of the BfitBwell Telehealth Program (Figure 1), with assessments at baseline, program completion (12 weeks), and 6-month follow-up. Data collection occurred between November 2021 and September 2023. This 12-week telehealth exercise program was adapted from the clinically supervised BfitBwell Cancer Exercise Program [19,20]. Initial adaptations were made internally by program and research staff and designed to (1) address known barriers to exercise participation in rural survivors of cancer and (2) replicate the effects of a supervised clinical exercise program via telehealth. The specific barriers addressed were distance and transportation, lack of program flexibility (in regard to time and location), and lack of access to a knowledgeable exercise provider [12,13]. Participation in the BfitBwell Telehealth Program was decentralized, with centralized program oversight, to remove distance and transportation barriers. Three telehealth technologies were used: videoconferencing software (Zoom Video Communications, Inc), a personal training smartphone or internet app (TrueCoach Inc), and automated emails (REDCap [Research Electronic Data Capture]; Vanderbilt University

[21,22]). The majority of exercise sessions were delivered asynchronously (ie, without real-time interaction between the participant and cancer exercise specialist [CES]) via the personal training app to provide program flexibility. Participants received scheduled (eg, recommended day of performance) and personalized exercise sessions via the app, but the timing of performance was determined by participants (eg, participants could complete sessions at any time of day, and "missed" sessions could be performed on later days). The personal training app also allowed embedded text communication between the prescribing CES and the participant following each individual exercise or exercise session. This within-app communication facilitated direct access to a knowledgeable exercise provider, as did regular engagement via automated emails for symptom monitoring.

Figure 1. The design of the BfitBwell Telehealth Program with the timing of program components delivery. CES: cancer exercise specialist; PRO: patient-reported outcome; REDCap: Research Electronic Data Capture.



Symptom monitoring was accomplished by plotting individual participant CRF scores on a validated reference chart of CRF improvements during a supervised exercise program [23]. The reference chart displays the typical course of CRF improvement for individuals in the supervised BfitBwell Cancer Exercise Program, against which the progress of individual participants can be easily monitored. The supervised BfitBwell Cancer Exercise Program has demonstrated effectiveness for improving CRF [19,20]. This type of reference chart has previously been proposed to inform personalized rehabilitation in other clinical populations [24,25]. In the BfitBwell Telehealth Program, CRF was monitored every 2 weeks and used to initiate symptom-triggered exercise sessions for participants failing to improve as predicted by the reference chart (see Symptom Tracking and Symptom-Triggered Sessions section). This served as the primary means of replicating the effects of a supervised program in a rural setting, as the reference chart allows individual participant progress to be compared to that of similar participants from a supervised program.

Finally, to promote participant safety in this remote telehealth context, a detailed safety and emergency response plan was developed. This plan included regular verification of the physical location of all videoconference sessions and the identification of a nearby "local support individual" (though this individual did not need to be present for assessments or sessions). Detailed plans of action were developed for emergent and nonemergent adverse events.

Participants

Eligibility criteria for participation were adults ≥18 and ≤80 years of age, a diagnosis of any cancer type and any stage, completion of medical cancer treatment with curative intent within the past 12 months, or currently receiving treatment with no planned changes for the next 4 months. These criteria are similar to the patient population in the supervised BfitBwell Cancer Exercise Program and the associated CRF reference chart. Additional eligibility criteria included current moderate to severe CRF (>3 on a 10-point scale per National Comprehensive Cancer Network definition [26]), high-speed home internet and videoconference capable device (smartphone or laptop with camera), and residence in a rural area (defined here as >1-hour commute to a major city in Colorado or surrounding states with a known exercise oncology program, based upon review of a national program directory) [27]. Exclusion criteria included medical conditions that would impact the safety of, or participation in, a telehealth exercise program (eg, significant pulmonary or cardiovascular conditions and

mobility-limiting orthopedic conditions). These conditions were self-reported on screening forms and individually reviewed by a licensed physical therapist (RJM) to determine eligibility. Safety was again assessed by the same therapist during the initial assessment (see Assessments section)).

A power analysis was not performed a priori, but a goal of 20 enrolled participants (with 15 completing the program) was set at study initiation. This recruitment goal was based primarily on the capacity of the supporting clinical exercise program combined with historical attrition rates of 25%.

Recruitment and Enrollment

Initial recruitment efforts were made through clinical staff at a large urban cancer center that serves rural Colorado and surrounding states. Incoming and recent participants in the supervised BfitBwell Cancer Exercise Program were screened for eligibility. Recruitment efforts were later adapted to include targeted internet and social media advertisements (BuildClinical, LLC). Participants recruited through these efforts completed screening evaluations via emailed surveys or phone calls with study personnel. Eligible and interested participants were then invited to live videoconference sessions to further discuss the study and provide written informed consent. Enrolled participants were mailed all necessary equipment including a resistance exercise band set, a commercial fitness tracker (Withings Move, Withings Health), a smartphone tripod, an aerobic step (adjustable height 2-8 inches), and other necessary assessment equipment (eg, a measured length of rope, tape measure, and a pulse oximeter). Participants were incentivized to enroll by being allowed to keep all mailed study equipment, and gift cards were distributed to facilitate program and within-program survey completion.

Primary Clinical Outcome

CRF was the primary clinical outcome and was assessed using the Functional Assessment of Chronic Illness Therapy—Fatigue Scale (FACIT-Fatigue) [28]. The FACIT-Fatigue is one of the most common measures of CRF with demonstrated reliability and validity in survivors of cancer [8,29]. It is a 13-item scale with scores ranging from 0 to 52, with higher scores indicating less CRF, asking participants to consider how they have felt during the past week. The minimal clinically important difference (MCID) of the FACIT-Fatigue scale has been identified as 3 points [30].

Assessments

Initial and final assessments were performed via videoconference by a licensed physical therapist (RJM, separate from program CESs). In addition to collecting study outcomes, the initial assessment served as a final assessment of participant safety (based on the physical therapist's clinical observations and judgment) and provided information used by the CES to personalize the exercise prescription. Demographic and cancer-related information were collected, and basic measures of physical fitness and function were performed similar to prior studies of telehealth assessments in survivors of cancer (adapted to be performed safely if the participant was alone) [31]. Physical assessment outcomes included single limb stance [32], gait speed [33], timed up and go [33], 30-second sit-to-stand [34], and a 3-minute step test (following the Tecumseh protocol [35]). Participants were interviewed about previous exercise experience, exercise preferences and goals, and available exercise resources (eg, home equipment and local gymnasiums). Patient-reported outcomes were then collected via emailed surveys. These included the FACIT-Fatigue [28], Functional Assessment of Cancer Treatment—General [36], Hospital Anxiety and Depression Scale [37], and the Godin Leisure-Time Exercise Questionnaire (GLTEQ, modified to estimate weekly moderate to vigorous physical activity and resistance exercise) [38]. Assessments occurred within the 2 weeks prior to intervention initiation and after completion. Patient-reported outcomes were again emailed to participants 6 months following program completion (follow-up assessment).

Exercise Intervention

Exercise prescription followed current recommendations for survivors of cancer from the ACSM, specifically targeting 2 resistance exercise sessions and three 30-minute aerobic exercise sessions per week [9]. The prescribed exercise plan was delivered by 2 CESs employed by the supervised BfitBwell Cancer Exercise Program and resembled typical sessions for this program, previously described [19,20]. Both CESs (IAM and JJS) had an undergraduate degree in exercise science (or a related field), an exercise training certification (ACSM Certified Exercise Physiologist or equivalent), and at least 4 years of experience working exclusively with survivors in the supervised BfitBwell Cancer Exercise Program, and the primary CES (IAM) had an ACSM-CES certification. Resistance exercises targeted large upper and lower extremity muscle groups (using resistance bands, participant equipment, household objects, and body weight), and aerobic exercise was based on participant preference and available equipment (typically outdoor walking, treadmill walking, or stationary cycling). All exercise plans were personalized based on participant abilities, preferences, and available resources (established during the initial assessment).

Exercise Sessions

One-hour, synchronous videoconference telehealth exercise sessions were scheduled with all participants in weeks 1 and 6. These sessions were performed in real time with the CES and participants interacting via live videoconference, mirroring an in-person supervised exercise session. Session content focused on providing education on exercise, demonstration and practice of proper exercise form, and supervised performance of exercises prescribed in the subsequent 6 weeks. All other exercise sessions (except the symptom-triggered sessions described in the next section) were delivered asynchronously via the personal training smartphone or internet app and included detailed individualized exercise content with embedded videos of a program CES performing the prescribed exercises. Participants had to indicate each exercise as completed within each session, creating a self-report measurement of asynchronous session completion. The embedded text communication between the prescribing CES and participant was regularly reviewed by the CES, who would then respond and adapt subsequent asynchronous exercise sessions, as

necessary. The app was available via both smartphone app and internet browser.

Symptom Tracking and Symptom-Triggered Sessions

Participants were emailed the FACIT-Fatigue 2 days prior to initiating the exercise program and every 2 weeks during the program, with daily reminders sent for up to 3 days. FACIT-Fatigue scores were monitored using the CRF reference chart to detect whether an individual's progress matched the typical progress of similar individuals in a supervised exercise program, based on percentile rank established by initial scores. A symptom-triggered, synchronous videoconference exercise session was scheduled in the week following a FACIT-Fatigue score ≥3 points lower than the projected percentile at a given measurement time. The threshold was based on the MCID of the FACIT-Fatigue [30] in an attempt to ensure that symptom-triggered sessions were initiated due to a true lack of progress rather than normal variation or measurement error. Symptom-triggered session length ranged from 15 to 60 minutes, and a CES discussed program progress and challenges with the participant, with adaptations made to improve program response. These sessions were designed to mirror what would occur in a supervised program for participants who expressed a poor exercise response (eg, stated they continued to have fatigue) or in whom the CES identified a poor response (eg, performance plateau or decline). Common adaptations included changes in various exercise prescription components (Frequency, Intensity, Timing, and Type, via the ACSM guidelines [39]) and were made based on CES judgment in each occurrence. Emailed surveys also included a simple form for participants to ask exercise-related questions and request an additional videoconference exercise session the following week, even if not triggered by FACIT-Fatigue scores.

Outcomes

Rationale

All outcomes were designed based on the recommended objectives, with guiding questions, for feasibility studies, as described by Orsmond and Cohn [18]. The current investigation focused on the use of collected objective data. While the ability of the research team to conduct the study and provide the intervention was not directly assessed, evaluation of other objectives allowed an indirect assessment (eg, successful data collection and attendance rates support the ability of the research team to perform these tasks). "Success" thresholds for feasibility outcomes were not set a priori, but rather outcomes were assessed holistically following the study as a means of assessing overall feasibility, providing the context of experience acquired while delivering the pilot program.

Recruitment

The number of participants screened, determined eligible, and ultimately enrolled were tracked, as was the recruitment rate (participants enrolled per month). The medium by which enrolled participants were recruited was recorded (eg, clinical referral or targeted internet and social media advertisements). The demographics of enrolled participants were summarized and separated by those who did and did not complete the program.

Data Collection

Completion rates of all clinical program outcomes from the initial, final, and follow-up assessments were calculated. Completion rates of the within-program emailed surveys (of 5 total) were calculated.

Intervention Acceptability and Suitability

Attendance rates were calculated for videoconference assessments and synchronous exercise sessions (including both the standard [2 sessions] and symptom-triggered sessions [up to 5 sessions]). Completion rates were calculated for asynchronous exercise sessions (based on downloaded session logs including previously described self-reported completion). Program safety was assessed by recording the number and nature of adverse events.

Preliminary Evaluation of Participant Responses

Participant responses to the intervention were evaluated by calculating changes in patient-reported and physical outcomes from initial to final assessments. Maintenance of participant responses following the intervention was investigated by calculating changes in patient-reported outcomes from final to follow-up assessments. FACIT-Fatigue change was also assessed on an individual participant level, with the number of participants achieving an MCID following the program determined. Within-program individual changes were visually investigated by plotting FACIT-Fatigue scores on the CRF reference chart, along with the occurrence of symptom-triggered sessions.

Statistical Analysis

As a feasibility study with a small sample size, the majority of outcomes were summarized with descriptive statistics (counts, percentages, means, and SDs) and separately quantified for each assessment time point (initial, final, and follow-up) when appropriate. Following recommendations [18], the preliminary evaluation of participant responses was assessed using multiple approaches. In addition to the descriptive statistics provided for clinical outcomes and their change scores, the significance of change scores was assessed by statistically comparing these scores to 0 (ie, representing no change) using the nonparametric Wilcoxon signed rank test, as recommended for smaller studies as it does not assume a normal data distribution [40]. Significance was set at $\alpha<.05$, established a priori, without correction for multiple tests, but with the presentation of individual *P* values facilitating further scrutiny. The clinical meaningfulness of changes in FACIT-Fatigue scores, the primary clinical outcome, was assessed by summarizing the participant-level outcomes and visual investigations (see Preliminary Evaluation of Participant Responses section). Only participants with available data were included in each analysis, with sample sizes reported for each analysis. All statistical analyses were performed using R statistical software (R Foundation for Statistical Computing).

Ethical Considerations

This study was reviewed and approved by the Colorado Multiple Institutional Review Board (COMIRB # 20-2015) and registered on ClinicalTrials.gov (NCT04533165). All participants provided

written informed consent prior to enrollment. Participants were incentivized to enroll by being allowed to keep all distributed assessment and exercise equipment (approximate US $200 value). Within-program survey completion was incentivized by providing a US $50 gift card if at least 3 were completed. Participation in the final assessment was incentivized with a US $50 gift card upon completion. Study data were stored securely in REDCap, and analyses used deidentified data.

Results

Recruitment

Figure 2 displays the flow of participants from screening to program completion, with attrition at each stage. Ultimately, 51 survivors of cancer were screened, 19 enrolled in the program, and 15 completed the program. All enrolled participants reported a “White” racial background. No participants reported Hispanic or Latino ethnicity. Of the enrolled participants, 14 (74%) were recruited through internet and social media advertisements, 4 (21%) were recruited from a registry of past participants in the supervised BfitBwell Cancer Exercise Program, and 1 person was recruited through clinical connections at a large urban cancer center. When all recruitment efforts were active, the maximum recruitment rate was 5 enrolled participants per month. Demographics and cancer-related data for enrolled participants, separated by those who did and did not complete the program, are shown in Table 1, and reasons for study withdrawal are shown in Figure 2 (note that most withdrawal reasons are unrelated to demographic and cancer-related data). All subsequent results include only participants who completed the program (and follow-up assessment when relevant).

One participant was withdrawn from the study due to the clinical determination that participation in a telehealth exercise program would not be safe due to previously unreported balance impairments, making the participant ineligible for the investigation. This safety determination was based primarily upon movement observation and objectively supported by poor performance on physical measures (single limb stance=3.5 seconds, gait speed=0.92 m/s, and inability to perform step test despite adaptation). Upon informing the participant of this decision, the study team facilitated a connection to a nearby facility providing supervised and skilled rehabilitation.

Figure 2. Illustration of screening, enrollment, and attrition during the study. CRF: cancer-related fatigue.

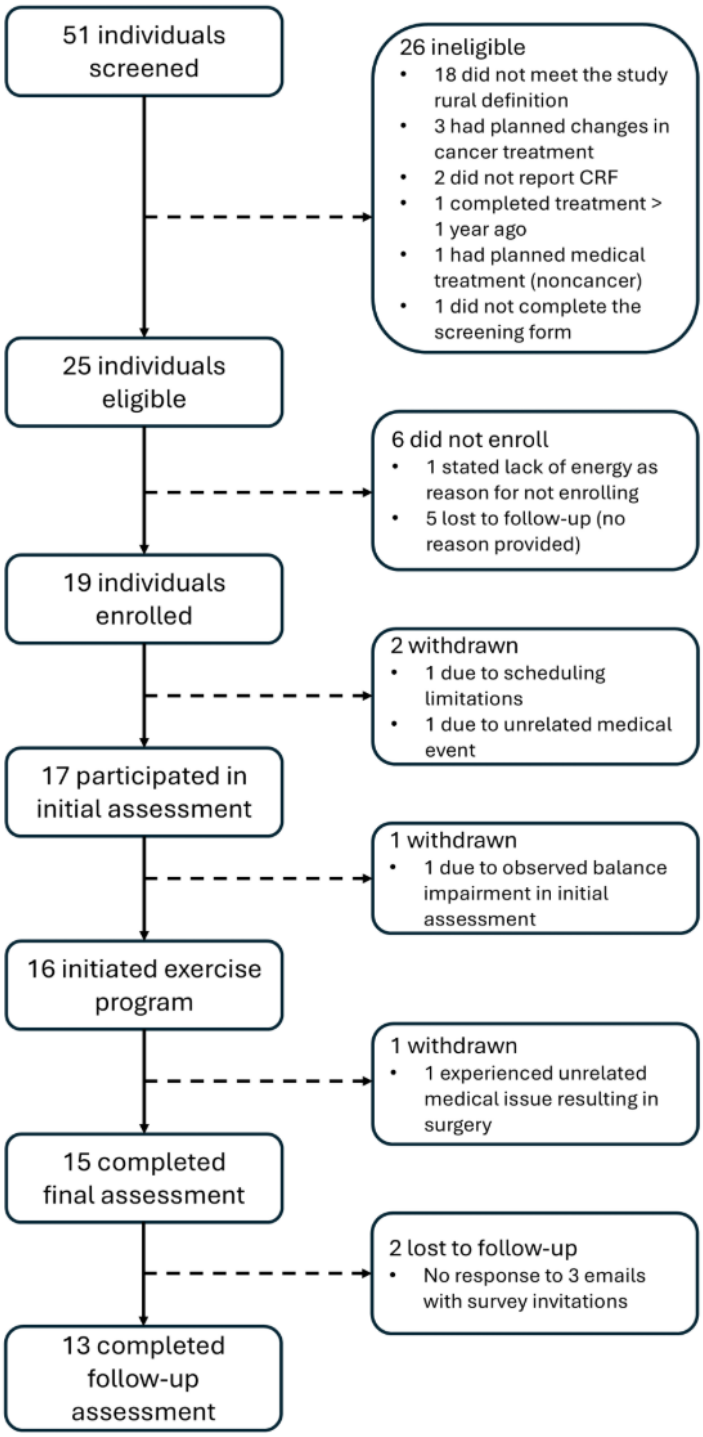


Table 1. Enrolled participant demographic and cancer-related information.

	Completed program (n=15)	Did not complete program (n=4)
Age (years)		
Mean (SD)	60.7 (6.7)	55.5 (8.1)
Range	49-71	44-63
Sex distribution (female), n(%)	11 (73)	4 (100)
BMI (kg/m^2)		
Mean (SD)	26.3 (4.1)	34.2 (11.9)
Range	20.1-33.7	20.8-43.6
Cancer diagnosis, n (%)		
Breast	6 (40)	3 (75)
Colorectal	1 (7)	1 (25)
Kidney	1 (7)	0 (0)
Lung	1 (7)	0 (0)
Multiple myeloma	2 (13)	0 (0)
Non-Hodgkin lymphoma	1 (7)	0 (0)
Ovarian	1 (7)	0 (0)
Prostate	1 (7)	0 (0)
Thyroid	1 (7)	0 (0)
Receiving current treatment, n (%)		
Chemotherapy	3 (20)	0 (0)
Immunotherapy	1 (7)	0 (0)
Hormonal therapy	5 (33)	0 (0)
None	6 (40)	4 (100)

Data Collection

All physical assessment outcomes were successfully completed during the initial and final assessments except the 3-minute step test. In total, 3 participants were unable to maintain the required step rate for the duration of the test, 1 participant did not perform the test due to concerns of exceeding a safe heart rate (85% age-predicted maximum), and 2 participants experienced technical difficulties in measuring heart rate following completion of the test. All patient-reported outcomes were successfully completed at the initial assessment, and only 1 participant failed to complete both the Hospital Anxiety and Depression Scale and GLTEQ (presented last in emailed surveys) at the final assessment. In total, 13 (87% of those completing the program) participants completed all patient-reported outcomes at the follow-up assessment. The average completion rate of the within-program emailed surveys was 93% (70/75 surveys across all participants), with 12 (80% of those completing the program) completing 100%.

Intervention Acceptability and Suitability

Attendance at initial and final videoconference assessments was 100% (30/30 possible assessments attended across all participants). Attendance at videoconference exercise sessions (weeks 1 and 6) was 97% (29/30 possible sessions attended), with only 1 participant missing 1 session. In total, 7 symptom-triggered sessions were initiated in 6 participants, with 6 (86%) of these sessions attended. A total of 3 participants requested 1 additional session, and 1 participant requested 2 additional sessions (non–CRF related). A total of 5 participants did not trigger or request any additional sessions. Participants received an average of 58 (SD 7) asynchronous sessions each, with an average of 49 (SD 11) indicated as complete (49/58, 84%, individual range 38%-100% [23/60 and 60/60 possible sessions completed, respectively]).

In total, 7 adverse events were reported in 6 participants. Of them, 2 were minor musculoskeletal issues (muscle strains) likely related to the exercise intervention. The remaining events were unrelated to the exercise intervention (minor musculoskeletal issues and illness).

Preliminary Evaluation of Participant Responses

Table 2 displays the averages and changes in all outcomes at initial, final, and follow-up assessments. Group changes in all patient-reported outcomes from initial to final assessments were significantly different from 0 (in directions demonstrating improvement; all *P* values <.05, see Table 2 for individual values). Change in 30-second sit-to-stand was significantly greater than 0 (*P*=.005), and change in timed up and go trended toward being less than 0 (*P*=.09). At the follow-up assessment,

only self-reported resistance exercise (GLTEQ-Resistance) significantly decreased from the final assessment (*P*=.01).

Table 2. Patient-reported and physical outcomes at initial, final, and 6-month follow-up assessments (n=15).

	Initial		Final				Follow-up			
	Values, n (%)	Mean (SD)	Values, n (%)	Mean (SD)	Change (SD)[a]	*P* value[b]	Values, n (%)	Mean (SD)	Change (SD)[a]	*P* value[b]
FACIT-Fatigue[c]	15 (100)	30.9 (7.4)	15 (100)	42.1 (8.0)	*11.2 (6.8)* [d]	*.001*	13 (87)	40.9 (9.7)	–1.1 (3.4)	.29
FACT-G[e]	15 (100)	71.9 (11.6)	15 (100)	85.3 (13.3)	*13.4 (8.7)*	*.001*	13 (87)	85.4 (16.7)	0.8 (6.2)	>.99
HADS-A[f]	14 (93)	8.4 (5.7)	15 (100)	5.1 (4.3)	*–3.1 (3.6)*	*.002*	13 (87)	5.3 (4.1)	0 (2.2)	>.99
HADS-D[g]	14 (93)	6.2 (3.6)	15 (100)	3.8 (3.8)	*–2.2 (1.2)*	*.001*	13 (87)	4.1 (4.1)	0.1 (1.6)	.86
GLTEQ[h]-MVPA[i] (minutes per week)	14 (93)	89.6 (105.4)	15 (100)	178.7 (171.1)	*95.4 (131.8)*	*.01*	13 (87)	148.1 (114.8)	–13.8 (129.0)	.66
GLTEQ-Resistance (minutes per week)	14 (93)	23.6 (37.3)	15 (100)	78.3 (32.1)	*50.7 (55.8)*	*.01*	13 (87)	21.9 (39.2)	*–51.2 (53.9)*	*.01*
Gait speed (m/s)	15 (100)	1.16 (0.15)	15 (100)	1.23 (0.10)	0.07 (0.17)	.12	—[j]	—	—	—
TUG[k] (seconds)	15 (100)	9.2 (1.1)	15 (100)	8.8 (1.0)	–0.4 (0.9)	.08	—	—	—	—
30 s StS[l] (repetitions)	15 (100)	11.7 (3.7)	15 (100)	13.3 (3.8)	*1.6 (1.7)*	*.005*	—	—	—	—
SLS-D[m] (seconds)	15 (100)	28.2 (5.2)	15 (100)	26.4 (7.9)	–1.8 (4.8)	.10	—	—	—	—
SLS-ND[n] (seconds)	15 (100)	29.0 (3.7)	15 (100)	28.2 (5.2)	–0.8 (6.7)	.79	—	—	—	—

[a]Change calculated from the previous assessment.

[b]Change statistically compared to 0 with the Wilcoxon rank sum test.

[c]FACIT-Fatigue: Functional Assessment of Chronic Illness Therapy—Fatigue Scale.

[d]Values in italics format emphasize *P*<.05, the a priori significance threshold.

[e]FACT-G: Functional Assessment of Cancer Treatment—General Scale.

[f]HADS-A: Hospital Anxiety and Depression Anxiety Scale—Anxiety Score.

[g]HADS-D: Hospital Anxiety and Depression Anxiety Scale—Depression Score.

[h]GLTEQ: Godin Leisure-Time Exercise Questionnaire.

[i]MVPA: moderate to vigorous physical activity.

[j]Not applicable.

[k]TUG: timed up and go.

[l]30 s StS: 30-second sit-to-stand.

[m]SLS-D: single leg stance—dominant (30 seconds maximum).

[n]SLS-ND: single leg stance—nondominant (30 seconds maximum).

Figure 3 displays the individual FACIT-Fatigue changes from the initial to final assessments. In total, 13 of 15 (87%) participants demonstrated an MCID in FACIT-Fatigue change. Figure 4 displays individual within-program FACIT-Fatigue scores, plotted on the CRF reference chart, for participants who did not require any symptom-triggered exercise sessions. Figure 5 displays individual within-program FACIT-Fatigue scores, plotted on the CRF reference chart, for participants requiring symptom-triggered exercise sessions, along with sessions attended. Of note, FACIT-Fatigue scores improved following symptom-triggered exercise sessions in 5 of 6 events, following previous declines.

Figure 3. Waterfall plot displaying individual participant (n=15) changes in FACIT-Fatigue from initial to final assessments. The sample mean and FACIT-Fatigue MCID are displayed as dotted lines. Note that an increase in the FACIT-Fatigue score indicates improved fatigue. FACIT-Fatigue: Functional Assessment of Chronic Illness Therapy—Fatigue Scale; MCID: minimal clinically important difference.

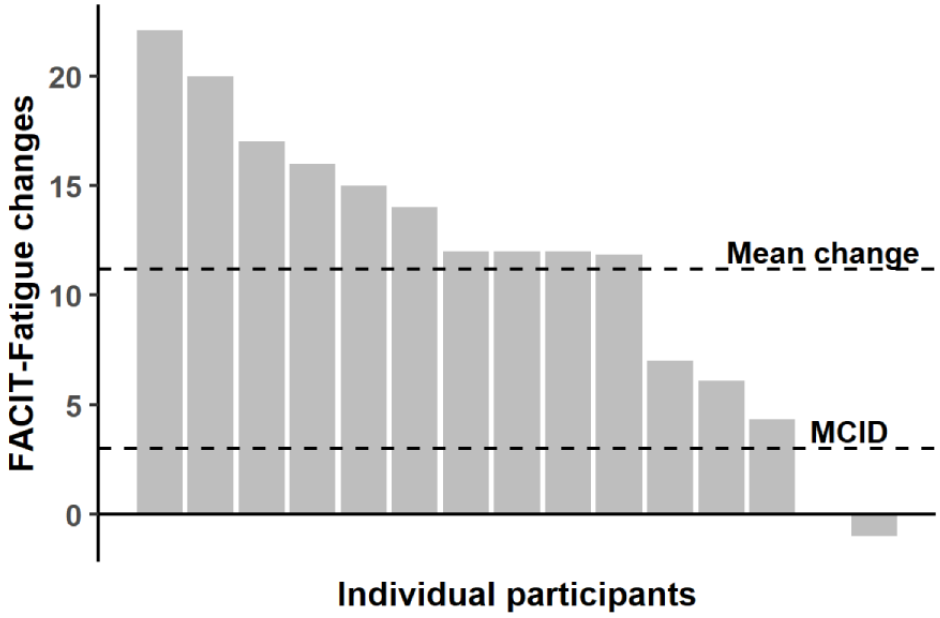


Figure 4. Within-program FACIT-Fatigue scores for participants with no required symptom-triggered exercise session (n=9) plotted on the cancer-related fatigue reference chart. FACIT-Fatigue: Functional Assessment of Chronic Illness Therapy—Fatigue Scale.

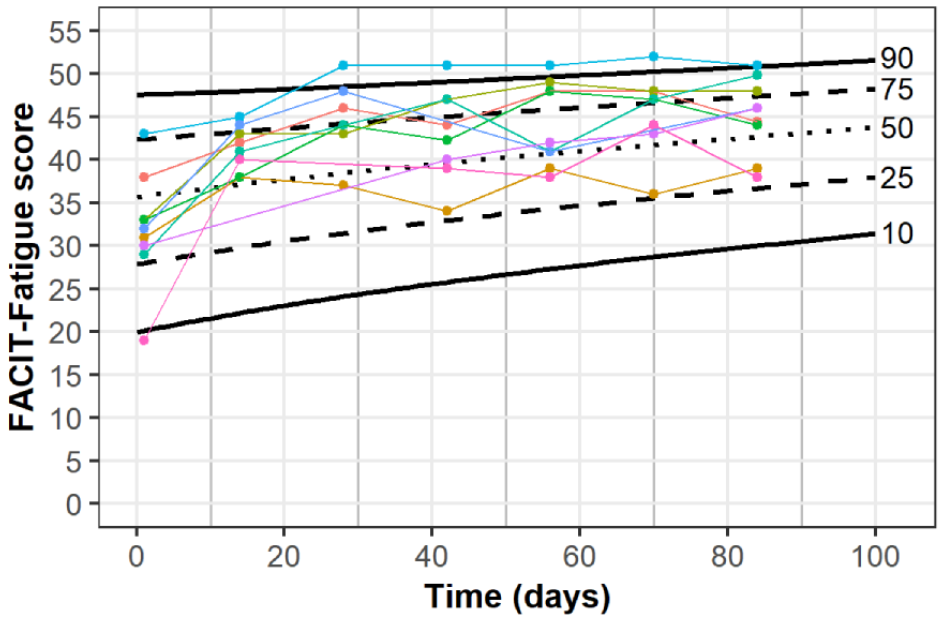
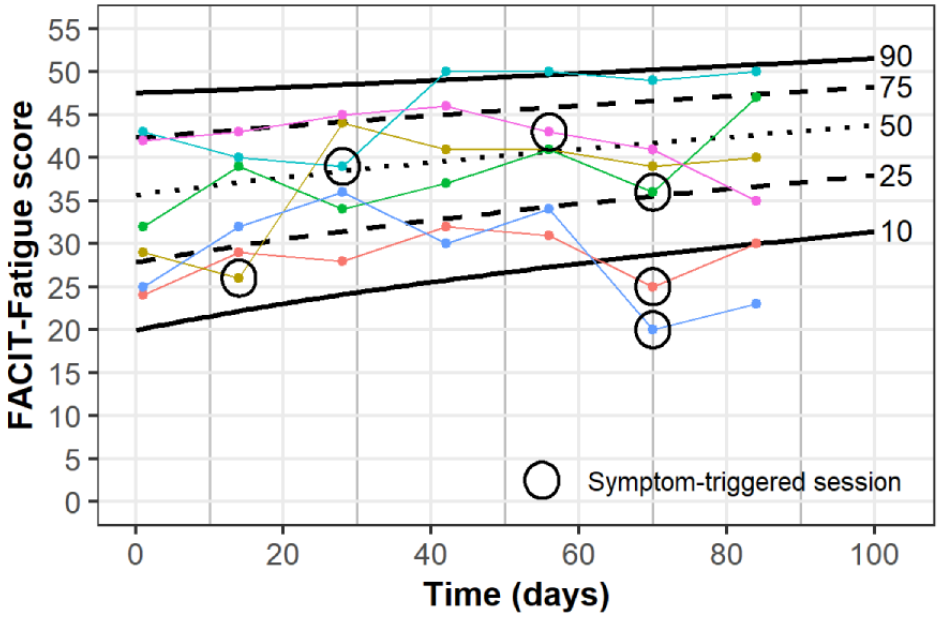


Figure 5. Within-program FACIT-Fatigue scores for participants requiring symptom-triggered exercise sessions (n=6) plotted on the cancer-related fatigue reference chart. Attended symptom-triggered sessions are indicated with circles. FACIT-Fatigue: Functional Assessment of Chronic Illness Therapy—Fatigue Scale.



Discussion

Principal Findings

This investigation assessed the feasibility of the novel BfitBwell Telehealth Program for rural survivors of cancer with CRF. Initial recruitment barriers were overcome through the adoption of alternative strategies (targeted internet advertisements). Outcome completion rates and intervention adherence rates were high, demonstrating the feasibility of data collection and intervention acceptability and suitability. Significant and meaningful improvements in CRF were observed at both the group and individual levels in the preliminary evaluation of participant responses. Based on these assessments of the recommended objectives [18], the BfitBwell Telehealth Program appears to be highly feasible, supporting the progression to a large efficacy trial.

Comparisons to Prior Work

One of the main goals of this feasibility study was to adjudicate avenues for recruitment of eligible rural-living individuals. Given the barriers in accessing this population, it is not surprising that little work has been done specifically investigating the recruitment of rural-living people with cancer into exercise programs. A recent review of recruitment rates and strategies in exercise trials for survivors of cancer (not rural specific) revealed a low overall recruitment rate of 38%, with only 11% of included trials using web-based advertisements [41]. Additionally, “passive” strategies (including web-based advertisement) resulted in lower rates than “active” strategies (including clinic-based recruitment). Another larger exercise trial in rural survivors also demonstrated the success of active strategies such as community and clinical engagement [16]. Our findings contrast with these previous studies, in that only 5% of participants were enrolled as a result of active recruitment strategies (ie, clinical connections) at a large urban cancer center. A much more productive approach involved the use of targeted internet and social media advertisements. These approaches may indeed be considered more “active” than previous web-based approaches, as advertisements are routed to individuals based on their past internet activity. The reach of these advertisements to rural communities is undoubtedly broader than what can be achieved through clinic-based recruitment approaches. In our case, web-based recruitment might also provide a good match for individuals seeking a web- or telehealth-based program. More investigation is required into how to best engage and recruit rural survivors in exercise programming and associated clinical trials.

A unique aspect of our particular program was the reference chart–based monitoring of participant response to exercise along with the standardized addition of symptom-triggered sessions when participant progress deviated from the expected response. Lower than expected CRF improvements triggered additional sessions in 6 (40%) participants, and subsequent CRF scores appeared to improve following these symptom-triggered sessions (Figure 5). The purpose of these sessions was to mirror within-program exercise adaptations that are common in supervised clinical exercise programs in response to provider observations. While this facet of personalized exercise prescription is a relatively emergent feature in contemporary clinical research, the approach adheres to recommended best practices published by the National Comprehensive Cancer Network [42]. The current combination of symptom monitoring and symptom-triggered intervention is similar to a previous study of “chemotherapy-periodized” exercise, which demonstrated improved attendance when exercise prescription was adapted in anticipation of changing chemotherapy-associated symptoms during consecutive chemotherapy cycles [43]. Adaptations in the current program were made in accordance with the ACSM Frequency, Intensity, Time, and Type criteria of exercise prescription [39] and were

based on CES expertise and information gathered during participant discussions.

Exercise attendance and program completion rates in our study were high. In total, 15 of 16 (94%) participants completed the program, and attendance and completion of all exercise sessions ranged from 84% (49/58 asynchronous sessions completed, on average) to 97% (29/30 videoconference sessions attended, across all participants). Additionally, within-program CRF monitoring was objectively successful, with an overall survey completion rate of 93% (70/75 surveys completed, across all participants). In total, 9 participants did not require symptom-triggered exercise sessions, completing the program largely autonomously and asynchronously. The lack of symptom-triggered sessions indicates similar CRF improvement compared to participants in a supervised exercise program, despite a reduced number of real-time (or synchronous) exercise sessions compared to many of these programs [19]. The asynchronous method of providing exercise programming and supervision used in this investigation may hold promise for improving both exercise efficacy and program efficiency in telehealth exercise programs.

The mechanisms underlying the observed CRF improvements remain uncertain, given the current investigation is a feasibility trial. Nonetheless, several strategies are likely to have contributed to engagement with the exercise prescription, facilitating exercise-associated CRF improvements in this cohort. The programmatic design specifically addressed three known barriers to exercise engagement in rural survivors of cancer, which are not frequently or consistently addressed in other programs (beyond reducing travel burden) [12,13]: (1) the need to travel was completely removed, (2) asynchronous programming provided flexibility for where and when exercise was performed, and (3) several methods of communication provided direct access to a knowledgeable exercise provider. For the involved CES, an unanticipated benefit of asynchronous programming was that the time commitment required may be lower than for a supervised program. While perhaps obvious, pragmatic and cost-effective strategies for improved exercise engagement may prove critical to overall program effectiveness, where personnel costs and reimbursement funds limit more intensive strategies. To this end, additional investigation of objective measures of program efficiency using asynchronous exercise programming is required.

Strengths and Limitations

The primary limitation of this investigation is the lack of a control group and small sample size. Particularly in evaluating preliminary participant responses, the observed improvements during the program or in response to symptom-triggered sessions cannot be solely attributed to the intervention. However, several outcomes support the plausibility of the program influencing a meaningful improvement in CRF. First, the observed within-program improvements in CRF were similar to those documented in the supervised exercise program used to develop the CRF reference chart [23]. Second, significant improvements in CRF were observed immediately following the program, but no significant changes were observed in the 6 months following the program, indicating a lack of change due to time as well as the potential maintenance of program effects. Statistical analyses of pilot and feasibility trials, however, should be interpreted with caution [18]. Finally, on the participant level, the majority of participants (13/15, 87%) experienced a clinically meaningful improvement in CRF.

The sample in the current investigation is also likely biased in several ways. Given that participants responded to advertisements for an exercise intervention, the current sample is likely to reflect individuals who are able and willing to exercise. Additionally, the current sample is not large enough to adequately assess the contribution of various demographic and clinical variables on program adherence and response. The use of web-based recruitment methods and telehealth technology may discourage the participation of survivors without or not comfortable with technology. While the incorporation of this self-selecting sample population may contribute to a current "digital divide," parallel efforts are actively reducing technological barriers through efforts to increase the availability and acceptance of adequate technology and internet connections. For example, the Broadband Equity, Access, and Deployment Program [44] seeks to expand high-speed internet access throughout the United States, facilitating a societal migration to technology-based health care interventions (including exercise). Given that the current investigation was performed during the COVID-19 pandemic, participants' perspectives of and adherence to telehealth exercise may have been positively influenced in a manner unrepresentative of their more general context.

The used definition of rurality was developed independently by the research team, designed to target a population in need of services in the local region (Colorado and surrounding states). While this emphasis on local context is appropriate for this small feasibility study, future studies targeting larger populations in larger regions should use standardized rurality definitions to facilitate broader generalizations [10]. The telehealth tools used to facilitate access to this rural population may also contribute to a decrease in the fidelity of the prescribed exercise prescription. Specifically, only self-report completion data were available for asynchronous exercise sessions, limiting the knowledge of actual completion and performance. Future work could integrate additional technology (eg, improved activity trackers) to objectively assess the performance of these sessions.

Finally, while this investigation supports the feasibility of the BfitBwell Telehealth Program, it does not represent a comprehensive assessment of feasibility. While one set of recommended assessments was followed, multiple alternative definitions of feasibility and associated measures exist. Specifically, the assessment of intervention acceptability can be complex, including both quantitative and qualitative evaluations [45]. Nonetheless, the presented feasibility assessments provide strong support for further investigation of this program and its methods.

Conclusions

The BfitBwell Telehealth Program used several telehealth modalities combined with regular within-program symptom monitoring and symptom-triggered intervention to deliver an exercise program to rural survivors of cancer with CRF. This

investigation demonstrated high program feasibility, supported by positive assessments of recruitment, data collection, intervention acceptability and suitability, and preliminary evaluation of participant responses. Novel methods used by the program also provide a potential model for improving exercise program efficiency by using asynchronous exercise prescription. Future work should pursue large-scale efficacy testing, objective assessments of program efficiency, and systematic investigations of the effects of within-program exercise adaptations.

Acknowledgments

This work was supported (in part) by a research grant from the Cancer League of Colorado, Inc. The BfitBwell Cancer Exercise Program is partially supported by the University of Colorado Cancer Center (5 P30 CA046934). REDCap (Research Electronic Data Capture) use is supported by the National Institutes of Health/National Center for Research Resources Colorado Clinical and Translational Science Award (grant UL1 TR002535). RJM is supported by K12CA086913. AJK is supported by the Center for Population Health Research at the University of Montana (National Institute of General Medical Sciences NIH P20GM130418). The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

RJM, AJK, JJS, and IAM contributed to conceptualization and methodology. RJM, JJS, and IAM contributed to project administration and data curation. RJM, JJS, and AJK contributed to the formal analysis. RJM, AJK, JJS, IAM, JCQ, and HJL contributed to writing the original draft as well as review and edits.

Conflicts of Interest

None declared.

References

1. Harrington CB, Hansen JA, Moskowitz M, Todd BL, Feuerstein M. It's not over when it's over: long-term symptoms in cancer survivors—a systematic review. Int J Psychiatry Med 2010;40(2):163-181. [doi: 10.2190/PM.40.2.c] [Medline: 20848873]
2. Ma Y, He B, Jiang M, Yang Y, Wang C, Huang C, et al. Prevalence and risk factors of cancer-related fatigue: a systematic review and meta-analysis. Int J Nurs Stud 2020;111:103707. [doi: 10.1016/j.ijnurstu.2020.103707] [Medline: 32920423]
3. Al Maqbali M, Al Sinani M, Al Naamani Z, Al Badi K, Tanash MI. Prevalence of fatigue in patients with cancer: a systematic review and meta-analysis. J Pain Symptom Manage 2021;61(1):167-189.e14 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.07.037] [Medline: 32768552]
4. Roila F, Fumi G, Ruggeri B, Antonuzzo A, Ripamonti C, Fatigoni S, et al. Prevalence, characteristics, and treatment of fatigue in oncological cancer patients in Italy: a cross-sectional study of the Italian Network for Supportive Care in Cancer (NICSO). Support Care Cancer 2019;27(3):1041-1047. [doi: 10.1007/s00520-018-4393-9] [Medline: 30084104]
5. Bower JE, Ganz PA, Desmond KA, Bernaards C, Rowland JH, Meyerowitz BE, et al. Fatigue in long-term breast carcinoma survivors: a longitudinal investigation. Cancer 2006;106(4):751-758 [FREE Full text] [doi: 10.1002/cncr.21671] [Medline: 16400678]
6. Hofman M, Ryan JL, Figueroa-Moseley CD, Jean-Pierre P, Morrow GR. Cancer-related fatigue: the scale of the problem. Oncologist 2007;12 Suppl 1:4-10 [FREE Full text] [doi: 10.1634/theoncologist.12-S1-4] [Medline: 17573451]
7. Meneses-Echávez JF, González-Jiménez E, Ramírez-Vélez R. Effects of supervised multimodal exercise interventions on cancer-related fatigue: systematic review and meta-analysis of randomized controlled trials. Biomed Res Int 2015;2015:328636 [FREE Full text] [doi: 10.1155/2015/328636] [Medline: 26167483]
8. van Vulpen JK, Sweegers MG, Peeters PHM, Courneya KS, Newton RU, Aaronson NK, et al. Moderators of exercise effects on cancer-related fatigue: a meta-analysis of individual patient data. Med Sci Sports Exerc 2020;52(2):303-314 [FREE Full text] [doi: 10.1249/MSS.0000000000002154] [Medline: 31524827]
9. Campbell KL, Winters-Stone KM, Wiskemann J, May AM, Schwartz AL, Courneya KS, et al. Exercise guidelines for cancer survivors: consensus statement from international multidisciplinary roundtable. Med Sci Sports Exerc 2019;51(11):2375-2390 [FREE Full text] [doi: 10.1249/MSS.0000000000002116] [Medline: 31626055]
10. Franks JA, Davis ES, Bhatia S, Kenzik KM. Defining rurality: an evaluation of rural definitions and the impact on survival estimates. J Natl Cancer Inst 2023;115(5):530-538 [FREE Full text] [doi: 10.1093/jnci/djad031] [Medline: 36762829]
11. Levit LA, Byatt L, Lyss AP, Paskett ED, Levit K, Kirkwood K, et al. Closing the rural cancer care gap: three institutional approaches. JCO Oncol Pract 2020;16(7):422-430. [doi: 10.1200/OP.20.00174] [Medline: 32574128]

12. Smith-Turchyn J, Allen L, Dart J, Lavigne D, Rooprai S, Dempster H, et al. Characterizing the exercise behaviour, preferences, barriers, and facilitators of cancer survivors in a rural Canadian community: a cross-sectional survey. Curr Oncol 2021;28(4):3172-3187 [FREE Full text] [doi: 10.3390/curroncol28040276] [Medline: 34436042]
13. Adams JL, Martin MY, Pisu M, Oster RA, Qu H, Shewchuk RM, et al. Determining patient needs to enhance exercise program implementation and uptake in rural settings for women after a cancer diagnosis. Support Care Cancer 2021;29(8):4641-4649 [FREE Full text] [doi: 10.1007/s00520-021-05990-8] [Medline: 33495849]
14. Smith-Turchyn J, Gravesande J, Sabiston CM. Exercise interventions for survivors of cancer living in rural or remote settings: a scoping review. Rehabil Oncol 1991;38(2):61-80. [doi: 10.1097/01.Reo.0000000000000208]
15. Culos-Reed N, Wagoner CW, Dreger J, McNeely ML, Keats M, Santa Mina D, et al. Implementing an exercise oncology model to reach rural and remote individuals living with and beyond cancer: a hybrid effectiveness-implementation protocol for project EXCEL (EXercise for Cancer to Enhance Living Well). BMJ Open 2022;12(12):e063953 [FREE Full text] [doi: 10.1136/bmjopen-2022-063953] [Medline: 36581419]
16. Wagoner CW, Dreger J, Keats MR, Santa Mina D, McNeely ML, Cuthbert C, et al. First-year implementation of the EXercise for Cancer to Enhance Living Well (EXCEL) study: building networks to support rural and remote community access to exercise oncology resources. Int J Environ Res Public Health 2023;20(3):1930 [FREE Full text] [doi: 10.3390/ijerph20031930] [Medline: 36767296]
17. Wonders KY, Gnau K, Schmitz KH. Measuring the feasibility and effectiveness of an individualized exercise program delivered virtually to cancer survivors. Curr Sports Med Rep 2021;20(5):271-276. [doi: 10.1249/JSR.0000000000000846] [Medline: 33908915]
18. Orsmond GI, Cohn ES. The distinctive features of a feasibility study: objectives and guiding questions. OTJR (Thorofare N J) 2015;35(3):169-177. [doi: 10.1177/1539449215578649] [Medline: 26594739]
19. Marker RJ, Cox-Martin E, Jankowski CM, Purcell WT, Peters JC. Evaluation of the effects of a clinically implemented exercise program on physical fitness, fatigue, and depression in cancer survivors. Support Care Cancer 2018;26(6):1861-1869. [doi: 10.1007/s00520-017-4019-7] [Medline: 29270829]
20. Marker RJ, Ostendorf DM, Leach HJ, Peters JC. Cancer-related fatigue mediates the relationships between physical fitness and attendance and quality of life after participation in a clinical exercise program for survivors of cancer. Qual Life Res 2022;31(11):3201-3210. [doi: 10.1007/s11136-022-03173-4] [Medline: 35895163]
21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
22. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
23. Marker RJ, Kittelson AJ, Jankowski CM, Scorsone JJ, Peters JC, Purcell WT. Development of a reference chart for monitoring cancer-related fatigue throughout a supervised exercise program. Rehabil Oncol 2022;40(2):82-88 [FREE Full text] [doi: 10.1097/01.reo.0000000000000285] [Medline: 35711819]
24. van Buuren S, Hulzebos EHJ, Valkenet K, Lindeman E, van Meeteren NLU. Reference chart of inspiratory muscle strength: a new tool to monitor the effect of pre-operative training. Physiotherapy 2014;100(2):128-133. [doi: 10.1016/j.physio.2013.08.007] [Medline: 24216046]
25. Kittelson AJ, Elings J, Colborn K, Hoogeboom TJ, Christensen JC, van Meeteren NLU, et al. Reference chart for knee flexion following total knee arthroplasty: a novel tool for monitoring postoperative recovery. BMC Musculoskelet Disord 2020;21(1):482 [FREE Full text] [doi: 10.1186/s12891-020-03493-x] [Medline: 32698900]
26. Berger AM, Mooney K, Alvarez-Perez A, Breitbart WS, Carpenter KM, Cella D, et al. Cancer-related fatigue, version 2.2015. J Natl Compr Canc Netw 2015;13(8):1012-1039 [FREE Full text] [doi: 10.6004/jnccn.2015.0122] [Medline: 26285247]
27. Exercise program directory. Exercise is Medicine. URL: https://www.exerciseismedicine.org/eim-in-action/moving-through-cancer/exercise-program-registry/ [accessed 2024-01-11]
28. Cella D, Lai JS, Stone A. Self-reported fatigue: one dimension or more? Lessons from the Functional Assessment of Chronic Illness Therapy—Fatigue (FACIT-F) questionnaire. Support Care Cancer 2011;19(9):1441-1450. [doi: 10.1007/s00520-010-0971-1] [Medline: 20706850]
29. Minton O, Stone P. A systematic review of the scales used for the measurement of cancer-related fatigue (CRF). Ann Oncol 2009;20(1):17-25 [FREE Full text] [doi: 10.1093/annonc/mdn537] [Medline: 18678767]
30. Cella D, Eton DT, Lai JS, Peterman AH, Merkel DE. Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. J Pain Symptom Manage 2002;24(6):547-561 [FREE Full text] [doi: 10.1016/s0885-3924(02)00529-8] [Medline: 12551804]
31. Hoenemeyer TW, Cole WW, Oster RA, Pekmezi DW, Pye A, Demark-Wahnefried W. Test/retest reliability and validity of remote vs. in-person anthropometric and physical performance assessments in cancer survivors and supportive partners. Cancers (Basel) 2022;14(4):1075 [FREE Full text] [doi: 10.3390/cancers14041075] [Medline: 35205823]

32. Marker RJ, Kakar RS, Scorsone JJ, Peters JC, Purcell WT. Single-leg stance times in a diverse group of survivors of cancer and the relationship to history of recent falls. Rehabil Oncol 2021;39(1):23-30. [doi: 10.1097/01.reo.0000000000000243]
33. Huang MH, Hile E, Croarkin E, Wampler-Kuhn M, Blackwood J, Colon G, et al. Academy of oncologic physical therapy EDGE Task Force: a systematic review of measures of balance in adult cancer survivors. Rehabil Oncol 2019;37(3):92-103. [doi: 10.1097/01.REO.0000000000000177]
34. Nilsen TS, Raastad T, Skovlund E, Courneya KS, Langberg CW, Lilleby W, et al. Effects of strength training on body composition, physical functioning, and quality of life in prostate cancer patients during androgen deprivation therapy. Acta Oncol 2015;54(10):1805-1813. [doi: 10.3109/0284186X.2015.1037008] [Medline: 25927504]
35. Hughes AD, Chaturvedi N. Estimation of maximal oxygen consumption and heart rate recovery using the Tecumseh sub-maximal step test and their relationship to cardiovascular risk factors. Artery Res 2017;18:29-35 [FREE Full text] [doi: 10.1016/j.artres.2017.02.005] [Medline: 28546848]
36. Brucker PS, Yost K, Cashy J, Webster K, Cella D. General population and cancer patient norms for the Functional Assessment of Cancer Therapy-General (FACT-G). Eval Health Prof 2005;28(2):192-211. [doi: 10.1177/0163278705275341] [Medline: 15851773]
37. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
38. Amireault S, Godin G, Lacombe J, Sabiston CM. The use of the Godin-Shephard Leisure-Time Physical Activity Questionnaire in oncology research: a systematic review. BMC Med Res Methodol 2015;15:60. [doi: 10.1186/s12874-015-0045-7] [Medline: 26264621]
39. American College of Sports Medicine, Liguori G. ACSM's Guidelines for Exercise Testing and Prescription. 10th Edition. Philadelphia, PA: Wolters Kluwer; 2018.
40. Fagerland MW. t-tests, non-parametric tests, and large studies—a paradox of statistical practice? BMC Med Res Methodol 2012;12:78 [FREE Full text] [doi: 10.1186/1471-2288-12-78] [Medline: 22697476]
41. Reynolds SA, O'Connor L, McGee A, Kilcoyne AQ, Connolly A, Mockler D, et al. Recruitment rates and strategies in exercise trials in cancer survivorship: a systematic review. J Cancer Surviv 2024;18(4):1233-1242. [doi: 10.1007/s11764-023-01363-8] [Medline: 37022641]
42. Denlinger CS, Ligibel JA, Are M, Baker KS, Demark-Wahnefried W, Dizon D, et al. Survivorship: healthy lifestyles, version 2.2014. J Natl Compr Canc Netw 2014;12(9):1222-1237 [FREE Full text] [doi: 10.6004/jnccn.2014.0121] [Medline: 25190692]
43. Kirkham AA, Bland KA, Zucker DS, Bovard J, Shenkier T, McKenzie DC, et al. "Chemotherapy-periodized" exercise to accommodate for cyclical variation in fatigue. Med Sci Sports Exerc 2020;52(2):278-286. [doi: 10.1249/MSS.0000000000002151] [Medline: 31490858]
44. Broadband Equity, Access, and Deployment (BEAD) Program. BroadbandUSA. URL: https://broadbandusa.ntia.doc.gov/funding-programs/broadband-equity-access-and-deployment-bead-program [accessed 2024-10-04]
45. Bucyibaruta JB, Peu D, Bamford L, van der Wath A. Closing the gaps in defining and conceptualising acceptability of healthcare: a qualitative thematic content analysis. Afr Health Sci 2022;22(3):703-709 [FREE Full text] [doi: 10.4314/ahs.v22i3.75] [Medline: 36910403]

Abbreviations

ACSM: American College of Sports Medicine
CES: cancer exercise specialist
CRF: cancer-related fatigue
FACIT-Fatigue: Functional Assessment of Chronic Illness Therapy—Fatigue Scale
GLTEQ: Godin Leisure-Time Exercise Questionnaire
MCID: minimal clinically important difference
REDCap: Research Electronic Data Capture

Edited by N Cahill; submitted 15.04.24; peer-reviewed by ST Jenna, C Rhéaume, MC Robertson, Z Pope; comments to author 13.09.24; revised version received 01.11.24; accepted 09.12.24; published 10.01.25.

Please cite as:
Marker RJ, Kittelson AJ, Scorsone JJ, Moran IA, Quindry JC, Leach HJ
A Novel Telehealth Exercise Program Designed for Rural Survivors of Cancer With Cancer-Related Fatigue: Single-Arm Feasibility Trial
JMIR Cancer 2025;11:e59478
URL: https://cancer.jmir.org/2025/1/e59478
doi:10.2196/59478
PMID:

Original Paper

Changes in Physical Activity Across Cancer Diagnosis and Treatment Based on Smartphone Step Count Data Linked to a Japanese Claims Database: Retrospective Cohort Study

Yoshihide Inayama[1,2], MD; Ken Yamaguchi[1], MD, PhD; Kayoko Mizuno[2,3], MD, PhD; Sachiko Tanaka-Mizuno[2,3,4], PhD; Ayami Koike[1], MD; Nozomi Higashiyama[1], MD; Mana Taki[1], MD, PhD; Koji Yamanoi[1], MD, PhD; Ryusuke Murakami[1], MD, PhD; Junzo Hamanishi[1], MD, PhD; Satomi Yoshida[2,5], PhD; Masaki Mandai[1], MD, PhD; Koji Kawakami[2], MD, PhD

[1]Department of Gynecology and Obstetrics, Graduate School of Medicine and Faculty of Medicine, Kyoto University, Kyoto, Japan

[2]Department of Pharmacoepidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan

[3]Department of Digital Health and Epidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan

[4]Laboratory of Epidemiology and Prevention, Kobe Pharmaceutical University, Kobe, Japan

[5]Institute of Medicine, University of Tsukuba, Tsukuba, Japan

Corresponding Author:
Ken Yamaguchi, MD, PhD
Department of Gynecology and Obstetrics
Graduate School of Medicine and Faculty of Medicine
Kyoto University
54 Shogoin Kawahara cho
Sakyo ku
Kyoto, 606-8507
Japan
Phone: 81 75 751 3269
Email: soulken@kuhp.kyoto-u.ac.jp

Abstract

Background: Although physical activity (PA) is recommended for patients with cancer, changes in PA across cancer diagnosis and treatment have not been objectively evaluated.

Objective: This study aimed to assess the impact of cancer diagnosis and treatment on PA levels.

Methods: This was a retrospective cohort study using a Japanese claims database provided by DeSC Healthcare Inc, in which daily step count data, derived from smartphone pedometers, are linked to the claims data. In this study, we included patients newly diagnosed with cancer, along with those newly diagnosed with diabetes mellitus for reference. We collected data between April 2014 and September 2021 and analyzed them. The observation period spanned from 6 months before diagnosis to 12 months after diagnosis. We applied a generalized additive mixed model with a cubic spline to describe changes in step counts before and after diagnosis.

Results: We analyzed the step count data of 326 patients with malignant solid tumors and 1388 patients with diabetes. Patients with cancer exhibited a 9.6% (95% CI 7.1%-12.1%; *P*<.001) reduction in step counts from baseline at the start of the diagnosis month, which further deepened to 12.4% (95% CI 9.5%-15.2%; *P*<.001) at 3 months and persisted at 7.1% (95% CI 4.2%-10.0%; *P*<.001) at 12 months, all relative to baseline. Conversely, in patients with diabetes, step counts remained relatively stable after diagnosis, with a slight upward trend, resulting in a change of +0.6% (95% CI –0.6% to 1.9%; *P*=.31) from baseline at 3 months after diagnosis. At 12 months after diagnosis, step counts remained decreased in the nonendoscopic subdiaphragmatic surgery group, with an 18.0% (95% CI 9.1%-26.2%; *P*<.001) reduction, whereas step counts returned to baseline in the laparoscopic surgery group (+0.3%, 95% CI –6.3% to 7.5%; *P*=.93).

Conclusions: The analysis of objective pre- and postdiagnostic step count data provided fundamental information crucial for understanding changes in PA among patients with cancer. While cancer diagnosis and treatment reduced PA, the decline may have already started before diagnosis. The study findings may help tailor exercise recommendations based on lifelog data for patients with cancer in the future.

(JMIR Cancer 2025;11:e58093) doi:10.2196/58093

KEYWORDS

cancer; lifelog data; physical activity; quality of life; step count; Japanese; database; smartphone; mobile app; exercise; mobile phone

Introduction

With the increasing number of cancer survivors, attention has been paid to not only cancer remission but also quality of life, which strongly correlates with physical activity (PA) in patients with cancer [1,2]. PA has been reported to be associated with a favorable prognosis in cancer survivors, with associations demonstrated in various cancer types, such as breast and colorectal cancers [1,3].

Although PA is a pivotal factor in patients with cancer, objective evidence on how it declines after cancer diagnosis and treatment is still lacking. Objective evaluation of prediagnostic PA is challenging. Previous studies that compared PA before and after cancer diagnosis relied on retrospective recall [4,5], which may introduce recall bias [6]. Therefore, research using objective precancer onset data is necessary to avoid this bias. However, this is challenging in studies that recruit participants after hospital visits.

Recently, many studies have leveraged real-world data collected during daily clinical practice and patients' daily lives [7]. With advancements in digital devices, access to patients' daily life data has improved [8]. Furthermore, the prevalent use of smartphone apps has facilitated the routine tracking of step counts, which strongly correlate with PA levels [9,10]. A daily step count of at least 10,000 steps is recommended for adults by the World Health Organization, while Japan's Ministry of Health, Labour and Welfare sets a more realistic target of 8000 steps per day for adults [11,12]. Integrating step count data obtained using smartphones with medical information allows for objective evaluation of changes in PA from the prediagnosis period. Thus, this study aimed to assess the effect of cancer diagnosis and treatment on PA levels by analyzing real-world step count data.

Methods

Data Source

This was a retrospective cohort study analyzing a database provided by DeSC Healthcare Inc, in which daily step count data are linked to a Japanese claims database [13,14]. Briefly, members of affiliated health insurance associations can access the Kencom smartphone app, developed by DeSC Healthcare Inc, free of charge. Daily step counts were measured using the Kencom app, synchronized with smartphone pedometers [13,14]. The Kencom database was integrated and anonymized with the Japanese claims database in affiliated health insurance associations under the opt-out agreement. Subsequently, DeSC Healthcare Inc merged data from various insurance associations, creating a large, longitudinal database for research use [13,14]. Diseases were classified according to *International Classification of Diseases, 10th edition* (*ICD-10*) codes. This database, which integrates step count information with medical information, mainly includes employees of large businesses and their dependents [13,14]. In this study, step count data were used to quantify PA levels.

Study Population

This study included patients newly diagnosed with malignant tumors (*ICD-10* code: C). In addition, patients newly diagnosed with diabetes mellitus (*ICD-10* code: E11-E14), a representative chronic disease in which the importance of PA is widely recognized [15], were included as a positive control, given the expectation that step counts would increase after diagnosis [16]. The inclusion of newly diagnosed patients also simplified the identification of index time [17]. Data between April 2014 and September 2021 were analyzed. The observation period spanned from 6 months before diagnosis to 12 months after diagnosis. The detailed eligibility criteria are described in Table S1 in Multimedia Appendix 1. Patients who did not receive treatment were excluded, and the Japanese Claims database has been validated as having over 98% specificity for identifying patients using this method [18].

Statistical Analysis

We applied a generalized additive mixed model with a cubic spline to describe changes in step counts. Individual-specific random effects, age at diagnosis, seasonal effects, day of the week (weekday, weekend, or holiday), and sex were included as covariates. Days with missing step count records were excluded from the analysis. A negative binomial distribution was assumed for step counts because of the considerable variability.

Subgroup analysis was performed on changes in step counts stratified by treatment method and cancer type. The definitions of treatment methods and cancer types are described in Tables S2 and S3 in Multimedia Appendix 1, respectively. Statistical analyses including data cleaning were conducted using R version 4.3.1 (R Foundation) with packages *mgcv* (version 1.8-42) and *ggplot2* (version 3.4.2).

Ethical Considerations

The Kyoto University Graduate School and Faculty of Medicine Ethics Committee approved this study (reference R3514). The need to obtain written informed consent from patients was waived because the data were anonymized.

Results

Study Population

A final cohort of 326 patients with malignant solid tumors and 1388 patients with diabetes met the eligibility criteria and were included in the analysis (Figure S1 in Multimedia Appendix 1). Table 1 shows patient characteristics and their respective cancer treatments. The median age of patients with cancer was 51 (IQR 45-56) years, and that of patients with diabetes was 50 (IQR

44-55) years. The median proportion of missing step count days during the observation period was 0.5% (IQR 0.0%-2.1%) among patients with cancer and 0.2% (IQR 0.0%-1.9%) in patients with diabetes. Further data on specific cancer diagnoses and the corresponding treatment for patients with cancer are described in Table S4 in Multimedia Appendix 1.

Table 1. Patient characteristics.

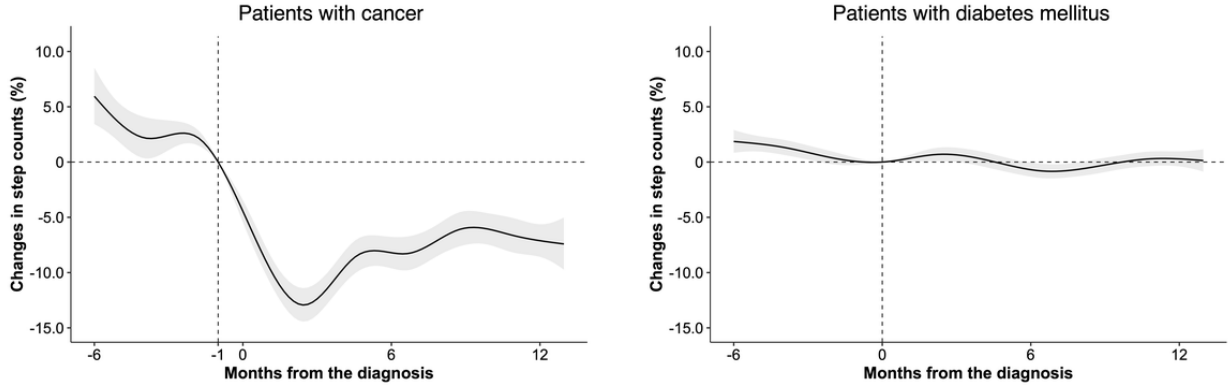
Characteristic	Patients with cancer (n=326)	Patients with diabetes mellitus (n=1388)	*P* values
Age (years), median (IQR)	51 (45-56)	50 (44-55)	.38
Male sex, n (%)	194 (59.5)	1144 (82.4)	<.001
Surgeries, n (%)	207 (63.5)	—[a]	—
Small interventions[b], n (%)	114 (35)	—	—
Radiation, n (%)	67 (20.6)	—	—
Conventional cytotoxic chemotherapies, n (%)	67 (20.6)	—	—
Molecular targeted therapy, n (%)	16 (4.9)	—	—
Immunotherapies, n (%)	3 (0.9)	—	—
Daily step counts for the reference month[c], median (IQR)	5754.9 (4219.2-7283.5)	6210.6 (4339.4-8015.8)	.01
Year of diagnosis, n (%)			
2015	6 (1.8)	49 (3.5)	—
2016	24 (7.4)	137 (9.9)	—
2017	52 (16)	204 (14.7)	.44
2018	81 (24.8)	324 (23.3)	—
2019	102 (31.3)	414 (29.8)	—
2020	61 (18.7)	260 (18.7)	—

[a]Not applicable

[b]The following procedures were classified as small interventions: cervical conization, diagnostic laparoscopy and thoracoscopy, endometrial curettage, excision of skin tumors, orchiectomy, transurethral resection of bladder tumors, and upper and lower gastrointestinal endoscopic surgeries (mucosal resection, polypectomy, and submucosal dissection).

[c]Reference month is month −1 for patients with cancer and month 0 for patients with diabetes mellitus (Figure 1).

Figure 1. Estimated daily step count changes before and after diagnosis. Daily step count changes in (left) patients diagnosed with cancer and (right) patients diagnosed with diabetes mellitus are described. Shading indicates the SE. Month 0 signifies the month of diagnosis. For patients with cancer, the baseline for step count change was defined as the first day of month −1, accounting for the time needed for pathology after biopsy or diagnostic surgery to confirm diagnosis.



Changes in Step Counts

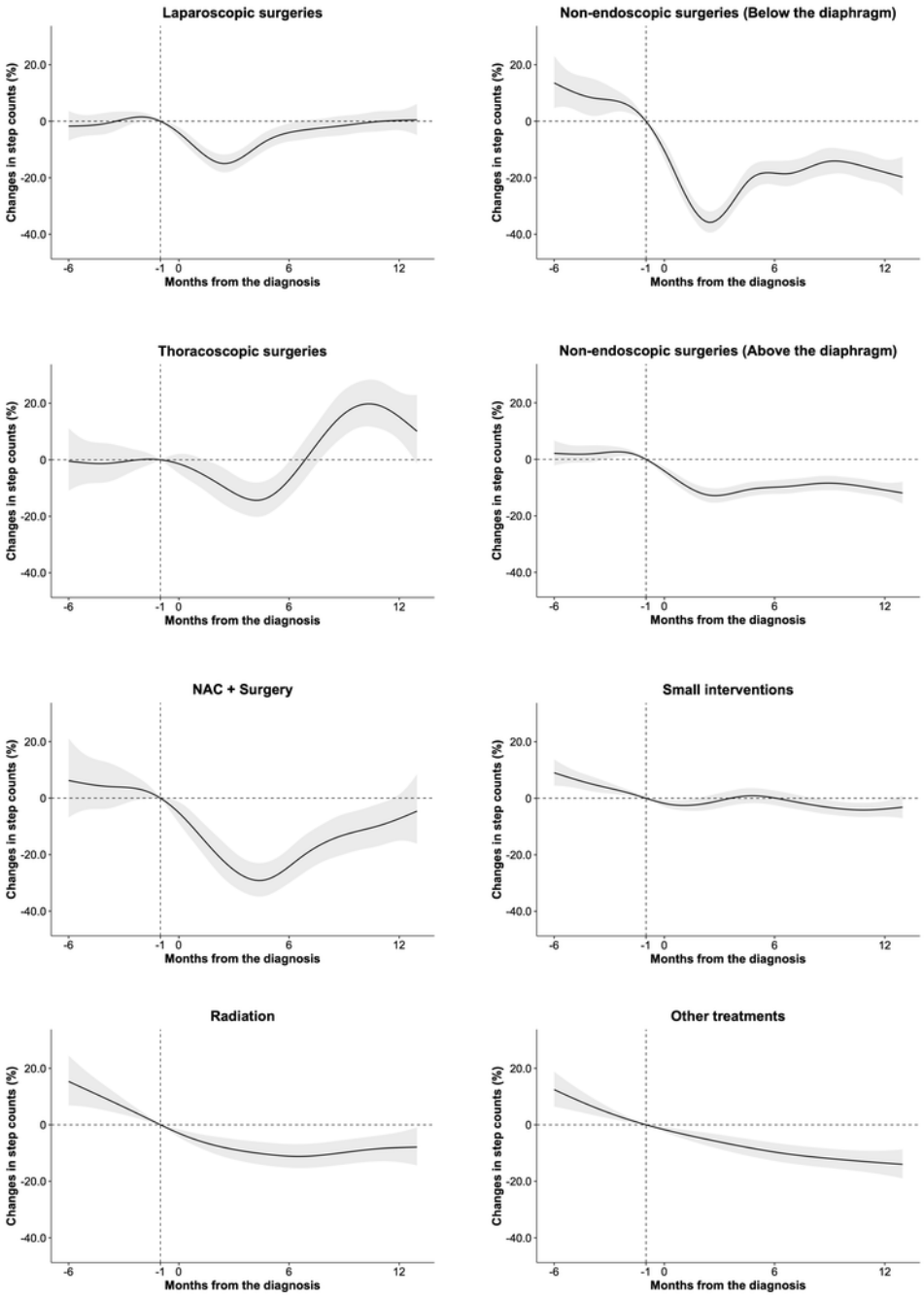
The results of the descriptive analysis of step count data are presented in Figure S2 in Multimedia Appendix 1. Step counts showed no significant difference between patients with cancer and those with diabetes 2 months before diagnosis (*P*=.17) but declined significantly in the former during the diagnosis month (*P*<.001). In subsequent analyses, the baseline for step counts in patients with cancer was therefore set at the beginning of the month before diagnosis. Figure 1 shows the percent change in step counts over time in relation to the time of diagnosis. Patients with cancer showed a 9.6% (95% CI 7.1%-12.1%; *P*<.001) reduction in step counts from baseline in the month of diagnosis, which deepened to 12.4% (95% CI 9.5%-15.2%;

P<.001) at 3 months and persisted at 7.1% (95% CI 4.2%-10.0%; *P*<.001) at 12 months, all relative to baseline. Conversely, in patients with diabetes, step counts remained relatively stable after diagnosis, with a slight upward trend, resulting in a change of +0.6% (95% CI –0.6% to 1.9%; *P*=.31) from baseline at 3 months after diagnosis. The change from baseline at 12 months after diagnosis was not significant (+0.3%, 95% CI –1% to 1.6%; *P*=.68).

Changes in Step Counts According to Treatment Modalities

Figure 2 shows the results of the subgroup analysis stratified by treatment modalities. Patients who received systemic chemotherapy without surgery were classified as the other treatment group because of the limited number of patients. There were no significant changes in the distribution of patient numbers by treatment group across the years (*P*=.69; Table S5 in Multimedia Appendix 1). At 3 months after diagnosis, the reduction in step counts from baseline was 14.3% (95% CI 8.0%-20.1%; *P*<.001) in the laparoscopic surgery group and 34.2% (95% CI 26.6%-41.1%; *P*<.001) in the nonendoscopic subdiaphragmatic surgery group. At 12 months after diagnosis, step counts remained decreased in the nonendoscopic subdiaphragmatic surgery group, with an 18.0% (95% CI 9.1%-26.2%; *P*<.001) reduction, whereas step counts returned to baseline in the laparoscopic surgery group (+0.3%, 95% CI –6.3% to 7.5%; *P*=.93). At 12 months after diagnosis, the nonendoscopic supradiaphragmatic surgery group showed a 10.9% (95% CI 5.7%-15.7%; *P*<.001) step count reduction, whereas this reduction was not seen in the thoracoscopic surgery group, and the step count change was not significant (+15.2%, 95% CI –0.4% to 33.1%; *P*=.06).

Figure 2. Estimated daily step count changes before and after diagnosis by treatment methods: Other treatments include conventional cytotoxic chemotherapy, immunotherapy, molecular targeted therapy, small interventions with chemotherapy or immunotherapy, and chemoradiotherapy. Shading indicates the SE. NAC: neoadjuvant chemotherapy.



Changes in Step Counts According to Cancer Type

Figure S3 and Table S6 in Multimedia Appendix 1 show the results of the subgroup analysis stratified by cancer type. A significant postoperative reduction in step counts was observed in patients with gynecologic cancers, and the majority of them underwent nonendoscopic surgeries (endometrial cancer: 4/5, 80%; cervical cancer: 3/6, 50%; and ovarian cancer: 3/4, 75%). Conversely, in gastrointestinal and urologic cancers, where abdominal surgeries were also performed, the proportion of patients who underwent small interventions or laparoscopic surgeries was high, and the postoperative reduction in step counts was not as pronounced.

Discussion

Principal Findings

The analysis of objective pre- and postdiagnostic lifelog data showed that step counts in patients with cancer decreased with cancer diagnosis and treatment, with step counts already below baseline at the beginning of the diagnosis month and failing to return to baseline 12 months later. Conversely, the change from baseline in those newly diagnosed with diabetes was not significant. The decline in step count varies with treatment and cancer type.

Comparison With Previous Work

To the best of our knowledge, this is the first study to objectively describe the decline in PA levels from precancer diagnosis using lifelog data on step counts. Previous studies that relied on retrospective recall have reported a decline in PA by 59% after treatment of breast cancer [4]. In colorectal cancer, the reduced PA during treatment improved after treatment completion [5]. Our analysis using objective real-world data provides a closer representation of real-life events. In patients with diabetes, while the increase in step counts was not significant, an upward trend was observed after diagnosis, consistent with previous reports [16].

This study showed possible variations in PA measured as step count based on the specific treatment modality used. Treatments for specific cancer types are typically standardized. However, by analyzing patients with varying cancer types, this study explored the impacts of different treatments on step counts. Interestingly, a prominent decrease in step counts was observed in patients who underwent nonendoscopic surgery, suggesting that opting for less invasive surgical approaches may help mitigate the decrease in step counts. However, this result should be carefully interpreted because patients requiring open surgery may have had more advanced stages with worse prognoses or require more intensive treatment. Further examination is needed to assess the long-term impact of surgical approaches on PA.

Measuring PA has become increasingly accessible through the use of smartphones and other wearable devices [8]. Previous studies have been conducted to categorize PA in patients with cancer using these devices. However, no established threshold has been identified [19]. This study adds new insights by focusing on changes in PA, which are recorded as step counts. Step counts serve as a valid and useful surrogate marker of PA, and a decrease in step counts is strongly associated with an increase in all-cause mortality [20,21]. Therefore, conducting close follow-ups on posttreatment PA levels using step counts has the potential to contribute to patient-centered care and personalized recommendations for PA, which remain insufficient [2]. Further studies are needed to address tailoring recommendations for PA during cancer treatment according to cancer type, treatment approach, and comorbidities.

Limitations

This study has some limitations. First, influences beyond cancer diagnosis and treatment, such as aging, societal changes, and changes in smartphone pedometer specifications due to device upgrades or replacements by users, may be involved. However, minimal changes in step counts in the diabetes group indicated a limited impact of these factors. Second, the proportion of patients managed with small interventions was high, indicating that a considerable number of patients were detected at an early stage. This might be because this study included relatively young patients with a median age of 51 (IQR 45-56) years, primarily employees of large businesses and their dependents, who were exclusively Kencom app users, indicative of a potential high health awareness demographic. Therefore, this population may not represent the general Japanese population. However, the decline in step count levels would be more pronounced in patients with a higher prevalence of advanced-stage cancer, who are likely to undergo more invasive treatments [22]. Third, the cancer stage was not determined in this study. However, because treatments are generally determined by cancer stage, results stratified by treatment may have higher generalizability to populations with different cancer stages. Fourth, the follow-up period in this study was 1 year, which did not allow for the analysis of step count data in relation to patient prognosis. PA may improve prognosis in cancer survivors [1,3]. Therefore, further studies are needed to assess its effect on patient prognosis. Finally, step count data may not fully reflect actual PA levels, as its accuracy can be affected by factors, including the possibility that individuals are not always carrying their smartphones.

Conclusions

This study used objective pre- and postdiagnostic step count data and showed that while PA decreased with cancer diagnosis and treatment, the decline may have already started before diagnosis. The decline varies with treatment and cancer type, with a notable decrease in patients who underwent nonlaparoscopic abdominal surgeries. Assessing PA using step count data has the potential to contribute to patient-centered care. This finding may help tailor exercise recommendations based on lifelog data for patients with cancer.

Acknowledgments

This study was conducted as part of a research project selected through an open call by DeSC Healthcare Inc. We would like to thank Enago for the English language review. This research did not receive any funding. The first draft was prepared in Japanese without the use of generative artificial intelligence and was translated into English with the assistance of ChatGPT-4o by OpenAI [23]. In addition, our paper underwent professional scientific editing (Enago) for language refinement.

Data Availability

The datasets analyzed during this study are not publicly available due to the contract with the data provider.

Authors' Contributions

YI, K Yamaguchi, and JH developed the study concept. YI, K Yamaguchi, KM, MM, and KK developed the study design. YI conducted the data extraction and curation. YI, KM, and ST-M conducted the analysis. YI, K Yamaguchi, AK, NH, MT, K Yamanoi, RM, and JH interpreted the study results. MM and KK supervised the study. YI, K Yamaguchi, KM, and ST-M wrote

the draft of the manuscript. AK, NH, MT, K Yamanoi, RM, JH, SY, MM, and KK contributed to the critical revision of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

K Yamaguchi and MM have received a donation from DUMSCO Inc. KK received research funds from AstraZeneca KK, Eisai Co, Ltd, Kyowa Kirin Co, Ltd, OMRON Corporation, and Toppan Inc; consulting fees from Advanced Medical Care Inc, JMDC Inc, Santen Pharmaceutical Co, Ltd, Shin Nippon Biomedical Laboratories Ltd, and Ubicom Holdings Inc; executive compensation from Cancer Intelligence Care Systems, Inc; and honoraria from Chugai Pharmaceutical Co, Ltd, Taisho Pharmaceutical Co, Ltd, and Pharma Business Academy.

Multimedia Appendix 1
Eligibility criteria, definition of subgroups by treatment method, definition of subgroups by cancer type, patient characteristics by cancer type, the annual number of patients in each treatment group, summary of step count changes by cancer type, selection of study samples, distribution of mean daily step counts in each month, and estimated daily step count changes before and after diagnosis by cancer type.
[PDF File (Adobe PDF File), 857 KB - cancer_v11i1e58093_app1.pdf]

References

1. Rock CL, Thomson CA, Sullivan KR, Howe CL, Kushi LH, Caan BJ, et al. American Cancer Society nutrition and physical activity guideline for cancer survivors. CA Cancer J Clin 2022;72(3):230-262 [FREE Full text] [doi: 10.3322/caac.21719] [Medline: 35294043]
2. Campbell KL, Winters-Stone KM, Wiskemann J, May AM, Schwartz AL, Courneya KS, et al. Exercise guidelines for cancer survivors: consensus statement from international multidisciplinary roundtable. Med Sci Sports Exerc 2019;51(11):2375-2390 [FREE Full text] [doi: 10.1249/MSS.0000000000002116] [Medline: 31626055]
3. Patel AV, Friedenreich CM, Moore SC, Hayes SC, Silver JK, Campbell KL, et al. American College of Sports Medicine roundtable report on physical activity, sedentary behavior, and cancer prevention and control. Med Sci Sports Exerc 2019;51(11):2391-2402 [FREE Full text] [doi: 10.1249/MSS.0000000000002117] [Medline: 31626056]
4. Hair BY, Hayes S, Tse CK, Bell MB, Olshan AF. Racial differences in physical activity among breast cancer survivors: implications for breast cancer care. Cancer 2014;120(14):2174-2182 [FREE Full text] [doi: 10.1002/cncr.28630] [Medline: 24911404]
5. Chung JY, Lee DH, Park J, Lee MK, Kang D, Min J, et al. Patterns of physical activity participation across the cancer trajectory in colorectal cancer survivors. Support Care Cancer 2013;21(6):1605-1612. [doi: 10.1007/s00520-012-1703-5] [Medline: 23292698]
6. Blome C, Augustin M. Measuring change in quality of life: bias in prospective and retrospective evaluation. Value Health 2015;18(1):110-115 [FREE Full text] [doi: 10.1016/j.jval.2014.10.007] [Medline: 25595241]
7. Baumfeld Andre E, Reynolds R, Caubel P, Azoulay L, Dreyer NA. Trial designs using real-world data: the changing landscape of the regulatory approval process. Pharmacoepidemiol Drug Saf 2020;29(10):1201-1212 [FREE Full text] [doi: 10.1002/pds.4932] [Medline: 31823482]
8. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
9. Amagasa S, Kamada M, Sasai H, Fukushima N, Kikuchi H, Lee IM, et al. How well iPhones measure steps in free-living conditions: cross-sectional validation study. JMIR Mhealth Uhealth 2019;7(1):e10418 [FREE Full text] [doi: 10.2196/10418] [Medline: 30626569]
10. Höchsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksäss A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. Scand J Med Sci Sports 2018;28(7):1818-1827. [doi: 10.1111/sms.13074] [Medline: 29460319]
11. WHO. Pacific Physical Activity Guidelines for Adults : Framework for Accelerating the Communication of Physical Activity Guidelines. Manila, the Philippines: WHO Regional Office for the Western Pacific; 2008.
12. Explanatory materials for promoting health Japan 21. Ministry of Health Labour and Welfare. URL: https://www.mhlw.go.jp/www1/english/wp_5/vol1/p2c6s4.html [accessed 2024-12-12]
13. Hamaya R, Mori M, Miyake K, Lee IM. Association of smartphone-recorded steps over years and change in cardiovascular risk factors among working-age adults. J Am Heart Assoc 2022;11(14):e025689 [FREE Full text] [doi: 10.1161/JAHA.121.025689] [Medline: 35861838]
14. Hamaya R, Fukuda H, Takebayashi M, Mori M, Matsushima R, Nakano K, et al. Effects of an mHealth app (Kencom) with integrated functions for healthy lifestyles on physical activity levels and cardiovascular risk biomarkers: observational study of 12,602 users. J Med Internet Res 2021;23(4):e21622 [FREE Full text] [doi: 10.2196/21622] [Medline: 33900203]

15. Kanaley JA, Colberg SR, Corcoran MH, Malin SK, Rodriguez NR, Crespo CJ, et al. Exercise/physical activity in individuals with type 2 diabetes: a consensus statement from the American College of Sports Medicine. Med Sci Sports Exerc 2022;54(2):353-368 [FREE Full text] [doi: 10.1249/MSS.0000000000002800] [Medline: 35029593]
16. Schneider KL, Andrews C, Hovey KM, Seguin RA, Manini T, Lamonte MJ, et al. Change in physical activity after a diabetes diagnosis: opportunity for intervention. Med Sci Sports Exerc 2014;46(1):84-91 [FREE Full text] [doi: 10.1249/MSS.0b013e3182a33010] [Medline: 23860414]
17. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol 2003;158(9):915-920. [doi: 10.1093/aje/kwg231] [Medline: 14585769]
18. de Luise C, Sugiyama N, Morishima T, Higuchi T, Katayama K, Nakamura S, et al. Validity of claims-based algorithms for selected cancers in Japan: results from the VALIDATE-J study. Pharmacoepidemiol Drug Saf 2021;30(9):1153-1161 [FREE Full text] [doi: 10.1002/pds.5263] [Medline: 33960542]
19. Aix SP, Núñez-Benjumea FJ, Cervera-Torres S, Flores A, Arnáiz, Fernández-Luque L. Data-driven personalized care in lung cancer: scoping review and clinical recommendations on performance status and activity level of patients with lung cancer using wearable devices. JCO Clin Cancer Inform 2023;7:e2300016 [FREE Full text] [doi: 10.1200/CCI.23.00016] [Medline: 37922433]
20. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017;47(7):1303-1315 [FREE Full text] [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
21. Jayedi A, Gohari A, Shab-Bidar S. Daily step count and all-cause mortality: a dose-response meta-analysis of prospective cohort studies. Sports Med 2022;52(1):89-99. [doi: 10.1007/s40279-021-01536-4] [Medline: 34417979]
22. Kwan ML, Sternfeld B, Ergas IJ, Timperi AW, Roh JM, Hong CC, et al. Change in physical activity during active treatment in a prospective study of breast cancer survivors. Breast Cancer Res Treat 2012;131(2):679-690 [FREE Full text] [doi: 10.1007/s10549-011-1788-4] [Medline: 21953007]
23. OpenAI. ChatGPT. URL: https://chatgpt.com/ [accessed 2024-11-18]

Abbreviations

ICD-10: International Classification of Diseases, 10th edition
PA: physical activity

Edited by N Cahill; submitted 05.03.24; peer-reviewed by Y Chu, L Horrell; comments to author 04.10.24; revised version received 24.10.24; accepted 16.12.24; published 20.01.25.

Please cite as:
Inayama Y, Yamaguchi K, Mizuno K, Tanaka-Mizuno S, Koike A, Higashiyama N, Taki M, Yamanoi K, Murakami R, Hamanishi J, Yoshida S, Mandai M, Kawakami K
Changes in Physical Activity Across Cancer Diagnosis and Treatment Based on Smartphone Step Count Data Linked to a Japanese Claims Database: Retrospective Cohort Study
JMIR Cancer 2025;11:e58093
URL: https://cancer.jmir.org/2025/1/e58093
doi:10.2196/58093
PMID:39726139

Original Paper

An App-Based Intervention With Behavioral Support to Promote Brisk Walking in People Diagnosed With Breast, Prostate, or Colorectal Cancer (APPROACH): Process Evaluation Study

Fiona Kennedy[1*], BSc, MSc, PhD; Susan Smith[2*], BA, MSc; Rebecca J Beeken[1], BA, MSc, PhD; Caroline Buck[2], BA, MSc; Sarah Williams[2], BSc, MSc; Charlene Martin[3], BA, MA, PhD; Phillippa Lally[4], BSc, MSc, PhD; Abi Fisher[2], BSc, PhD

[1]Leeds Institute of Health Sciences, University of Leeds, Leeds, United Kingdom

[2]Department of Behavioural Science and Health, University College London, London, United Kingdom

[3]Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, Sheffield, United Kingdom

[4]School of Psychology, University of Surrey, Guildford, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Abi Fisher, BSc, PhD
Department of Behavioural Science and Health
University College London
1-19 Torrington Place
London, WC1E7HB
United Kingdom
Phone: 44 2076791722
Email: abigail.fisher@ucl.ac.uk

Abstract

Background: The APPROACH pilot study explored the feasibility and acceptability of an app (NHS Active 10) with brief, habit-based, behavioral support calls and print materials intended to increase brisk walking in people diagnosed with cancer.

Objective: Following UK Medical Research Council guidelines, this study assessed the implementation of the intervention, examined the mechanisms of impact, and identified contextual factors influencing engagement.

Methods: Adults (aged ≥18 y) with breast, prostate, or colorectal cancer who reported not meeting the UK guidelines for moderate-to-vigorous physical activity (≥150 min/wk) were recruited from a single hospital site in Yorkshire, United Kingdom. They were randomly assigned to the intervention or control (usual care) arm and assessed via quantitative surveys at baseline (time point 0 [T0]) and 3-month follow-up (time point 1 [T1]) and qualitative exit interviews (36/44, 82%) at T1. The process evaluation included intervention participants only (n=44). Implementation was assessed using data from the T1 questionnaire exploring the use of the intervention components. The perceived usefulness of the app, leaflet, and behavioral support call was rated from 0 to 5. Behavioral support calls were recorded, and the fidelity of delivery of 25 planned behavior change techniques was rated from 0 to 5 using an adapted Dreyfus scale. Mechanisms of impact were identified by examining T0 and T1 scores on the Self-Reported Behavioural Automaticity Index and feedback on the leaflet, app, call, and planner in the T1 questionnaire and qualitative interviews. Contextual factors influencing engagement were identified through qualitative interviews.

Results: The implementation of the intervention was successful: 98% (43/44) of the participants received a behavioral support call, 78% (32/41) reported reading the leaflet, 95% (39/41) reported downloading the app, and 83% (34/41) reported using the planners. The mean perceived usefulness of the app was 4.3 (SD 0.8) in participants still using the app at T1 (n=33). Participants rated the leaflet (mean 3.9, SD 0.6) and the behavioral support call (mean 4.1, SD 1) as useful. The intended behavior change techniques in the behavioral support calls were proficiently delivered (overall mean 4.2, SD 1.2). Mechanisms of impact included habit formation, behavioral monitoring, and support and reassurance from the intervention facilitator. Contextual factors impacting engagement included barriers, such as the impact of cancer and its treatment, and facilitators, such as social support.

Conclusions: The APPROACH intervention was successfully implemented and shows promise for increasing brisk walking, potentially through promoting habit formation and enabling self-monitoring. Contextual factors will be important to consider when interpreting outcomes in the larger APPROACH randomized controlled trial.

International Registered Report Identifier (IRRID): RR2-10.1186/s40814-022-01028-w

(JMIR Cancer 2025;11:e64747) doi:10.2196/64747

KEYWORDS

cancer; physical activity; process evaluation; randomized controlled trial; intervention; app; habit

Introduction

Background

The number of people being diagnosed with cancer is continuing to increase in the United Kingdom, with an estimated 4 million adults living with and beyond cancer (LWBC) by 2030 [1]. Many people LWBC experience challenges related to cancer and its treatment, including increased fatigue, pain, psychological distress, and reduced physical capacity [2-5]. These challenges can significantly impact their quality of life and well-being [6]. Therefore, the importance of improving outcomes for those LWBC is vital [7]. A large body of trial data demonstrates that physical activity (PA) improves many outcomes after a cancer diagnosis, including reduced fatigue, pain, anxiety, depression, and sleep problems and an overall improvement in quality of life [8-11]. Observational data suggest that PA is also associated with improvements in survival [12-14]. In light of this ever-growing evidence base, the World Cancer Research Fund recommends that people LWBC follow the guidelines for healthy populations in achieving at least 150 minutes of moderate-to-vigorous PA (MVPA) per week and recommend limiting the amount of time spent sedentary [15]. Despite this, many people LWBC are physically inactive, with Macmillan Cancer Support (United Kingdom) estimating this to be as high as 80% of those LWBC not meeting recommended PA levels [16]. This is supported by the systematic review of 41 studies conducted by Wong et al [17] that indicated that only around a third of people LWBC were meeting PA guidelines, although this ranged from 16% to 88% across studies. Brisk walking is a form of MVPA that may be more appropriate for people LWBC due to its accessibility and achievability. In a systematic scoping review of 98 studies in people LWBC, walking was cited as the preferred type of PA across all cancer types and treatment stages [18].

The APPROACH Intervention

APPROACH is an app-based, multicomponent intervention informed by extensive development work with individuals with breast, prostate, or colorectal cancer and cancer nurse specialists [19,20]. It focuses on promoting and monitoring brisk walking using a publicly available mobile phone app, alongside brief behavioral support in the form of a specially designed leaflet, walking planner cards, and 2 phone or video calls with a trained researcher (CB) [21]. In line with the Medical Research Council framework for intervention development and evaluation, the pilot trial explored the feasibility and acceptability of conducting a complex PA intervention trial with people LWBC. A total of 90 people diagnosed with breast, prostate, or colorectal cancer were recruited for the pilot randomized controlled trial (RCT), with 49% (n=44) randomly assigned to the intervention group and 51% (n=46) to the control group. APPROACH pilot results demonstrated a high retention rate (97%) and high assessment completion rates (>86%), indicating that the trial procedures were feasible and acceptable to be carried out as intended in a confirmatory, phase-3, larger trial [22]. In addition, results showed that the intervention was delivered successfully with 98% receiving at least 1 behavioral support call and 95% of participants downloading the app [22].

Conducting a Process Evaluation

The importance of conducting process evaluations within RCTs has been emphasized to explore the way in which any complex intervention is implemented [23,24]. This can help uncover why interventions are successful or unsuccessful, determine why they may have unexpected consequences, as well as explore how an intervention that is effective can be optimized [23]. Moore et al [25] provided specific guidance for carrying out process evaluations, which highlighted the importance of exploring the implementation (eg, the fidelity of intervention calls), the mechanisms of impact (eg, views on the different components of the intervention), and the contextual factors influencing use and outcomes (eg, barriers or facilitators to engagement) [25].

Therefore, this paper extends the published APPROACH feasibility results [22] with the following aims: (1) to evaluate the implementation of the APPROACH intervention and the fidelity of the delivery of intended behavior change techniques (BCTs), (2) to identify the potential mechanisms of impact that underlie behavior changes attributed to the APPROACH intervention, and (3) to better understand how contextual factors influence engagement with the APPROACH intervention.

Methods

Design

This was an embedded design, mixed methods study where qualitative and quantitative data were collected simultaneously with equal priority [26]. Data were collected as part of the APPROACH pilot RCT [21,22]. The pilot RCT compared an app-based, brisk walking intervention delivered alongside usual care with a control arm (usual care alone) in people diagnosed with breast, prostate, or colorectal cancer at a single hospital site. The trial was registered on the International Standard Randomized Controlled Trial Number registry on April 16, 2021 (ISRCTN 18063498). The primary outcome for the pilot and future RCT is weekly minutes spent brisk walking (a cadence of 100 steps per min or more [27) measured by an activPAL accelerometer (PAL Technologies Ltd). Following baseline assessments, participants were individually randomly assigned (1:1 allocation) using minimization to either the control or intervention arm, stratified by cancer type (breast, prostate, or colorectal) and disease status (metastatic vs not).

Participants

The pilot RCT included 90 participants: 49% (n=44) in the intervention arm and 51% (n=46) in the control arm. All participants had a confirmed diagnosis of breast, prostate, or colorectal cancer (localized or metastatic). At the point of screening, localized participants had to be within 6 months of completion of radical treatment. This criterion was not applied to participants with metastatic disease. All participants required a clinician's sign-off that their life expectancy was >6 months. All participants self-reported achieving <150 minutes of MVPA weekly. Full inclusion and exclusion criteria have been published previously [21].

Procedure and Description of Intervention

The recruitment procedure and trial data collection have previously been described [21,22]. Participants completed assessments at baseline (time point 0 [T0]) and 3 months (time point 1 [T1]; operationalized as 12-16 wk after randomization). The intervention included an endorsement letter from a member of the participant's clinical team, alongside the provision of a leaflet with information about the importance of PA after cancer and a recommendation and instructions on how to download the freely available NHS Active 10 app. The NHS Active 10 app promotes brisk walking in bouts of 10 minutes, called "Active 10s." This was augmented with 2 behavioral support phone or video calls with the intervention facilitator (CB). The app was chosen after previous qualitative work with people LWBC, and clinicians identified key features that were important to be offered within the app and highlighted the importance of the app being supported by a professional organization, such as the NHS [19,20]. These behavioral support calls were underpinned by habit theory [28] and BCTs shown to be effective in promoting PA [29-31] and involved supporting participants in downloading and using the app and discussions around setting PA goals. The intervention facilitator was trained in the principles of behavior change theories and in the application of BCTs [32], with a thorough understanding of habit theory [33]. This training, alongside previous experience in delivering health behavior change interventions, allowed them to conduct conversations with patients closely mirroring those a health care professional might have with a patient in a routine care situation. The intervention calls took place via Zoom (Zoom Video Communications) [34] or telephone and were recorded by the intervention facilitator with the participants' permission. The first call took place approximately 1 week after randomization, and the second call took place approximately 4 weeks after the first call to check in with the participant about their goals and recap any information required. Participants were asked during the first intervention call if they had downloaded the NHS Active 10 app, and the intervention facilitator noted this in their records. Participants were also given 12 copies of a walking planner card that was designed to enable them to plan how many "Active 10s" they were aiming for and how they were going to achieve these, including where and when they would complete them.

Implementation of the APPROACH Intervention

Delivery of the Intervention

The implementation of the intervention was explored by looking at whether each component was delivered as intended and the participants' use of each intervention component. Between 12 and 16 weeks after randomization participants completed the T1 questionnaire. Participants were asked about the intervention, including the following: whether they downloaded the app (yes or no), their self-reported app use if still using the app (less than monthly, monthly, fortnightly, weekly, 3-4 times a week, almost every day, or every day), how long they used the app for if they had stopped using it (never, once, less than monthly, fortnightly, weekly, 3-4 times per week, almost every day, or every day), perceived accuracy of the app in recording their time spent walking (5-point scale from not accurate to very accurate), whether they read the leaflet (all, some, or did not read), used the walking planner cards (yes or no), and received either behavioral support call (yes or no).

The Usefulness of Intervention Components

Participants rated the usefulness of the call for going through the leaflet information, downloading the app, and thinking about ways to use the app to increase their brisk walking (5-point scale from not at all useful to extremely useful). Using the same scale, they rated the usefulness of the app and sections of the leaflet for supporting their walking.

Delivery Fidelity of Behavioral Support Calls

The recorded intervention calls were coded by 1 researcher (SW) to assess delivery fidelity. All calls were listened to and coded according to a 25-item checklist of BCTs [31] as presented in the study's protocol paper [21]. Each item represented a BCT paired with the intended delivery technique (Multimedia Appendix 1). If a participant received 2 calls, these were combined when coding delivery of the BCT. A 5-point rating scale was applied to the fidelity checklist using an adaptation of the Dreyfus scale [35,36] ranging from low fidelity (0), indicating that the facilitator did not mention the intended BCT at all, to expert (5), indicating that the facilitator delivered the BCT to an exceptional standard (Multimedia Appendix 2). A value of ≥3 represented competent delivery of an individual BCT, thus presenting successful delivery. A second researcher (SS) coded a subset of interviews (n=5). It was agreed that if there was a discrepancy of over 20% in the coding, then the transcript would be discussed among the researchers. This occurred for 1 transcript that was double coded. This iterative process enabled SW to incorporate any learnings from the discussion into the coding of all transcripts and allowed a more consistent coding of the data.

Mechanisms of Impact and Contextual Factors Influencing Engagement

T0 and T1 Questionnaires

Habit strength for walking ("going for a walk" and "walking briskly") was assessed using the Self-Report Behavioural Automaticity Index (SRBAI) [37] in the T0 and T1 questionnaires. Participants responded on a 7-point scale ranging from disagree to agree for 4 statements on their perceived

automaticity of performing the behavior. An average score across items was calculated, representing the level of automaticity for the behavior being measured. Higher average scores indicate stronger habit or greater automaticity [37]. The SRBAI is presented in Multimedia Appendix 3. Mechanisms of impact and contextual factors impacting engagement were also identified by examining responses to the T1 questionnaire about the delivery of intervention components and their perceived usefulness.

Qualitative Interviews

Participants were asked in the initial study consent form if they agreed to be invited to participate in a semistructured interview at the end of the study. After the completion of all other data collection at T1, all participants who agreed were invited to be interviewed. Two members of the research team (SS and FK) carried out the interviews. SS and FK were involved in organizing assessments with participants throughout the pilot RCT. The interviews followed a topic guide exploring trial procedures, and participants were asked to give feedback on the intervention components (Multimedia Appendix 4). Interviews took place over the phone and, with the participants' permission, were audio recorded and transcribed verbatim. Contextual factors were explored in the interviews and were described in terms of barriers and facilitators of engaging with the intervention.

Data Analysis

Implementation of the APPROACH Intervention

The T1 questionnaire responses on intervention components were explored descriptively by calculating percentage frequencies and, where relevant, measures of central tendency. Mean scores were calculated for the delivery of each BCT in the intervention calls as well as an overall mean fidelity score for each call.

Mechanisms of Impact and Contextual Factors Influencing Engagement

The data from the questionnaires and interviews were pooled during interpretation to investigate the mechanisms of impact and contextual factors of intervention engagement. Qualitative and quantitative data were analyzed separately. The SRBAI results from the T0 and T1 questionnaires were explored descriptively using medians and IQRs due to the skewness of the data. The T1 questionnaire responses on intervention components were also used to identify mechanisms of impact.

Three authors (SW, FK, and SS) analyzed the data from the qualitative interviews using reflexive thematic analysis [38,39]. Inductive coding was undertaken, with these codes then used to develop themes, and early and final themes were discussed throughout the coding process among multiple authors (FK, SS, and PL). While initial coding was inductive and focused on identifying commonalities across the transcripts, final theme development was also organized by focusing on the outlined process evaluation aims on exploring the delivery of the intervention, the mediating processes (mechanisms of impact), and the barriers and facilitators to engagement (contextual factors). All interview transcripts were managed in NVivo (version 12; Lumivero) to facilitate analysis and data management.

Data were integrated using a complementarity approach where the interpretation of quantitative and qualitative results together allowed a more holistic interpretation of the findings [40].

Ethical Considerations

This pilot study was approved by the Yorkshire & The Humber-South Yorkshire Research Ethics Committee (21/YH/0029, Health Research Authority and the local hospital. All participants gave informed consent and the data reported are anonymized.

Results

Overview

Table 1 presents the sample characteristics of the 44 participants in the intervention arm. Most (42/44, 95%) of the participants were of White ethnicity, comprising an equal number of male participants (22/44, 50%) and female participants (22/44, 50%), with a mean age of 63 (SD 11; range 40-85) years. Participants had received a diagnosis of breast cancer (18/44, 41%), prostate cancer (18/44, 41%), or colorectal (8/44, 18%) cancer. The CONSORT (Consolidated Standards of Reporting Trials) diagram and the flow of participants through the study have previously been reported [22]. After eligibility screening and assessment of interest in taking part, the study information sheet was sent to 148 patients, and 63% (n=93) consented to participate, with 61% (n=90) being randomized.

Table 1. Characteristics of the APPROACH intervention group (N=44).

Characteristics	Participants, n (%)
Sex	
Male	22 (50)
Female	22 (50)
Age range (y)	
40-50	7 (16)
51-60	10 (23)
61-70	14 (32)
71-80	12 (27)
>81	1 (2)
Ethnic group	
Asian	1 (2)
Black	0 (0)
Mixed	0 (0)
Other	1 (2)
White	42 (95)
Cancer type	
Breast	18 (41)
Prostate	18 (41)
Colorectal	8 (18)
Localized or metastatic	
Localized	41 (93)
Metastatic	3 (7)
Relationship status	
Married or in a relationship	37 (84)
Single, divorced or separated	3 (7)
Widowed	4 (9)
Employment	
Full time	8 (18)
Part time	9 (20)
Unemployed	2 (5)
Retired	22 (50)
Unable or too ill to work	3 (7)
Living arrangements	
Alone	5 (11)
With partner	25 (57)
With family	14 (32)
Index of Multiple Deprivation quintile	
1 (most deprived)	8 (18)
2	6 (14)
3	9 (20)
4	16 (36)
5 (least deprived)	5 (11)

Implementation of the APPROACH Intervention

Overview

In total, 2 (5%) of the 44 participants withdrew from the intervention group for reasons unrelated to the intervention (frustration with the accelerometer and increased caring responsibilities). Most (41/42, 98%) participants who remained in the study answered the section of the T1 questionnaire on intervention feedback. Moreover, 1 (2%) of the 42 participants did not complete this section on intervention feedback.

Delivery of the Intervention

Leaflet

In the T1 questionnaire, 78% (32/41) of the intervention participants reported reading the entire intervention leaflet, while 10% (4/41) reported reading some of it, and 12% (5/41) reported not reading it at all. Of those who did not read it at all, 80% (4/5) stated that they did not remember receiving the leaflet, and 20% (1/5) stated that it was not relevant to them.

NHS Active 10 App

At the time of the first behavioral support phone call, the intervention facilitator recorded that 95% (42/44) participants had downloaded the NHS Active 10 app, with 93% (39/42) independently downloading it before the first intervention call and 7% (3/42) downloading it during the call. However, 2% (1/43) of the participants left call 1 not having downloaded it. In the T1 questionnaire, 95% (39/41) of the intervention participants self-reported successfully downloading the app. In total, 5% (2/41) of the participants were not asked about their use of the app as they reported not downloading the app earlier in the questionnaire. Most (33/39, 85%) participants reported still using the app. Of these, 82% (27/33) reported using it "almost every day or every day," and 18% (6/33) reported that they used it "3-4 times per week." A few (5/41, 12%) participants reported using the app during the study but were no longer using it. When asked how long they had used the app, the participants reported using it for "1 week," "2 weeks," "1 month," "2 months," and "3 months." In addition, 2% (1/41) of the participants reported not using the app at all despite downloading it.

Planner Cards

In the T1 questionnaire, 83% (34/41) of the participants reported using the walking planner cards, whereas 17% (7/41) did not, including 1 participant who said they did not receive any cards. Other nonuse was mainly explained in terms of not finding it helpful/not needing to plan (5/41, 12%) or having a more physical job (1/41, 2%). Of those who used the planners, 65% (22/34) reported using the planners for the full 3 months, whereas others reported using them for 2 weeks (4/34, 12%), 1 month (4/34, 12%), or 2 months (4/34, 12%).

The Usefulness of Intervention Components

Overview

The perceived usefulness of the intervention components is presented in Table 2.

Table 2. Perceived usefulness of the APPROACH intervention components (n=41).

Intervention components[a]	Values, mean (SD)	Respondents, n (%)
Behavioral support call	4.1 (1.0)[b]	40[c] (98)
Leaflet sections		
Physical activity and cancer	3.8 (0.9)[b]	36[d] (88)
Walking	4.0 (0.8)[b]	36[d] (88)
Information about Active 10	3.9 (0.8)[b]	36[d] (88)
Instructions on how to download Active 10	4.0 (0.8)[b]	36[d] (88)
Walking habits	4.1 (0.7)[b]	36[d] (88)
Walking websites	3.8 (1.5)[b]	36[d] (88)
Mean usefulness of leaflet sections	3.9 (0.6)[b]	36[d] (88)
App usefulness in participants still using the app	4.3 (0.8)[b]	33[e] (80)
App usefulness in participants who had stopped using the app	2.6 (0.9)[b]	5[e] (12)
App accuracy in participants still using the app	3.9 (1.2)[f]	33[e] (80)
App accuracy in participants who had stopped using the app	2.2 (0.8)[f]	5[e] (12)

[a]The perceived usefulness of the walking planner card was not explored in the time point 1 questionnaire (n=41).

[b]A 5-point scale from not at all useful to extremely useful.

[c]One person reported not receiving a behavioral support call and was not shown this question.

[d]Five people reported that they had not read the leaflet and were not shown these questions.

[e]One participant reported downloading but never using the app to track their walking. Two participants self-reported not downloading the app earlier in the questionnaire. These participants were not shown this question.

[f]A 5-point scale from not accurate to very accurate.

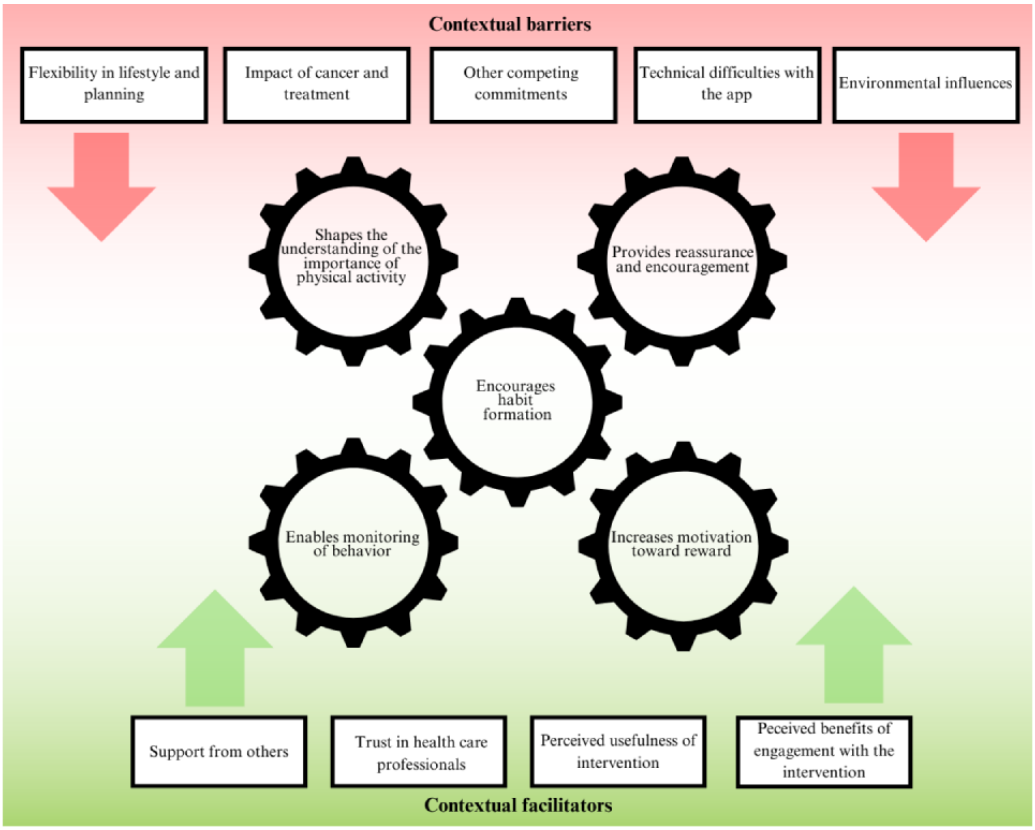
Delivery Fidelity of Behavioral Support Calls

Most (43/44, 98%) of the participants received the first behavioral support call. The mean time from randomization to the first intervention call was 11.6 (SD 9.8; range 5-57) days, and the mean time from randomization to the second intervention call was 39.2 (SD 9.0; range 33-78) days. In total, 31 (72%) of the 43 calls were conducted on Zoom and 12 (28%) via telephone. Most (40/41, 91%) of the participants received the second call. Some (22/40, 55%) of these calls were conducted on Zoom and 45% (18/40) via telephone. In total, 81 intervention calls from 42 participants were included in the analysis (n=42, 52% first calls and n=39, 48% second calls). One intervention participant did not receive any calls, and another participant was removed from the analysis due to a recording issue with the first call, so neither of their calls was included in the fidelity results. The overall mean delivery fidelity score across all BCTs and all participants was 4.2 (SD 1.2), which demonstrates overall proficient delivery. Some (18/25, 72%) of the BCTs had a rating of >4, 16% (4/25) had a rating of 3 to 4, and 12% (3/25) had a rating <3. The BCT called provide information on health consequences had the highest fidelity (4.98). This was followed by action planning (4.90) and habit formation (4.88). The BCT called framing/reframing had the lowest fidelity (1.31). Nonspecific reward (2.71) and nonspecific incentive also displayed low fidelity. The delivery fidelity of each BCT that was intended to be delivered during the calls is presented in Multimedia Appendix 1.

Mechanisms of Impact and Contextual Factors Influencing Engagement

Figure 1 presents the theme diagram showing the identified mechanisms of impact and the contextual barriers and facilitators that affect these mechanisms and intervention engagement.

Figure 1. Theme diagram presenting the identified mechanisms of impact and the contextual barriers and facilitators that affect these mechanisms.



T0 and T1 Questionnaires

All (44/44, 100%) intervention participants completed the SRBAI at baseline. All (42/42, 100%) participants who remained in the study completed the SRBAI at T1. As mentioned earlier, most (41/42, 98%) of the participants who remained in the study answered the section of the T1 questionnaire on intervention feedback.

Qualitative Interviews

Of the 42 participants who remained in the study, 86% (n=36) took part in the qualitative interviews. A few (n=3, 7%) participants did not give a reason for declining to participate, and 2 (5%) participants consented to the interview but then did not respond to the interview invitation. Moreover, 1 (2%) participant did not feel up to taking part in the interview due to illness-related side effects.

Mechanisms of Impact

Identified mechanisms of impact are outlined in the subsequent sections with exemplar quotes indicating the participants' self-identified sex (male or female) and age.

Shapes Understanding of PA and Its Importance

Many participants reported gaining information about brisk walking and its benefits as well as information on how to use the app through the intervention call and the leaflet:

> *...the lady basically went through everything. That was probably the most helpful thing. How to use things and everything.* [Female participant; aged 42 y]

The delivery of comprehensive and meaningful information enhanced participants' understanding of the target behavior and provided them with a clear purpose for implementing it:

> *She went thoroughly through the app with me...Then when she started to explain it, I thought yes, that makes sense.* [Male participant; aged 60 y]

Enables Monitoring of Behavior

Participants reported that having 2 intervention calls was motivating as it helped them to reflect on their progress between the calls:

> *...the follow up call halfway through was touching base and seeing how I was getting on which obviously encouraged me to do it.* [Male participant; aged 73 y]

The feedback on behavior received via the app showing daily, weekly, and monthly minutes of brisk walking was considered effective. Participants reported that they found the tracking feature motivated them to continue their walking efforts:

> *And actually, when I realised I wasn't doing enough, when I felt able to, I extend my walk to get the thirty minutes.* [Female participant; aged 60 y]

Several participants reported using the planner cards to record their walks afterward rather than planning with them upfront. However, this sense of accountability through recording their activity engaged them to keep walking:

> *I wrote down what was on my app, every day, how many minutes walking I did everyday.* [Female participant; aged 62 y]

Increases Motivation Toward Rewards

Many participants reported that the app was the primary intervention component that kept them most motivated and engaged, particularly through the trophies or cups awarded (for every 10 min of brisk walking):

> *I did enjoy getting them cups every day, I thought that were great.* [Female participant; aged 61 y]

Participants often reported walking a few more minutes to achieve the next reward or cup on the app, some even referring to being obsessed or addicted to achieving their goals:

> *...if I get to say 28 minutes, I'll just do the extra two to make it thirty.* [Female participant; aged 49 y]
>
> *30 has been the minimum goal I've gone for. So even if it's not been a nice day or if I'm tired...I still go out...it's addictive.* [Male participant; aged 75 y]

The achievability of these rewards influenced engagement, with several participants reporting exceeding their targets and wishing that more rewards were available:

> *That's another downside, you can't set your goal to any more than three.* [Female participant; aged 61 y]

Participants also reported a sense of satisfaction when being able to tick off completing their walks in their planners:

> *...you can see something, you're achieving something.* [Female participant; aged 47 y]

Encourages Habit Formation

Many participants reported feeling that they had formed habits throughout the intervention period, and this enabled them to establish and maintain their walking and brisk walking habits:

> *It's part of it now, it's part of your day, it's part of your walk so its not I'm going I've got to do this, I've got to do that...It's just a normal day for us going for a walk. And you get back and you think, mmm, I didn't realise I was doing that quick.* [Female participant; aged 65 y]

This is also supported by the SRBAI results for "walking," where total SRBAI scores in the intervention group increased from baseline (mean 4.1, SD 1.6) to T1 (mean 4.7, SD 1.9). Similarly, total SRBAI scores for "brisk walking" in the intervention group increased from baseline (mean 4.0, SD 1.7) to T1 (mean 5.1, SD 1.8).

Providing Reassurance and Encouragement

Many participants recalled how helpful and friendly the facilitator was in the intervention calls:

> *...they were lovely, caring and friendly.* [Female participant; aged 61 y]

This positive rapport helped participants adhere to their brisk walking, and they recalled feeling encouraged and supported throughout the intervention period:

> *...it's useful because it makes you feel as if you're not forgotten.* [Male participant; aged 73 y]

Contextual Factors

The contextual factors influencing engagement are summarized according to barriers and facilitators with exemplar quotes indicating the participants' self-identified sex (male or female) and age.

Barriers

Flexibility in Lifestyle and Planning

Data from the qualitative interviews and the T1 questionnaire indicated that participants felt the planners, in their intended use (to plan walks each week), were not flexible enough and did not fit their lifestyles:

> *...it didn't work for me at all because every day is different.* [Female participant; aged 42 y]

Related to this, some participants suggested making the planners daily, rather than weekly, which would allow more nuanced plans to be made:

> *...it wasn't going to be at the same time every day so it just needed breaking down into a daily thing.* [Female participant; aged 57 y]

The Impact of Cancer and Its Treatment

Several participants reported in the interviews and the T1 questionnaire that their cancer diagnosis and treatment sometimes made it difficult to engage with the intervention:

> *...this was just after my treatment and I was tired a bit.* [Female participant; aged 67 y]
>
> *I couldn't plan, because I was in and out of different appointment times.* [Female participant; aged 57 y]

This was also seen when participants were asked about the appropriateness of the app for people LWBC, with several participants reflecting on how differing experiences may help or hinder engagement:

> *I had some days where I felt like I didn't want to see daylight, not talk to anybody...I think on those bad days I wouldn't have wanted to be bothered with it.* [Female participant; aged 67 y]

Many participants discussed that the appropriateness of the timing of the intervention would be dependent on where the patient was in their cancer care. Although many participants noted that taking part in the intervention during treatment would have been too difficult for them due to the side effects, several also suggested that using the intervention during treatment was suitable as it gave them something else to focus on:

> *No I wouldn't have been able to do it when I was having chemo, I could barely even walk round the garden.* [Female participant; aged 58 y]
>
> *This gave me something else to think about you know something on a daily basis which took me mind off you know that three-week cycle as much as it could.* [Male participant; aged 65 y]

Other Competing Commitments

Several participants reported not having time to walk due to factors, such as having caring responsibilities, working hours, and appointments:

> *...well the kids can't walk with me they have only got little legs, who is going to watch them?* [Female participant; aged 41 y]

Technical Difficulties With the App

In the interviews, several participants reported experiencing difficulties with the recording of walks on the app if there was a lack of signal or due to the phone's positioning. For example, sometimes participants found the recording differed depending on what pocket their phone was in:

> *...if I had it in my trouser pocket by my leg it worked all fine, ok, no problem at all, if I had it in my shirt pocket it didn't register anything.* [Male participant; aged 73 y]

This is supported by the T1 questionnaire results indicating that perceived app accuracy was an issue that likely influenced use and engagement with the app over time:

> *Didn't find the app as accurate as it could be and doubted its recordings on occasions.* [Female participant; aged 41 y]

Environmental Influences

Participants described how the weather played an influential role in their ability and motivation to go on their walks:

> *What it's going to be like when it starts raining and it's really awful weather I don't know.* [Female participant; aged 40 y]

As well as hindering their ability to go out on a walk, some participants also described feeling guilty if they did not go out and walk due to the weather:

> *There were some days where it was absolutely throwing it down outside…and I thought oh no and I really felt guilty that I'd not actually done it.* [Female participant; aged 61 y]

Facilitators

Support From Others

Many participants talked about telling people of their involvement in the trial and their efforts to increase their walking, describing a sense of accountability from sharing the experience with others:

> *...so now it's a case of, "Have you got your minutes in yet dad?" and it, everyone's sort of like joining in with it.* [Male participant; aged 75 y]

In addition, having someone to go on a walk with and even having family members and friends also use the app was described as encouraging and gave participants a sense of comradery in changing their behavior:

> *It made us all as a family go for a walk. It not just helped me with that. It helped all of us.* [Female participant; aged 42 y]

Personal contact from the study team also facilitated engagement, with several participants reporting that they felt supported and that they had the opportunity to get in touch if needed:

> *...it were good to know that somebody you know, I'd phone and they'd take an interest.* [Female participant; aged 61 y]

Trust in Health Care Professionals

Several participants noted that endorsement from their health care professional facilitated their engagement with the intervention and willingness to change their behavior, as their medical team is seen as a credible source of information:

> *So the fact it had come from the doctor made me want to do it even more.* [Female participant; aged 42 y]

Perceived Usefulness of the Intervention

The perceived usefulness of the intervention components appeared to influence engagement in the target behavior, with participants reporting that some components (eg, app) were more helpful than others (eg, planners).

Overall, the behavioral support call was rated as useful (mean 4.1, SD 1; Table 2). In the qualitative interviews, the participants reported finding the calls useful as a source of information and motivation as well as helping them to regain their focus:

> *You know, if you've got an issue you can talk to somebody about it. But also it keeps you motivated.* [Female participant; aged 56 y]

The leaflet was generally rated as useful (mean 3.9, SD 0.6), particularly the sections on "walking," "downloading the app," and "walking habits" (Table 2). The qualitative data suggested that although this was useful for reading initial information about the benefits of brisk walking and particularly for downloading the app, some participants reported limited recollection of the leaflet at the follow-up point:

> *I did have a quick flick through it. And obviously the bit about finding the app.* [Female participant; aged 49 y]

T1 questionnaire results indicated a mean usefulness of 4.1 (SD 1) for the app, but with higher ratings among the 33 participants still using it compared to the 5 who had ceased. Most (28/38, 74%) of the users found it extremely useful or very useful. Participants who reported still using the app reported higher perceived accuracy of the app in recording their time spent walking compared to those not using the app anymore (Table 2). This is supported by the qualitative findings, with participants reporting still using the app and finding it enjoyable to monitor their progress on it:

> *I use it all the time now...I still want to make sure I have got at least two cups.* [Male participant; aged 74 y]

The usefulness of the planner explored in the qualitative interviews highlighted that although the planners were useful to get started and to form habits, their use ceased over time due to factors such as not finding planning as helpful long-term,

finding the app sufficient to motivate them, and just forgetting to use them:

> *I think you need the planner thing for the first week or so but after that I don't think you do.* [Female participant; aged 56 y]

Perceived Benefits of Engagement With the Intervention

Where participants felt they were benefiting from taking part, they reported feeling motivated to continue their brisk walking, reporting feelings of enjoyment, better mood and well-being, and improvements in physical health and fitness:

> *I definitely felt better in myself, there's no question about that. I felt fitter as well.* [Male participant; aged 60 y]

As well as the direct impact of the intervention, many participants also reported feeling more able to engage in activities of daily living, such as going shopping, socializing, and doing housework because of their improved fitness and well-being:

> *I have started going out with more friends...getting out a bit more and feel better and everything.* [Female participant; aged 67 y]

Discussion

Summary of Findings

This study combined data collected as part of the APPROACH pilot RCT to assess the implementation of a multicomponent, app-based behavioral intervention to promote brisk walking in people LWBC. The findings of this process evaluation demonstrated proficient implementation of the intervention and suggest that there are several mechanisms of impact underlying the efficacy of the intervention as well as contextual factors that can be barriers or facilitators to engagement.

Successful Implementation of the APPROACH Intervention

The successful delivery of intervention components and intended BCTs is essential for attributing any changes in behavior to the intervention in question [41]. This study demonstrated proficient implementation, with most participants reporting engagement with the intervention components, including downloading the app, reading the leaflet, receiving the behavioral support call, and using the planners. In addition, fidelity is rarely reported in evaluations of PA interventions [29]. In a systematic review of 21 studies assessing the quality of measuring delivery fidelity in PA interventions, Lambert et al [42] reported considerable heterogeneity in evaluating delivery fidelity. This study reports high delivery fidelity of the intended BCTs in the behavioral support calls [21]. While interventions with only a single intervention facilitator often exhibit higher fidelity, investigating how various intervention facilitators across different contextual settings deliver BCTs is essential for understanding the real-world application and scalability of interventions [43]. Particularly where multiple intervention facilitators are involved in the delivery, future research should consider factors such as personal characteristics and individual context when evaluating intervention delivery [44,45].

The BCTs called information on health consequences, action planning, and habit formation showed the highest fidelity, and these BCTs appear to influence key mechanisms of impact highlighted by participants in the qualitative interviews. Empowering participants with information on why they should go walking and providing them with information about its benefits appeared to enhance their engagement and adherence to their planned walks. The delivery of and engagement with the action planning BCT is promising, as this BCT was shown to be associated with larger effect sizes in a systematic review of 26 studies evaluating BCT effectiveness in PA interventions in adults who are healthy and inactive [29]. This study extends this finding and suggests that this BCT is suitable and appropriate for people LWBC and should be included in future intervention designs with this population. The BCTs that were not successfully delivered included those called nonspecific reward, nonspecific incentive, and framing or reframing. Despite the BCT called nonspecific reward receiving a low fidelity rating in the recorded behavioral support calls, the qualitative interviews suggested that this BCT was effectively covered in other aspects of the intervention, with participants reporting enjoying working toward the rewards and trophies in the NHS Active 10 app. The low delivery fidelity of the BCT called framing/reframing could be attributed to participants' voluntary enrollment in a PA trial. It is likely that they already recognized the importance of PA even without fully recognizing its role in life beyond a cancer diagnosis [46,47].

BCTs Underlying Change

The mechanisms of impact identified by participants in the qualitative interviews and follow-up questionnaire were in line with previous research. For instance, self-monitoring of behavior is one of the most frequently used components in complex PA interventions [48], and multiple systematic reviews have demonstrated its effectiveness in increasing PA that is maintained long-term [49-51]. In this intervention, the NHS Active 10 app allowed participants to track their brisk walking and total walking, and participants reported that being able to see their improvement over time motivated them to continue their behaviors. In a similar pilot RCT using a self-monitoring app alone, Ormel et al [52] reported that participants in the intervention group increased their activity more from baseline to 6 weeks, but this difference was not maintained at 12 weeks. The authors attributed this to a potential loss of novelty and interest in the app. Although not powered to detect differences in PA, the findings of this study suggest that the second behavioral support call was important in continuing encouragement of monitoring of behavior and reviewing of goals, with participants describing a sense of accountability with the later call. This feedback indicates that a light-touch intervention call to check in with participants can be beneficial in consolidating commitment to PA goals in people LWBC. While technology can help reduce resource demands, future research should consider the benefits of a low-burden behavioral support call to augment an app as an already powerful and scalable intervention component.

Participants extensively discussed the app as the driving component of the intervention, supporting our previous reports of high engagement with the NHS Active 10 app [22]. The

discussions revealed that another mechanism underlying behavior change in this context was the ability of the app to increase motivation toward reward using gamification techniques [53]. The performance of the desired behavior (brisk walking) was reinforced by the positive feelings of encouragement and dedication resulting from these cups and trophies in the app.

Encouraging Habit Formation

The APPROACH intervention was informed by habit theory [21], and the identification of habit formation as a mechanism of impact in both the qualitative interviews and questionnaire data demonstrates that this BCT was delivered effectively in the intervention. Participants reported that the leaflet, app, behavioral support call, and planners helped to establish sustainable walking habits. Gardner et al [54] define habit formation as "learning cue–behaviour associations, that when cued, automatically generate action impulses." This sense of automaticity was described by participants in the qualitative interviews, whereby consistent repetition of their walking daily led to the enactment of brisk walking. This meant that they engaged in brisk walking even during activities that previously would not have involved this exercise intensity. These findings are also supported by the increased SRBAI results from baseline to follow-up, which showed an increase in the initiation of walks as well as the way walking was executed (ie, briskly) [55]. The importance of encouraging long-term engagement in PA after the intervention period is highlighted in systematic reviews of PA maintenance in cancer populations that report only modest improvements at longer follow-up time points [30,56]. The results of this study endorse the integration of habit theory into future interventions aimed at increasing PA to overcome the challenge of sustaining behavioral changes over time [56].

To reinforce the idea of habit, the walking planner cards were designed to promote habit formation, facilitate planning, and prompt participants to engage in PA [21,54]. However, both questionnaire data and interview data showed that the structured design of the planner cards was not compatible with the day-to-day changing schedules and lifestyles of some participants, highlighting the importance of conducting this process evaluation to account for and reconsider this contextual aspect of the intervention and future similar interventions. The need for flexibility in lifestyle and planning was reported as a barrier by participants and further confirmed by the reflections of the intervention facilitator (CB) after discussing their use with participants.

Barriers to Engagement: Cancer Impact and Competing Commitments

Other contextual barriers included the impact of cancer and its treatment, having other competing commitments, technical difficulties with the app, and environmental influences. The side effects of cancer and its treatment have previously been identified as a key barrier to PA participation in systematic reviews [18] as well as in our own preparatory work for this pilot RCT [19]. Participants in this study reported that the impact of cancer and its treatment inhibited their ability to engage in some elements of the intervention due to different physiological, structural, and psychological factors. Cancer-related fatigue is the most reported symptom in people LWBC who have undergone treatment with prevalence estimates of up to 90% of those treated with radiotherapy and 80% of those treated with chemotherapy [57,58]. In this study, participants discussed fatigue symptoms and felt that engaging with the intervention during treatment would have been difficult. Beyond this physiological barrier, the structural barrier of having any appointments for their cancer care also reduced their ability to engage with some components, including the planner card, as there were many hospital appointments that they had to attend and plan around. We have previously reported on the perceived suitability of the timing of the APPROACH intervention, with most participants feeling that it was reasonable [22]. While some participants felt that engaging during treatment would be difficult, others felt that it was useful to have something else to focus on and have control over [22]. Previous reviews in this area have also reported discrepancies in the preferred timing of PA intervention delivery within the cancer care pathway [18,59]. Involvement in PA at an earlier stage has been associated with improved treatment response, tolerance, and quality of life [13,60,61]. Considering, the delivery of the APPROACH intervention during and after the treatment for cancer is still endorsed while recognizing the contextual barriers, such as this, when interpreting APPROACH intervention outcomes. These findings highlight the importance of involving patient perspectives in future intervention design with this population, with the acknowledgment that different stages of the treatment pathway can facilitate or inhibit PA participation and should be accounted for when assessing intervention delivery and engagement.

Driving Engagement: Support, Trust, and Perceived Benefits

Facilitators to engagement included having support from others, trust in health care professionals, the perceived usefulness of the intervention, and the perceived benefits of engagement with the intervention. The BCTs called social support (practical) and social support (emotional) were competently delivered in line with the protocol [21,31]. Accordingly, participants recognized how having support from others, such as family members and partners, enhanced their engagement with the intervention and how they felt supported by the intervention facilitator. Social support has previously been identified as highly important in PA engagement [62], with reasons, such as accountability, being cited as helping to facilitate and promote engagement [63]. In our preparatory work for this RCT, trust in health care professionals emerged as a crucial factor influencing engagement [19], and this was highlighted again in this study, where incorporating an endorsement letter from the clinical care team enhanced app credibility.

Limitations

Limitations of this study include that only 1 intervention facilitator delivered all behavioral support calls, which may explain the high-fidelity ratings. It is crucial to consider the transferability of the intervention across individuals, particularly when envisioning integration into routine NHS care [20,22]. In addition, there may be some recall bias influencing results, as most of the questionnaire data were collected at the 3-month

follow-up [64]. Participants may have had difficulty in answering questions on earlier components (eg, leaflet information) compared to components they were still using (eg, app). Due to the pilot nature of this study, where feasibility and acceptability were the main outcomes of interest, we were unable to examine how engagement with each intervention component influenced the primary outcome of brisk walking. For future research, the application of the Multiphase Optimization Strategy with a factorial design could offer more insights into how differing engagement with each component can impact the main outcomes and help inform intervention optimization for larger efficacy RCTs [65,66]. Finally, despite its recognition as a potential mechanism, it is difficult to assess how the intervention helps to establish longer-term habits, which are key for PA maintenance [67], as this pilot RCT only examined outcomes at 3 months.

Conclusions

This study extends our previously published findings on the APPROACH pilot RCT [22] by demonstrating that the intervention was delivered as intended with high levels of engagement from participants. In addition, this paper highlighted the potential mechanisms through which change occurs, such as habit formation and behavioral monitoring, which are in line with the intended BCTs used in this intervention. The process evaluation also highlighted important contextual factors to consider when progressing to the APPROACH main trial, including facilitators, such as social support, which played a significant role in promoting adherence to the intervention. The protocol for the definitive RCT will report on adaptations made to APPROACH based on the feedback gathered in this study. This process evaluation provides strong support for the progression to the stage-3, definitive RCT to evaluate the effectiveness of the APPROACH intervention (began in November 2023) and enables a more nuanced understanding of how the APPROACH intervention works and the contextual factors to consider with implementation.

Acknowledgments

The authors would like to thank Yorkshire Cancer Research for funding this pilot study and the participants for dedicating the time to participate and complete the assessments. They also thank the extended APPROACH team members Rithu Anand and Gaby Heuchan for their help with carrying out trial tasks.

Conflicts of Interest

None declared.

Multimedia Appendix 1
APPROACH intervention call behavior change technique checklist, intended delivery technique, and mean delivery fidelity score.
[DOCX File , 31 KB - cancer_v11i1e64747_app1.docx]

Multimedia Appendix 2
Adapted Dreyfus rating scale for the intervention behavior change technique calls.
[DOCX File , 28 KB - cancer_v11i1e64747_app2.docx]

Multimedia Appendix 3
Self-Report Behavioural Automaticity Index.
[DOCX File , 21 KB - cancer_v11i1e64747_app3.docx]

Multimedia Appendix 4
Exit interview guide.
[DOCX File , 32 KB - cancer_v11i1e64747_app4.docx]

References

1. Maddams J, Utley M, Møller H. Projections of cancer prevalence in the United Kingdom, 2010-2040. Br J Cancer 2012 Sep 25;107(7):1195-1202 [FREE Full text] [doi: 10.1038/bjc.2012.366] [Medline: 22892390]
2. Frikkel J, Götte M, Beckmann M, Kasper S, Hense J, Teufel M, et al. Fatigue, barriers to physical activity and predictors for motivation to exercise in advanced cancer patients. BMC Palliat Care 2020 Mar 31;19(1):43 [FREE Full text] [doi: 10.1186/s12904-020-00542-z] [Medline: 32234027]
3. Maass SW, Boerman LM, Verhaak PF, Du J, de Bock GH, Berendsen AJ. Long-term psychological distress in breast cancer survivors and their matched controls: a cross-sectional study. Maturitas 2019 Dec;130:6-12. [doi: 10.1016/j.maturitas.2019.09.003] [Medline: 31706438]

4. Neo J, Fettes L, Gao W, Higginson IJ, Maddocks M. Disability in activities of daily living among adults with cancer: a systematic review and meta-analysis. Cancer Treat Rev 2017 Dec;61:94-106 [FREE Full text] [doi: 10.1016/j.ctrv.2017.10.006] [Medline: 29125982]
5. Snijders RA, Brom L, Theunissen M, van den Beuken-van Everdingen M. Update on prevalence of pain in patients with cancer 2022: a systematic literature review and meta-analysis. Cancers (Basel) 2023 Jan 18;15(3):591 [FREE Full text] [doi: 10.3390/cancers15030591] [Medline: 36765547]
6. Annunziata MA, Muzzatti B, Flaiban C, Gipponi K, Carnaghi C, Tralongo P, et al. Long-term quality of life profile in oncology: a comparison between cancer survivors and the general population. Support Care Cancer 2018 Feb;26(2):651-656. [doi: 10.1007/s00520-017-3880-8] [Medline: 28918552]
7. Maher J, Petchey L, Greenfield D, Levitt G, Fraser M. Implementation of nationwide cancer survivorship plans: experience from the UK. J Cancer Policy 2018 May;15:76-81. [doi: 10.1016/j.jcpo.2018.01.002]
8. Mustian KM, Alfano CM, Heckler C, Kleckner AS, Kleckner IR, Leach CR, et al. Comparison of pharmaceutical, psychological, and exercise treatments for cancer-related fatigue: a meta-analysis. JAMA Oncol 2017 Jul 01;3(7):961-968 [FREE Full text] [doi: 10.1001/jamaoncol.2016.6914] [Medline: 28253393]
9. Salam A, Woodman A, Chu A, Al-Jamea LH, Islam M, Sagher M, et al. Effect of post-diagnosis exercise on depression symptoms, physical functioning and mortality in breast cancer survivors: a systematic review and meta-analysis of randomized control trials. Cancer Epidemiol 2022 Apr;77:102111. [doi: 10.1016/j.canep.2022.102111] [Medline: 35091272]
10. Gerritsen JK, Vincent AJ. Exercise improves quality of life in patients with cancer: a systematic review and meta-analysis of randomised controlled trials. Br J Sports Med 2016 Jul 30;50(13):796-803. [doi: 10.1136/bjsports-2015-094787] [Medline: 26719503]
11. Plinsinga ML, Singh B, Rose GL, Clifford B, Bailey TG, Spence RR, et al. The effect of exercise on pain in people with cancer: a systematic review with meta-analysis. Sports Med 2023 Sep;53(9):1737-1752 [FREE Full text] [doi: 10.1007/s40279-023-01862-9] [Medline: 37213049]
12. Cariolou M, Abar L, Aune D, Balducci K, Becerra-Tomás N, Greenwood DC, et al. Postdiagnosis recreational physical activity and breast cancer prognosis: global Cancer Update Programme (CUP Global) systematic literature review and meta-analysis. Int J Cancer 2023 Feb 15;152(4):600-615 [FREE Full text] [doi: 10.1002/ijc.34324] [Medline: 36279903]
13. Rock CL, Thomson CA, Sullivan KR, Howe CL, Kushi LH, Caan BJ, et al. American Cancer Society nutrition and physical activity guideline for cancer survivors. CA Cancer J Clin 2022 May;72(3):230-262 [FREE Full text] [doi: 10.3322/caac.21719] [Medline: 35294043]
14. Friedenreich CM, Neilson HK, Farris MS, Courneya KS. Physical activity and cancer outcomes: a precision medicine approach. Clin Cancer Res 2016 Oct 01;22(19):4766-4775. [doi: 10.1158/1078-0432.CCR-16-0067] [Medline: 27407093]
15. Clinton SK, Giovannucci EL, Hursting SD. The World Cancer Research Fund/American Institute for Cancer Research third expert report on diet, nutrition, physical activity, and cancer: impact and future directions. J Nutr 2020 Apr 01;150(4):663-671 [FREE Full text] [doi: 10.1093/jn/nxz268] [Medline: 31758189]
16. Move More: physical activity the underrated ‘wonder drug’. Macmillan Cancer Support. 2011. URL: https://cdn.ps.emap.com/wp-content/uploads/sites/3/2011/08/Physical-activity-campaign-report-lores.pdf [accessed 2024-04-29]
17. Wong JN, McAuley E, Trinh L. Physical activity programming and counseling preferences among cancer survivors: a systematic review. Int J Behav Nutr Phys Act 2018 Jun 07;15(1):48 [FREE Full text] [doi: 10.1186/s12966-018-0680-6] [Medline: 29879993]
18. Elshahat S, Treanor C, Donnelly M. Factors influencing physical activity participation among people living with or beyond cancer: a systematic scoping review. Int J Behav Nutr Phys Act 2021 Apr 06;18(1):50 [FREE Full text] [doi: 10.1186/s12966-021-01116-9] [Medline: 33823832]
19. Roberts AL, Potts HW, Koutoukidis DA, Smith L, Fisher A. Breast, prostate, and colorectal cancer survivors' experiences of using publicly available physical activity mobile apps: qualitative study. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e10918 [FREE Full text] [doi: 10.2196/10918] [Medline: 30609982]
20. Roberts AL, Potts HW, Stevens C, Lally P, Smith L, Fisher A. Cancer specialist nurses' perspectives of physical activity promotion and the potential role of physical activity apps in cancer care. J Cancer Surviv 2019 Oct 02;13(5):815-828 [FREE Full text] [doi: 10.1007/s11764-019-00801-w] [Medline: 31475306]
21. Lally P, Miller N, Roberts A, Beeken RJ, Greenfield DM, Potts HW, et al. An app with brief behavioural support to promote physical activity after a cancer diagnosis (APPROACH): study protocol for a pilot randomised controlled trial. Pilot Feasibility Stud 2022 Mar 29;8(1):74 [FREE Full text] [doi: 10.1186/s40814-022-01028-w] [Medline: 35351187]
22. Lally P, Kennedy F, Smith S, Beeken RJ, Buck C, Thomas C, et al. The feasibility and acceptability of an app-based intervention with brief behavioural support (APPROACH) to promote brisk walking in people diagnosed with breast, prostate and colorectal cancer in the UK. Cancer Med 2024 Mar;13(6):e7124 [FREE Full text] [doi: 10.1002/cam4.7124] [Medline: 38529687]
23. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]

24. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
25. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350(mar19 6):h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
26. Creswell JW, Clark VL. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: SAGE Publications; 2017.
27. Tudor-Locke C, Aguiar EJ, Han H, Ducharme SW, Schuna JM, Barreira TV, et al. Walking cadence (steps/min) and intensity in 21-40 year olds: CADENCE-adults. Int J Behav Nutr Phys Act 2019 Jan 17;16(1):8 [FREE Full text] [doi: 10.1186/s12966-019-0769-6] [Medline: 30654810]
28. Lally P, Gardner B. Promoting habit formation. Health Psychol Rev 2013 May;7(sup1):S137-S158. [doi: 10.1080/17437199.2011.603640]
29. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. Transl Behav Med 2019 Jan 01;9(1):147-157. [doi: 10.1093/tbm/iby010] [Medline: 29506209]
30. Turner RR, Steed L, Quirk H, Greasley RU, Saxton JM, Taylor SJ, et al. Interventions for promoting habitual exercise in people living with and beyond cancer. Cochrane Database Syst Rev 2018 Sep 19;9(9):CD010192 [FREE Full text] [doi: 10.1002/14651858.CD010192.pub3] [Medline: 30229557]
31. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
32. Behaviour change interventions: introductory principles and practice. Future Learn. 2021. URL: https://www.futurelearn.com/courses/behaviour-change-interventions [accessed 2024-04-29]
33. Gardner B, Rebar AL, Lally P. A matter of habit: recognizing the multiple roles of habit in health behaviour. Br J Health Psychol 2019 May;24(2):241-249. [doi: 10.1111/bjhp.12369] [Medline: 30945793]
34. Zoom security guide 2016. Zoom Video Communications Inc. URL: https://d24cgw3uvb9a9h.cloudfront.net/static/81625/doc/Zoom-Security-White-Paper.pdf [accessed 2024-04-29]
35. Dreyfus SE. The five-stage model of adult skill acquisition. Bull Sci Technol Soc 2004 Jun 01;24(3):177-181. [doi: 10.1177/0270467604264992]
36. Cross R, Greaves CJ, Withall J, Rejeski WJ, Stathi A. Delivery fidelity of the REACT (REtirement in ACTion) physical activity and behaviour maintenance intervention for community dwelling older people with mobility limitations. BMC Public Health 2022 Jun 03;22(1):1112 [FREE Full text] [doi: 10.1186/s12889-022-13496-z] [Medline: 35658857]
37. Gardner B, Abraham C, Lally P, de Bruijn GJ. Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. Int J Behav Nutr Phys Act 2012 Aug 30;9:102 [FREE Full text] [doi: 10.1186/1479-5868-9-102] [Medline: 22935297]
38. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
39. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Greene JC, Caracelli VJ, Graham WF. Toward a conceptual framework for mixed-method evaluation designs. Educ Eval Policy Anal 1989;11(3):255. [doi: 10.2307/1163620]
41. Elliott DS, Mihalic S. Issues in disseminating and replicating effective prevention programs. Prev Sci 2004 Mar;5(1):47-53. [doi: 10.1023/b:prev.0000013981.28071.52] [Medline: 15058912]
42. Lambert JD, Greaves CJ, Farrand P, Cross R, Haase AM, Taylor AH. Assessment of fidelity in individual level behaviour change interventions promoting physical activity among adults: a systematic review. BMC Public Health 2017 Oct 02;17(1):765 [FREE Full text] [doi: 10.1186/s12889-017-4778-6] [Medline: 28969669]
43. James K, Quirk A, Patterson S, Brennan G, Stewart D. Quality of intervention delivery in a cluster randomised controlled trial: a qualitative observational study with lessons for fidelity. Trials 2017 Nov 17;18(1):548 [FREE Full text] [doi: 10.1186/s13063-017-2189-8] [Medline: 29149915]
44. Thompson TP, Lambert JD, Greaves CJ, Taylor AH. Intervention delivery fidelity assessment of a counseling-based intervention for promoting smoking reduction and increasing physical activity. Health Psychol 2018 Jul;37(7):627-637 [FREE Full text] [doi: 10.1037/hea0000613] [Medline: 29708387]
45. Campbell BK, Buti A, Fussell HE, Srikanth P, McCarty D, Guydish JR. Therapist predictors of treatment delivery fidelity in a community-based trial of 12-step facilitation. Am J Drug Alcohol Abuse 2013 Sep 09;39(5):304-311 [FREE Full text] [doi: 10.3109/00952990.2013.799175] [Medline: 23837717]
46. Nuzzo J. Volunteer bias and female participation in exercise and sports science research. Quest 2021 Feb 23;73(1):82-101. [doi: 10.1080/00336297.2021.1875248]

47. de Souto Barreto P, Ferrandez AM, Saliba-Serre B. Are older adults who volunteer to participate in an exercise study fitter and healthier than nonvolunteers? The participation bias of the study population. J Phys Act Health 2013 Mar;10(3):359-367. [Medline: 23619184]
48. McEwan D, Beauchamp MR, Kouvousis C, Ray CM, Wyrough A, Rhodes RE. Examining the active ingredients of physical activity interventions underpinned by theory versus no stated theory: a meta-analysis. Health Psychol Rev 2019 Mar 20;13(1):1-17. [doi: 10.1080/17437199.2018.1547120] [Medline: 30412685]
49. Chaudhry UA, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020 Oct 09;17(1):129 [FREE Full text] [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]
50. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
51. Compernolle S, DeSmet A, Poppe L, Crombez G, de Bourdeaudhuij I, Cardon G, et al. Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Aug 13;16(1):63 [FREE Full text] [doi: 10.1186/s12966-019-0824-3] [Medline: 31409357]
52. Ormel HL, van der Schoot GG, Westerink NL, Sluiter WJ, Gietema JA, Walenkamp AM. Self-monitoring physical activity with a smartphone application in cancer patients: a randomized feasibility study (SMART-trial). Support Care Cancer 2018 Nov;26(11):3915-3923 [FREE Full text] [doi: 10.1007/s00520-018-4263-5] [Medline: 29785635]
53. Robson K, Plangger K, Kietzmann JH, McCarthy I, Pitt L. Is it all a game? Understanding the principles of gamification. Bus Horiz 2015 Jul;58(4):411-420. [doi: 10.1016/j.bushor.2015.03.006]
54. Gardner B, Rebar AL, de Wit S, Lally P. What is habit and how can it be used to change real‐world behaviour? Narrowing the theory‐reality gap. Soc Personal Psychol Compass 2024 May 31;18(6):e12975. [doi: 10.1111/spc3.12975]
55. Gardner B. Habit and behavioural complexity: habitual instigation and execution as predictors of simple and complex behaviours. Curr Res Behav Sci 2022;3:100081. [doi: 10.1016/j.crbeha.2022.100081]
56. Grimmett C, Corbett T, Brunet J, Shepherd J, Pinto BM, May CR, et al. Systematic review and meta-analysis of maintenance of physical activity behaviour change in cancer survivors. Int J Behav Nutr Phys Act 2019 Apr 27;16(1):37 [FREE Full text] [doi: 10.1186/s12966-019-0787-4] [Medline: 31029140]
57. Brown LF, Kroenke K. Cancer-related fatigue and its associations with depression and anxiety: a systematic review. Psychosomatics 2009;50(5):440-447 [FREE Full text] [doi: 10.1176/appi.psy.50.5.440] [Medline: 19855028]
58. Hofman M, Ryan JL, Figueroa-Moseley CD, Jean-Pierre P, Morrow GR. Cancer-related fatigue: the scale of the problem. Oncologist 2007;12 Suppl 1:4-10 [FREE Full text] [doi: 10.1634/theoncologist.12-S1-4] [Medline: 17573451]
59. Phillips SM, Conroy DE, Keadle SK, Pellegrini CA, Lloyd GR, Penedo FJ, et al. Breast cancer survivors' preferences for technology-supported exercise interventions. Support Care Cancer 2017 Oct;25(10):3243-3252 [FREE Full text] [doi: 10.1007/s00520-017-3735-3] [Medline: 28470368]
60. Bland KA, Zadravec K, Landry T, Weller S, Meyers L, Campbell KL. Impact of exercise on chemotherapy completion rate: a systematic review of the evidence and recommendations for future exercise oncology research. Crit Rev Oncol Hematol 2019 Apr;136:79-85. [doi: 10.1016/j.critrevonc.2019.02.005] [Medline: 30878132]
61. Yang L, Morielli AR, Heer E, Kirkham AA, Cheung WY, Usmani N, et al. Effects of exercise on cancer treatment efficacy: a systematic review of preclinical and clinical studies. Cancer Res 2021 Oct 01;81(19):4889-4895 [FREE Full text] [doi: 10.1158/0008-5472.CAN-21-1258] [Medline: 34215623]
62. McDonough MH, Beselt LJ, Daun JT, Shank J, Culos-Reed SN, Kronlund LJ, et al. The role of social support in physical activity for cancer survivors: a systematic review. Psychooncology 2019 Oct;28(10):1945-1958. [doi: 10.1002/pon.5171] [Medline: 31278800]
63. Lesser IA, Prystupa J, Belanger L, Thomson C, Nienhuis CP. A mixed-methods evaluation of a group based trail walking program to reduce anxiety in cancer survivors. Appl Cancer Res 2020 Oct 19;40(1):10. [doi: 10.1186/s41241-020-00094-x]
64. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016;9:211-217 [FREE Full text] [doi: 10.2147/JMDH.S104807] [Medline: 27217764]
65. Collins LM, Kugler KC. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: Advanced Topics. Cham, Switzerland: Springer; 2018.
66. Collins LM, Nahum-Shani I, Guastaferro K, Strayhorn JC, Vanness DJ, Murphy SA. Intervention optimization: a paradigm shift and its potential implications for clinical psychology. Annu Rev Clin Psychol 2024 Jul;20(1):21-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-080822-051119] [Medline: 38316143]
67. Rhodes RE, Sui W. Physical activity maintenance: a critical narrative review and directions for future research. Front Psychol 2021 Sep 6;12:725671 [FREE Full text] [doi: 10.3389/fpsyg.2021.725671] [Medline: 34552537]

Abbreviations

BCT: behavior change technique
CONSORT: Consolidated Standards of Reporting Trials

LWBC: living with and beyond cancer
MVPA: moderate-to-vigorous physical activity
PA: physical activity
RCT: randomized controlled trial
SRBAI: Self-Report Behavioural Automaticity Index
T0: time point 0
T1: time point 1

Edited by J Bender; submitted 26.07.24; peer-reviewed by E Arden-Close, A Campbell; comments to author 11.11.24; revised version received 22.11.24; accepted 30.12.24; published 10.02.25.

Please cite as:
Kennedy F, Smith S, Beeken RJ, Buck C, Williams S, Martin C, Lally P, Fisher A
An App-Based Intervention With Behavioral Support to Promote Brisk Walking in People Diagnosed With Breast, Prostate, or Colorectal Cancer (APPROACH): Process Evaluation Study
JMIR Cancer 2025;11:e64747
URL: https://cancer.jmir.org/2025/1/e64747
doi:10.2196/64747
PMID:

Development of a Mobile App to Support Head and Neck Cancer Caregiving: Mixed Methods Study

Katherine Sterba[1], MPH, PhD; Evan Graboyes[1,2], MPH, MD; Jessica Burris[3], PhD; Megan Scallion[1], MSW, MPH; Hannah Kinder[1], MPH; Jama Olsen[1], MPA; Benjamin Toll[1], PhD; Kent Armeson[1], MS; Terry Day[2], MD; Bhishamjit Chera[4], MD; Kenneth Ruggiero[5], PhD

[1]Department of Public Health Sciences, Medical University of South Carolina, 86 Jonathan Lucas Street, Charleston, SC, United States

[2]Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, Charleston, SC, United States

[3]Department of Psychology, University of Kentucky, Lexington, KY, United States

[4]Department of Radiation Oncology, Medical University of South Carolina, Charleston, United States

[5]Department of Behavioral Sciences and Psychiatry, College of Nursing, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Katherine Sterba, MPH, PhD
Department of Public Health Sciences, Medical University of South Carolina, 86 Jonathan Lucas Street, Charleston, SC, United States

Abstract

Background: Survivors with head and neck cancer (HNC) face challenging treatment consequences that can lead to severe disruptions in swallowing and result in weight loss, malnutrition, and feeding tube dependence. Caregivers (family or friends who provide support), therefore, often encounter distressing nutritional caregiving burdens and feel unprepared to provide adequate support at home.

Objective: The purpose of this mixed methods study was to develop a mobile support app to support HNC caregiving with an emphasis on nutritional support following treatment.

Methods: We assessed perspectives on nutritional recovery challenges and mobile support app preferences in (1) a national panel of oncology dietitians using a web-based cross-sectional survey and (2) survivors with HNC completing treatment within the past 24 months and their nominated caregivers using dyadic semistructured interviews. Descriptive statistics for survey data were synthesized with thematic analysis of interview data to characterize nutrition-related perceptions and intervention preferences; results were integrated, and themes were translated to high-priority main menu domains and subdomains for a mobile app for caregivers.

Results: Surveys were completed by dietitians (n=116, 100%; female n=87, 50%, with >10 years practice experience). Interviews included survivors with HNC (n=15; 12/15, 80% male, and 6/15, 40% with oropharynx cancer) and their caregivers (n=13; 11/13, 85% female, and 10/13, 77% spouses). Dietitians, survivors, and caregivers perceived that the majority of nutritional concerns assessed (eg, swallowing, feeding tube management, weight maintenance, and caregiver distress about nutrition) were very or extremely important to caregiving in the 6 months following treatment conclusion. The caregiving tasks rated highest in importance by dietitians included tracking nutritional concerns (n=113, 97%), working together as a team on nutritional concerns (n=104, 90%), and making care decisions (n=102, 88%). Five themes emerged from dyadic interviews, including types of nutritional challenges faced, that competing symptoms were difficult to separate from nutritional challenges, the emotional challenges related to nutrition and recovery, the diverse set of medical and support tasks taken on by caregivers, and information and resource needs in caregivers. Qualitative interview and survey themes guided the content of the Healthy Eating and Recovery Together (HEART) app with an intake tracker and sections for nutrition recovery support, other competing caregiving tips, peer support, and caregiver self-care.

Conclusions: Results pinpointed optimal content for a mobile app for caregivers of individuals with HNC and support the acceptability of implementing the HEART app following HNC treatment.

(JMIR Cancer 2025;11:e66471) doi:10.2196/66471

KEYWORDS

head and neck cancer; cancer survivorship; caregiving; nutrition; mobile health; app development; mixed methods

Introduction

Approximately 522,846 people were living with oral cavity, pharynx, and larynx cancers in the United States in 2021 [1,2]. With increasing numbers of survivors with head and neck cancer (HNC), it is imperative that survivorship concerns are addressed [3,4]. Survivors with HNC face extremely difficult treatment consequences that impair their nutritional well-being [5]. Specifically, HNC and its multimodal treatments [6] can result in oral health problems related to swallowing, speech, mucositis, and dry mouth [7] that disrupt nutritional intake during and after treatment. Most survivors with HNC experience weight loss during treatment [8-11]. Population-level analyses using Surveillance, Epidemiology, and End Results data have estimated that approximately half of patients with HNC need a feeding tube and 40% - 45% of survivors with HNC experience dysphagia-related morbidities up to 2 years following treatment [12]. These nutritional recovery challenges impair the quality of life substantially [13-17].

Nutritional challenges in survivors with HNC also impact their caregivers [18,19], family members, and friends who provide cancer-related support. Relative to caregivers for people with other cancer types, caregivers of people with HNC confront unique support tasks such as feeding tube management, meal preparation, and speech support [6,19-22]. Caregivers report feeling unprepared for their roles in nutritional caregiving, sometimes experiencing a disconnect between survivors' goals and their own, and experiencing significant unmet needs as caregivers [23-29]. While interventions for caregivers caring for a loved one with cancer have been designed and tested [30-33], few evidence-based interventions are available to support caregivers of people with HNC with a focus on nutritional caregiving tasks in the early posttreatment period. Challenges in providing high-quality comprehensive support to caregivers of people with cancer include cost (financial and time) and competing demands while caring for a loved one [34-36], complexity of survivors' needs [20], and high prevalence and extent of emotional concerns among both survivors and caregivers [37,38].

Digital health strategies may overcome some of these barriers and offer a promising way to reach and support caregivers of survivors with HNC, particularly during the transition from cancer care to home, a critical point at which in-person interventions may not be feasible [39]. Research is growing on the feasibility, acceptability, and efficacy of digital health interventions for caregivers of people with cancer, with encouraging results for interventions designed to decrease burden and improve mood [39-45]. To guide the development of a mobile app to support nutrition-related caregiving among caregivers of people with HNC, this mixed methods study characterized nutritional challenges and caregiving tasks and intervention preferences in HNC oncology dietitians and survivor-caregiver dyads.

Methods

Study Design

Using a concurrent parallel mixed methods design [46], this study included a cross-sectional, web-based survey of a national panel of oncology dietitians and semistructured interviews with survivors with HNC and caregivers. We selected a mixed methods approach to facilitate gathering the perspectives of three groups, including dietitians, survivors with HNC, and caregivers [46]. A mixed methods approach also allowed for a more comprehensive understanding of the unmet needs of caregivers of people with HNC to drive the selection of content in a supportive care app to address those needs. Our team included researchers with expertise and training in HNC, cancer survivorship, oncology caregiving, and mobile health. We used a team approach to reflexivity with discussion and attention to the potential of the research team's background to influence research findings.

The purpose of this mixed methods study was to characterize nutritional challenges and caregiving tasks and intervention preferences in HNC oncology dietitians (quantitative) and survivor-caregiver dyads (qualitative and quantitative). We concurrently collected survey data from dietitians. Both quantitative and qualitative data were considered of equal priority and were analyzed separately and then integrated using the merging technique [47] as described below. Data were collected between April and September 2018. The GRAMMS (Good Reporting of A Mixed Methods Study) checklist was used to guide the mixed methods approach and reporting [48] (Checklist 1). Multimedia Appendix 1 provides the interview guide and surveys.

Dietitian Surveys

A convenience sample of dietitians was recruited by posting a study notice on the listserv for the Oncology Nutrition Dietetic Practice Group of the Academy of Nutrition and Dietetics. Dietitians were eligible for the 15-minute web-based survey if they reported providing care for patients with HNC in the past 6 months; the survey was hosted on REDCap (Research Electronic Data Capture; Vanderbilt University). The 67-item survey was developed by the study team. All items were optional, and participants could return to the survey over time if requested. The survey assessed demographic (race, ethnicity, sex, and age) and practice (credentials, years practicing as a dietitian, practice setting, and number of patients with HNC seen per week) characteristics. In addition, perceptions of the importance of posttreatment concerns in caregiving (0=not at all important to 4=extremely important) and perceived importance and difficulty (0=not important or difficult at all to 4=extremely important or difficult) of a variety of support tasks for caregivers of people with HNC, guided by the transactional model of caregiving were assessed (eg, tracking nutritional intake, changes and patterns in symptoms, and making care decisions [49]). Other measures included ratings of nutritional support resource needs (0=not at all to 4=extremely) in caregivers of people with HNC (eg, screening process to identify caregiver nutritional concerns, assessment tool to identify caregiver distress, and educational materials) and barriers to

addressing support needs (1=not a barrier at all to 4=major barrier) in caregivers of people with HNC (eg, time, caregiver interest, lack of evidence about the value of caregiver programs, and leadership). Participants then reviewed example app screens and then answered questions about the preferred focus of app content (eg, increasing caregivers' awareness of the importance of addressing nutritional challenges, changing caregivers' attitudes about improving nutritional status, and encouraging help-seeking for nutritional support) using an adapted version of the app-specific subscale in the Mobile App Rating Scale [50] (1=strongly disagree to 6=strongly agree) and an open-ended question. Finally, participants completed an open-ended question asking them to describe any specific suggestions they had for the development of a mobile support app for HNC caregivers.

Dyad Interviews

Survivors with HNC who completed treatment with curative intent 6 - 24 months prior to enrollment and were free of disease, and their caregivers, were recruited at the Medical University of South Carolina Hollings Cancer Center, with initial screening for eligibility by chart review. Inclusion criteria included being 18 years or older, completing treatment for stage I-IVA HNC (mucosal squamous cell carcinoma of the oral cavity, oropharynx, hypopharynx, and larynx), and experiencing nutritional challenges at the end of treatment as assessed by a 6-item screen. Survivors were excluded if they were unable to identify a primary caregiver, and survivors and caregivers were excluded if they either did not speak English or were cognitively impaired. After mailing a study letter and determining eligibility via phone screen, dyadic interviews were scheduled. Informed consent documents were signed, and dyadic interviews were conducted in person by 2 female investigators with training in qualitative methods (KS and MS). Interviewers did not know the participants and were not involved in their clinical care. Interviews were continued until we reached saturation in themes [51]. They were conducted in a private room, audio-recorded, and lasted approximately 45 minutes. Field notes were taken to provide interview observations.

We developed a semistructured interview guide using a cancer survivorship quality of life framework [52] to examine participants' reflections on the physical, emotional, and social challenges they experienced at the end of treatment and in the 6 months following treatment conclusion, with an emphasis on nutritional challenges and caregiving. Survivors and caregivers were asked to describe their emotional and physical well-being at the end of treatment. They were asked specifically about nutritional challenges, expectations they had about intake abilities, and the caregiver's role in nutritional recovery. Finally, they were asked about caregiver needs and suggestions for resources to better meet their needs at the end of treatment. Participants then viewed a set of example app screens (eg, nutrition tips, recipes, and stress management advice) on a tablet, after which they provided feedback on the content and format of a future app. After the interview, survivors and caregivers completed separate brief paper-based surveys assessing demographic (age, race, sex, and education), clinical (stage, treatment type, and nutritional status at the end of treatment), and technology access (home computer and smartphone) characteristics. They also completed ratings of the importance of caregiving tasks (0=not important to 4=extremely important) and ratings of agreement (1=strongly disagree to 6=strongly agree) about the benefits of checking in with caregivers after treatment, providing support messages to caregivers, and the importance of providing practical information to help with patients' nutritional recovery.

Ethical Considerations

Study procedures were approved by the Medical University of South Carolina Institutional Review Board (Pro00066211). A waiver of written informed consent was granted for dietitian surveys; after reading a study statement, participants advancing to the survey implied consent. Dyads completed written informed consent before completing interviews and received a copy of the signed consent for their records. All screening and survey data were stored in password-protected databases. The underlying databases were hosted in a secure data center. All data were identified only by code number (participant IDs). Dietitians completing surveys were entered into a lottery to receive a US $25 gift card to thank them for their time. Survivors with HNC and caregiver participants each received a US $50 gift card.

Statistical Analysis and App Development

Descriptive statistics were used to characterize survey data on dietitian and survivor-caregiver dyad demographic characteristics, perspectives about posttreatment, and caregiving challenges, and app preferences using R (R Core Team) [53]. Interviews were transcribed and analyzed using rigorous content analysis methods for systematic theme identification [54] in NVivo software (QSR International) [55]. Transcripts were coded by pairs of independent coders (KS, JO, and HK) and regrouped and reorganized until the investigators agreed on categories. This initial inductive theme identification process was followed by team meetings to finalize themes and guide implications for the intervention design. We sought trustworthiness in the qualitative data analysis approach by including prolonged engagement with the data, triangulating data, and using an audit trail to finalize themes [56]. Quantitative and qualitative data were analyzed separately, and then a data synthesis integration step was used to guide app development [57]. We selected the mixed methods integration approach of merging [47] and brought our quantitative and qualitative results together for elaboration. For example, the quantitative results (eg, perceptions and preferences of dietitians, survivors, and caregivers) were merged with key themes identified in interviews used to offer a more in-depth understanding of appropriate content for the app. Our interdisciplinary team of HNC clinicians, researchers, and developers completed a set of planning meetings to translate study results into app content by discussing the meaning of themes, identifying potential similarities and differences across themes, and mapping the themes to content using a consensus-based approach. The app development method was an agile approach, specifically rapid application development or rapid application building, which focuses on timely delivery in a fast-paced environment with the use of prototyping and iterative development [58]. The research team worked closely with the development team to review and

test evolving prototypes; a final prototype was pretested with 2 caregiver volunteers who were not involved in interviews or development activities.

Results

Participant Characteristics

Dietitians

All dietitians (n=116, 100%) were female, and the majority were registered dietitians (n=115, 99%) and White (n=110, 95%). Most practiced more frequently in the outpatient (n=107, 92%) versus inpatient (n=23, 20%) setting; options not mutually exclusive. Half had more than 10 years of experience working with patients with HNC, and most (n=87, 75%) cared for 1 - 10 patients with HNC per week.

HNC Dyads

A total of 50 survivors mailed a study letter. Several (n=15, 30%) were ineligible due to exclusion criteria, while others declined (n=20, 40%) due to scheduling conflicts prohibiting attendance at the in-person interview, lacking interest or reporting being too ill, or overwhelmed to participate. In total, 15 survivors enrolled in the study and nominated a caregiver. All survivors completed the in-person interview and 11 survivors were accompanied by their caregivers (8 spouses and 3 children). Scheduling conflicts precluded interview completion for 4 caregivers in person, 2 of whom were interviewed independently by phone (n=13 caregivers overall).

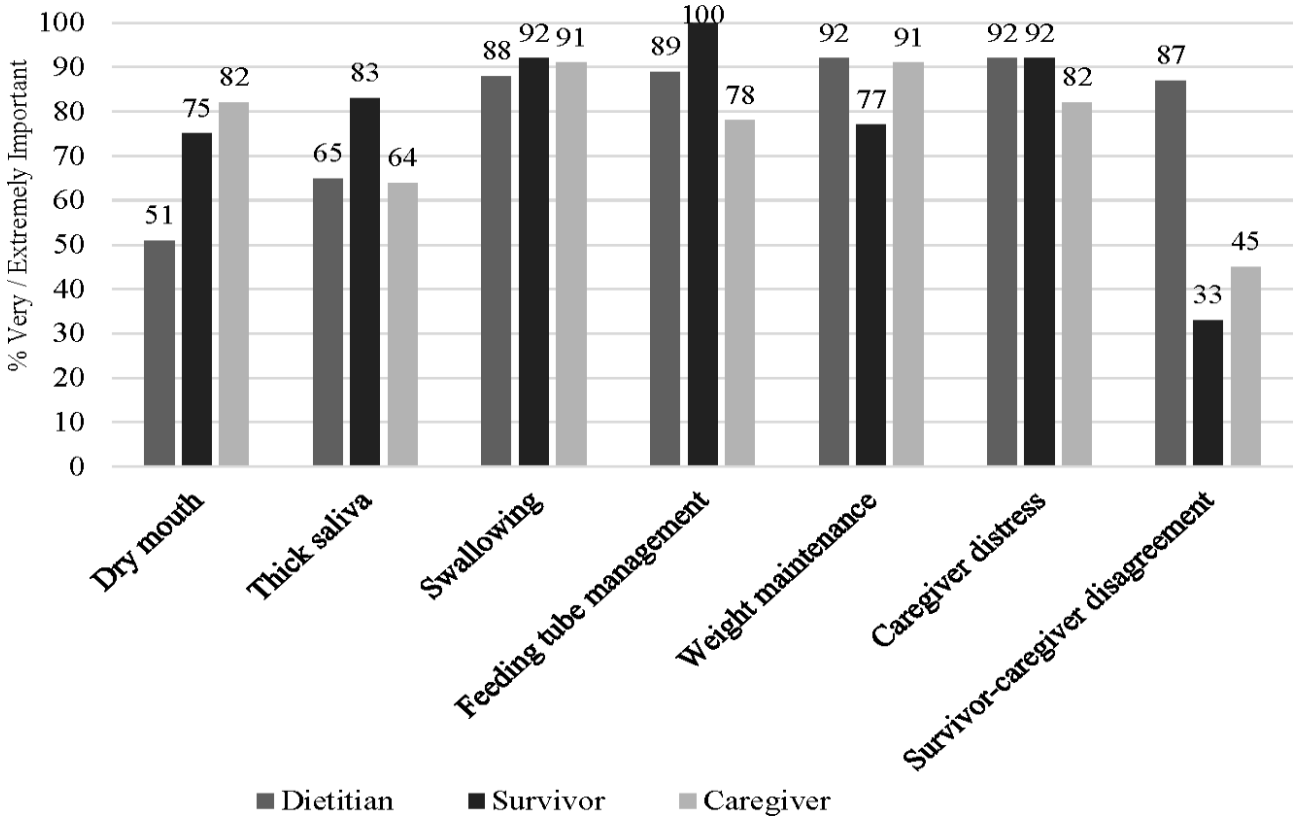
Most survivors (n=12, 80%) were male, while most caregivers (n=10, 77% spouse) were female (n=11, 85%). A total of 83% (n=19) of survivors and 87% (n=20) of caregivers were White, and age varied from 28 to 79 years (mean age 66, SD 15.1 for survivors and 61, SD 16.6 for caregivers). A total of 20% (n=3) of survivors and 27% (n=3) of caregivers had a high school or lower education. The most common cancer types included oropharynx (n=6, 40%) and oral cavity (n=3, 20%). Most survivors had surgery (n=14, 93%) and radiation (n=12, 80%); not mutually exclusive. One-third (n=5, 33%) had a liquid diet and approximately half (n=7, 47%) had a feeding tube at the end of treatment. Finally, most participants had a home computer (n=13, 87% survivors; n=13, 100% caregivers) but fewer survivors than caregivers had a smartphone (n=11, 73% vs n=10, 90%).

Posttreatment Challenges

Perceptions of HNC Nutritional Concerns in Caregiving

In the 6 months following treatment completion, dietitians, survivors, and caregivers perceived that the majority of nutritional concerns assessed were important to caregiving (Figure 1). The caregiving concerns rated most important (ie, very or extremely) by dietitians included weight maintenance (n=107, 92%), feeding tube management (n=103, 89%), caregiver distress about nutrition (n=103, 89%), and swallowing (n=102, 88%). The caregiving concerns rated as most important by survivors included feeding tube management (n=8, 100%), swallowing (n=14, 92%), and caregiver distress about nutrition (n=14, 92%). The caregiving concerns rated most important by caregivers themselves included weight maintenance (n=10, 91%), swallowing (n=10, 91%), caregiver distress about nutrition (n=9, 82%), and dry mouth (n=9, 82%).

Figure 1. Dietitian, survivor, and caregiver perceptions of the nutritional concerns in caregiving after treatment. This figure reports data from dietitian (n=116), survivor (n=15), and caregiver (n=11) surveys. Each caregiving concern endorsed at the very or extremely important level is graphed for dietitians, survivors, and caregivers.



Perceptions of HNC Caregiving Tasks

Most dietitians perceived all evaluated caregiving tasks to be very or extremely important to survivors' recovery. The highest-rated caregiving tasks included tracking nutritional concerns (n=113, 97%), working together as a team on nutritional concerns (n=104, 90%), and making care decisions (n=102, 88%). Relative to importance, most dietitians did not perceive that the measured caregiving tasks were overly difficult for caregivers to perform. The caregiving tasks most frequently endorsed as very or extremely difficult included navigating the health system (n=45, 39%), making care decisions (n=44, 38%), and interpreting changes in nutritional status (n=40, 34%).

HNC Recovery Challenges

Dyadic interviews identified 5 unique themes, which are described below with example quotes from survivors and caregivers shown in Textbox 1.

Textbox 1. Themes identified in interviews with survivors with head and neck cancer and their caregivers.

Nutritional challenges were expected but extremely frustrating with weight loss and disrupted intake caused by texture/taste, swallowing, dry mouth and sticky saliva concerns

Exemplary survivor quotes:

- Well, the issue for me is that, I mean I couldn't taste food and water tasted really weird. So I basically ate because I knew I was supposed to. Not because I enjoyed it.
- [My wife] said, “Baby, please eat, eat something.” Cause I mean, I'm doing my best...I'm trying, I'm trying, I'm trying, and I'm trying but I wish I could just eat something.
- Peg tube blockages [happened because] protein powder didn’t dissolve well, gurgling, irritation. Part of lining of my wound started sticking out and ached. After hospital discharge, I was given a nutrient solution from a medical supply company and then had diarrhea, night sweats, and projectile vomiting. Lost 26 pounds.

Exemplary caregiver quotes:

- I kept weighing him and you know in the back of your mind you know that losing weight is a sign of cancer and you really, in your mind you know it's gone, you think it's gone. But there's that worry that maybe it's not really gone and that's why his weight is still going down. So that really confused me about his weight.
- But what bothered me was that he won't eat or couldn't eat and he lost a lot of weight. He looked terrible, and he didn't have energy. I didn't know what to fix when we get home. I tried to do mostly liquid stuff but there's only so much liquid stuff you can do. The thing is, they had pulled all of his molars before the surgery, because they figure if he had a radiation, it would have to have his teeth fit. Well, with no teeth, you can't chew.
- It did affect him quickly and he dropped the weight like a ton of bricks. It was a challenge to get anything down him and it was like I wanted to do things on his schedule but it was just really, really difficult. We tried this and we tried that and we tried you know, tried several different things.

Other competing symptoms were common and difficult to separate from nutritional challenges

Exemplary survivor quotes:

- Well, the pain had a lot to do with that. Your neck, shoulders, arms and hands are in lots of pain and discomfort.
- When I finished everything, I think the last two weeks for me were the worst, where the radiation had really taken on where it had in my jaw because I could probably open my mouth the width of my finger.
- I had two really bad sores in my mouth that didn't make it fun to eat anything. When I started again occasionally, I would get some irritation from eating crunchy foods.
- I couldn't raise my arms up above my shoulders… I think my taste was coming back and the physical therapy really, I mean, that really turned me around. I started feeling a lot better. They really rehabbed my shoulders.

Exemplary caregiver quotes:

- And the sores that was on his neck kind of bothered me. Some nights he couldn't sleep normally. You know, you just go to sleep but he couldn't.
- It was more helplessness than anything. There's not anything you can do for him. You can't force him to eat because the pain was there.
- His stomach was upset a lot. After the radiation was over for a period of time after the radiation was over. It was just miserable. And also he is having [GI] issues.
- But as far as some of the other issues as far as being nauseous, we had to deal with that. The doctor gave her medicines for that, another big one is constipation. Oh, that one is a tough one.

Emotional challenges related to nutrition and recovery were common

Exemplary survivor quotes:

- Oh, yeah, confused, I mean the cancer just takes the world, just takes a lot out of yourself and I wondered if I'm going back to work and how I'm going to feel at the end of the day, for how long I would be able to see my brand new granddaughter.
- I'm still at the point where anything - I'm still afraid…because you know I still do have a lump down there, I know every time I swallow my saliva I know it’s there and I can feel it and it's - that's one thing that's always going back of my mind.
- Emotionally, I was depressed especially after the radiation specifically. Physically, I couldn't swallow. I couldn't eat. It was the most egregious treatment I ever had my life and if I knew what the outcome was going to be I would have never done it.

Exemplary caregiver quotes:

- I was with her through her depression. I think, I was depressed along with her. Seeing my mother not being able to eat or any of that and on Sundays we have Sunday dinners at her house so, we went ooh, about a good two, three months without having Sunday dinner at mom's unless one of us had to go over and cook.
-

Depression lingered, it's always in the back of your mind - is it coming back? Or did they get it all? Or those kinds of things and then what's going to be next?

- At the end of treatment I was a basket case. I checked on him constantly when he slept.

Caregivers take on a diverse set of medical and support tasks

Exemplary survivor quotes:

- Well my wife made up a chart and we had everything, I want to say laminated, definitely a plastic cover over it which was a good thing. We had it all laid out there [to monitor intake].
- But I mean-- patience. Because we're going to fight back, not meaning to be mean. But we're hurting. We're just trying to get back—we're trying to get better. I would just say, frustration. That'd be a better word. Because it is frustrating. And you seem to take it out on the ones you love the most.
- I know I relied on them a lot to get me and to help me get to the tub, to get me out of the tub. At that time, it wasn't easy. If I was this big, they wouldn't have been-- it would take a couple of them.

Exemplary caregiver quotes:

- I went to the grocery store and I just went up and down the aisle trying to find something that he could eat.
- You wouldn't believe what I've learned to do from the last year. Changing IV's. Trained to do those things. There's nothing I wouldn't do for him anyway, but I was nervous I would maybe do something wrong and harm him accidentally.
- Knowing that he's like a glass half empty kind of guy, I tried to keep everything - I didn't... not that I didn't think about it, but I just kept thinking positive. That they got it all.
- You need someone who have their best interest at heart. There's times that she's not in her right state of mind with everything that is going on so, we go to these doctor's appointments or we go to certain things, you have to have someone that is there that's going to ask the doctors the question.

Caregivers need information about what to expect with nutritional challenges and recovery

Exemplary survivor quotes:

- [My caregiver] was given a calorie intake document for a day and that was what we're trying to keep at, but there's no real specifics given us - as to how we should accomplish that.
- Yeah, I think seeing a nutritionist would have been a plus. That was never approached to us. Maybe while I was in the hospital. Given some kind of program to try to keep certain calories or what kind of foods and how to build up to.
- I think that we should have been counseled about … a feeding tube. There was no plan. There was no nutritional plan. So, full disclosure and a plan to support that disclosure would have been great. And early on in the planning – you need it in the treatment planning phase.

Exemplary caregiver quotes:

- It would be nice to have a list of do's and don'ts, a list of things that have worked for patients in the past like, you know. Puddings, Jell-O's, more common-sense and anything but having a list would be really nice when you're in a grocery store and going, okay, he's gonna need blank, blank, blank and blank.
- Well, I think they should give you a folder with all the instructions for about everything. Nutrition, food to eat, everything that needs to be done on our way. Or even we didn't know you had to have this port flushed.
- Okay, I think what could be helpful...it seems to me, if somebody can be pre-briefed as to what difficulties they may have with the-- like the chewing aspects and the saliva aspects, and what things that people have found in general that might be worth avoiding, that would be a helpful thing.

Theme 1: Nutritional Challenges Were Expected but Extremely Frustrating With Weight Loss and Disrupted Intake Caused by Texture or Taste, Swallowing, Dry Mouth, and Sticky Saliva Concerns

Both survivors and caregivers reported an array of nutritional challenges at the end of treatment that required special diets and feeding tubes. With challenges in swallowing, texture, and taste during recovery, dyads reported difficulty finding diets that were satisfying and tolerable, and often caregivers experienced distress in encouraging their loved ones to eat. Shopping, cooking, and communicating and negotiating with one's loved one around intake were also commonly reported challenges. Weight loss was reported by all survivors and caused significant distress, particularly in caregivers. The routine around feeding was reported to be very tedious.

Theme 2: Other Competing Symptoms Were Difficult to Separate From Nutritional Challenges

Survivors experienced many symptoms in addition to and related to nutritional concerns. For example, participants reported fatigue, mobility challenges, nausea, sores, and pain, and recovery challenges that often exacerbated nutritional and caregiving concerns. The array of symptoms caused frustration in survivors and worry in caregivers.

Theme 3: Emotional Challenges Related to Nutrition and Recovery Were Common

The emotional challenges faced by survivors and caregivers included frustration, fear, uncertainty, confusion, and depression. They reported frustration associated with eating challenges and the persistent focus on feeding and symptom management. Some participants reported worry and fear about survivors not regaining functional abilities. Relatedly, confusion and uncertainty about the future were commonly described by both survivors and caregivers. Caregivers expressed fear about weight loss and choking, as well as their ability to care for their loved ones. Both survivors and caregivers reported depression and other psychosocial concerns, and both also emphasized the importance of caregiver well-being.

Theme 4: Caregivers Take on a Diverse Set of Medical and Support Tasks After Treatment Completion

Caregivers focused on practical, nutritional, and emotional support tasks. Common practical tasks included providing transportation to and attending health care appointments, helping a loved one get around, and managing medications and stoma and IV care. Common nutritional support tasks included monitoring weight loss and intake, grocery shopping, researching recipes, preparing meals, and caring for feeding tubes. Emotional support tasks included supporting survivors' frustration with recovery and nutritional challenges and trying to keep a positive attitude. It was common for caregivers to report distress about support challenges, and they sometimes faced resistance from their loved ones around eating. Caregivers described a dynamic process of being persistent, creative, and patient.

Theme 5: Caregivers Need Information and Resources About What to Expect and How to Cope With Nutritional Challenges

Caregivers described feeling unprepared to support nutritional recovery and said they would have benefited from additional resources and support at the end of treatment. They emphasized the importance of early education during the treatment planning process to provide a better understanding of what to expect, resources and tools to support food preparation, and tips to help monitor intake. Survivors and caregivers both highlighted interest in meeting with a dietitian, yet also raised concerns about information overload.

Caregiver Needs and App Recommendations to Support Caregivers

Dietitians endorsed caregivers' time constraints (100/116, 86%) and caregivers being overwhelmed (102/116, 88%) as major barriers to meeting caregivers' needs. They also endorsed oncology clinics lacking designated staff to coordinate caregiver resources (70/116, 60%), higher priority care issues (n=72/116, 62%), and clinical team time constraints (n=70/116, 60%) as barriers. In light of unmet needs and to better support caregivers' provision of quality nutritional support to survivors with HNC, dietitians rated the importance of a variety of services and resources. While all 8 proposed services were rated by the majority (≥60%) of dietitians as very or extremely important, the highest ranked services or resources included a clinic referral process to link caregivers to appropriate nutritional resources (104/116, 90%), educational materials about diet (92/116, 79%), one-on-one counseling about nutrition (103/116, 89%), and training in nutritional support and symptom management (90/116, 78%).

Dietitians' responses to an open-ended question after reviewing example app screens yielded suggestions for new app content, including recipe recommendations, intake tracker, encouragement about caregiving tasks, and support for the caregiver's own well-being. Exemplary quotes include:

> *Caregivers want recipes! Tips for sore mouths, mucositis and dry mouth ... trouble-shooting enteral tube issues, constipation tips.*
>
> *Specific tips on adding high-calorie foods for weight maintenance (low in acid, soft/liquid). Same for high protein foods for tissue healing and muscle mass maintenance/recovery. Tips for frequent eating until appetite improves or side effects diminish (nausea, early satiety).*
>
> *Having exercises listed that can maintain muscle strength would be great. Also, a tracking device for the number of tube feedings completed and fluid intake would be helpful.*
>
> *It is important for the caregiver to have one set number to call for info. So many times, they try to get info from the internet which isn't always helpful.*
>
> *Support for caregivers themselves knowing they are not alone and such a big part of success moving forward.*

Survivors and caregivers' survey responses indicated that dyads were in strong agreement that checking in with caregivers after survivors complete treatment (12/15, 80% and 9/11, 82%, respectively) and providing support messages to caregivers (11/15, 73%, and 10/11, 91%, respectively) would be helpful. In addition, dyads were in strong agreement that it is important to provide practical information to caregivers to help survivors' nutritional recovery (15/15, 100%, and 10/11, 91%, respectively).

Feedback from dyads' responses to open-ended questions after reviewing example app screens included keeping the content simple, providing tips about what foods to avoid, providing recipes and dynamic nutrition information as needs change, including tips from other survivors with HNC and caregivers, and emphasizing support for caregiver well-being and self-care. Examples of exemplary quotes include:

> *This is like your personal resource right here at your fingertips - you're not alone. Let them know you're here if they need you. Simplicity is important.* [Caregiver]
>
> *Provide more advice about foods you can eat and problems with specific types of food groups…practical information with step-by-step directions. Food suggestions based on symptoms.* [Survivor]

Prepare caregivers for the possibility that the patient may not like the same dishes caregivers prepared before and not to take this personally. [Caregiver]

Encourage caregivers to ask for support from others: Doesn't mean you're less of a person [if you ask for help]. You need all the help you can get and a lot of the time you don't want to ask for it. [Caregiver]

It is important to ask about how the caregiver is doing; I feel guilty for not asking how she was doing. In general, we are in the dark; anything that brings some light into the room is helpful. [Survivor]

Data Synthesis: App Development

Building on results from all surveys and interviews underscoring the high need and interest in a comprehensive caregiver app, we designed the Healthy Eating and Recovery Together (HEART) app. The overall goal of the app was to support caregivers of survivors with HNC as they transition to the home setting after completing treatment and decrease their unmet needs and caregiver burden. The study investigators evaluated qualitative themes from dyad interviews side by side with quantitative themes observed in dietitian and dyad surveys. Team meetings were used to identify similarities in themes that were translated to high-priority main menu domains and subdomains for the app. The integration of survey and qualitative data resulted in an expansion of the findings, as the qualitative themes provided a detailed understanding of quantitative findings about caregivers' unmet needs and responsibilities as caregivers. The research team worked closely with the development team to review and test evolving prototypes, and a final prototype was pretested with 2 caregivers. The app was updated with feedback over the course of these steps (eg, we updated icons, modified menu choices, added instructions to components, reordered messages, and modified color choices). HEART includes educational information, caregiving tips and encouragement, and resources, with 4 elements, including survivor nutritional support, intake tracker, caregiver toolkit, and support videos (Table 1 provides more detailed descriptions and sample screenshots). The app also provides caregivers with notifications twice a week to check in with them about their concerns and deliver real-time resources mapped to current concerns. Two main areas in the app's content included a focus on survivor-caregiver teamwork and support of caregivers' own well-being.

Table . Content of the HEART[a] app.

App section	Content	Results guiding content	Screenshot
Nutritional support	Tips and encouragement for supporting a loved one with nutritional intake	A need for coverage of a broad array of topics to support the dynamic nutritional recovery process with content on common issues, recipes, and oral care support.	
Intake trackers	Supports tracking of feeding tube, liquid diet, and solid food in real time to monitor quantity, tolerability, and preferences, with the ability to share the intake journal with others.	Caregivers experience distress in monitoring intake and need a simple, convenient way to support monitoring.	
Caregiver toolkit	Emphasis on caregiving tasks and how caregivers can take care of their own physical and emotional well-being with tips and relaxation techniques	Caregivers feel unprepared for their roles as caregivers and face significant burdens, often overlooking their own well-being.	
Support videos	Survivor, caregiver, and clinician videos to support nutritional recovery and well-being	Interest in social support and increased interaction with dietitians and other survivors and caregivers.	
My resources	Stores biweekly prompts and tailored resources	Participants desire real-time connections and resources that are dynamically matched to their changing needs during recovery and caregiving.	

[a]HEART: Healthy Eating and Recovery Together.

Discussion

Principal Results

Survivors with HNC and their caregivers face exceptionally difficult posttreatment concerns that negatively impact their quality of life [28,59-61]. Caregivers are often tasked with addressing survivors' nutritional challenges in the home setting, yet many are unprepared for these caregiving tasks [20,24]. Getting adequate nutritional intake, maintaining a healthy weight, and managing physical symptoms and emotional distress are imperative for survivors of HNC during the posttreatment period; yet, there is a paucity of tools available to support survivors and their caregivers in meeting these goals [31,32]. This study addressed this important survivorship care gap, specifically the availability and accessibility of high-quality interventions to support nutritional caregiving. Recognizing the

value of stakeholder input in intervention development [62-64], we used mixed methods to assess the perspectives of survivors, and then also used responses from caregivers and oncology dietitians to guide the development of an app to support caregivers at the end of treatment. Previous studies have emphasized the acceptability of apps for cancer caregivers while also calling for more formative research to ensure their suitability [43,44,65,66]. While we intended to focus this tool on nutritional caregiving support, study results led to a more comprehensive product focused on nutrition and other survivor recovery concerns, plus caregiver self-care. The final HEART app includes nutrition support with an intake tracker, along with tips and encouragement for other caregiving areas (daily support, emotional support, and medical support), support videos from peers and clinicians, and a caregiver toolkit (taking care of yourself and relaxation exercises).

To guide app development, dietitians provided their perspectives on caring for and supporting HNC dyads. They endorsed a broad array of important nutritional concerns for HNC caregiving, including swallowing, feeding tube management, weight maintenance, and caregiver distress about nutrition. They also perceived that there were multiple nutritional care tasks that were important to survivors' recovery, including monitoring nutrition, making care decisions, and working together with their survivors as a team to manage nutritional concerns. While most dietitians did not rate these care tasks as extremely difficult for caregivers to manage, the number of tasks endorsed was substantial, and the importance to survivors' recovery was rated highly, indicating that an app intervention would likely need to be complex and cover a comprehensive set of nutritional concerns. These results suggest that an app should help caregivers be flexible and skilled in multiple nutritional care tasks. Additionally, dietitians' perceptions regarding potential survivor-caregiver mismatch on nutritional goals were consistent with previous studies [18,26], which may suggest that a focus on teamwork in dyads would be beneficial in an app [67,68].

HNC survivor-caregiver dyads in this study confirmed the challenges reported by dietitians in both surveys and interviews. First, survey results highlighted similar nutritional caregiving concerns after treatment in survivors, caregivers, and dietitians. Second, interviews confirmed the types of nutritional challenges experienced and the emotional toll they take on dyads, consistent with previous studies [5,6,69,70]. Results also highlighted that nutritional caregiving tasks were compounded by other competing HNC recovery concerns (eg, support for pain, mobility, pain, fatigue, and emotional challenges); it was difficult for caregivers to focus on nutrition without considering these additional concerns. As highlighted in previous research, caregivers take on a multitude of caregiving tasks and need resources matched to these responsibilities [14,19,20], again supporting the coverage of a broad array of HNC caregiving [20] content in an app.

Dietitians also endorsed high barriers to meeting caregiver needs, including clinician time constraints and a lack of designated staff to coordinate resources. While increasing research has prioritized interventions to meet caregiver needs [31,71], cancer care settings are not adequately resourced to address their needs [72,73]. Technology tools such as the HEART app may be a promising approach to address these barriers and complement other interventions to support caregivers [39,74]. Dietitians in this study endorsed a broad variety of strategies and content to include in an app to support caregivers. Dietitians tended to prefer high-resource strategies for inclusion in an app, such as intake tracking, training, screening processes, educational resources, and counseling. Survivors and caregivers confirmed interest in an app for caregivers with check-ins, support, and practical tips. Both dietitians and dyads recommended the provision of recipes and support for caregivers' own well-being. Dietitians uniquely recommended an intake tracker, while dyads uniquely recommended tips from peers.

While mobile health intervention development and testing to support caregivers of people with cancer is growing and shows evidence of promising acceptability, adherence, and some improvements in short-term outcomes such as caregiver burden [41-43,45,65,75,76], more research is needed in this area to better understand caregiver adoption and engagement in these interventions, impacts on caregiver psychosocial and health outcomes, and best practices for disseminating such intervention in practice using rigorous methods. Of key importance, few studies have focused on HNC caregiving and nutritional support. Apps to support caregivers of people with HNC also have the potential for supporting and addressing HNC dietitian time constraints by supporting their recommendations outside of the clinic; more research is needed in this area.

Strengths and Limitations

The strengths of this study include its mixed methods approach with data collection from multiple perspectives, including those who had experienced HNC recovery, served as an HNC caregiver after treatment, and provided clinical care for HNC survivor-caregiver dyads. The integration of survey data and qualitative data allowed for the examination of similarities and unique findings across qualitative and quantitative themes, ultimately allowing a deeper understanding of perspectives to guide the HEART app. Dietitians were recruited from across the United States, and qualitative interviews with dyads were used to supplement surveys and provide a more in-depth understanding of survey results. The app's focus on nutritional support, particularly after the end of treatment, is innovative and would help address an area of major concern for HNC providers, survivors, and caregivers. In the context of changing digital health use patterns over time, it is important to note that an important limitation of this study is that the data were collected in 2018. However, it is notable that digital health use and engagement rates have increased over time, and this expanded reach and growing societal acceptance of these tools are therefore encouraging [77]. Other limitations of this study include recruitment of dyads from only one medical center, recruitment of a convenience sample of dieticians, and a modest sample size for interviews, all of which limit the transferability of findings. In addition, we experienced a lack of diversity in clinical and sociodemographic factors for surveys and interviews. We selected a parallel convergent approach for data collection and analysis and followed with an integration approach (merging) to synthesize our mixed methods data; while a sequential approach may have allowed more iterative app

development and testing, this approach was selected to facilitate rapid technology development.

Conclusions

In summary, this study identified the optimal content for a mobile support app for caregivers of people with HNC and supported the acceptability of implementing this intervention at the end of treatment. Future steps include evaluating the implementation of the HEART app and its impact on survivor and caregiver outcomes. It will be important for a future study to rigorously test the HEART app in a prospective clinical trial.

Acknowledgments

This work was supported by the National Cancer Institute at the National Institutes of Health (R21CA215557), the Biostatistics Shared Resource, Hollings Cancer Center, Medical University of South Carolina (P30 CA138313), and the National Center for Advancing Translational Sciences (UL1 TR000062).

Authors' Contributions

KS and KR conceptualized and designed the study. MS supported the recruitment and data collection study activities. All authors supported the data analysis, interpretation, and writing of the paper. All authors approved the final product.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study Interview Guide and Suveys.
[PDF File, 558 KB - cancer_v11i1e66471_app1.pdf]

Checklist 1
GRAMMS (Good Reporting of A Mixed Methods Study) checklist.
[DOCX File, 13 KB - cancer_v11i1e66471_app2.docx]

References

1. Cancer stat factors: oral cavity and pharynx cancer. National Cancer Institute Surveillance, Epidemiology and End Results Program. 2021. URL: http://seer.cancer.gov/statfacts/html/oralcav.html [accessed 2024-09-13]
2. Cancer stat factors: larynx cancer. National Cancer Institute Surveillance, Epidemiology and End Results Program. URL: http://seer.cancer.gov/statfacts/html/laryn.html [accessed 2024-09-13]
3. Cohen EEW, LaMonte SJ, Erb NL, et al. American Cancer Society head and neck cancer survivorship care guideline. CA Cancer J Clin 2016 May;66(3):203-239. [doi: 10.3322/caac.21343] [Medline: 27002678]
4. Goyal N, Day A, Epstein J, et al. Head and neck cancer survivorship consensus statement from the American Head and Neck Society. Laryngoscope Investig Otolaryngol 2022 Feb;7(1):70-92. [doi: 10.1002/lio2.702] [Medline: 35155786]
5. Kuhn MA, Gillespie MB, Ishman SL, et al. Expert consensus statement: management of dysphagia in head and neck cancer patients. Otolaryngol Head Neck Surg 2023 Apr;168(4):571-592. [doi: 10.1002/ohn.302] [Medline: 36965195]
6. National Comprehensive Cancer Network practice guidelines in oncology: head and neck cancers. NCCN. URL: http://www.nccn.org/professionals/physician_gls/PDF/head-and-neck.pdf [accessed 2024-01-02]
7. Jensen SB, Pedersen AML, Vissink A, et al. A systematic review of salivary gland hypofunction and xerostomia induced by cancer therapies: prevalence, severity and impact on quality of life. Support Care Cancer 2010 Aug;18(8):1039-1060. [doi: 10.1007/s00520-010-0827-8]
8. Jager-Wittenaar H, Dijkstra PU, Vissink A, van der Laan B, van Oort RP, Roodenburg JLN. Critical weight loss in head and neck cancer—prevalence and risk factors at diagnosis: an explorative study. Support Care Cancer 2007 Sep;15(9):1045-1050. [doi: 10.1007/s00520-006-0212-9] [Medline: 17277925]
9. Jin S, Lu Q, Sun Y, et al. Nutrition impact symptoms and weight loss in head and neck cancer during radiotherapy: a longitudinal study. BMJ Support Palliat Care 2021 Mar;11(1):17-24. [doi: 10.1136/bmjspcare-2019-002077] [Medline: 32019753]
10. Langius JAE, Twisk J, Kampman M, et al. Prediction model to predict critical weight loss in patients with head and neck cancer during (chemo)radiotherapy. Oral Oncol 2016 Jan;52:91-96. [doi: 10.1016/j.oraloncology.2015.10.021] [Medline: 26564309]
11. Singh GK, Patil VM, Noronha V, et al. Weight loss and its impact on outcome in head and cancer patients during chemo-radiation. Oral Oncol 2021 Nov;122:105522. [doi: 10.1016/j.oraloncology.2021.105522] [Medline: 34571463]

12. Hutcheson KA, Nurgalieva Z, Zhao H, et al. Two-year prevalence of dysphagia and related outcomes in head and neck cancer survivors: an updated SEER-Medicare analysis. Head Neck 2019 Feb;41(2):479-487. [doi: 10.1002/hed.25412] [Medline: 30536748]
13. Barrios R, Tsakos G, García-Medina B, Martínez-Lara I, Bravo M. Oral health-related quality of life and malnutrition in patients treated for oral cancer. Support Care Cancer 2014 Nov;22(11):2927-2933. [doi: 10.1007/s00520-014-2281-5] [Medline: 24848576]
14. Bressan V, Bagnasco A, Aleo G, et al. The life experience of nutrition impact symptoms during treatment for head and neck cancer patients: a systematic review and meta-synthesis. Support Care Cancer 2017 May;25(5):1699-1712. [doi: 10.1007/s00520-017-3618-7] [Medline: 28204992]
15. Crowder SL, Douglas KG, Yanina Pepino M, Sarma KP, Arthur AE. Nutrition impact symptoms and associated outcomes in post-chemoradiotherapy head and neck cancer survivors: a systematic review. J Cancer Surviv 2018 Aug;12(4):479-494. [doi: 10.1007/s11764-018-0687-7] [Medline: 29556926]
16. Langius JAE, van Dijk AM, Doornaert P, et al. More than 10% weight loss in head and neck cancer patients during radiotherapy is independently associated with deterioration in quality of life. Nutr Cancer 2013;65(1):76-83. [doi: 10.1080/01635581.2013.741749] [Medline: 23368916]
17. Yuwanati M, Gondivkar S, Sarode SC, et al. Oral health-related quality of life in oral cancer patients: systematic review and meta-analysis. Future Oncol 2021 Mar;17(8):979-990. [doi: 10.2217/fon-2020-0881] [Medline: 33541115]
18. Mayre-Chilton KM, Talwar BP, Goff LM. Different experiences and perspectives between head and neck cancer patients and their care-givers on their daily impact of a gastrostomy tube. J Hum Nutr Diet 2011 Oct;24(5):449-459. [doi: 10.1111/j.1365-277X.2011.01165.x] [Medline: 21649745]
19. Penner JL, McClement S, Lobchuk M, Daeninck P. Family members' experiences caring for patients with advanced head and neck cancer receiving tube feeding: a descriptive phenomenological study. J Pain Symptom Manage 2012 Oct;44(4):563-571. [doi: 10.1016/j.jpainsymman.2011.10.016] [Medline: 22699088]
20. Bond SM, Schumacher K, Sherrod A, et al. Development of the head and neck cancer caregiving task inventory. Eur J Oncol Nurs 2016 Oct;24:29-38. [doi: 10.1016/j.ejon.2016.08.004]
21. Castellanos EH, Dietrich MS, Bond SM, et al. Impact of patient symptoms and caregiving tasks on psychological distress in caregivers for head and neck cancer (HNC). Psychooncology 2019 Mar;28(3):511-517. [doi: 10.1002/pon.4968] [Medline: 30578599]
22. Sterba KR, Zapka J, Armeson KE, et al. Physical and emotional well-being and support in newly diagnosed head and neck cancer patient-caregiver dyads. J Psychosoc Oncol 2017;35(6):646-665. [doi: 10.1080/07347332.2017.1323817] [Medline: 28459393]
23. Nund RL, Scarinci NA, Cartmill B, Ward EC, Kuipers P, Porceddu SV. Third-party disability in carers of people with dysphagia following non-surgical management for head and neck cancer. Disabil Rehabil 2016;38(5):462-471. [doi: 10.3109/09638288.2015.1046563] [Medline: 25981671]
24. Nund RL, Ward EC, Scarinci NA, Cartmill B, Kuipers P, Porceddu SV. Carers' experiences of dysphagia in people treated for head and neck cancer: a qualitative study. Dysphagia 2014 Aug;29(4):450-458. [doi: 10.1007/s00455-014-9527-8]
25. Hiatt J, Young A, Brown T, Banks M, Segon B, Bauer J. A qualitative comparison of the nutrition care experiences of carers supporting patients with head and neck cancer throughout surgery and radiation treatment and survivorship. Support Care Cancer 2022 Nov;30(11):9359-9368. [doi: 10.1007/s00520-022-07348-0]
26. Hiatt JS, Young A, Brown TE, Banks M, Bauer J. Patient and carer experience of nutrition care throughout and beyond treatment for head and neck cancer: a qualitative longitudinal study. Support Care Cancer 2022 Jan;30(1):813-824. [doi: 10.1007/s00520-021-06484-3]
27. Giuliani M, Milne R, McQuestion M, et al. Partner's survivorship care needs: an analysis in head and neck cancer patients. Oral Oncol 2017 Aug;71:113-121. [doi: 10.1016/j.oraloncology.2017.06.011]
28. Balfe M, O'Brien K, Timmons A, et al. The unmet supportive care needs of long-term head and neck cancer caregivers in the extended survivorship period. J Clin Nurs 2016 Jun;25(11-12):1576-1586. [doi: 10.1111/jocn.13140] [Medline: 27139171]
29. van Hof KS, Hoesseini A, Dorr MC, et al. Unmet supportive care needs among informal caregivers of patients with head and neck cancer in the first 2 years after diagnosis and treatment: a prospective cohort study. Support Care Cancer 2023 May;31(5). [doi: 10.1007/s00520-023-07670-1]
30. Chow R, Mathews JJ, Cheng EY, et al. Interventions to improve outcomes for caregivers of patients with advanced cancer: a meta-analysis. J Natl Cancer Inst 2023 Aug 8;115(8):896-908. [doi: 10.1093/jnci/djad075]
31. Ferrell B, Wittenberg E. A review of family caregiving intervention trials in oncology. CA Cancer J Clin 2017 Jul 8;67(4):318-325. [doi: 10.3322/caac.21396] [Medline: 28319263]
32. Secinti E, Fischer IC, Brennan EA, Christon L, Balliet W. The efficacy of psychosocial interventions for cancer caregiver burden: a systematic review and meta-analysis of randomized controlled trials. Clin Psychol Rev 2023 Feb;99:102237. [doi: 10.1016/j.cpr.2022.102237] [Medline: 36516641]
33. Ugalde A, Gaskin CJ, Rankin NM, et al. A systematic review of cancer caregiver interventions: appraising the potential for implementation of evidence into practice. Psychooncology 2019 Apr;28(4):687-701. [doi: 10.1002/pon.5018] [Medline: 30716183]

34. Shi Y, Gao Y, Huai Q, Du Z, Yang L. Experiences of financial toxicity among caregivers of cancer patients: a meta-synthesis of qualitative studies. Support Care Cancer 2024 Mar;32(3):146. [doi: 10.1007/s00520-024-08349-x]
35. Coumoundouros C, Brahim LO, Lambert SD, McCusker J. The direct and indirect financial costs of informal cancer care: a scoping review. Health Soc Care Community 2019 Sep;27(5):e622-e636. [doi: 10.1111/hsc.12808] [Medline: 31293013]
36. Reblin M, Ketcher D, Vadaparampil ST. Care for the cancer caregiver: a qualitative study of facilitators and barriers to caregiver integration and support. J Cancer Educ 2022 Dec;37(6):1634-1640. [doi: 10.1007/s13187-021-02001-6] [Medline: 33783762]
37. Gascon B, Panjwani AA, Mazzurco O, Li M. Screening for distress and health outcomes in head and neck cancer. Curr Oncol 2022 May 24;29(6):3793-3806. [doi: 10.3390/curroncol29060304] [Medline: 35735413]
38. Bedaso A, Dejenu G, Duko B. Depression among caregivers of cancer patients: updated systematic review and meta-analysis. Psychooncology 2022 Nov;31(11):1809-1820. [doi: 10.1002/pon.6045] [Medline: 36209385]
39. Kent EE, Rowland JH, Northouse L, et al. Caring for caregivers and patients: research and clinical priorities for informal cancer caregiving. Cancer 2016 Jul 1;122(13):1987-1995. [doi: 10.1002/cncr.29939] [Medline: 26991807]
40. Badr H, Carmack CL, Diefenbach MA. Psychosocial interventions for patients and caregivers in the age of new communication technologies: opportunities and challenges in cancer care. J Health Commun 2015;20(3):328-342. [doi: 10.1080/10810730.2014.965369] [Medline: 25629218]
41. DuBenske LL, Gustafson DH, Namkoong K, et al. CHESS improves cancer caregivers' burden and mood: results of an eHealth RCT. Health Psychol 2014 Oct;33(10):1261-1272. [doi: 10.1037/a0034216] [Medline: 24245838]
42. DuBenske LL, Gustafson DH, Shaw BR, Cleary JF. Web-based cancer communication and decision making systems: connecting patients, caregivers, and clinicians for improved health outcomes. Med Decis Making 2010;30(6):732-744. [doi: 10.1177/0272989X10386382] [Medline: 21041539]
43. Heynsbergh N, Heckel L, Botti M, Livingston PM. Feasibility, useability and acceptability of technology-based interventions for informal cancer carers: a systematic review. BMC Cancer 2018 Mar 2;18(1):244. [doi: 10.1186/s12885-018-4160-9] [Medline: 29499663]
44. Shin JY, Kang TI, Noll RB, Choi SW. Supporting caregivers of patients with cancer: a summary of technology-mediated interventions and future directions. Am Soc Clin Oncol Educ Book 2018 May 23;38:838-849. [doi: 10.1200/EDBK_201397] [Medline: 30231412]
45. Slev VN, Mistiaen P, Pasman HRW, Verdonck-de Leeuw IM, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: a meta-review. Int J Med Inform 2016 Mar;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]
46. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research, 3rd edition: SAGE; 2018.
47. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
48. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
49. Schumacher KL, Beidler SM, Beeber AS, Gambino P. A transactional model of cancer family caregiving skill. ANS Adv Nurs Sci 2006;29(3):271-286. [doi: 10.1097/00012272-200607000-00009] [Medline: 17139208]
50. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015;3(1):e27. [doi: 10.2196/mhealth.3422]
51. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907. [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
52. Ferrell BR, Hassey Dow K. Quality of life among long-term cancer survivors. Oncology (Williston Park) 1997 Apr;11(4):565-568. [Medline: 9130276]
53. The R project for statistical computing. R-project. URL: https://www.R-project.org [accessed 2025-04-01]
54. Corbin J, Strauss A. Basics of Qualitative Research: Grounded Theory Procedures and Techniques: Sage; 1990.
55. NVivo qualitative data analysis software. Lumivero. URL: https://lumivero.com/products/nvivo/ [accessed 2025-05-20]
56. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. Ann Fam Med 2015 Nov;13(6):554-561. [doi: 10.1370/afm.1865] [Medline: 26553895]
57. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017;16(1):1-13. [doi: 10.1177/1609406917733847]
58. Wilson K, Bell C, Wilson L, Witteman H. Agile research to complement agile development: a proposal for an mHealth research lifecycle. NPJ Digital Med 2018;1:46. [doi: 10.1038/s41746-018-0053-1] [Medline: 31304326]
59. Giuliani M, McQuestion M, Jones J, et al. Prevalence and nature of survivorship needs in patients with head and neck cancer. Head Neck 2016 Jul;38(7):1097-1103. [doi: 10.1002/hed.24411] [Medline: 26894614]
60. Miller MC, Shuman AG, Head A. Survivorship in head and neck cancer: a primer. JAMA Otolaryngol Head Neck Surg 2016 Oct 1;142(10):1002-1008. [doi: 10.1001/jamaoto.2016.1615] [Medline: 27414519]
61. Murphy BA, Deng J. Advances in supportive care for late effects of head and neck cancer. J Clin Oncol 2015 Oct 10;33(29):3314-3321. [doi: 10.1200/JCO.2015.61.3836] [Medline: 26351334]

62. Forsythe LP, Carman KL, Szydlowski V, et al. Patient engagement in research: early findings from the patient-centered outcomes research institute. Health Aff (Millwood) 2019 Mar;38(3):359-367. [doi: 10.1377/hlthaff.2018.05067] [Medline: 30830822]
63. Frank L, Forsythe L, Ellis L, et al. Conceptual and practical foundations of patient engagement in research at the patient-centered outcomes research institute. Qual Life Res 2015 May;24(5):1033-1041. [doi: 10.1007/s11136-014-0893-3] [Medline: 25560774]
64. Maurer M, Mangrum R, Hilliard-Boone T, et al. Understanding the influence and impact of stakeholder engagement in patient-centered outcomes research: a qualitative study. J Gen Intern Med 2022 Apr;37(Suppl 1):6-13. [doi: 10.1007/s11606-021-07104-w] [Medline: 35349017]
65. Li Y, Li J, Zhang Y, Ding Y, Hu X. The effectiveness of e-Health interventions on caregiver burden, depression, and quality of life in informal caregivers of patients with cancer: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2022 Mar;127:104179. [doi: 10.1016/j.ijnurstu.2022.104179]
66. Lorca-Cabrera J, Grau C, Martí-Arques R, Raigal-Aran L, Falcó-Pegueroles A, Albacar-Riobóo N. Effectiveness of health web-based and mobile app-based interventions designed to improve informal caregiver's well-being and quality of life: a systematic review. Int J Med Inform 2020 Feb;134:104003. [doi: 10.1016/j.ijmedinf.2019.104003] [Medline: 31790857]
67. Revenson TA, Kayser K, Bodenmann G. Couples Coping with Stress: Emerging Perspectives on Dyadic Coping: American Psychological Association; 2005.
68. Traa MJ, De Vries J, Bodenmann G, Den Oudsten BL. Dyadic coping and relationship functioning in couples coping with cancer: a systematic review. Br J Health Psychol 2015 Feb;20(1):85-114. [doi: 10.1111/bjhp.12094] [Medline: 24628822]
69. Elad S, Cheng KKF, Lalla RV, et al. MASCC/ISOO clinical practice guidelines for the management of mucositis secondary to cancer therapy. Cancer 2020 Oct 1;126(19):4423-4431. [doi: 10.1002/cncr.33100] [Medline: 32786044]
70. Lalla RV, Brennan MT, Gordon SM, Sonis ST, Rosenthal DI, Keefe DM. Oral mucositis due to high-dose chemotherapy and/or head and neck radiation therapy. J Natl Cancer Inst Monogr 2019 Aug 1;2019(53):lgz011. [doi: 10.1093/jncimonographs/lgz011] [Medline: 31425601]
71. Odom JN, Young HM, Sterba K, et al. Developing a national implementation strategy to accelerate uptake of evidence-based family caregiver support in U.S. cancer centers. Psychooncology 2024 Jan;33(1):e6221. [doi: 10.1002/pon.6221] [Medline: 37743780]
72. Nightingale CL, Snavely AC, McLouth LE, et al. Processes for identifying caregivers and screening for caregiver and patient distress in community oncology: results from WF-1803CD. J Natl Cancer Inst 2024 Feb 8;116(2):324-333. [doi: 10.1093/jnci/djad198] [Medline: 37738445]
73. Odom JN, Applebaum A, Bakitas MA, et al. Availability of family caregiver programs in US cancer centers. JAMA Netw Open 2023 Oct 2;6(10):e2337250. [doi: 10.1001/jamanetworkopen.2023.37250] [Medline: 37819661]
74. Clauser SB, Wagner EH, Aiello Bowles EJ, Tuzzio L, Greene SM. Improving modern cancer care through information technology. Am J Prev Med 2011 May;40(5 Suppl 2):S198-S207. [doi: 10.1016/j.amepre.2011.01.014] [Medline: 21521595]
75. Oakley-Girvan I, Yunis R, Fonda SJ, et al. A novel smartphone application for the informal caregivers of cancer patients: usability study. PLOS Digital Health 2023 Mar;2(3):e0000173. [doi: 10.1371/journal.pdig.0000173] [Medline: 36867639]
76. Yunis R, Fonda SJ, Aghaee S, et al. Mobile app activity engagement by cancer patients and their caregivers informs remote monitoring. Sci Rep 2024 Feb 9;14(1):3375. [doi: 10.1038/s41598-024-53373-w] [Medline: 38336943]
77. Portz JS, Moore S, Bull S. Evolutionary trends in the adoption, adaptation, and abandonment of mobile health technologies: viewpoint based on 25 Years of research. J Med Internet Res 2024;26:e62790. [doi: 10.2196/62790]

Abbreviations

GRAMMS: Good Reporting of A Mixed Methods Study
HEART: Healthy Eating and Recovery Together
HNC: head and neck cancer
REDCap: Research Electronic Data Capture

Edited by N Cahill; submitted 14.09.24; peer-reviewed by A Shi, R Pozzar; revised version received 11.04.25; accepted 14.04.25; published 10.06.25.

Please cite as:
Sterba K, Graboyes E, Burris J, Scallion M, Kinder H, Olsen J, Toll B, Armeson K, Day T, Chera B, Ruggiero K
Development of a Mobile App to Support Head and Neck Cancer Caregiving: Mixed Methods Study
JMIR Cancer 2025;11:e66471
URL: https://cancer.jmir.org/2025/1/e66471
doi:10.2196/66471

Supporting Medication Adherence in Pediatric Patients Undergoing Hematopoietic Stem Cell Transplant Using the BMT4me mHealth App: Mixed Methods Usability Study

Mariam Kochashvili[1], BA; Parishma Guttoo[1], MBBS; Emre Sezgin[1,2], PhD; Ahna Pai[1,2], PhD; Rajinder Bajwa[2,3], MD; Wendy Landier[4], PhD; Cynthia Gerhardt[1,2], PhD; Micah Skeens[1,2,3], PhD

[1]Center for Biobehavioral Health, The Abigail Wexner Research Institute, Nationwide Children's Hospital, 431 South 18th Street, Columbus, OH, United States

[2]The Ohio State University, Columbus, OH, United States

[3]Nationwide Children's Hospital, Columbus, OH, United States

[4]Division of Pediatric Hematology Oncology, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Micah Skeens, PhD
Center for Biobehavioral Health, The Abigail Wexner Research Institute, Nationwide Children's Hospital, 431 South 18th Street, Columbus, OH, United States

Abstract

Background: Due to multifaceted outpatient regimens, children receiving hematopoietic stem cell transplants (HCTs) are at high risk of medication nonadherence, leading to life-threatening complications. Mobile health (mHealth) interventions have proven effective in improving adherence in various pediatric conditions; however, adherence intervention literature on HCT is limited.

Objective: This study aimed to assess the usability of a mHealth intervention (BMT4me) designed to serve as a real-time, personalized tool for medication management or adherence, symptom tracking, and journal keeping.

Methods: Following a mixed methods approach, 14 caregivers (n=11, 79% female; n=10, 71% White) of children aged 2 - 18 (mean age 8.51, SD 5.18) years in the acute phase (first 100 d) post-HCT were recruited. Caregivers were asked to use the BMT4me app for 100 days or until weaning of the immunosuppressant medications to measure usability. The System Usability Scale (assessing functionality and acceptability), reaction cards (assessing desirability), caregiver satisfaction (assessing satisfaction) with the app, and semistructured interviews (assessing participant experience using the app and feedback regarding features) were conducted at two time points, at enrollment and study completion.

Results: The mean System Usability Scale score was 86.15 (SD 12.81) at enrollment and 73.13 (SD 16.13) at study completion, with most participants reporting the app easy to use and accepable during both time points. At enrollment, 80% (n=12) of caregivers reported that the app was effective in motivating them to stay on schedule, and 87% (n=13) indicated they would recommend it to others. At study completion, 75% (n=6) of caregivers found the app helpful for tracking their child's medication schedule, and 64% (n=5) would recommend it to others. Caregivers described the app as "accessible," "useful," and "valuable." Qualitative interviews during both time points revealed caregivers' positive reactions to the app, particularly regarding medication reminders, tracking symptoms, and notes features, while also providing suggestions for improvements, such as integrating the BMT4me app with electronic medical records, incorporating educational content, adding fields for recording vital signs, and important phone numbers.

Conclusions: The BMT4me app demonstrated promising usability as a mHealth intervention among pediatric patients undergoing HCT. Caregivers considered the app user-friendly and valuable, with positive feedback on its features, such as medication reminders and symptom tracking. Despite minor reported issues with app functionality, the overall acceptance of the app suggests its potential to support families in managing complex treatment. The findings from this study will inform the feasibility of testing in larger randomized controlled trials.

Trial Registration: ClinicalTrials.gov NCT04976933; https://clinicaltrials.gov/ct2/show/NCT04976933

International Registered Report Identifier (IRRID): RR2-10.2196/39098

(JMIR Cancer 2025;11:e66847) doi:10.2196/66847

KEYWORDS

mHealth; pediatric transplant; digital health; medication adherence; usability; hematopoietic stem cell transplant; bone marrow transplant; pediatric; children; hematopoietic stem cell; HSC; smartphone; mobile health; BMT4me; digital health intervention; descriptive statistics; thematic analysis; usability study; mixed method; social support; health outcomes; medication management; symptom tracking; electronic medical records; user-friendly

Introduction

Pediatric hematopoietic stem cell transplants (HCTs) are an intensive life-saving treatment for several malignant and nonmalignant disorders [1]. However, HCT often requires a long hospital stay that is stressful for the children and their caregivers [2]. Symptom and medication management are important components of the HCT experience, and it is critical for caregivers to adhere to recommendations and communicate with the care team [3]. After discharge, caregivers must follow complex medication regimens with various dosages and frequent dose adjustments, increasing the risk for nonadherence [4,5]. In the pediatric HCT population, 52% to 73% [2,6,7] of patients do not take medications as they are prescribed during the treatment course. Therefore, medication adherence is a primary concern for health care providers and caregivers after pediatric HCT [8,9]. Many factors impact medication adherence rates, including patient-related factors, forgetfulness, therapy side-effects, complexity and length of treatment, and route of administration [10-13]. Although medication adherence in pediatric HCT is understudied and interventions are limited, research in other pediatric chronic conditions has demonstrated the potential of mobile health (mHealth) interventions in improving medication adherence [14-16].

As smartphones become nearly ubiquitous in daily life, mHealth interventions can improve families' ability to manage their child's medical care [17,18]. Recent estimates show that over 5 billion people have access to mobile phone services around the world [18]. Additionally, a study of adults with chronic diseases suggests that mobile apps as mHealth intervention tools are more effective for improving medication adherence than non-mHealth interventions [19]. mHealth interventions have resulted in better clinical outcomes (eg, increase in health-related quality of life, symptom management, and decrease in readmissions and treatment anxiety) [20-22] through behavior change and enhancement of adherence to treatment [23-25]. mHealth interventions allow individuals and caregivers to track medication doses and symptoms, make notes of discussion points with their health care team, and find educational resources and support networks [26-29]. However, such interventions have yet to be tested to promote medication adherence among children in the acute phase post-HCT (ie, hospital discharge to day 100) [30].

This paper reports a longitudinal mixed methods pilot study examining the usability of a mHealth intervention (BMT4me) with caregivers of children in the acute phase following HCT. This intervention helps caregivers to record and track their child's medications, set reminders, report symptoms, and take notes on their child's progress. The goal is to inform future refinement of the intervention, a feasibility trial, and a pilot randomized controlled trial examining efficacy.

Methods

Study Design

Data are from a longitudinal mixed methods study to assess the usability of a newly developed mHealth intervention for pediatric post-HCT medication management. The study was conducted at a large Midwestern children's hospital from September 2021 to January 2023. Eligible caregivers were identified from the HCT clinic schedule and inpatient HCT unit based on the following eligibility criteria: (1) English-speaking, (2) 18 years of age or older, (3) having a child between 2 and 18 years of age undergoing allogeneic HCT, and (4) having a smartphone (either Android or iPhone) at recruitment and during the study period. All 20 caregivers of children who received HCT during the study period and met the eligibility criteria were approached for participation. A total of 15 caregivers consented to participate in the study, while 5 caregivers declined due to reasons including being busy with caring for the child and not being comfortable using the apps in general. One caregiver withdrew after initial consenting, resulting in a final cohort of 14 caregivers.

Ethical Considerations

The study was approved by the Nationwide Children's Hospital Institutional Review Board (approval STUDY00000910) and was designed in accordance with the ethical standards laid out by the Declaration of Helsinki. Eligible interested participants provided informed consent prior to enrollment and were assigned ID numbers for confidentiality. Recordings of qualitative interviews were destroyed after completion of the transcription, and identifying information in the transcripts was removed. All participant information has been anonymized in this paper, including the text, tables, and figures. Upon completion of the study, all participants received a US $20 gift card as compensation.

Measures

Demographic Data Form

Caregivers self-reported the child's and their own background characteristics, including age, sex, race, ethnicity, education level, and family income at the time of study enrollment.

System Usability Scale

Caregivers rated 10 items on a 5-point Likert scale to evaluate the functionality and acceptability of the BMT4me app. Total scores range from 0 to 100, with scores >68% considered above average [31]. Internal consistency (Cronbach α=0.91) and convergent validity (r=0.81 with a 7-point scale of "user-friendliness") have been well established [32,33].

Caregiver Satisfaction

Investigators developed a 9-item survey to obtain feedback regarding caregiver app use, benefits, burdens, barriers, suggested modifications, and overall satisfaction. Questions were rated on a 1 to 4 scale, with higher total scores indicating greater caregiver satisfaction.

Reaction Card

Reaction cards were developed by Microsoft as part of a "desirability toolkit" to elicit immediate reactions, thoughts, or opinions from individuals regarding a particular tool or technology [34]. Using a reaction card of 55 listed words, caregivers provided feedback on the desirability of the BMT4me app. The words and the number of times they were chosen were summarized to indicate the overall attitude toward the app. Higher frequencies of positive words indicated greater usability, while higher frequencies of negative words indicated lower usability [35].

Qualitative Interview

Qualitative nterview guides were developed using a combination of literature review, expert consultation, and pilot testing to ensure relevance and clarity. Questions were designed to elicit in-depth perspectives on key study themes while allowing for flexibility in participant responses. A semistructured format was used to balance consistency across interviews, with the opportunity for participants to elaborate. The semistructured interviews were transcribed verbatim and analyzed using NVivo software (QSR International).

BMT4me App

The development of the BMT4me app (Figures 1 and 2) followed a user-centered, multiphase iterative approach. This method actively involved patients, caregivers, and health care providers at every stage of the app's creation. Initially, a wireframe (Figure 1) was designed, serving as a simple visual representation of the app's structure and content [36].

Feedback was then collected from caregiver and child dyads, which informed the creation of the BMT4me app prototype (Figure 2). Afterward, health care providers assessed the prototype, and feedback led to further refinement of the app.

Figure 1. Version 1 of mobile health app wireframes.

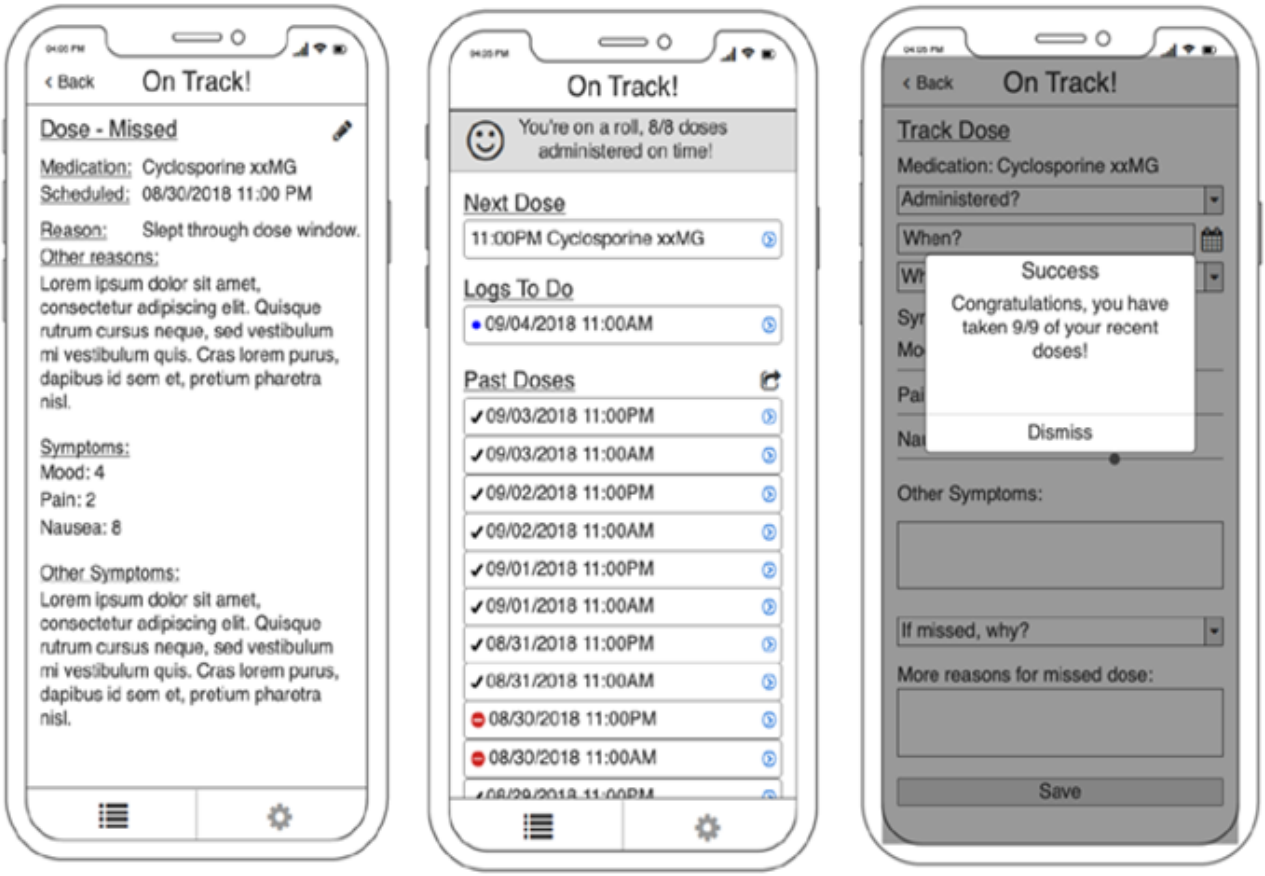


Figure 2. Revised BMT4me app prototype.

The BMT4me app was developed for both iOS and Android devices as a real-time, personalized medication management or adherence tool, and to track medication, symptoms, and side effects. Medications, doses, and schedules can be manually typed into the app by the user or entered using the image-to-text feature, which converts the medication label into text. Through pop-up notifications, the BMT4me app reminds caregivers of the medication doses and schedule, while also recording the time medication was taken and reasons for missed doses. Symptom ratings are represented by emojis on a slider scale of 1 to 10. Additional features include a note section to upload pictures and for recording any details of care to communicate with the provider. In the case of shared caregiver responsibility, the app can be installed on separate devices using the same sign-in code to sync information in real-time between caregivers. Amazon Web Services were used for data storage and app database management.

Procedures

If eligible, recruitment occurred either before or on the day of the child's discharge following HCT or at the first HCT clinic visit postdischarge. Trained research staff introduced the study, and interested caregivers provided written informed consent prior to participation.

To facilitate app login and use, participants received a unique QR code for activating the BMT4me app on their smartphones, and the research team helped with installation. Understanding and engagement with the app were evaluated in the following steps (further explained in [37]).

Step 1: Unobtrusive Observation to Measure the Intuitiveness of the Interface (10 - 15 Min)

Ease of Navigation

This was how quickly and accurately participants were able to find features and complete tasks, as well as any moments of hesitation or confusion when navigating the menu and buttons.

Task completion

This was how quickly and accurately participants found features or completed tasks such as inputting medications, symptoms, or notes.

Error frequency

This included mistakes made while interacting with the app, such as tapping the wrong button or misunderstanding the instructions.

Flow and Progression

This was how naturally participants moved through the app interface without guidance.

Caregivers were encouraged to interact with the app independently, without guidance from the study staff, to evaluate the intuitiveness of the app's interface. During this phase, research staff observed and recorded participant progress related to the other steps described below.

Step 2: Interactive Observation (10 - 15 Min)

After observing independent app use, the research team started to interrupt the participants and ask questions regarding observed cues. Caregivers were encouraged to share their thoughts and criticisms of the app. The discussion was based on individual participant cues, and preprescribed questions were not possible.

Step 3: Debriefing (15 - 20 Min)

Caregivers were asked to share their experience with the app, including the app's interface and content, as well as thoughts on incorporating the app into their daily routine. Feedback regarding the strengths and weaknesses of the app was also collected.

Step 4: Passive Use Observation

To ensure caregivers understood the app, research staff provided a detailed introduction, answered participant questions, and had the caregivers practice adding medications and tracking symptoms. Participants were invited to continue using the BMT4me app at home for 100 days or until the child had been weaned off the immunosuppressant medications. Participants' passive use data on app use and phone activity was digitally logged

Caregivers then evaluated the usability of the BMT4me app at the beginning and end of the study, using the System Usability Scale (SUS), Caregiver Satisfaction, and Reaction Card assessments. SUS, Caregiver Satisfaction, and Reaction cards were completed after participants independently interacted with the app for 5 - 10 minutes during enrollment. Participants completed enrollment measures in the hospital prior to discharge, and they had an option to complete them either electronically via REDCap (Research Electronic Data Capture; Vanderbilt University) on a study iPad or using paper and pencil. Exit measures were completed electronically, where participants were emailed the survey link. Participants were invited to continue using the BMT4me app at home in the acute phase post-HCT (100 d or until the child had been weaned off the immunosuppressant medications) because the risk of nonadherence and complications such as GVHD is highest during that time. The BMT4me app collected daily data on medication-taking, the time medication was taken, reasons for missed doses and barriers, symptoms, and notes. Passive use data on phone activity and app use were digitally recorded and sent to the research team by software developers each week. During the study, research staff followed up with participants weekly to provide technical app support if needed.

Upon completion of surveys, semistructured interviews were also conducted in-person, both at the beginning and end of the study, to explore app use experiences, helpful features, suggestions for future improvements, and barriers encountered while using the app at home. To schedule the interviews, participants were contacted by research staff via phone call or email. During their child's clinic visit, semistructured interviews, lasting approximately 15 - 20 minutes, were conducted by research staff trained in qualitative interview methods. The study investigator conducted fidelity checks to ensure consistency across the research staff's qualitative interview techniques. All interviews were audio-recorded for analysis.

Data Management and Analysis

Data cleaning and verification were completed using Excel (Microsoft Corp), and the data were then analyzed using SPSS software (version 26; IBM Corp). Descriptive statistics (frequencies, means, and SDs) summarized quantitative data, including demographic characteristics, BMT4me app activity, and survey responses. During the passive observation period, passive data modules recorded phone activity and caregivers' app use (such as time, date, and duration of use). Descriptive statistics were applied to analyze the phone activity, and correlation analysis was conducted to examine use patterns over time. Usability and acceptability of the BMT4me app were assessed by averaging total scores from the SUS. The proportion of participants who enrolled and completed the study was examined to assess the feasibility of the intervention.

Semistructured interviews were transcribed verbatim for content analysis and were organized and coded using NVivo software. Initially, the study team read the transcripts to familiarize themselves with the data. Afterward, the team generated an initial list of codes to align with study questions, and later, the codes were sorted to create a thematic framework. For consistency and accuracy, themes and code groups were then revised systematically by the team's identified coders (MS, MK, and Kathryn A Vannatta), and the thematic framework was adjusted to reflect any changes. Once the review was complete, final codes and themes were reviewed by the study's principal investigator (MS), and any disagreements were resolved through consensus with the coding team. Finally, findings were interpreted and reported in relation to existing literature.

Results

Participants

Initially, 15 caregivers were enrolled in the study (Table 1). Among them, 10 (67%) were approached at discharge and 5 (33%) at the first BMT follow-up visit postdischarge. One caregiver opted out after consenting for the study, and 2 were lost to follow-up, one at week 2 and the other at week 3, both due to transfer of care to another institution. Most caregivers were female (n=11, 79%), White (n=10, 71%), non-Hispanic (n=13, 93%), married (n=9, 64%), and had a college-level education (n=8, 57%). More than half of caregivers reported an annual family income of ≤US $75,000 (n=8, 57%). All caregivers were biological parents of the children. The sample of children was mostly male (n=10, 71%), White (n=10, 71%), and non-Hispanic (n=13, 93%).

Table . Caregiver and child demographic characteristics.

Characteristics	Value (n=14)
Caregiver demographic characteristics	
Caregiver age, years	
Mean (SD)	37.92 (7.62)
Median (IQR)	39.0 (25 - 43)
Sex, n (%)	
Male	3 (21)
Female	11 (79)
Marital status, n (%)	
Single	1 (7)
Married	9 (64)
Divorced	1 (7)
Separated	3 (21)
Highest grade of school completed, n (%)	
College	8 (57)
High school	2 (14)
Graduate or professional	1 (7)
Post-secondary high school (technical or trade school)	2 (14)
Not reported	1 (7)
Annual family income (US $), n (%)	
Under 25,00	3 (21)
25,001-50,000 per year	4 (29)
50,001-75,000 per year	1 (7)
75,001-100,000 per year	2 (14)
100,001-150,000 per year	2 (14)
150,001 or more	1 (7)
Not reported	1 (7)
Caregiver's race, n (%)	
Asian	1 (7)
Black or African American	3 (21)
White	10 (71)
Caregiver's ethnicity, n (%)	
Hispanic or Latin	1 (7)
Not Hispanic	13 (93)
Caregiver's device used, n (%)	
iOS mobile phone	11 (79)
Android mobile phone	3 (21)
Child demographic characteristics	
Child age (years)	
Mean (SD)	8.51 (5.19)
Median (IQR)	7.98 (2.01 - 11.36)
Child's sex, n (%)	
Male	10 (71)

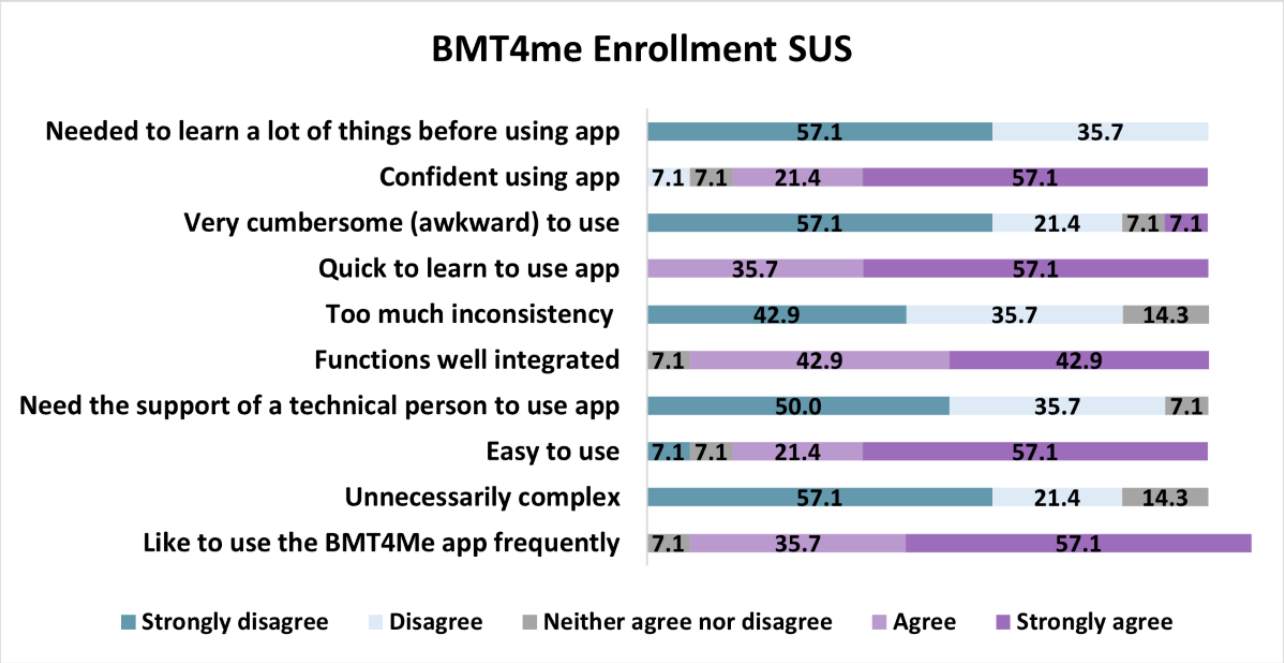
Characteristics	Value (n=14)
Female	4 (29)
Child's race, n (%)	
Asian	1 (7)
Black or African American	2 (14)
Other, please specify	1 (7)
White	10 (71)
Child's ethnicity, n (%)	
Non-Hispanic	13 (93)
Not reported	1 (7)
Child education, n (%)	
Preschool	1 (7)
Grade 1	2 (14)
Grade 3	1 (7)
Grade 4	3 (21)
Grade 10	2 (14)
Graduated high school	1 (7)
Not reported	4 (29)

Usability

SUS

At enrollment, the average SUS score was 86.15 (SD 12.81), and the median was 87.5 (IQR 62.5 - 97.5). On individual items, caregivers reported that the app was quick to learn (n=13, 93%), they were confident using the app (n=11, 79%), the app was easy to use (n=11, 79%) and that they would like to use the BMT4me app (n=13, 93%; Figure 3).

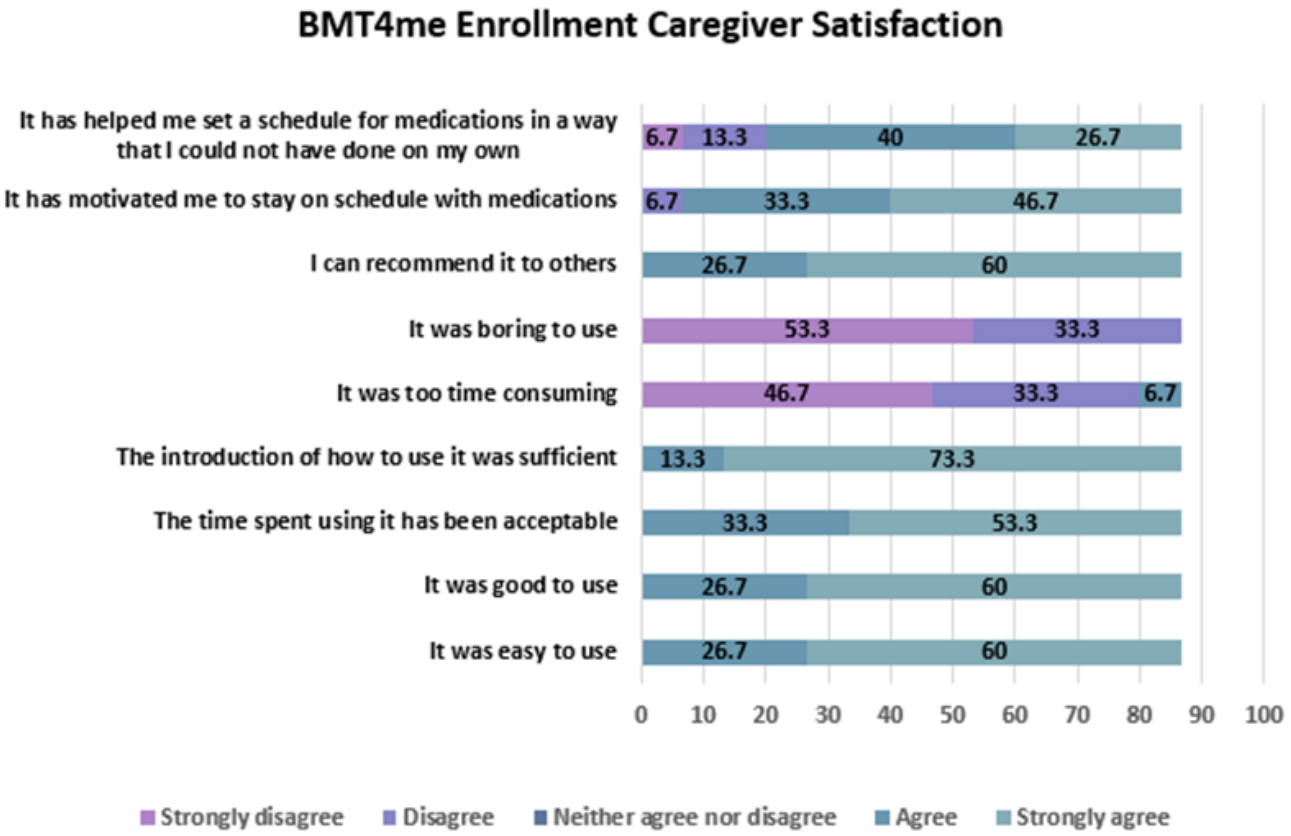
Figure 3. Enrollment SUS survey responses from caregivers. SUS: System Usability Scale.



Caregiver Satisfaction

Participants reported that the BMT4me app was easy to use (n=13, 87%) and effective in motivating them to stay on schedule with medications (n=12, 80%). Most did not find it time-consuming (n=12, 80%) or boring (n=13, 87%). Caregivers indicated they would recommend the app to others (n=13, 87%) and felt it helped them maintain their child's medication schedule in ways they could not have managed on their own (n=10, 67%; Figure 4).

Figure 4. Enrollment caregiver satisfaction survey responses.



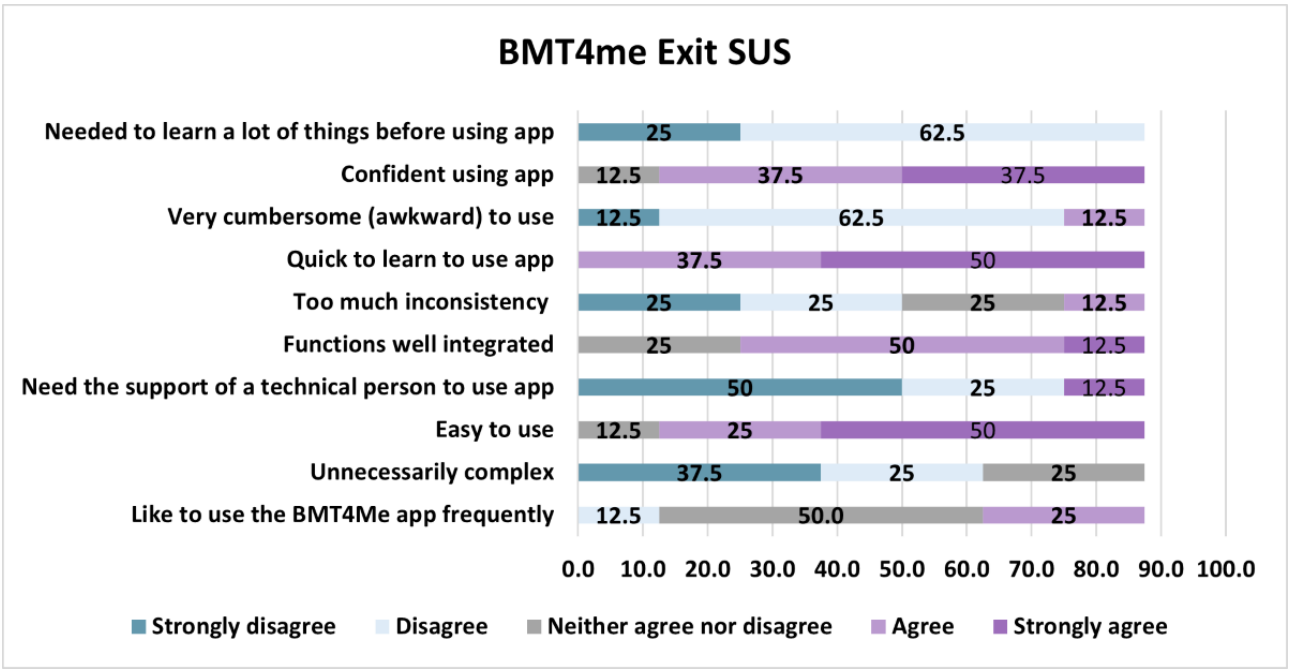
Reaction Cards

Most caregivers expressed a positive reaction after their initial use of the app with the top three endorsed reactions being the app was accessible (n=11, 73%), useful (n=11, 73%), and easy to use (n=10, 67%), which received the highest percentage of responses (Multimedia Appendix 1).

SUS

At study completion, the mean SUS score was 73.13 (SD 16.13), and the median was 75.0(IQR 50 - 86.3). On individual items, caregivers reported the app was quick to learn (n=7, 88%), they were confident in using the app (n=6, 75%), app was easy to use (n=6, 75%), and that features were well-integrated (n=5, 65%; Figure 5).

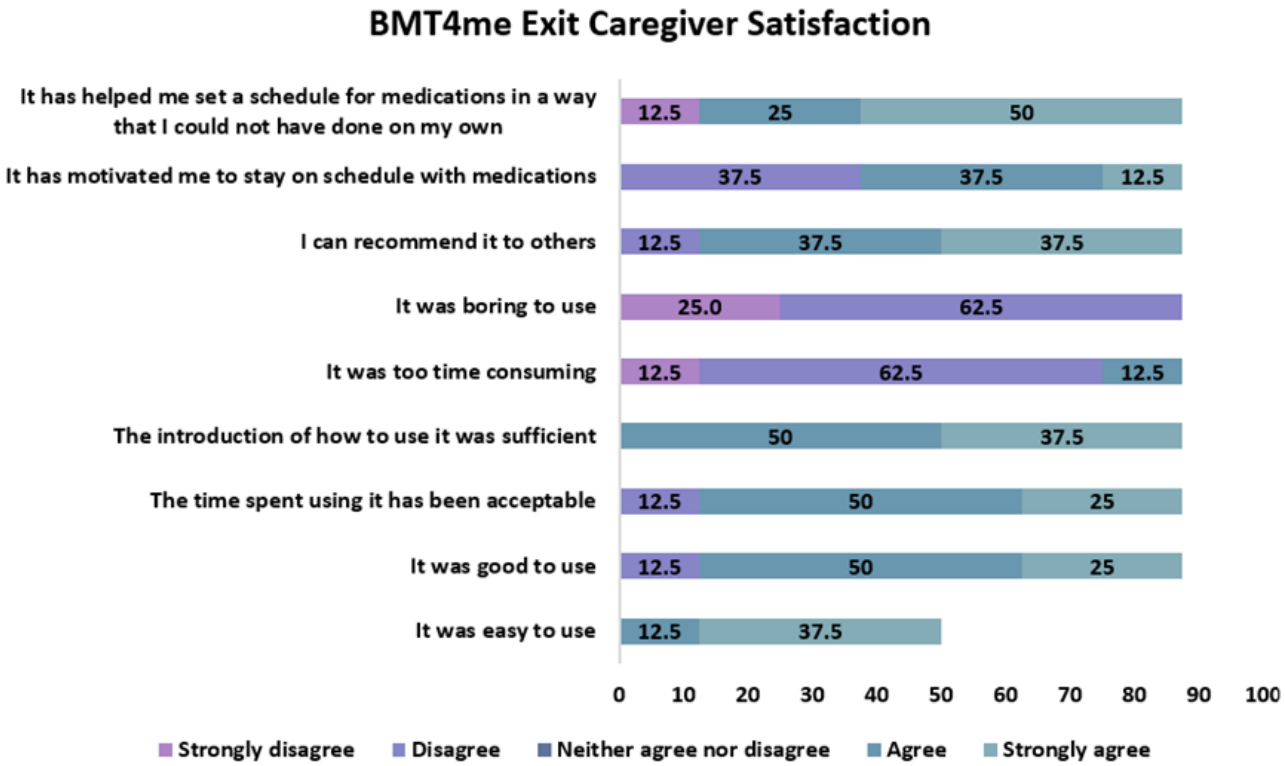
Figure 5. Exit SUS responses from caregivers. SUS: System Usability Scale.



Caregiver Satisfaction

Caregivers reported that the app was helpful in keeping track of their child's medication schedule (n=6, 75%), while also being good to use (n=6, 75%) and not boring (n=7, 88%). A total of 6 caregivers indicated that they would recommend the app to others (n=6, 75%; Figure 6).

Figure 6. Exit caregiver satisfaction survey responses.



Reaction Cards

After a few weeks of interaction with the app at home, most caregivers expressed positive reactions to the app. The top three endorsed reactions were accessible (n=7, 88%), useful (n=7, 88%), and valuable (n=7, 88%), which received the highest percentage of responses (Multimedia Appendix 2).

Feasibility

Of the 14 caregivers who successfully installed the app, 7 (50%) caregivers did not use the app after the initial sign-in, 2 (14%) caregivers added their child's regimen, and 5 (36%) caregivers used the app for at least 1 week. In total, 2 (14%) caregivers used the app for 1 week, while the other 3 (21%) caregivers used the app until study weeks 2, 6, and 7, respectively (Figures 7 and 8).

Descriptive analyses were conducted for the app data (Multimedia Appendix 3) . A total of 1579 app engagement activities were recorded, of which 286 (18%) were for "loading the app," suggesting the number of times enrolled caregivers either attempted logging in or opening the app. The "creating a medication" feature, indicating medication entry into the app, was used 172 (11%) times. The feature "create a dose" was used 552 (35%) times, reflecting the number of times a medication was either entered or edited. A total of 100 (6%) responses were registered for the "take-all-doses." The "take-all-doses" feature pertained to the number of times enrolled caregivers registered on the app that their child took all their prescribed medications. The note feature was used 24 (2%) times. Caregivers used the app to enter their child's symptoms 22 (1%) times. Four of the enrolled caregivers used the note feature at least once.

Figure 7. Weekly app engagement.

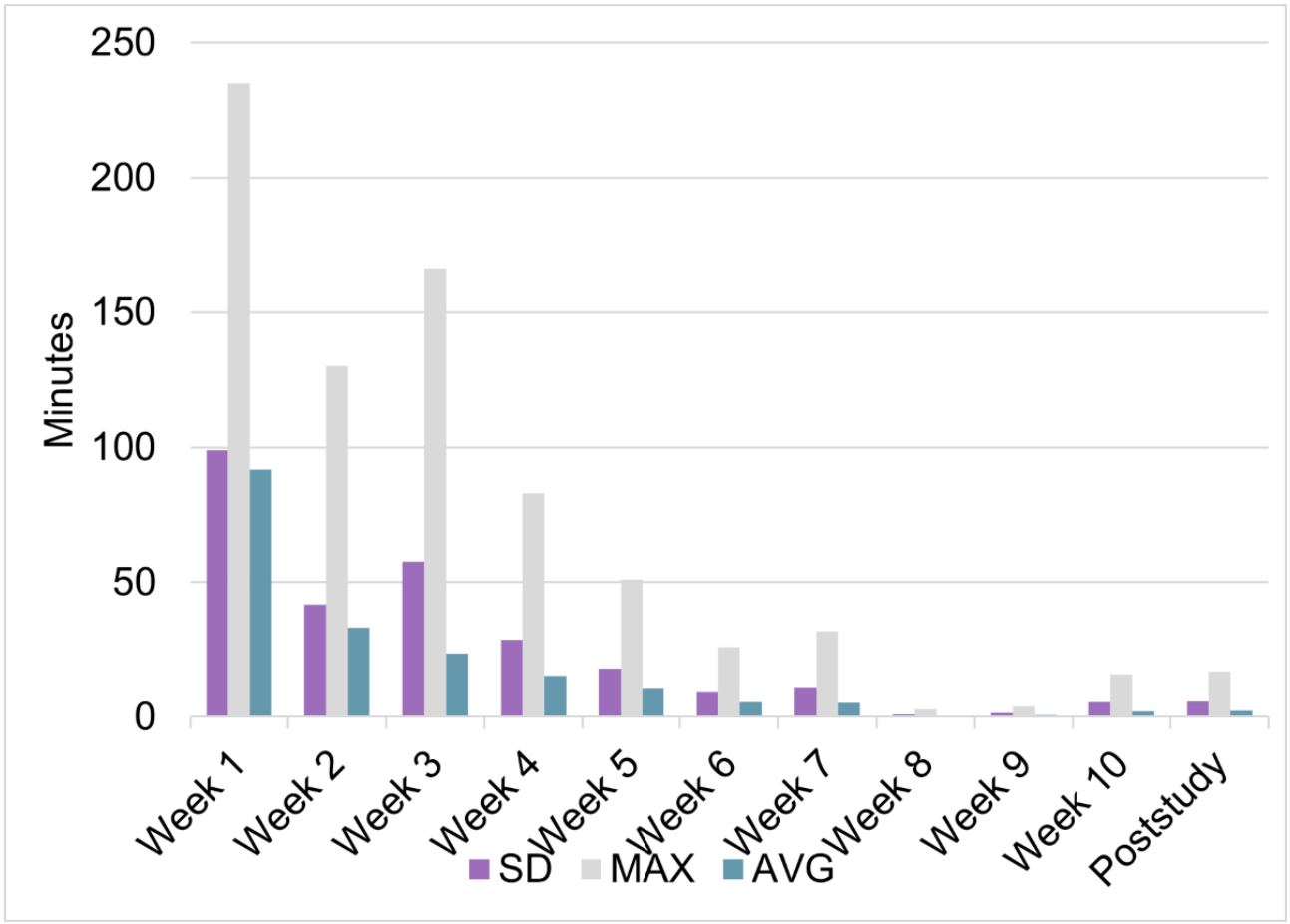
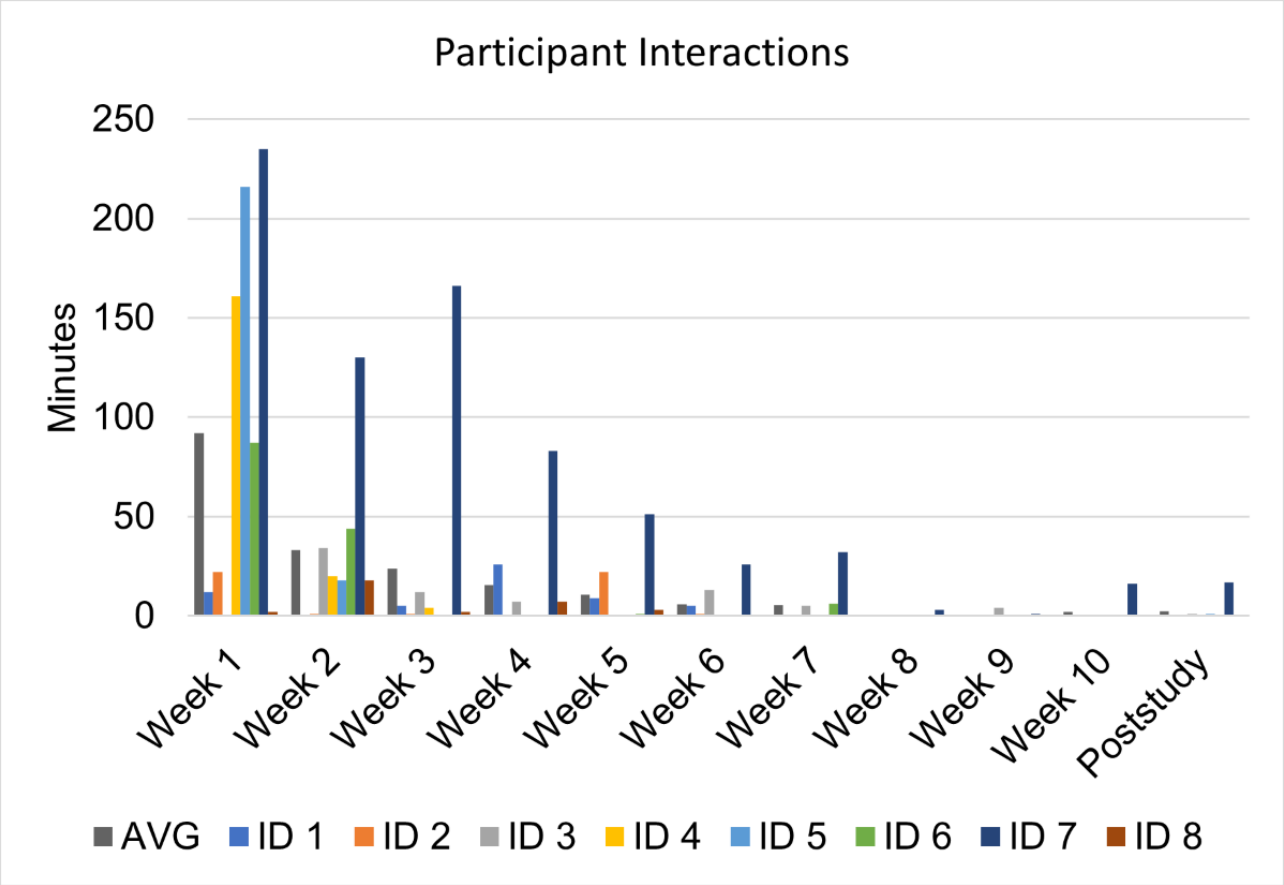


Figure 8. Weekly app engagement by individual user.



Semistructured Interviews

Our qualitative analysis of both enrollment and exit semistructured interviews found that medication reminders were highly favored by 43% (n=6) of families, while recording side effects and symptoms were reported by 21% (n=3) of participants, and writing notes was favored by 14% (n=2). Participants also suggested several new features for the app (Table 2). Half of the families (n=7, 50%) recommended integrating the app with electronic medical record (EMR) to automatically add prescribed medications and appointments and to allow providers to review app activity for pattern identification. EMR is a secure member website that allows patients to access their health information, view appointments, medical test results, and medication therapies, and communicate with their providers [38]. Additionally, 14% (n=2) of families suggested including educational content about common symptoms posttransplant and a search field for medications and their uses. Another 14% (n=2) recommended a field for recording vital signs, such as temperature and blood pressure. Finally, 14% (n=2) proposed adding important contact numbers such as pharmacies, providers, and support groups to the app.

Table . Features suggested by participants.

Features families suggested	Value, n (%)
Integrate with EMR[a] to automatically add prescribed medications and appointments to the app, as well as allowing the provider to review app activity to identify patterns.	7 (50)
Including educational content such as frequently experienced symptoms after transplant, and a field to look up medications and for what they are used.	2 (14)
A field for recording vitals such as fever and blood pressure on the app.	2 (14)
Include important numbers in the app, such as pharmacy, provider, and support groups.	2 (14)

[a]EMR: electronic medical record.

During the enrollment interviews, 3 key themes regarding usability and features of the app were identified (Table 3). Ease of use was the most frequently recurring theme. Participants repeatedly highlighted the intuitive design, which made navigation and operation simple, especially for those who were less technologically skilled. Several participants remarked that grandparents, who lack digital proficiency, could use the app without any issues. Another significant theme was the app's

ability to keep families "on track" with medication schedules and doses. Caregivers expressed enthusiasm for using the app at home to manage a variety of medications and doses and record symptoms. Participants appreciated the ability to mark off administered medications and track notes, finding the digital logging of all information in one place helpful.

Table . Participants' reflections on the use of the BMT4me app.

Themes identified from qualitative interviews	Quotes from enrollment interviews	Quotes from exit interviews
Easy to use	• "It's very simple. Even if my mom had to do that, I think it would be simple for her to use it." [ID 5] • "It looks straightforward and easy to use." [ID 3]	• "It was pretty easy to use. It's pretty intuitive, I think. I think it's easy to pick it up and just start entering things, which is very helpful. Like, it's pretty straightforward." [ID 3] • "It's really pretty self-explanatory. So easy usage which is always nice because if it's complicated, you kind of don't want to deal with it, you know, human nature kind of thing." [ID 11]
Keep us on track	• "I think it'd be useful to help keep us on track. Hopefully not forget any doses." [ID 6] • It seems like it could help to keep on schedule and just like simple reminders of what to do, certain things especially get put and medications." [ID 13]	• "I think having something to be able to keep track of things for you is the number one way to get to have success with them after care because when I left the hospital, like the first time after he had his treatment, I was completely overwhelmed. It was like, how am I going to keep track of all of this stuff?" [ID 3] • "It was nice because if you just didn't realize what time was, it would ring and kind of let you know like, hey, it's time to take medications that so just getting into the swing and kind of getting used to taking all that medication too so." [ID 2]
Helpful	• "I think it would be helpful. You know, just having it. To mark off you've taken it or having the option to have notes to look back on. Have it all on one place and you always have your phone with you, so." [ID 2] • "Yeah, I mean, especially if she's taking a lot of meds I don't remember, like knowing like getting a reminder would probably help you." [ID 8]	_[a]
Technical difficulties	–	• "The other thing I noticed that sometimes it reminded me, sometimes it didn't." [ID 3] • "It says symptoms recorded and it says what time when the dose was taken at nine p.m. But then you can go back and look, it says it was given at eight p.m." [ID 1]
Nontech preferences	–	• "You got a pencil and the journal, you know I can easily, you know, erase. And I just like to have everything in front of me on one page. I'm a visual person and then like going next to next page, you know?" [ID 5] • "I have it in my head, I know what it is to do, so I don't really need it." [ID 13]

[a]Not applicable.

Upon study completion, families reiterated the app's ease of use and its effectiveness in maintaining medication routines (Table 3). Participants described the app as easy and self-explanatory, allowing them to quickly enter information and interact with it. They emphasized that simplicity was highly desirable, as a complicated interface would deter use. Additionally, participants noted that the app's ability to help them stay on track with medication regimens was helpful, particularly when life's demands could interfere with their child's medication schedule. However, new themes also emerged during the exit interviews. One notable theme was technical difficulties. Some families reported experiencing glitches or

issues with the app's functionality, which occasionally hindered their ability to fully use the app. Another emerging theme was a preference for nondigital methods. A subset of families expressed a preference for traditional, paper-based methods of managing health care routines. These participants indicated that while the app offered useful features, they were more comfortable with paper-based tracking and reminders.

Discussion

Principal Findings

The BMT4me app was developed in collaboration with health care providers, caregivers, and patients to aid in medication management, improve adherence monitoring, and track symptoms or medication side effects in real-time [37]. This pilot study aimed to evaluate the usability of the BMT4me app among caregivers of children during the acute phase post-HCT. The caregiver-reported mean SUS score of 73.13 (SD 16.13) at study completion indicated a favorable perception among caregivers, surpassing the threshold of 68 and demonstrating above-average usability. Most caregivers found the app easy to learn and use, with well-integrated features, fostering confidence for independent use at home. These findings are consistent with prior research emphasizing the significance of intuitive design and user-friendly interfaces in mHealth apps [39,40].

The positive usability feedback suggests that the app's design and features were well-received, enhancing caregivers' confidence in managing their child's complex medication regimen [41]. Engagement with the app generally depends on user motivation, perceived value, and satisfaction. As highlighted by Kim et al [42], technology must be both useful and enjoyable, with perceived value and satisfaction stemming from the overall user experience [42]. Caregiver satisfaction emerges as a critical factor for sustainability. Similar to existing literature, out of the participants who use the app regularly, the satisfaction with the BMT4me app was high [43,44]. Most caregivers in this study reported the app as helpful for tracking their child's medication schedule, finding it easy to use, and expressing willingness to recommend it to others. These sentiments were further echoed in qualitative interviews, with caregivers affirming the app's ease of use and effectiveness in navigating their child's post-HCT journey.

Current literature highlights a preference among users for mHealth apps that allow communication with health care providers [45]. This aligns with this study, where 50% (n=7) of caregivers recommended integrating EMR with the BMT4me app and including important contact numbers to facilitate easy communication with health care providers through the app. Including educational content in the mHealth apps has been shown to be preferred among individuals with chronic illnesses [46-48]. This trend was similarly observed in this study, with 14% (n=2) of our participants suggesting the inclusion of information on frequently experienced symptoms after HCT, along with details about medications and their uses.

While initial reactions to the app were positive, the sustainability of long-term use among our sample was constrained. Several caregivers stopped using the app beyond the initial sign-in phase, highlighting potential barriers such as technical challenges, time constraints, feeling overwhelmed, inconvenience, or perceived lack of necessity. Identifying and addressing these barriers is crucial for optimizing app design and implementation strategies to foster sustained engagement among caregivers. Notably, a significant proportion of participants used the app for the first few weeks postdischarge, establishing medication-giving routines before stopping the use. This suggests that the BMT4me app may serve as a valuable resource in assisting families with establishing medication administration habits, particularly during the initial phase of treatment. Some caregivers also expressed the potential value of the BMT4me app in the broader oncology population and wished to have access to it earlier in their child's treatment journey. The app's ability to accommodate the transition from a simple to a complex medication regimen, characterized by frequent changes in medications and dosages, underscores its potential utility beyond the HCT context.

Caregiver feedback on the BMT4me app indicated a generally positive experience from enrollment to study completion, however, some areas showed declines. The mean SUS score decreased from 86.15 to 73.13, reflecting a decrease in overall usability satisfaction; nevertheless, it surpassed the 68% cut-off for usability. Initially, 73% (n=11) of caregivers found the app accessible and useful, and 67% (n=10) found it easy to use [31]. These ratings improved with continued interaction, with 88% (n=7) later describing the app as accessible, useful, and valuable. Caregiver satisfaction measure showed minimal change from enrollment to completion, demonstrating high caregiver satisfaction throughout the study.

Analyzing app engagement activities provided insights into caregivers' usability patterns and preferences. For instance, some features demonstrated more active engagement than others, such that creating and administering doses suggested active engagement in medication management, while note-taking and symptom tracking exhibited low utilization. Therefore, there are potential areas for improvement to better address caregivers' diverse needs and preferences. This observation aligns with existing literature, which often highlights the importance of user-centered design in health apps to enhance engagement and adherence [49]. Studies have shown that while medication management tools are frequently used due to their direct impact on patient care, features like symptom tracking and note-taking require more intuitive interfaces and clear benefits to encourage consistent use. By addressing these gaps, developers can create more effective and comprehensive health management tools that cater to the varied needs of caregivers, ultimately improving patient outcomes [23,50,51].

Overall, this study underscores the importance of user-centered design with a mixed methods approach. By using multiple methods of data collection and data sources, we gained an in-depth understanding of patients with HCT and their families' complex needs. The results suggest that while the BMT4me app has significant potential, there are critical areas that require attention for sustained engagement and effectiveness. Future development should focus on overcoming technical challenges, enhancing the perceived value of lesser-used features, and ensuring the app remains a supportive tool throughout the entire treatment journey. By addressing these gaps, the BMT4me app

can be optimized to better meet the evolving needs of caregivers and patients and improve their health outcomes and quality of life.

Limitations

In addition, this study is not without limitations. First, the use of a single pilot recruitment site may restrict the generalizability of our findings. Although we included a diverse group of children who received HCT at this institution, the small number of caregivers may not represent the broader population. The demographic characteristics of our caregiver (female, White, and non-Hispanic) and child (male, White, and non-Hispanic) cohorts are consistent with other studies in this population [52-54]. However, while our cohort reflects the typical demographics of the pediatric HCT population and their caregivers, it is important to acknowledge that our results may be limited in their broader generalizability. Future research should aim to include a more diverse sample to ensure that findings are relevant across different demographic groups. Furthermore, our recruitment was limited to English-speaking caregivers, due to the app only being available in English. Future versions of the app should be developed in multiple languages to allow for more inclusive recruitment and to examine the app's usability across diverse linguistic and cultural groups. Including caregivers who speak other languages could have enriched the diversity and applicability of our findings. Another limitation is the variability in the timing of app distribution. Participants received the app at different stages of their HCT journey, some on the day of discharge and others after discharge. This inconsistency could have influenced their established medication routines, potentially affecting the outcomes such that participants who were recruited after discharge were less likely to use the app due to already having a system in place for medication management. Additionally, technical issues with the app were noted, which could have impacted the user experience and the app's usability and sustained engagement. Addressing these barriers is crucial for optimizing the app's design and ensuring sustained use. Future improvements should focus on enhancing technical stability, expanding user-centered features, and increasing accessibility to diverse populations.

Conclusions

Despite these limitations, we demonstrated the usability of the BMT4me app among caregivers of children undergoing HCT. This study extends the literature by providing insights into caregiver technology expectations and needs and highlights the potential of mHealth tools to manage medication adherence at home. Notably, 50% (n=7) of caregivers expressed that integrating the app with EMR would be helpful, benefiting both caregivers and health care providers by streamlining access to medication-taking patterns and updating medication regimens. Furthermore, there is potential to adapt this work for children with other chronic conditions, extending the benefits of the BMT4me app beyond the HCT context. The study findings offer valuable insights into the feasibility of conducting a larger randomized controlled trial, potentially leading to significant improvements in adherence and clinical outcomes in children post-HCT.

Acknowledgments

We would like to thank the families who generously participated in this work. We are also grateful for the support of the Information Technology Research and Innovation group at the Nationwide Children's Hospital, specifically Brannon Oiler, Brandon Abbott, and John Luna, for app design and development. We would also like to thank Dr Kathryn A Vannatta for her assistance in qualitative coding. This research was supported by a grant from the National Institute of Nursing Research of the National Institutes of Health (K99 NR019115). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. No generative artificial intelligence was used to write the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

MS contributed toward conceptualization, methodology, funding acquisition, supervision, writing—original draft, writing-review, and editing. MK contributed toward formal analysis, visualization, writing—original draft, writing—review, and editing. PG contributed toward formal analysis, visualization, writing—original draft. ES contributed toward formal analysis, and writing—original draft. AP, RB, WL, and CG contributed to writing—original draft. All authors read and approved the final version of the manuscript.

Conflicts of Interest

ES serves on the editorial board of JMIR Publications.

Multimedia Appendix 1
Enrollment reaction card responses from caregivers.
[DOCX File, 20 KB - cancer_v11i1e66847_app1.docx]

Multimedia Appendix 2
Exit reaction card responses from caregivers.
[DOCX File, 20 KB - cancer_v11i1e66847_app2.docx]

Multimedia Appendix 3
Frequency of app use across different features.
[DOCX File, 16 KB - cancer_v11i1e66847_app3.docx]

References

1. Bazinet A, Popradi G. A general practitioner's guide to hematopoietic stem-cell transplantation. Curr Oncol 2019 Jun;26(3):187-191. [doi: 10.3747/co.26.5033] [Medline: 31285665]
2. Chardon ML, Klages KL, Joffe NE, Pai ALH. Caregivers' experience of medication adherence barriers during pediatric hematopoietic stem cell transplant: a qualitative study. J Pediatr Psychol 2022 Jun 7;47(6):685-695. [doi: 10.1093/jpepsy/jsab138] [Medline: 35066587]
3. Kazak AE, Swain AM, Pai ALH, et al. Caregiver perspectives on psychosocial care in pediatric hematopoietic stem cell transplantation (HCT). Clin Pract Pediatr Psychol 2020 Mar;8(1):67-78. [doi: 10.1037/cpp0000278]
4. Chardon ML, Klages KL, Joffe NE, Pai ALH. Pediatric hematopoietic stem cell transplant medication adherence facilitators and strategies: a qualitative study. J Pediatr Psychol 2023 May 20;48(5):415-424. [doi: 10.1093/jpepsy/jsac078] [Medline: 36173365]
5. Coleman CI, Limone B, Sobieraj DM, et al. Dosing frequency and medication adherence in chronic disease. J Manag Care Pharm 2012 Sep;18(7):527-539. [doi: 10.18553/jmcp.2012.18.7.527] [Medline: 22971206]
6. Morrison CF, Martsolf DM, Wehrkamp N, Tehan R, Pai ALH. Medication adherence in hematopoietic stem cell transplantation: a review of the literature. Biol Blood Marrow Transplant 2017 Apr;23(4):562-568. [doi: 10.1016/j.bbmt.2017.01.008] [Medline: 28087458]
7. Pai ALH, Rausch J, Drake S, et al. Poor adherence is associated with more infections after pediatric hematopoietic stem cell transplant. Biol Blood Marrow Transplant 2018 Feb;24(2):381-385. [doi: 10.1016/j.bbmt.2017.10.033] [Medline: 29102720]
8. Hoegy D, Bleyzac N, Rochet C, et al. Medication adherence after pediatric allogeneic stem cell transplantation: barriers and facilitators. Eur J Oncol Nurs 2019 Feb;38:1-7. [doi: 10.1016/j.ejon.2018.11.006] [Medline: 30717930]
9. Skeens MA, Gerhardt CA, Bajwa R, Akard TF. Toward a better understanding: an exploration of provider perceptions in pediatric hematopoietic stem cell transplant adherence. Pediatr Transplant 2020 Sep;24(6):e13786. [doi: 10.1111/petr.13786] [Medline: 32678483]
10. Goh H, Kwan YH, Seah Y, Low LL, Fong W, Thumboo J. A systematic review of the barriers affecting medication adherence in patients with rheumatic diseases. Rheumatol Int 2017 Oct;37(10):1619-1628. [doi: 10.1007/s00296-017-3763-9] [Medline: 28681249]
11. Shemesh E, Shneider BL, Savitzky JK, et al. Medication adherence in pediatric and adolescent liver transplant recipients. Pediatrics 2004 Apr;113(4):825-832. [doi: 10.1542/peds.113.4.825] [Medline: 15060234]
12. Quine L, Steadman L, Thompson S, Rutter DR. Adherence to anti-hypertensive medication: proposing and testing a conceptual model. Br J Health Psychol 2012 Feb;17(1):202-219. [doi: 10.1111/j.2044-8287.2011.02034.x] [Medline: 22107150]
13. Adherence to health-related regimens in pediatric chronic disorders. Clin Psychol Rev 1984 Jan;4(5):585-596. [doi: 10.1016/0272-7358(84)90046-1]
14. Israni A, Dean C, Kasel B, Berndt L, Wildebush W, Wang CJ. Why do patients forget to take immunosuppression medications and miss appointments: can a mobile phone app help? JMIR Public Health Surveill 2016;2(1):e15. [doi: 10.2196/publichealth.5285] [Medline: 27227150]
15. Heneghan MB, Hussain T, Barrera L, et al. Access to technology and preferences for an mHealth intervention to promote medication adherence in pediatric acute lymphoblastic leukemia: approach leveraging behavior change techniques. J Med Internet Res 2021 Feb 18;23(2):e24893. [doi: 10.2196/24893] [Medline: 33599621]
16. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: a systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
17. Paglialonga A, Mastropietro A, Scalco E, et al. The mhealth. In: m_Health Current and Future Applications: Springer; 2019:5-17.
18. Taylor K, Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. 2019 URL: https://coilink.org/20.500.12592/ffd2v5 [accessed 2025-05-08]
19. Peng Y, Wang H, Fang Q, et al. Effectiveness of mobile applications on medication adherence in adults with chronic diseases: a systematic review and meta-analysis. J Manag Care Spec Pharm 2020 Apr;26(4):550-561. [doi: 10.18553/jmcp.2020.26.4.550] [Medline: 32223596]

20. Osborn J, Ajakaiye A, Cooksley T, Subbe CP. Do mHealth applications improve clinical outcomes of patients with cancer? A critical appraisal of the peer-reviewed literature. Support Care Cancer 2020 Mar;28(3):1469-1479. [doi: 10.1007/s00520-019-04945-4] [Medline: 31273501]
21. Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
22. Jibb LA, Stevens BJ, Nathan PC, et al. Implementation and preliminary effectiveness of a real-time pain management smartphone app for adolescents with cancer: a multicenter pilot clinical study. Pediatr Blood Cancer 2017 Oct;64(10):10. [doi: 10.1002/pbc.26554] [Medline: 28423223]
23. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]
24. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4. [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
25. Park LG, Howie-Esquivel J, Chung ML, Dracup K. A text messaging intervention to promote medication adherence for patients with coronary heart disease: a randomized controlled trial. Patient Educ Couns 2014 Feb;94(2):261-268. [doi: 10.1016/j.pec.2013.10.027] [Medline: 24321403]
26. Tinschert P, Jakob R, Barata F, Kramer JN, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 2;5(8):e113. [doi: 10.2196/mhealth.7177] [Medline: 28768606]
27. Tabi K, Randhawa AS, Choi F, et al. Mobile apps for medication management: review and analysis. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e13608. [doi: 10.2196/13608] [Medline: 31512580]
28. Mueller EL, Cochrane AR, Campbell ME, Nikkhah S, Miller AD. An mHealth app to support caregivers in the medical management of their child with cancer: co-design and user testing study. JMIR Cancer 2022 Mar 16;8(1):e33152. [doi: 10.2196/33152] [Medline: 35293867]
29. Nievas Soriano BJ, Uribe-Toril J, Ruiz-Real JL, Parrón-Carreño T. Pediatric apps: what are they for? A scoping review. Eur J Pediatr 2022 Apr;181(4):1321-1327. [doi: 10.1007/s00431-021-04351-1] [Medline: 35048178]
30. Ralph JE, Sezgin E, Stanek CJ, et al. Improving medication adherence monitoring and clinical outcomes through mHealth: a randomized controlled trial protocol in pediatric stem cell transplant. PLoS ONE 2023;18(8):e0289987. [doi: 10.1371/journal.pone.0289987] [Medline: 37590237]
31. Lewis JR, Sauro J. The factor structure of the system usability scale. In: HCD 09: Proceedings of the 1st International Conference on Human Centered Design: Held as Part of HCI International 2009: Springer; 2009:94-103. [doi: 10.1007/978-3-642-02806-9_12]
32. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
33. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
34. Barnum CM. Preparing for usability testing. In: Barnum CM, editor. Usability Testing Essentials: Elsevier; 2011:157-197. [doi: 10.1016/B978-0-12-375092-1.00006-4]
35. Merčun T. Evaluation of information visualization techniques: analysing user experience with reaction cards. Presented at: BELIV '14: Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization; Nov 10, 2014; Paris, France. [doi: 10.1145/2669557.2669565]
36. Morson S. Using wireframes to design your app. In: Morson S, editor. Learn Design for iOS Development Apress 2014:69-84. [doi: 10.1007/978-1-4302-6365-4_6]
37. Skeens M, Sezgin E, Stevens J, Landier W, Pai A, Gerhardt C. An mHealth app to promote adherence to immunosuppressant medication and track symptoms in children after hematopoietic stem cell transplant: protocol for a mixed methods usability study. JMIR Res Protoc 2022 Jul 21;11(7):e39098. [doi: 10.2196/39098] [Medline: 35862184]
38. Serrato CA, Retecki S, Schmidt DE. MyChart—a new mode of care delivery: 2005 personal health link research report. Perm J 2007;11(2):14-20. [doi: 10.7812/TPP/07-002] [Medline: 21461088]
39. Griffin L, Lee D, Jaisle A, et al. Creating an mHealth app for colorectal cancer screening: user-centered design approach. JMIR Hum Factors 2019 May 8;6(2):e12700. [doi: 10.2196/12700] [Medline: 31066688]
40. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
41. Costa Stutzel M, Filippo MP, Sztajnberg A, et al. Multi-part quality evaluation of a customized mobile application for monitoring elderly patients with functional loss and helping caregivers. BMC Med Inform Decis Mak 2019 Jul 22;19(1):140. [doi: 10.1186/s12911-019-0839-3] [Medline: 31331309]
42. Kim YH, Kim DJ, Wachter K. A study of mobile user engagement (MoEN): engagement motivations, perceived value, satisfaction, and continued engagement intention. Decis Support Syst 2013 Dec;56:361-370. [doi: 10.1016/j.dss.2013.07.002]

43. Klee P, Bussien C, Castellsague M, et al. An intervention by a patient-designed do-it-yourself mobile device app reduces hba1c in children and adolescents with type 1 diabetes: a randomized double-crossover study. Diabetes Technol Ther 2018 Dec;20(12):797-805. [doi: 10.1089/dia.2018.0255] [Medline: 30403495]
44. Whittemore R, Jaser SS, Faulkner MS, et al. Type 1 diabetes eHealth psychoeducation: youth recruitment, participation, and satisfaction. J Med Internet Res 2013 Jan 29;15(1):e15. [doi: 10.2196/jmir.2170] [Medline: 23360729]
45. Collado-Borrell R, Escudero-Vilaplana V, Calles A, et al. Oncology patient interest in the use of new technologies to manage their disease: cross-sectional survey. J Med Internet Res 2018 Oct 23;20(10):e11006. [doi: 10.2196/11006] [Medline: 30355554]
46. Conway N, Campbell I, Forbes P, Cunningham S, Wake D. mHealth applications for diabetes: user preference and implications for app development. Health Informatics J 2016 Dec;22(4):1111-1120. [doi: 10.1177/1460458215616265] [Medline: 26635324]
47. Bengtsson U, Kasperowski D, Ring L, Kjellgren K. Developing an interactive mobile phone self-report system for self-management of hypertension. Part 1: patient and professional perspectives. Blood Press 2014 Oct;23(5):288-295. [doi: 10.3109/08037051.2014.883203] [Medline: 24564289]
48. Gatwood J, Balkrishnan R, Erickson SR, An LC, Piette JD, Farris KB. Addressing medication nonadherence by mobile phone: development and delivery of tailored messages. Res Social Adm Pharm 2014;10(6):809-823. [doi: 10.1016/j.sapharm.2014.01.002] [Medline: 24603134]
49. Kujala S. User involvement: a review of the benefits and challenges. Behav Inf Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
50. Schnall R, Rojas M, Bakken S, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
51. Messner EM, Probst T, O'Rourke T, et al. MHealth applications: potentials, limitations, current quality and future directions. In: Digital Phenotyping and Mobile Sensing: Springer; 2019:235-248. [doi: 10.1007/978-3-030-31620-4_15]
52. Olson TL, Pollack MM, Dávila Saldaña BJ, Patel AK. Hospital survival following pediatric HSCT: changes in complications, ICU therapies and outcomes over 10 years. Front Pediatr 2023;11:1247792. [doi: 10.3389/fped.2023.1247792] [Medline: 37900687]
53. Kaziunas E, Buyuktur AG, Jones J, Choi SW, Hanauer DA, Ackerman MS. Transition and reflection in the use of health information: the case of pediatric bone marrow transplant caregivers. Presented at: CSCW '15: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; Feb 28, 2015; Vancouver BC Canada. [doi: 10.1145/2675133.2675276]
54. Yoo J, Halley MC, Lown EA, et al. Supporting caregivers during hematopoietic cell transplantation for children with primary immunodeficiency disorders. J Allergy Clin Immunol 2019 Jun;143(6):2271-2278. [doi: 10.1016/j.jaci.2018.10.017] [Medline: 31178019]

Abbreviations

EMR: electronic medical record
HCT: hematopoietic stem cell transplant
mHealth: mobile health
REDCap: Research Electronic Data Capture
SUS: System Usability Scale

Edited by N Cahill; submitted 24.09.24; peer-reviewed by C Abrams; revised version received 13.02.25; accepted 24.02.25; published 29.05.25.

Please cite as:
Kochashvili M, Guttoo P, Sezgin E, Pai A, Bajwa R, Landier W, Gerhardt C, Skeens M
Supporting Medication Adherence in Pediatric Patients Undergoing Hematopoietic Stem Cell Transplant Using the BMT4me mHealth App: Mixed Methods Usability Study
JMIR Cancer 2025;11:e66847
URL: https://cancer.jmir.org/2025/1/e66847
doi:10.2196/66847

Digital Health Intervention to Reduce Malnutrition Among Individuals With Gastrointestinal Cancer Receiving Cytoreductive Surgery Combined With Hyperthermic Intraperitoneal Chemotherapy: Feasibility, Acceptability, and Usability Trial

Yu Chen Lin[1*], MS, PhD; Ryan Hagen[1*], MS; Benjamin D Powers[2], MS, MD; Sean P Dineen[3], MHPE, MD; Jeanine Milano[4], RDN, MS; Emma Hume[1], MPH; Olivia Sprow[1], MPH; Sophia Diaz-Carraway[1], BA; Jennifer B Permuth[3,5], MS, PhD; Jeremiah Deneve[6], DO, FACS; Amir Alishahi Tabriz[1,3], MPH, MD, PhD; Kea Turner[7,8], MPH, MA, PhD

[1]Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, United States

[2]Department of Surgery, University of Maryland School of Medicine, Baltimore, MD, United States

[3]Department of Gastrointestinal Oncology, Moffitt Cancer Center, Tampa, FL, United States

[4]Department of Rehabilitation Services, Moffitt Cancer Center, Tampa, FL, United States

[5]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, United States

[6]Department of Surgery, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, United States

[7]Division of Health Systems, Policy, and Innovations, University of North Carolina at Chapel Hill School of Nursing, Chapel Hill, NC, United States

[8]UNC Lineberger Comprehensive Cancer Center, Chapel Hill, United States

[*]these authors contributed equally

Corresponding Author:
Kea Turner, MPH, MA, PhD
Division of Health Systems, Policy, and Innovations, University of North Carolina at Chapel Hill School of Nursing, Chapel Hill, NC, United States

Abstract

Background: Cytoreductive surgery combined with hyperthermic intraperitoneal chemotherapy (CRS-HIPEC) can improve survival outcomes for individuals with gastrointestinal (GI) cancer and peritoneal disease (PD). Individuals with GI cancer and PD receiving CRS-HIPEC are at increased risk for malnutrition. Despite the increased risk for malnutrition, there has been limited study of nutritional interventions for individuals receiving CRS-HIPEC.

Objective: We aimed to test the feasibility, acceptability, and usability of Support Through Remote Observation and Nutrition Guidance (STRONG), a multilevel digital health intervention to improve nutritional management among individuals with GI cancer and PD receiving CRS-HIPEC. We also assessed patient-reported outcomes, including malnutrition risk, health-related quality of life, and weight-related measures.

Methods: STRONG is a 12-week digital intervention in which participants received biweekly nutritional counseling with a dietitian, logged food intake using a Fitbit tracker, and reported nutrition-related outcomes. Dietitians received access to a web-based dashboard and remotely monitored patients' reported food intake and nutrition-impact symptoms. Implementation outcomes were assessed against prespecified benchmarks consistent with benchmarks used in prior studies. Changes in patient-reported outcomes at baseline and follow-up were assessed using linear and ordered logistic regressions.

Results: Participants (N=10) had a median age of 57.5 (IQR 54-69) years. Feasibility benchmarks were achieved for recruitment (10/17, 59% vs benchmark: 50%), study assessment completion (9/10, 90% vs benchmark: 60%), dietitian appointment attendance (7/10, 70% vs benchmark: 60%), daily food intake logging adherence (6/10, 60% vs benchmark: 60%), and participant retention (10/10, 100% vs benchmark: 60%). Most participants rated the intervention as acceptable (8/10, 80% vs benchmark: 70%) and reported a high level of usability for dietitian services (10/10, 100%). The benchmark usability for the Fitbit tracker to log food intake was not met. Compared to baseline, participants saw on average a 6.0 point reduction in malnutrition risk score (P=.01), a 20.5 point improvement in general health-related quality of life score (P=.002), and a 5.6 percentage point increase in 1-month weight change (P=.04) at the end of the study.

Conclusions: The STRONG intervention demonstrated to be feasible, acceptable, and usable among individuals with GI cancer and PD receiving CRS-HIPEC. A fully powered randomized controlled trial is needed to test the effectiveness of STRONG for reducing malnutrition and improving patient outcomes.

Trial Registration: ClinicalTrials.gov NCT05649969; https://clinicaltrials.gov/study/NCT05649969

(JMIR Cancer 2025;11:e67108) doi:10.2196/67108

KEYWORDS

gastrointestinal cancer; peritoneal disease; cytoreductive surgery combined with hyperthermic intraperitoneal chemotherapy; digital health intervention; nutrition; feasibility

Introduction

Malnutrition is commonly observed among individuals with gastrointestinal (GI) cancer and can severely affect disease prognosis, quality of life, and survival [1,2]. Individuals with GI cancer are at high risk of developing peritoneal disease (PD), the metastasis of cancer to the abdominal cavity, which occurs in about 40% of patients with GI cancer [3]. Cytoreductive surgery combined with hyperthermic intraperitoneal chemotherapy (CRS-HIPEC) can offer survival benefits for individuals with GI cancer and PD [4,5]. CRS-HIPEC is a 2-step approach that removes all visible cancerous tumors in the abdomen through a surgical procedure, followed by heated chemotherapy during surgery [5]. Due to the invasive nature of this extensive operation, postoperative morbidities are common, including weight loss, which occurs in more than 90% of individuals receiving CRS-HIPEC [6]. Malnutrition, arising from loss of appetite and malabsorption, occurs in about 50% to 60% of individuals receiving CRS-HIPEC, which can negatively affect postoperative outcomes (eg, length of hospital stay, hospital readmission, and mortality) [7-10]. After CRS-HIPEC, patients often experience a decline in nutritional status, heightening the importance of adequate nutritional support in the postoperative period [6,11].

Medical nutrition therapy (MNT), which includes dietitian-led nutritional counseling and additional dietary interventions, has been shown to improve nutritional outcomes for individuals with GI cancer [12,13]. However, multilevel barriers hinder access to MNT and its effectiveness. At the system level, there may be limited outpatient services for nutritional counseling, fragmented oncology and nutritional care, and inconsistent nutrition-screening procedures across clinics [14,15]. For example, a survey among surgical oncologists who specialize in CRS-HIPEC showed that only one-third of providers reported the availability of malnutrition screening at their practice [16]. At the provider level, available dietitians may be lacking, and MNT is not routinely provided to individuals with cancer [14,17,18]. There are also many barriers at the patient level, including the lack of adherence to nutritional programs due to clinical factors (eg, difficulties swallowing, fatigue, nausea, and pain) and nonclinical factors (eg, lack of motivation and time constraint) [19]. Digital nutritional interventions, such as remote monitoring, can help patients overcome barriers to accessing and adhering to traditional nutritional interventions and can improve patient outcomes [20,21]. However, research on digital nutritional interventions for individuals with cancer is limited [22]. There is a need to develop and test digital nutritional interventions, particularly for individuals receiving CRS-HIPEC who are at high risk for malnutrition.

To address this gap, the goal of this study is to pilot test the Support Through Remote Observation and Nutrition Guidance (STRONG) intervention, a multilevel digital health intervention to improve nutritional outcomes. The study aims (1) to assess the feasibility, acceptability, and usability of the STRONG intervention for individuals with GI cancer and PD undergoing CRS-HIPEC and (2) to evaluate patient-reported outcomes, including malnutrition risk, health-related quality of life, and weight-related measures. To the best of our knowledge, this is the first digital nutritional intervention conducted among individuals receiving CRS-HIPEC, who are at high risk of postoperative malnutrition and face unique barriers to accessing and using MNT [7,20,21]. Findings from this study will inform broader interventions to manage cancer-related malnutrition and guide a future randomized controlled trial to evaluate the impact of the STRONG intervention.

Methods

Study Design

We conducted a single-arm feasibility trial of STRONG, a 12-week digital intervention to improve postoperative nutrition. Guided by the Obesity-Related Behavioral Intervention Trials model, the goal of the single-arm study was to identify potential technical issues with digital health delivery, assess the optimal length of intervention delivery, and gather participant feedback on acceptability to inform intervention refinement prior to larger testing in a randomized trial [23]. The intervention was developed based on the Theoretical Domains Framework, a theory used to understand and address multilevel behavior change (ie, patient and clinician behavior) in health care settings [24]. Participants received biweekly MNT (6 sessions) that included nutritional counseling with a registered dietitian and continuous remote monitoring of participants' dietary needs by the dietitian. In addition, participants logged daily food intake using a Fitbit device (Inspire 2) and completed 5 study assessments related to patient malnutrition, nutrition-related symptoms, and quality-of-life outcomes (at baseline and 4, 8, 12, and 16 weeks after study enrollment). Participants provided feedback on the intervention's acceptability and usability (at week 12).

Participants

Individuals who met the following criteria were eligible to participate in the study: (1) older than 18 years, (2) diagnosed with primary GI cancer, (3) diagnosed with PD, (4) underwent curative-intent CRS-HIPEC at Moffitt Cancer Center (Moffitt; with cytoreduction completeness score of 0 - 1), (5) transitioned to a postoperative oral diet, (6) were able to speak and read English, and (7) provided informed consent. Individuals were excluded from the study if they met any of the following criteria: (1) had documented or observable psychiatric or neurological condition that would inhibit with study participation, (2) were undergoing treatment for another primary cancer, and (3) received postoperative parenteral or enteral nutrition.

Recruitment

Potential participants were identified through a collaboration between Moffitt's GI clinic staff and the study coordinators. In addition, we screened the patients' electronic health records (EHRs) to determine their eligibility. Eligible participants were contacted by phone, unencrypted email, videoconference, or in-person meetings to introduce them to the study and determine their interest in participating. The participants who provided informed consent were given study materials and equipment during patient visits or by mail, including (1) a welcome packet and checklist describing the study components, instructions on using and caring for the Fitbit tracker and tablet, and brief instructions on estimating food portion size; (2) a Fitbit tracker; and (3) a study loaned tablet to log daily food intake with the Fitbit application already downloaded and synced. Participants also had the choice of downloading the Fitbit application on a personal device if preferred. Within 3 to 5 days of the participant receiving the Fitbit tracker, one of the study coordinators (RH, OS, and SD-C) contacted the participant to confirm that they were able to use the device. Participants received an introduction to the Fitbit tracker before undergoing CRS-HIPEC. Recruitment occurred from December 2022 to July 2023.

Intervention

Dietitians reviewed the participants' food intake and nutritional assessments and conducted 6 biweekly telehealth or in-person counseling sessions with them. During these visits, the dietitian established individualized dietary plans that included a calorie goal, discussed challenges to dietary intake, and made recommendations for improving nutrition. If a participant did not record food intake for 5 days or more, a study coordinator contacted the participant to discuss barriers to using the Fitbit tracker and to encourage continued tracking. Study assessments were completed on REDCap (Research Electronic Data Capture; Vanderbilt University), a web-based software platform [25,26], on a paper survey, or in person using a tablet during clinic visits.

Measures

Sociodemographic Characteristics

Participants' sociodemographic characteristics were obtained from the EHR and the baseline survey. Information collected included age, sex at birth, race or ethnicity, marital status, primary language preference, whether the participant resided in an urban area (defined by matching the participant's zip code using the 2010 US Department of Agriculture rural-urban commuting area codes) [27], 2022 Area Deprivation Index (ADI; an area-level measure of socioeconomically disadvantaged neighborhoods ranging from 0 to the 100th percentile nationally, with higher percentiles indicating more disadvantaged neighborhoods) [28], insurance type, highest educational attainment, and annual household income.

Clinical Characteristics

Clinical characteristics were obtained from the EHR and included tobacco use, BMI, Charlson Comorbidity Index, and cancer type or histology. The peritoneal cancer index was also measured, which grades the extent of PD on a scale from 0 to 39, with higher scores indicating a more extensive disease [29]. Eastern Cooperative Oncology Group performance status was measured, which captures the extent to which the disease affects a patient's activities of daily living; the grades included in this study ranged from 0=fully active to 4=completed disabled [30]. Cytoreduction completeness score was measured, which captures the extent of the residual tumor, and was used to determine whether the patient underwent CRS-HIPEC for curative intent [31]. The patient's nutritional status was measured by the Patient-Generated Subjective Global Assessment (PG-SGA) Short Form, with scores ranging from 0=no risk to 36=highest risk [32].

Implementation Outcomes

The data on feasibility, acceptability, and usability of the intervention were collected through objective intervention data or measured by a participant survey at the end of the intervention (week 12). Implementation outcomes were assessed against prespecified benchmarks consistent with benchmarks (60% - 70%) used in previously reported single-arm digital health interventions for patients with cancer (Multimedia Appendix 1) [33,34]. The feasibility benchmarks of successful implementation of the intervention within the GI clinic included recruitment rate (≥50%), percentage of participants who completed baseline study assessment (≥70%), percentage of participants who completed 4 of 5 study assessments (≥60%), participant retention at the end of the intervention (≥70%), participant retention at the end of the study period (≥60%), percentage of participants who attended at least 4 of 6 dietitian appointments (≥60%), and percentage of participants who logged food intake for 63 of 90 days (≥60%).

Acceptability, defined as the participant's level of satisfaction with the intervention, was measured by the Acceptability of the Intervention Measure, a 4-item scale (score ranges 0 - 20) [35,36]. A ≥70% response rate with a score >12 on the Acceptability of the Intervention Measure was used as the cutoff for establishing acceptability, indicating that participants on average had a positive experience with the intervention [35].

Usability was assessed in 2 ways. Usability, defined as the extent to which individuals were able to use the Fitbit tracker and application to log food intake, was measured by the 10-item System Usability Scale (SUS; score ranges 0 - 100) [37]. A ≥65% response rate with a score >68 on the SUS was used as the cutoff, indicating that participants on average perceived the Fitbit tracker and application to be easy to use [37]. Usability of the clinical dietitian services, including the dietitian's interpersonal skills and patient-perceived health benefits of the dietitian service, was measured by a validated 8-item scale (score ranges 0 - 24) that has been used in outpatient MNT interventions for patients with cancer [38,39]. A ≥70% response rate with a score >12 was used as the cutoff for establishing acceptability, indicating that participants on average had a positive experience with the dietitian services [39].

Patient Outcomes

To evaluate the secondary aim of this study, patient outcomes were obtained from their EHR or study assessments and included malnutrition risk measured by the PG-SGA, health-related quality of life measured by the Functional Assessment of Cancer Therapy—General [40] and the Functional Assessment of

Anorexia/Cachexia Treatment—Anorexia/Cachexia Scale [41], BMI, weight, and 1-month weight change.

Analyses

Descriptive statistics were computed to describe the study sample and assess whether prespecified benchmarks for feasibility, acceptability, and usability were met at the end of the intervention. Given the small sample size, continuous variables were summarized using median and IQR, and categorical variables were summarized using frequency and percentage. Changes in patient outcomes at baseline and the end of the study period (week 16) were assessed using linear regressions for continuous outcomes and ordered logistic regressions for ordinal outcomes. Models included participant fixed effects to obtain within-participant estimates, and SEs were robust and clustered by the participant. We adhered to the CONSORT (Consolidated Standards of Reporting Trials) guidelines for pilot and feasibility studies and study reporting (Checklist 1) [42]. All analyses were performed in Stata (version 18; StataCorp).

Ethical Considerations

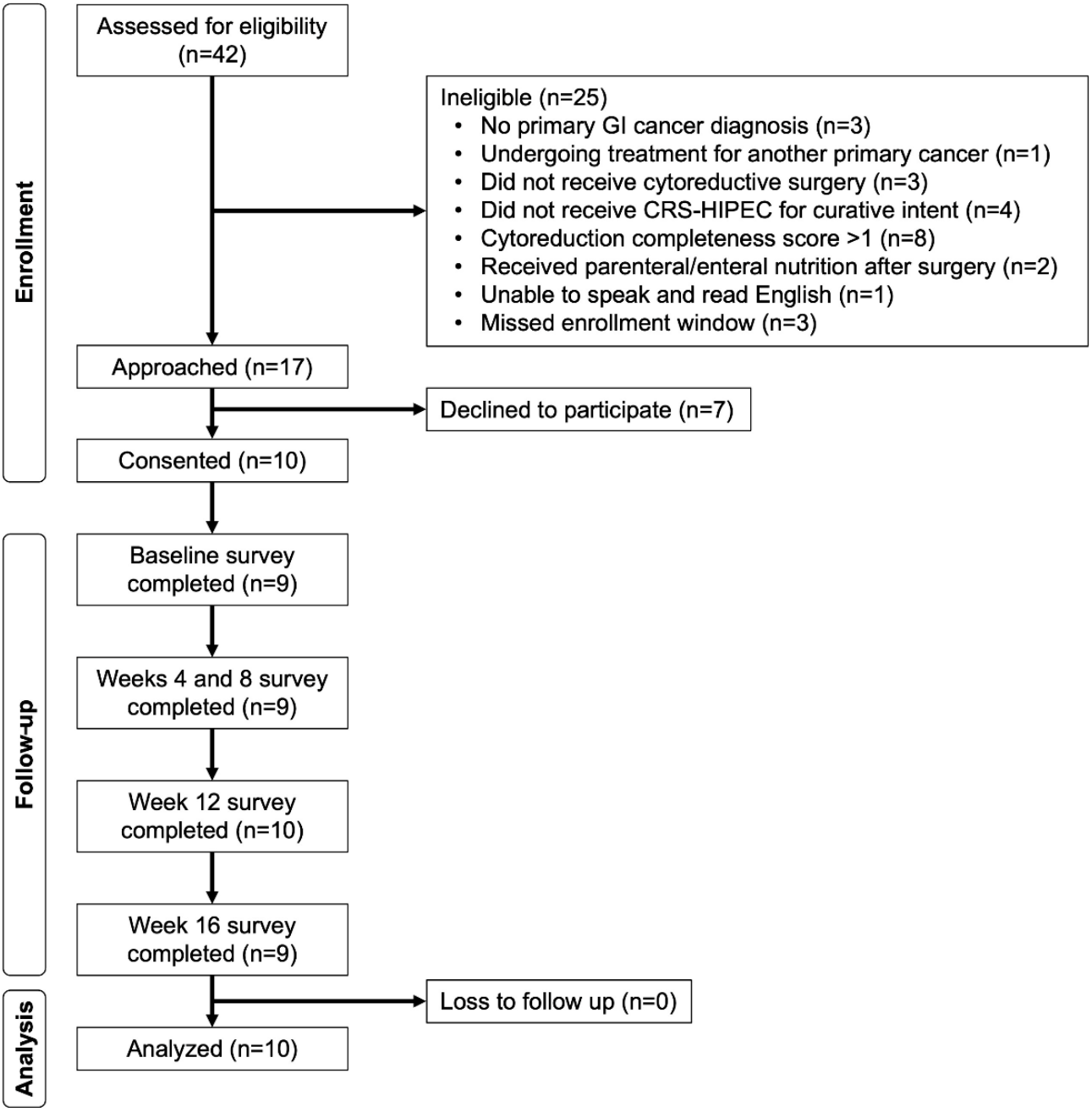
The trial (ClinicalTrials.gov NCT05649969) was conducted at a single site, Moffitt, a National Cancer Institute (NCI)–designated comprehensive cancer center. The study was approved by Moffitt's Institutional Review Board of Record, Advarra (protocol Pro00066098). Informed consent was obtained from all participants. To protect participants' confidentiality, deidentified information and pseudonym IDs (eg, Participant 1) were entered into participants' Fitbit profiles. The study staff maintained a separate password-protected file behind Moffitt's firewall linking participant IDs to patient identifiers (eg, name and medical record number). Paper questionnaires were stored in a locked file cabinet in an office with a locked door. Only the study team had access to participant research data, and only trained staff with appropriate approvals had access to patient medical records. Participants were compensated with a US $25 gift card for completing each of the 5 study assessments; participants who completed all 5 assessments received an additional US $25 gift card.

Results

Sample Characteristics

Among the patients (n=42) screened for eligibility, 25 patients were deemed ineligible (Figure 1). Of the 17 patients approached, 10 patients consented to participate in the study. All 10 participants completed at least 1 assessment, and no participants were lost to follow-up.

Figure 1. CONSORT diagram of the single-arm feasibility trial of the STRONG intervention. CONSORT: Consolidated Standards of Reporting Trials; CRS-HIPEC: cytoreductive surgery combined with hyperthermic intraperitoneal chemotherapy; GI: gastrointestinal; STRONG: Support Through Remote Observation and Nutrition Guidance.



Study participants had a median age of 57.5 (IQR 54.0 - 69.0) years; 8 (80%) participants were female, and 10 (100%) were non-Hispanic White (Table 1). In total, 6 (60%) participants were married, 1 (10%) had never been married, and 3 (30%) were divorced. All participants spoke English as their primary language and resided in an urban area. The median ADI of participants was in the 30th percentile nationally (IQR 24-32), with a higher ADI indicating a more disadvantaged neighborhood. In total, 4 (40%) participants received health insurance through employment, 4 (40%) received public insurance, 1 (10%) had direct-purchase insurance, and 1 (10%) did not provide insurance-type information. The highest education attained by participants was some college for 3 (30%) participants, a bachelor's degree for 3 (30%) participants, or a graduate or professional degree for 4 (40%) participants. Participants' annual household income ranged from US $35,000 to ≥US $100,000.

In total, 8 (80%) participants never smoked tobacco, and 2 (20%) were former smokers (Table 1). The median BMI of participants was 28 (IQR 27.3-29.5), and participants had a median Charlson Comorbidity Index of 7 (IQR 7-9). A total of 6 (60%) participants were diagnosed with appendiceal cancer, 2 (20%) with colorectal cancer, and 2 (20%) with peritoneal mesothelioma. The median peritoneal cancer index was 21.5 (IQR 13.0-25.5). Participants were all fully active or able to carry out light or sedentary work, with a median Eastern Cooperative Oncology Group performance status of 0 (IQR 0-0). No participants at baseline were assessed to be at low

malnutrition risk, 2 (20%) were at medium risk, and 7 (70%) were at high risk.

Table . Participant characteristics at baseline of the STRONG[a] intervention.

Characteristic	Participant (N=10)
Sociodemographic characteristics	
Age (years), median (IQR)	57.5 (54.0-69.0)
Sex at birth, n (%)	
Male	2 (20)
Female	8 (80)
Race or ethnicity, n (%)	
Non-Hispanic White	10 (100)
Marital status, n (%)	
Never married	1 (10)
Now married	6 (60)
Divorced	3 (30)
Primary language was English, n (%)	10 (100)
Resided in an urban area, n (%)	10 (100)
Area Deprivation Index national percentile, median (IQR)	30 (24-32)
Insurance type, n (%)	
Employment-based	4 (40)
Direct purchase	1 (10)
Public (eg, Medicare, Tricare, and Veterans Affairs)	4 (40)
Unknown	1 (10)
Highest educational attainment, n (%)	
Some college, vocational training, or associate degree	3 (30)
Bachelor’s degree	3 (30)
Graduate or professional degree	4 (40)
Income (US $), n (%)	
<$35,000	0 (0)
$35,000-$49,999	1 (10)
$50,000-$74,999	1 (10)
$75,000-$99,999	2 (20)
≥$100,000	3 (30)
Unknown	3 (30)
Clinical characteristics	
Tobacco use, n (%)	
Never smoker	8 (80)
Former smoker	2 (20)
BMI (kg/m^2), median (IQR)	28 (27.3 - 29.5)
Charlson Comorbidity Index, median (IQR)	7 (7-9)
Cancer type or histology, n (%)	
Appendiceal mucinous neoplasm	6 (60)
Colorectal adenocarcinoma	2 (20)
Peritoneal mesothelioma	2 (20)
Peritoneal cancer index, median (IQR)	21.5 (13-25.5)

Characteristic	Participant (N=10)
ECOG[b] performance status, median (IQR)	0 (0 - 0)
Cytoreduction completeness score, median (IQR)	1 (0 - 1)
Patient-Generated Subjective Global Assessment Short Form score	
Low malnutrition risk (score 0 - 3), n (%)	0 (0)
Medium malnutrition risk (score 4 - 8), n (%)	2 (20)
High malnutrition risk (score≥9), n (%)	7 (70)
Unknown, n (%)	1 (10)
Median (IQR)	12 (9 - 15)

[a]STRONG: Support Through Remote Observation and Nutrition Guidance.

[b]ECOG: Eastern Cooperative Oncology Group.

Feasibility

Among the eligible patients (n=17) who were approached to participate in the study, 10 (59%) consented to participate (Table 2). In total, 9 (90%) participants completed the baseline assessment and completed 4 of 5 assessments. All participants completed the assessment at the end of the intervention (week 12), and 9 (90%) participants completed the follow-up assessment at week 16. A total of 7 (70%) participants attended at least 4 of 6 dietitian appointments, and 6 (60%) participants logged food intake for at least 63 of the 90 days (median logged 76, IQR 15 - 87 days). Adherence to logging food intake decreased slightly over the span of the intervention, from 7 (70%) participants meeting the benchmark number of days logging food intake in the first 30 days (median logged 29, IQR 15 - 30 days) to 5 (50%) participants meeting the benchmark in the last 30 days (median logged 15, IQR 0 - 27 days; not shown in table).

Table . Feasibility outcomes of the STRONG[a] intervention.

Outcome	Benchmark (%)	STRONG intervention, n/N (%)
Recruitment		
Eligible patients who consented	≥50	10/17 (59)
Study assessment completion		
Participants who completed baseline assessment	≥70	9/10 (90)
Participants who completed 4 of 5 study assessments	≥60	9/10 (90)
Retention		
Participants retained at the end of the intervention (week 12)	≥70	10/10 (100)
Participants retained at the end of the study period (week 16)	≥60	10/10 (100)
Intervention adherence		
Participants who attended at least 4 of 6 dietitian appointments	≥60	7/10 (70)
Participants who logged food intake for 63 of 90 days	≥60	6/10 (60)

[a]STRONG: Support Through Remote Observation and Nutrition Guidance.

Acceptability and Usability

Among the 10 participants who completed the week 12 assessment, 8 (80%) rated the intervention as acceptable (benchmark score >12), with a median score of 18 (IQR 16 - 20; Table 3). In total, 5 (50%) participants rated the Fitbit tracker and application as usable for logging food intake (benchmark score >68), with a median score of 68.8 (IQR 54.4 - 90). All participants (100%) were satisfied with the dietitian services (benchmark score >12), with a median score of 23.5 (IQR 17.8 - 24.0). One participant reflected on the high acceptability of the dietitian services and said, “The program added a lot of value. It helped with my recovery, especially with getting the right nutrition. It has been really great!” Another participant conveyed the value of the digital nature of the STRONG intervention, expressing “The ZOOM meetings were wonderful because I am over an hour from the hospital. The [nutritionist]

was also a great encourager and contributed to my healing process."

Table . Acceptability and usability outcomes of the STRONG[a] intervention (N=10).

Outcome	Benchmark	STRONG intervention
Acceptability		
Acceptability of the Intervention Measure	≥70% response rate with score >12	8 (80)
Usability		
Fitbit	≥65% response rate with score >68	5 (50)
Dietitian services	≥70% response rate with score >12	10 (100)

[a]STRONG: Support Through Remote Observation and Nutrition Guidance.

Patient Outcomes

Compared to baseline, average PG-SGA malnutrition scores saw a decrease of 6 points (*P*=.01), with a corresponding reduction in patients with high malnutrition risk (*P*=.03; Table 4). Functional Assessment of Cancer Therapy—General scores increased by an average of 20.5 points (*P*=.002), and Functional Assessment of Anorexia/Cachexia Treatment—Anorexia/Cachexia Scale scores increased by an average of 7.4 points (*P*=.03), indicating an improvement in participants' health-related quality of life. There was no change in participants' average BMI or weight, suggesting a stabilization of weight loss. This is supported by a 5.6 percentage point increase in the average 1-month weight change (*P*=.04), in which participants saw a slight weight gain compared to the previous month at week 16.

Table . Changes in patient outcomes between baseline and the end of the intervention.

Patient outcomes (n=9)	Baseline	Week 16	*P* value[a]
Patient-Generated Subjective Global Assessment Short Form score, mean (SE)	11.7 (3.8)	5.7 (4.3)	.01
Low risk, n (%)	0 (0)	3 (33)	.03
Medium risk, n (%)	2 (22)	5 (56)	.03
High risk, n (%)	7 (78)	1 (11)	.03
Functional Assessment of Cancer Therapy—General score, mean (SE)	70.4 (15.4)	90.9 (10.5)	.002
Functional Assessment of Anorexia/Cachexia Treatment—Anorexia/Cachexia Scale score, mean (SE)	28.9 (4.3)	36.3 (7.8)	.03
BMI (kg/m^2), mean (SE)	27.5 (5.5)	26.3 (4.2)	.09
Weight (lb), mean (SE)	172.3 (29.1)	165.3 (22.0)	.09
Weight change since 1 month ago (%), mean (SE)	−5.2 (4.8)	0.4 (3.1)	.04

[a]*P* values of continuous outcomes were computed from linear regression models. *P* values of ordinal outcomes were computed from ordered logistic regression models. All models included participant fixed effects. SEs were robust and clustered by the participant.

Discussion

Principal Findings

The goal of this trial was to evaluate the feasibility and acceptability of the STRONG intervention for individuals with GI cancer and PD undergoing CRS-HIPEC. Our findings demonstrated that the STRONG intervention was feasible to be implemented with high participant recruitment, adherence, and retention to the intervention. Participants rated the intervention favorably and found the dietitian services to be both acceptable and usable. This rating is consistent with previous pilot studies assessing the implementation of mobile phone–based nutritional intervention for individuals with cancer [43,44]. Patient outcomes, including malnutrition risk, health-related quality of life, and 1-month weight change, saw marked improvement. Given that this study was a single-arm intervention without a comparison group, we were unable to attribute the changes in patient outcomes to the intervention alone without considering the effects of cancer treatment and disease progression. The successful implementation of STRONG in this study, positive feedback from participants, and promising improvements in patient outcomes suggest that a future, fully powered trial with a comparison group is warranted. Our team is currently evaluating potential improvements to STRONG and assessing alternative food logging approaches in preparation for a randomized controlled trial of STRONG.

Comparison With Prior Work

Malnutrition screening, counseling, and related interventions remain underused in cancer care [45,46]. Nutritional counseling has been shown to improve the nutritional status, quality of life, and survival for individuals with GI cancer [47,48]. Digital nutritional interventions show clear benefits over traditional MNT, including efficiency, accessibility (eg, reduced transportation barriers), and the ability to remotely monitor patients outside of a traditional clinic visit [22]. Digital tools can also help individuals maintain adherence to nutritional interventions [14]. In our study, despite undergoing a complex surgical procedure, participants were able to adhere to the dietitian visits and food logging in the postoperative period. We hypothesize that this may be facilitated by strong support from the clinic team. Members of the surgical and dietitian teams encouraged patients to participate in the intervention and periodically checked on participants to monitor their progress through the intervention. This study is innovative in that individualized and remote monitoring of dietary needs bridges the gap between the clinical need for close, in-person patient follow-up and the substantial barriers for this patient population to access nutritional support.

Prior studies have shown that digital nutritional interventions are feasible and effective for achieving weight loss among survivors of cancer [49,50]. However, there has been limited research on the use of digital health interventions for individuals with malnutrition to maintain weight or to prevent weight loss. To the best of our knowledge, this is the first digital nutritional intervention conducted among individuals receiving CRS-HIPEC, who have increased risk for malnutrition and face unique barriers to accessing and using traditional MNT [7,20,21].

One study that assessed food-intake tracking with a Fitbit device among individuals with colorectal cancer undergoing surgery found decreased acceptability of the intervention in the postoperative period due to the complexity of the Fitbit application [51]. In our study, we also found declining adherence to tracking food intake over time. The benchmark for tracking food intake was not achieved and was driven primarily by 3 participants who logged 0, 3, and 12 days over the course of the intervention. The decline in adherence to tracking food intake was driven primarily by 2 participants who logged 12 and 24 days over the first 30 days of the intervention, followed by no additional days logged over the rest of the intervention period. It was unclear why these participants were disengaged with tracking food intake, as they did not provide any qualitative feedback.

Further research is needed to investigate why the usability of Fitbit may be low (eg, differences in digital literacy) and what strategies could be used to improve usability. One potential explanation is the choice of the usability measure. Our study team used a generic SUS that was not targeted to mobile health specifically [52]. In future studies, we plan to include mobile health–specific measures of usability to see how that may affect usability ratings. Another potential explanation is that study participants did not have sufficient education on how to use the Fitbit. Since this pilot study, our team has added teach-back sessions to the Fitbit training. Additionally, our team has created a paper version of the food log as an alternative for patients who cannot manage electronic food logging even after training. A third hypothesis is that participants may only need to food log for a certain amount of time before learning enough about their dietary patterns to manage their nutrition. Further study is needed on the optimal time needed for food logging for malnutrition self-management.

Future digital health interventions should assess the eHealth literacy of participants and make efforts to address participant concerns about the potential complexity of digital interventions [53,54]. One approach may be to target participants with low eHealth literacy with additional technical support and resources to ensure equal and inclusive participation in the intervention [53]. Another approach to improve adherence to food tracking may be to adopt a tracking system designed for patients with cancer (eg, the Automated Self-Administered 24-Hour Dietary Recall developed in collaboration with the NCI) [55]. Recent developments in image processing technology can also be leveraged to reduce the burden on participants in tracking food intake [56]. For example, artificial intelligence–enabled applications such as MyFitnessPal, Fastic, and Noom enable food tracking through capturing photos of the foods via a smartphone [56]. However, the feasibility of these technologies for individuals with cancer has yet to be fully explored.

Limitations

Our study should be considered in light of the following limitations. First, this is a single-arm feasibility study conducted at an NCI-designated comprehensive cancer center. Therefore, the study findings may not be generalizable to other settings. Second, the study sample is small and consisted only of non-Hispanic White participants, limiting generalizability to other patient populations. There is a critical need to test the intervention in a more diverse population, across a wider range of settings, and on a larger sample. Additionally, our study focused on extending nutritional support in the postoperative setting, which has been previously recommended for individuals receiving CRS-HIPEC [57,58]. Further research is needed to test nutritional interventions for this population prior to surgery, which may improve CRS-HIPEC tolerance and reduce the likelihood of nutrition-related, postoperative complications [57,58]. Finally, this study was a single-arm trial without a comparison group. Findings on changes in patient outcomes over time cannot be solely attributed to the effect of the intervention without considering the effects of cancer treatment and disease progression. A larger, fully powered randomized controlled trial is needed to rigorously evaluate the impact of STRONG on patient outcomes.

Conclusions

Our study demonstrated that STRONG, a digital health intervention aimed at improving nutritional management for individuals with GI cancer and PD receiving CRS-HIPEC, is feasible, acceptable, and usable. Future studies are needed to establish the effectiveness of the STRONG intervention and to evaluate its implementation in more diverse patient populations and settings.

Acknowledgments

This study was funded by a Moffitt Cancer Center Innovation Award that was awarded to BDP and KT. Generative artificial intelligence tools were not used in any portion of the manuscript writing.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

YCL drafted the manuscript and conducted the data analyses. RH served as a study coordinator for the trial. BDP and SPD helped conceptualize the study and assisted with patient recruitment for the trial. JM served as the study dietitian for the trial. EH served as the research project manager for the trial. OS and SDC served as a study coordinator for the trial. JBP, JD, and AAT helped conceptualize the study. KT helped conceptualize the study, obtained funding for the study, and provided overall scientific direction for the study. All authors reviewed the final draft of the manuscript and provided feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Feasibility, acceptability, and usability scale computations and benchmarks.
[DOCX File, 37 KB - cancer_v11i1e67108_app1.docx]

Checklist 1
CONSORT (Consolidated Standards of Reporting Trials)-eHEALTH checklist.
[PDF File, 1116 KB - cancer_v11i1e67108_app2.pdf]

References

1. Gupta D, Lis CG, Granick J, Grutsch JF, Vashi PG, Lammersfeld CA. Malnutrition was associated with poor quality of life in colorectal cancer: a retrospective analysis. J Clin Epidemiol 2006 Jul;59(7):704-709. [doi: 10.1016/j.jclinepi.2005.08.020] [Medline: 16765273]
2. Poulia KA, Antoniadou D, Sarantis P, Karamouzis MV. Pancreatic cancer prognosis, malnutrition risk, and quality of life: a cross-sectional study. Nutrients 2022 Jan 19;14(3):442. [doi: 10.3390/nu14030442] [Medline: 35276801]
3. Sirody J, Kaji AH, Hari DM, Chen KT. Patterns of gastric cancer metastasis in the United States. Am J Surg 2022 Jul;224(1 Pt B):445-448. [doi: 10.1016/j.amjsurg.2022.01.024] [Medline: 35144812]
4. Yang XJ, Huang CQ, Suo T, et al. Cytoreductive surgery and hyperthermic intraperitoneal chemotherapy improves survival of patients with peritoneal carcinomatosis from gastric cancer: final results of a phase III randomized clinical trial. Ann Surg Oncol 2011 Jun;18(6):1575-1581. [doi: 10.1245/s10434-011-1631-5] [Medline: 21431408]
5. Foster JM, Sleightholm R, Patel A, et al. Morbidity and mortality rates following cytoreductive surgery combined with hyperthermic intraperitoneal chemotherapy compared with other high-risk surgical oncology procedures. JAMA Netw Open 2019 Jan 4;2(1):e186847. [doi: 10.1001/jamanetworkopen.2018.6847] [Medline: 30646202]
6. Dineen SP, Robinson KA, Roland CL, et al. Feeding tube placement during cytoreductive surgery and heated intraperitoneal chemotherapy does not improve postoperative nutrition and is associated with longer length of stay and higher readmission rates. J Surg Res 2016 Jan;200(1):158-163. [doi: 10.1016/j.jss.2015.08.003] [Medline: 26342838]
7. Reece L, Dragicevich H, Lewis C, et al. Preoperative nutrition status and postoperative outcomes in patients undergoing cytoreductive surgery and hyperthermic intraperitoneal chemotherapy. Ann Surg Oncol 2019 Aug;26(8):2622-2630. [doi: 10.1245/s10434-019-07415-7] [Medline: 31123932]
8. Martin AS, Abbott DE, Hanseman D, et al. Factors associated with readmission after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for peritoneal carcinomatosis. Ann Surg Oncol 2016 Jun;23(6):1941-1947. [doi: 10.1245/s10434-016-5109-3] [Medline: 26842489]
9. Gearing PF, Hawke JA, Mohan H, et al. Perioperative nutritional assessment and interventions in patients undergoing cytoreductive surgery and hyperthermic intraperitoneal chemotherapy (HIPEC): a systematic review. Eur J Surg Oncol 2023 May;49(5):902-917. [doi: 10.1016/j.ejso.2023.02.015] [Medline: 36872111]
10. Ryu SW, Kim IH. Comparison of different nutritional assessments in detecting malnutrition among gastric cancer patients. World J Gastroenterol 2010 Jul 14;16(26):3310-3317. [doi: 10.3748/wjg.v16.i26.3310] [Medline: 20614488]

11. Nordhausen K, Solass W, Demtroeder C, Tempfer CB, Reymond M. Cachexia-anorexia syndrome in patients with peritoneal metastasis: an observational study. Pleura Peritoneum 2016 Mar 1;1(1):57-63. [doi: 10.1515/pp-2016-0003] [Medline: 30911608]
12. Vashi P, Popiel B, Lammersfeld C, Gupta D. Outcomes of systematic nutritional assessment and medical nutrition therapy in pancreatic cancer. Pancreas 2015 Jul;44(5):750-755. [doi: 10.1097/MPA.0000000000000336] [Medline: 25872172]
13. Lee JLC, Leong LP, Lim SL. Nutrition intervention approaches to reduce malnutrition in oncology patients: a systematic review. Support Care Cancer 2016 Jan;24(1):469-480. [doi: 10.1007/s00520-015-2958-4] [Medline: 26404858]
14. Deftereos I, Kiss N, Brown T, et al. Awareness and perceptions of nutrition support in upper gastrointestinal cancer surgery: a national survey of multidisciplinary clinicians. Clin Nutr ESPEN 2021 Dec;46:343-349. [doi: 10.1016/j.clnesp.2021.09.734] [Medline: 34857218]
15. Platek ME, Johnson J, Woolf K, Makarem N, Ompad DC. Availability of outpatient clinical nutrition services for patients with cancer undergoing treatment at comprehensive cancer centers. J Oncol Pract 2015 Jan;11(1):1-5. [doi: 10.1200/JOP.2013.001134] [Medline: 25185213]
16. Naffouje SA, De La Cruz K, Berard D, Guy S, Salti GI. Knowledge, attitudes and practice of surgeons regarding nutritional support in CRS and HIPEC patients: are we missing something? Eur J Cancer Care (Engl) 2019 Jan;28(1):e12930. [doi: 10.1111/ecc.12930] [Medline: 30298962]
17. Trujillo EB, Claghorn K, Dixon SW, et al. Inadequate nutrition coverage in outpatient cancer centers: results of a national survey. J Oncol 2019;2019(1):7462940. [doi: 10.1155/2019/7462940] [Medline: 31885583]
18. Furness K, Huggins C, Croagh D, Haines T. Exploring the attitudes of health professionals providing care to patients undergoing treatment for upper gastrointestinal cancers to different models of nutrition care delivery: a qualitative investigation. Nutrients 2021 Mar 22;13(3):1020. [doi: 10.3390/nu13031020] [Medline: 33809826]
19. Nasrah R, Van Der Borch C, Kanbalian M, Jagoe RT. Defining barriers to implementation of nutritional advice in patients with cachexia. J Cachexia Sarcopenia Muscle 2020 Feb;11(1):69-78. [doi: 10.1002/jcsm.12490] [Medline: 31436033]
20. Gong Y, Jiang X, Chen X, et al. Effectiveness of mHealth diet interventions in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Asia Pac J Oncol Nurs 2023 Mar;10(3):100196. [doi: 10.1016/j.apjon.2023.100196] [Medline: 37124242]
21. Cruz C, Prado CM, Punja S, Tandon P. Use of digital technologies in the nutritional management of catabolism-prone chronic diseases: a rapid review. Clin Nutr ESPEN 2021 Dec;46:152-166. [doi: 10.1016/j.clnesp.2021.10.020] [Medline: 34857190]
22. Lazar DE, Postolica R, Hanganu B, Mocanu V, Ioan BG. Web-based nutrition: a useful resource for cancer patients? Front Nutr 2023;10:1134793. [doi: 10.3389/fnut.2023.1134793] [Medline: 37457987]
23. Czajkowski SM, Powell LH, Adler N, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982. [doi: 10.1037/hea0000161] [Medline: 25642841]
24. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77. [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
26. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
27. Rural-urban commuting area codes. US Department of Agriculture. 2010. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes [accessed 2023-09-25]
28. Neighborhood Atlas. University of WIsconsin SoMaPH. URL: https://www.neighborhoodatlas.medicine.wisc.edu [accessed 2025-03-21]
29. Harmon RL, Sugarbaker PH. Prognostic indicators in peritoneal carcinomatosis from gastrointestinal cancer. Int Semin Surg Oncol 2005 Feb 8;2(1):3. [doi: 10.1186/1477-7800-2-3] [Medline: 15701175]
30. ECOG Performance Status Scale. ECOG-ACRIN Cancer Research Group. URL: https://ecog-acrin.org/resources/ecog-performance-status [accessed 2025-03-21]
31. Mehta SS, Bhatt A, Glehen O. Cytoreductive surgery and peritonectomy procedures. Indian J Surg Oncol 2016 Jun;7(2):139-151. [doi: 10.1007/s13193-016-0505-5] [Medline: 27065704]
32. Balstad TR, Bye A, Jenssen CR, Solheim TS, Thoresen L, Sand K. Patient interpretation of the Patient-Generated Subjective Global Assessment (PG-SGA) Short Form. Patient Prefer Adherence 2019;13:1391-1400. [doi: 10.2147/PPA.S204188] [Medline: 31496666]
33. Gudenkauf LM, Li X, Hoogland AI, et al. Feasibility and acceptability of C-PRIME: a health promotion intervention for family caregivers of patients with colorectal cancer. Support Care Cancer 2024 Feb 28;32(3):198. [doi: 10.1007/s00520-024-08395-5] [Medline: 38416143]
34. Müller F, van Dongen S, van Woezik R, et al. A web-based mindfulness-based cognitive therapy for couples dealing with chronic cancer-related fatigue: protocol for a single-arm pilot trial. JMIR Res Protoc 2023 Nov 6;12(1):e48329. [doi: 10.2196/48329] [Medline: 37930767]

35. Weiner BJ, Lewis CC, Stanick C, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017 Aug 29;12(1):108. [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
36. Lewis JR, Sauro J. The factor structure of the System Usability Scale. Presented at: Human Centered Design: First International Conference, HCD 2009, Held as Part of HCI International 2009; Jul 19-24, 2009; San Diego, CA, United States.
37. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
38. Vivanti A, Ash S, Hulcombe J. Validation of a satisfaction survey for rural and urban outpatient dietetic services. J Hum Nutr Diet 2007 Feb;20(1):41-49. [doi: 10.1111/j.1365-277X.2007.00743.x] [Medline: 17241192]
39. Isenring E, Cross G, Kellett E, Koczwara B. Preliminary results of patient satisfaction with nutrition handouts versus dietetic consultation in oncology outpatients receiving chemotherapy. Nutr Diet 2008 Mar;65(1):10-15. [doi: 10.1111/j.1747-0080.2007.00170.x]
40. Functional Assessment of Cancer Therapy—General. FACIT. URL: https://www.facit.org/measures/fact-g [accessed 2025-03-21]
41. Functional Assessment of Anorexia/Cachexia Treatment. FACIT. URL: https://www.facit.org/measures/faact [accessed 2025-03-21]
42. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239. [doi: 10.1136/bmj.i5239] [Medline: 27777223]
43. Orlemann T, Reljic D, Zenker B, et al. A novel mobile phone app (OncoFood) to record and optimize the dietary behavior of oncologic patients: pilot study. JMIR Cancer 2018 Nov 20;4(2):e10703. [doi: 10.2196/10703] [Medline: 30459139]
44. Buchan ML, Goel K, Schneider CK, Steullet V, Bratton S, Basch E. National implementation of an artificial intelligence-based virtual dietitian for patients with cancer. JCO Clin Cancer Inform 2024 Jun;8:e2400085. [doi: 10.1200/CCI.24.00085] [Medline: 38832697]
45. Caccialanza R, Goldwasser F, Marschal O, et al. Unmet needs in clinical nutrition in oncology: a multinational analysis of real-world evidence. Ther Adv Med Oncol 2020;12:1758835919899852. [doi: 10.1177/1758835919899852] [Medline: 32110247]
46. Kleckner AS, Magnuson A. The nutritional needs of older cancer survivors. J Geriatr Oncol 2022 Jun;13(5):738-741. [doi: 10.1016/j.jgo.2021.12.007] [Medline: 34906443]
47. Ravasco P, Monteiro-Grillo I, Vidal PM, Camilo ME. Dietary counseling improves patient outcomes: a prospective, randomized, controlled trial in colorectal cancer patients undergoing radiotherapy. J Clin Oncol 2005 Mar 1;23(7):1431-1438. [doi: 10.1200/JCO.2005.02.054] [Medline: 15684319]
48. Ravasco P, Monteiro-Grillo I, Camilo M. Individualized nutrition intervention is of major benefit to colorectal cancer patients: long-term follow-up of a randomized controlled trial of nutritional therapy. Am J Clin Nutr 2012 Dec;96(6):1346-1353. [doi: 10.3945/ajcn.111.018838] [Medline: 23134880]
49. Befort CA, Klemp JR, Austin HL, et al. Outcomes of a weight loss intervention among rural breast cancer survivors. Breast Cancer Res Treat 2012 Apr;132(2):631-639. [doi: 10.1007/s10549-011-1922-3]
50. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HWW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec;11(6):704-719. [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
51. van der Linden MJW, Nahar van Venrooij LMW, Verdaasdonk EGG. Personal devices to monitor physical activity and nutritional intake after colorectal cancer surgery: feasibility study. JMIR Perioper Med 2022 Dec 13;5(1):e40352. [doi: 10.2196/40352] [Medline: 36512385]
52. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
53. Cheng C, Beauchamp A, Elsworth GR, Osborne RH. Applying the electronic health literacy lens: systematic review of electronic health interventions targeted at socially disadvantaged groups. J Med Internet Res 2020 Aug 13;22(8):e18476. [doi: 10.2196/18476] [Medline: 32788144]
54. Kim K, Shin S, Kim S, Lee E. The relation between eHealth literacy and health-related behaviors: systematic review and meta-analysis. J Med Internet Res 2023;25:e40778. [doi: 10.2196/40778]
55. Subar AF, Kirkpatrick SI, Mittl B, et al. The Automated Self-Administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. J Acad Nutr Diet 2012 Aug;112(8):1134-1137. [doi: 10.1016/j.jand.2012.04.016] [Medline: 22704899]
56. Li X, Yin A, Choi HY, Chan V, Allman-Farinelli M, Chen J. Evaluating the quality and comparative validity of manual food logging and artificial intelligence-enabled food image recognition in apps for nutrition care. Nutrients 2024;16(15):2573. [doi: 10.3390/nu16152573]
57. Laval G, Marcelin-Benazech B, Guirimand F, et al. Recommendations for bowel obstruction with peritoneal carcinomatosis. J Pain Symptom Manage 2014 Jul;48(1):75-91. [doi: 10.1016/j.jpainsymman.2013.08.022] [Medline: 24798105]
58. Raspé C, Flöther L, Schneider R, Bucher M, Piso P. Best practice for perioperative management of patients with cytoreductive surgery and HIPEC. Eur J Surg Oncol 2017 Jun;43(6):1013-1027. [doi: 10.1016/j.ejso.2016.09.008]

Abbreviations

ADI: Area Deprivation Index
CONSORT: Consolidated Standards of Reporting Trials
CRS-HIPEC: cytoreductive surgery combined with hyperthermic intraperitoneal chemotherapy
EHR: electronic health record
GI: gastrointestinal
MNT: medical nutrition therapy
NCI: National Cancer Institute
PD: peritoneal disease
PG-SGA: Patient-Generated Subjective Global Assessment
REDCap: Research Electronic Data Capture
STRONG: Support Through Remote Observation and Nutrition Guidance
SUS: System Usability Scale

Edited by N Cahill; submitted 03.10.24; peer-reviewed by M Al-Jumayli, SP Somashekhar; revised version received 20.12.24; accepted 19.02.25; published 07.04.25.

Please cite as:
Lin YC, Hagen R, Powers BD, Dineen SP, Milano J, Hume E, Sprow O, Diaz-Carraway S, Permuth JB, Deneve J, Alishahi Tabriz A, Turner K
Digital Health Intervention to Reduce Malnutrition Among Individuals With Gastrointestinal Cancer Receiving Cytoreductive Surgery Combined With Hyperthermic Intraperitoneal Chemotherapy: Feasibility, Acceptability, and Usability Trial
JMIR Cancer 2025;11:e67108
URL: https://cancer.jmir.org/2025/1/e67108
doi:10.2196/67108

Original Paper

Mobile Electronic Patient-Reported Outcomes and Interactive Support During Breast and Prostate Cancer Treatment: Health Economic Evaluation From Two Randomized Controlled Trials

Marie-Therése Crafoord[1], RN, PhD; Joakim Ekstrand[2], PhD; Kay Sundberg[1], RN, PhD; Marie I Nilsson[3,4], MSW, PhD; Maria Fjell[1], RN, OCN, PhD; Ann Langius-Eklöf[1], RN, PhD

[1]Division of Nursing, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

[2]Faculty of Health Science, Kristianstad University, Kristianstad, Sweden

[3]Function Area Social Work in Health Care, Karolinska University Hospital, Stockholm, Sweden

[4]Academic Primary Care Centre, Region Stockholm, Stockholm, Sweden

Corresponding Author:
Marie-Therése Crafoord, RN, PhD
Division of Nursing
Department of Neurobiology, Care Sciences and Society
Karolinska Institutet
Alfred Nobels Alle 23
Stockholm, 141 83
Sweden
Phone: 46 524 837 ext 49
Email: marie-therese.crafoord@ki.se

Abstract

Background: Digital interventions for supportive care during cancer treatment incorporating electronic patient-reported outcomes (ePROs) can enhance early detection of symptoms and facilitate timely symptom management. However, economic evaluations are needed.

Objective: This study aims to conduct a cost-utility analysis of an app for ePRO and interactive support from the perspective of the payer (Region Stockholm Health Care Organization) and to explore its impact on patient health care utilization and costs.

Methods: Two open-label randomized controlled trials (RCTs) were conducted, including patients undergoing neoadjuvant chemotherapy for breast cancer (B-RCT; N=149) and radiotherapy for prostate cancer (P-RCT; N=150), recruited from oncology clinics at 2 university hospitals in Stockholm, Sweden. EORTC QLQ-C30 scores were mapped to EQ-5D-3L to calculate quality-adjusted life years (QALYs). Intervention and implementation costs and health care costs, obtained from an administrative database, were used to calculate incremental cost-effectiveness ratios (ICERs) in 3 ways: including all health care costs (ICERa), excluding nonacute health care costs (ICERb), and excluding health care costs altogether (ICERc). Nonparametric bootstrapping was used to explore ICER uncertainty. Health care costs were analyzed by classifying them as disease-related or acute.

Results: In both RCT intervention groups, fewer QALYs were lost compared with the control group (P<.001). In the B-RCT, the mean intervention cost was €92 (SD €2; €1=US $1.03). The mean cost for the intervention and all health care was €36,882 (SD €1032) in the intervention group and €35,427 (SD €959) in the control group (P<.001), with an ICERa of €202,368 (95% CI €152,008-€252,728). The mean cost for the intervention and acute health care was €3585 (SD €480) in the intervention group and €3235 (SD €494) in the control group (P<.001). ICERb was €49,903 (95% CI €37,049-€62,758) and ICERc was €13,213 (95% CI €11,145-€15,281); 22 out of 74 (30%) intervention group patients and 24 out of 75 (32%) of the control group patients required acute inpatient care for fever. In the P-RCT, the mean intervention cost was €43 (SD €0.2). The mean cost for the intervention and all health care was €3419 (SD €739) in the intervention group and €3537 (SD €689) in the control group (P<.001), with an ICERa of –€1,092,136 (95% CI –€3,274,774 to €1,090,502). The mean cost for the intervention and acute health care was €1219 (SD €593) in the intervention group and €802 (SD €281) in the control group (P<.001). ICERb was €745,987 (95% CI –€247,317 to €1,739,292) and ICERc was €13,118 (95% CI –68,468 to €94,704). As many as 10 out of the 75 (13%) intervention group patients had acute inpatient care, with the most common symptom being dyspnea, while 9 out of the 75 (12%) control group patients had acute inpatient care, with the most common symptom being urinary tract infection.

Conclusions: ePRO and interactive support via an app generated a small improvement in QALYs at a low intervention cost and may be cost-effective, depending on the costs considered. Considerable variability in patient health care costs introduced uncertainty around the estimates, preventing a robust determination of cost-effectiveness. Larger studies examining cost-effectiveness from a societal perspective are needed. The study provides valuable insights into acute health care utilization during cancer treatment.

Trial Registration: ClinicalTrials.gov NCT02479607; https://clinicaltrials.gov/ct2/show/NCT02479607, ClinicalTrials.gov NCT02477137; https://clinicaltrials.gov/ct2/show/NCT02477137

International Registered Report Identifier (IRRID): RR2-10.1186/s12885-017-3450-y

(JMIR Cancer 2025;11:e53539) doi:10.2196/53539

KEYWORDS

cost-effectiveness; ePRO; mHealth; disease monitoring; cancer; RCT; randomized controlled trial; controlled trials; digital intervention; patient-reported outcomes; management; payers' perspective; health care costs; apps; prostate cancer; breast cancer

Introduction

Managing symptoms during cancer treatment is essential for patients' quality of life, workability, and performance [1]. Symptoms such as pain, fatigue, and gastrointestinal problems commonly lead to emergency department visits [2]. Emergency department visit rates appear to be higher among patients with cancer than in the general population, although the magnitude or underlying reasons for this remain understudied [3].

Electronic patient-reported outcome (ePRO) interventions have gained recognition as convenient and safe tools for promoting the early detection of symptoms and adverse events [4,5]. Collecting ePROs has demonstrated high acceptance [6-8], long-term feasibility [9], and positive outcomes related to physical and psychological symptoms [10-15], as well as increased survival [16]. ePROs are also suggested to help mitigate unplanned acute care and unnecessary hospitalizations during cancer treatment; however, this assertion requires more robust empirical confirmation [17,18]. In our studies, the use of the interactive app Interaktor was associated with a decreased symptom burden during radiotherapy (RT) for prostate cancer [19], neoadjuvant chemotherapy (NACT) for breast cancer [11], and up to 6 months after surgery for pancreatic cancer [20].

Health economic evaluations are essential for supporting the implementation of cost-effective interventions [21,22] and guiding decision-makers [23]. A cost-utility analysis (CUA) is one type of health economic evaluation that compares the costs and health outcomes of medical treatments or care by estimating the cost required to improve a unit of health outcome [24]. Quality-adjusted life years (QALYs) is a generic measure of disease burden that accounts for both life quality and quantity. One QALY corresponds to 1 year of perfect health, while 0 represents death [25]. In Sweden, the National Board of Health and Welfare (NBHW) has defined a cost per QALY of €9685 (€1=US $1.03) as low, more than €48,423 as high, and more than €96,846 as very high [26].

Most health economic evaluations of ePRO interventions have focused on patients with advanced or metastatic cancer [27]. Lizée et al [28] demonstrated the cost-effectiveness of ePRO from a national health insurance perspective, despite increased costs, due to associated survival benefits. Velikova et al [29] evaluated the cost-effectiveness of ePRO for patients undergoing systemic treatment for colorectal, breast, or gynecological cancer, comparing the cost per additional QALY gained at 18 weeks after randomization from both health care and societal perspectives. The analysis considered costs for the intervention manual, software maintenance, and patient time off work but excluded intervention development costs. No significant cost differences were observed between the intervention and usual care groups. The study indicated a 55% likelihood of cost-effectiveness at the National Institute for Health and Care Excellence cost-per-QALY threshold.

This study was conducted alongside 2 randomized controlled trials (RCTs) of the ePRO intervention Interaktor during NACT for breast cancer (B-RCT) and RT for prostate cancer (P-RCT). The primary aim is to evaluate the cost utility of the app for ePRO and interactive support from the health care provider's perspective (Region Stockholm Health Care Organization). Additionally, the study examines the impact on patients' health care utilization and associated costs.

Methods

Ethical Considerations

The research was approved by the Swedish Ethical Review Authority (permit numbers 2013/1652-31/2 and 2017/2519-32). Written informed consent was obtained from all patients at the time of study inclusion. Data were deidentified to protect participants' privacy. Patients received written and verbal information about their right to opt out without affecting their subsequent care. No compensation or payment was provided for participation.

Study Design

Between 2016 and 2019, Interaktor was evaluated through 2 parallel prospective open-label RCTs, with symptom burden as the primary endpoint, measured using the EORTC QLQ-C30 version 3.0 [30]. Patients were consecutively recruited from oncology clinics at 2 university hospitals in Stockholm, Sweden. Eligible and interested patients met with a researcher who provided detailed information about the trial. Refer to the previously published study protocol and clinical effectiveness article [11,31] for details on the eligibility criteria, intervention design, and randomization process. No changes were made to

the methods after the protocol was registered (NCT02479607 and NCT02477137) and the trials commenced.

Samples

One RCT included a sample of patients with breast cancer treated with NACT (B-RCT), and the other included a sample of patients with prostate cancer treated with RT (P-RCT). In both RCTs, patients were randomly allocated to the intervention or control group. In the B-RCT, there were 74 patients in the intervention group and 75 patients in the control group. Of these, 69 (93.2%) in the intervention group and 71 (94.7%) in the control group completed the follow-up and were considered complete cases (Multimedia Appendices 1 and 2). In the P-RCT, a total of 150 patients were randomly allocated to the intervention (n=75) or control (n=75) group. Of these, 58 (77%) in the intervention group and 56 (75%) in the control group completed the follow-up questionnaires and were considered complete cases [32] (Multimedia Appendices 2-4). The sample size for both RCTs was estimated based on an effect study conducted with patients receiving RT for prostate cancer [19], with symptom distress as the primary outcome. The effect size difference (Cohen d=0.54) indicated that, for 90% power at P<.05, 71 patients were required in each group.

Intervention and Standard Care

The Interaktor smartphone and tablet app is an ePRO intervention designed for daily symptom reporting and interactive support during cancer treatment. It includes a symptom questionnaire, graphs of symptom reporting history, self-care advice related to disease and treatment-associated symptoms, and links to websites with additional information. Oncology ward nurses are alerted via SMS text messages when severe symptom levels are reported. Nurses can access patients' reports through a web interface, which facilitates patient-clinician communication. Depending on the alert, nurses contact the patient within 1 hour or 1 day. The Interaktor versions used in this study did not include any institutional affiliation display or logo.

Patients in the intervention group reported daily via the Interaktor app on weekdays, starting from their first day of treatment and continuing until 2 weeks after treatment in the B-RCT (mean treatment duration: 15 weeks in both groups) and, until 3 weeks after treatment in the P-RCT (mean treatment duration: 5 weeks in both groups). In the intervention groups, registered nurses at the patients' oncology units responded to the symptom report alerts. Additionally, a researcher was available to assist with any technical questions or issues. Outside office hours, patients were advised to contact health care personnel according to the standard procedure of their oncology clinic. The intervention and app content remained unchanged during the evaluation process. Patients received daily reminders if a report had not been submitted. A comprehensive description, including screenshots, has been published previously [7].

All patients, in both the intervention and control groups, received standard care, which included an assigned contact nurse and a visit with the physician before treatment.

Data Collection

Before randomization, patients self-reported sociodemographic characteristics, including education level, marital and occupational status, and baseline outcomes via questionnaires. In the B-RCT, follow-up (via postal questionnaires) occurred 2 weeks after the end of NACT or the day before surgery, whichever came first. In the P-RCT, follow-up was 3 weeks after the conclusion of RT. Medical history and clinical treatment data were obtained from the patients' medical records, including comorbid conditions, tumor histopathology, cancer stage, and prostate-specific antigen score before treatment initiation, as well as the type and number of cancer treatments planned and completed, and reasons for discontinuing or altering treatment. Data on mortality and cause of death were obtained from the Stockholm and Gotland Regional Cancer Centre and the Swedish NBHW (Multimedia Appendix 5).

Health Care Utilisation and Costs

Administrative data on each patient's health care utilization and costs, from the first day of treatment and for 6 months thereafter, were obtained from the Stockholm Region Council administrative database (VAL). The database includes variables on primary care and emergency department visits (VAL-OVR) and hospitalizations (VAL-SLV) for Stockholm Region Council patients [33]. Health care costs were estimated using a variable (SIMKOST [simulerad kostnad/simulated cost]) that calculates the cost of visits based on the profit and loss account for the respective care branch. SIMKOST reflects approximately 90% of the costs for individuals' visits to outpatient care and 99% of the costs for inpatient care (Multimedia Appendix 6). Intervention costs were based on a fiscal estimate provided by the company that developed the app, expressed as a 1-time implementation/startup cost of €5212, with weekly licensing costs per capita of €39 for nurses and €2.25 for patients (Multimedia Appendix 6).

Data Analysis

Statistical Analysis

Data were handled using Microsoft Excel 2016 with the add-in XL-STAT, IBM SPSS Statistics version 27, and STATA 16 (StataCorp LP). Clinical trials with an RCT design should be analyzed using the intention-to-treat (ITT) principle [34], so missing values were imputed as the mean per group and time [32]. Health care utilization and cost values were imputed for 2 patients in each intervention group (B-RCT and P-RCT). In the P-RCT, EORTC dimension scores were imputed at baseline for 2 patients per group, and at follow-up for 15 patients in the intervention group and 20 patients in the control group. In the B-RCT, follow-up values were imputed for 5 patients in the intervention group and 4 patients in the control group. Distribution normality was assessed using skewness and kurtosis. All costs were adjusted for inflation from 2019 to 2022 [35] (×1.0764) and converted from Swedish kronor (SEK) to Euros (€) using the average exchange rate for April 2022 of 10.3257 SEK=€1 [36]. Nonparametric bootstrapping (1000 replications) was used to test nonnormally distributed variables, calculate the incremental cost-effectiveness ratios (ICERs), and explore sample uncertainty regarding the mean ICERs [37].

Health Outcome

EORTC QLQ-C30 [30] dimension scores were mapped onto EQ-5D-3L [38] health state utilities using a response mapping algorithm [38,39] (Multimedia Appendix 4). The original algorithm includes British utility weights [40], which were replaced with Swedish weights according to Burström et al [41] for this study. The mean predicted EQ-5D value (EQ-5DP) before treatment minus after treatment was used to measure effectiveness, with a smaller reduction in mean EQ-5DP indicating better outcomes.

Intervention Costs

The overall startup cost was divided by the total number of patients diagnosed and treated with the respective treatment regimens in the Stockholm Regional Council and Gotland Region for the years 2016-2018 (518 patients with breast cancer treated with NACT and 683 patients with prostate cancer treated with RT). Given that system updates may incur additional costs beyond the license fees, a time frame of 3 years was considered reasonable. The estimate assumed 5 nurses per 100 patients, with no additional costs for nurses to handle symptom alerts. Based on each patient's number of weeks in treatment (wt), the intervention costs were calculated per equations (1) and (2) for B-RCT and P-RCT, respectively:

$$(5212/518) + ([39 \times 5/100] \times [wt]) + (2.25 \times wt) \quad \textbf{(1)}$$

$$(5212/683) + ([39 \times 5/100] \times [wt]) + (2.25 \times wt) \quad \textbf{(2)}$$

Cost-Utility Analysis

Stochastic CUAs [42] were conducted for each RCT by calculating ICERs in 3 different ways. For ICERa, each patient's intervention cost, along with all health care costs from randomization through 6 months, was included. For ICERb, each patient's intervention costs, plus acute health care costs from randomization and the subsequent 6 months, were considered. Given the considerable variation in patient health care costs, which introduced substantial uncertainty in the cost-effectiveness estimates, a third ICER (ICERc) was calculated by dividing the intervention group's intervention costs minus the control group's intervention costs by the difference in QALYs lost between the 2 groups. The rationale behind this approach is that patients' health care utilization during cancer treatment is influenced by multiple factors, and a much larger study would be necessary to demonstrate a significant reduction in health care costs. Therefore, it was deemed appropriate to assess cost-effectiveness under the assumption that health care costs are not substantially affected. To capture the gradual change in the quality of life during treatment, QALYs lost were calculated linearly as follows: ([EQ-5DP after treatment minus EQ-5DP before treatment]/2) × (individual treatment duration in weeks/52). For visualization, bootstrap values of the incremental intervention costs and incremental health outcomes (QALYs) were plotted on cost-effectiveness planes.

In the P-RCT, the RT treatment was standardized with minimal variation between patients, so RT costs were excluded from both CUAs. By contrast, the B-RCT did not allow for standardized subtraction of treatment costs, so all health care costs were included. The analysis was conducted from the payer perspective (Stockholm Region Council) and focused on the patient's treatment duration (less than 1 year), meaning that no discounting of costs or results was applied. The cost per QALY, as defined by the Swedish NBHW, was used to evaluate cost-effectiveness.

ICERa=[(intervention costs + IG total health care costs) – (CG total health care costs)]/(IG change in QALY – CG change in QALY)

ICERb=[(intervention costs + IG acute health care costs) – (CG acute health care costs)]/(IG change in QALY – CG change in QALY)

ICER=[(IG intervention costs) – (IG intervention costs)]/(IG change in QALY – CG change in QALY)

Exploration of Health Care Utilization and Costs

Within each RCT, variables for total and acute health care visits and costs were generated by summing each participant's visits and costs, conditional on the VAL variable AKUT (acute) being marked as yes or no. Additionally, variables for health care utilization related to the respective cancer treatments were created through a qualitative analysis of the International Statistical Classification of Diseases and Related Health Problems (ICD) codes, using a conventional and summative approach [43]. All ICD codes for acute outpatient and inpatient visits within each RCT were compiled in Excel sheets, and the occurrence of all unique codes was counted. These ICD codes were either grouped or coded based on similarities into predefined and emerging categories. Examples of these categories include fever/neutropenia (D709C, R502, R508, and R509), gastroenteritis/colitis (K521 and A047), anemia (D649), urinary tract infection (N390), and urinary problems (R339, N390, R301, N390X, N304, N109, T830, R391, R319, and N300; Multimedia Appendix 7).

Each patient's visits and costs, according to the categories, were calculated to create variables used as dependent outcomes in multivariate regression analysis. Depending on the level of overdispersion, Poisson, negative binomial, or binary logistic models with a log-link function were fitted [44]. The variable "Group" was coded as control=0 and intervention=1. Prior studies have suggested an association between diminished performance status [45,46], the presence of multiple chronic diseases in older individuals [47], and increased costs. Therefore, the continuous variables—age at inclusion, Charlson Comorbidity Score, and Baseline EQ-5DP score—were included as covariates. The reference category was arranged in ascending order. For the B-RCT, each patient's number of NACT cycles was included as an independent variable. By contrast, for the prostate cancer trial, treatment was standardized, and all patients underwent a similar number of treatments.

Results

B-RCT

Health Outcome

The mean EQ-5DP before treatment was 0.86 in the intervention group and 0.87 in the control group. After treatment, the mean EQ-5DP was 0.84 in the intervention group and 0.80 in the

control group (*P*=.036, effect size=0.099). A statistically significant difference was observed in the mean changes in EQ-5DP from before to after treatment between the intervention and control groups (*P*=.012, effect size=0.042). The greatest difference in change was observed in the Anxiety/Depression dimensions (Multimedia Appendix 8). The CONSORT (Consolidated Standards of Reporting Trials) checklist is presented in Multimedia Appendix 9 (also see Multimedia Appendices 1 and 3).

Cost-Utility Analysis

The intervention group patients had a mean QALY loss of –0.004 (SD 0.002) from before treatment to after treatment, while the corresponding figure for patients in the control group was –0.012 (SD 0.002; *P*<.001). The mean cost for the Interaktor app per patient was €92 (SD €2). The mean total cost for the intervention and all health care was €36,882 (SD €1032) for patients in the intervention group and €35,427 (SD €959) for control group patients (*P*<.001). The ICERa was €202,368 (€SD 811,136; 95% CI €152,008-€252,728). The mean cost for the intervention and acute health care was €3585 (SD €480) in the intervention group and €3235 (SD €494) in the control group (*P*<.001). ICERb was €49,903 (SD €207,042; 95% CI €37,049-€62,758; Table 1). Lastly, when health care costs were excluded from the analysis, the ICERc was €13,213 (SD €33,327; 95% CI €11,145-€15,281; Table 1).

Table 1. Breast cancer trial cost-utility analysis.

	Intervention group (n=74)	Control group (n=75)			
	Mean (SD)	Mean (SD)	*P* value	*t* test (*df*)[a]	95% CI
Health utility					
QALYs[b,c]	–0.004 (0.002)	–0.012 (0.002)	<.001	80 (1998)	0.0074-0.0078
Incremental QALYs[b]	0.0076 (0.003)	N/A[d]	N/A	N/A	N/A
Costs (€[e])					
Intervention costs	92 (2)	N/A	N/A	N/A	N/A
All health care costs (€)					
Outpatient	27,571 (6392)	26,348 (5800)	N/A	N/A	N/A
Inpatient	9207 (5254)	9093 (5460)	N/A	N/A	N/A
Total	36,882 (1032)	35,427 (959)	<.001	33 (1987)	33,530-37,351
Incremental[b]	1454 (1386)	N/A	N/A	N/A	N/A
Acute health care costs (€)					
Outpatient	554 (597)	562 (606)	N/A	N/A	N/A
Inpatient	2932 (4023)	2665 (3992)	N/A	N/A	N/A
Total	3585 (480)	3235 (494)	<.001	16 (1998)	2242-4214
Incremental[b]	353 (676)	N/A	N/A	N/A	N/A
ICERa[b,f]	202,368 (811,136)	N/A	N/A	N/A	152,008-252,728
ICERb[b]	49,903 (207,042)	N/A	N/A	N/A	37,049-62,758
ICERc[b]	13,213 (33,327)	N/A	N/A	N/A	11,145-15,281

[a]Independent unpaired samples Student *t* test (2-tailed).

[b]Based on bootstrap.

[c]QALY: quality-adjusted life year.

[d]N/A: not applicable.

[e]€1=US $1.03.

[f]ICER: incremental cost-effectiveness ratio.

Figure 1 presents a cost-effectiveness plane depicting the bootstrapped values of the intervention group's joint incremental costs and incremental QALYs compared with the control group, as per ICERa and ICERb. Figure 2 illustrates the cost-effectiveness plane with the corresponding values based on ICERc.

Figure 1. Breast cancer cost-effectiveness plane ICERa and ICERb. ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year.

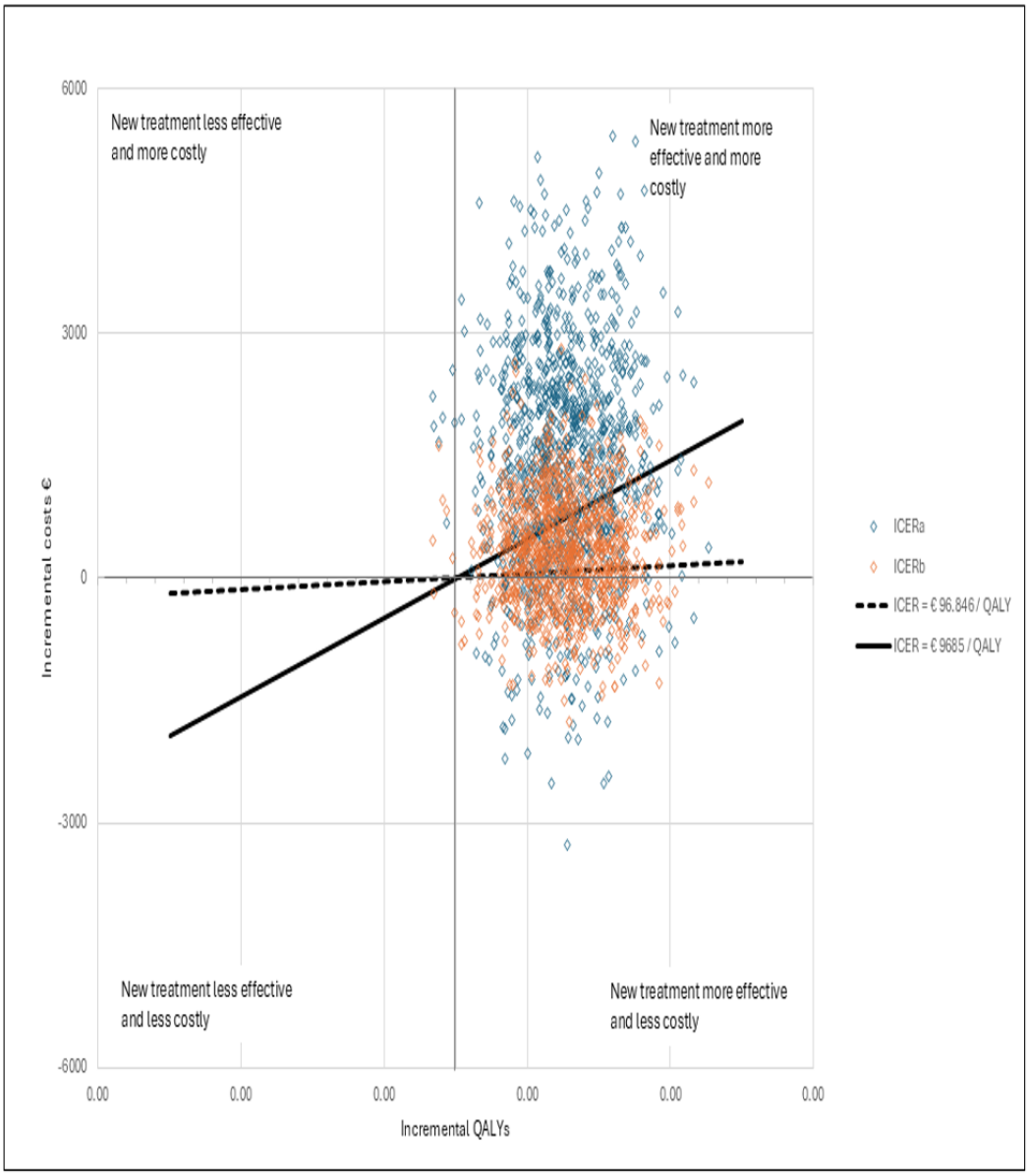
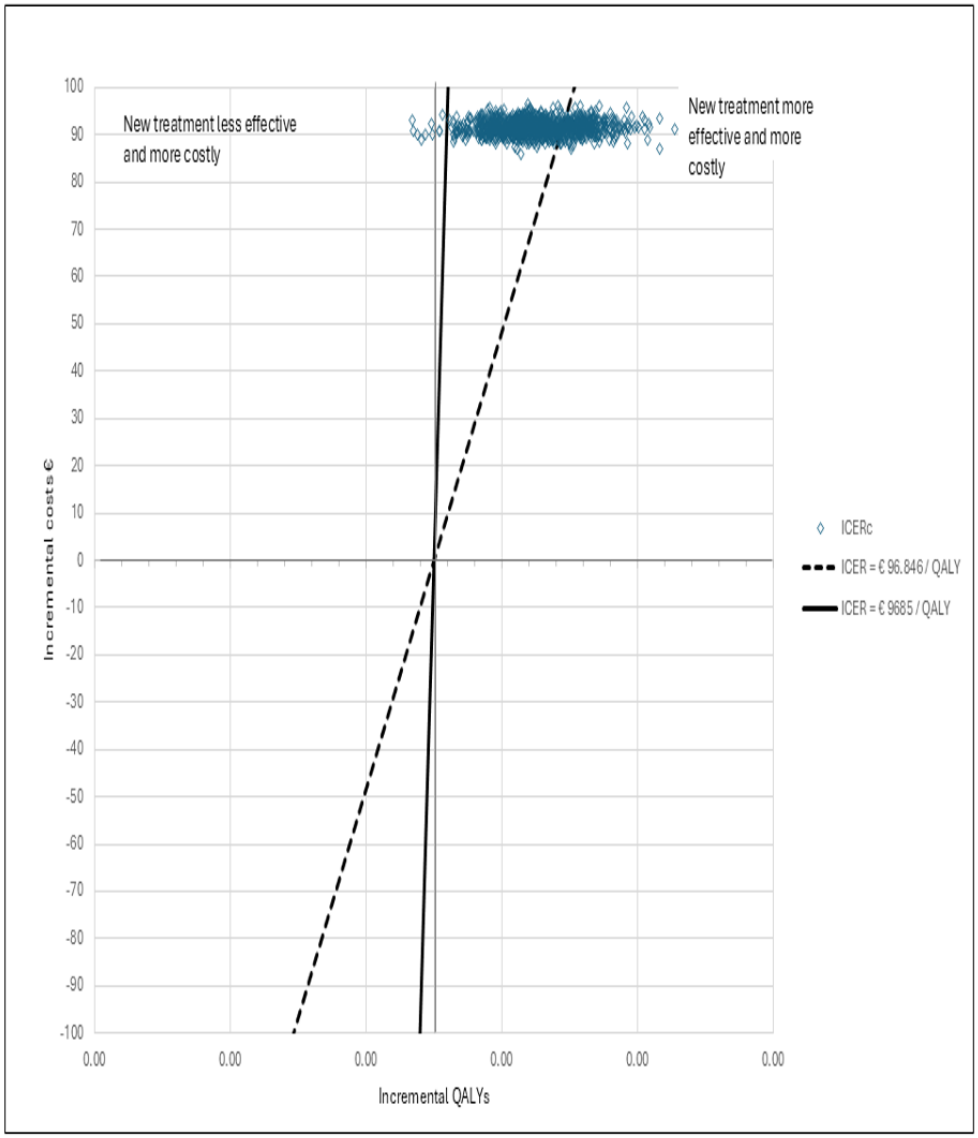


Figure 2. Breast cancer cost-effectiveness plane ICERc. ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year.



Exploration of Health Care Utilization and Health Care Costs

The mean outpatient cost for patients in the intervention group was €27,571 (SD €6392), while in the control group, it was €26,348 (SD €5800). In both groups, approximately 2% of the outpatient costs were attributable to acute care: €554 out of €29,321 (1.9%) in the intervention group and €562 out of €26,348 (2.13%) in the control group.

In the intervention group, 13 out of 74 (18%) patients had an acute outpatient visit for fever, with a total of 34 visits. In the control group, the corresponding proportion was 9 out of 75 (12%), with a total of 21 visits. Additionally, 7 out of 74 (9%) patients in the intervention group had an unplanned admission from outpatient to inpatient care, accounting for 37 unplanned admissions. In the control group, 6 out of 75 (8%) had an unplanned admission from outpatient to inpatient care, totaling 29 unplanned admissions.

The mean inpatient cost per patient was €9207 (SD €5254) in the intervention group and €9093 (SD €5460) in the control group. Approximately one-third of all inpatient care cost was acute in both groups (€2932/€9207, 31.85% in the intervention group and €2665/€9093, 29.31% in the control group). The most common diagnoses during acute inpatient care episodes in both groups were fever, gastroenteritis/colitis, anemia, and urinary tract infection. The variable group (intervention/control) was not associated with the number of visits for fever, gastroenteritis/colitis, anemia, or urinary tract infection, nor were age, health-related quality of life (HRQOL) before treatment, comorbidities, or the number of NACT (Table 2).

Table 2. Breast cancer trial multivariate regression analysis of predictors for acute healthcare visits for chemotherapy-related symptoms

	Intervention group (n=74)	Control group (n=75)	B	SE	Standardized coefficient [Exp(B)]	95% CI	*P* value	χ^2/df[a]
Visits								
Acute outpatient								
Fever[b] M (SD)	0.46 (0.86)	0.28 (0.63)	N/A[c]	N/A	N/A	N/A	N/A	N/A
Dependent variable: total, n	34 (n=13)	21 (n=9)	N/A	N/A	N/A	N/A	N/A	N/A
Independent variables:	N/A	N/A	N/A	N/A	N/A	N/A	.49[d]	1.141
Group[e]	N/A	N/A	–0.46	0.33	0.63	0.331-1.202	.16[f]	N/A
Age	N/A	N/A	–0.01	0.02	0.992	0.949-1.037	.72[f]	N/A
HRQOL	N/A	N/A	–0.70	1.93	0.498	0.011-22.106	.72[f]	N/A
Comorbidity	N/A	N/A	0.15	0.20	1.156	0.773-1.727	.48[f]	N/A
NACT[g]	N/A	N/A	–0.08	0.07	0.92	0.800-1.058	.24[f]	N/A
Unplanned admissions M (SD)	0.50 (71)	0.39 (0.72)	N/A	N/A	N/A	N/A	N/A	N/A
Dependent variable: total, n	37 (n=7)	29 (n=6)	N/A	N/A	N/A	N/A	N/A	N/A
Independent variables	N/A	N/A	N/A	N/A	N/A	N/A	.52[d]	.877
Group[e]	N/A	N/A	–0.220	0.302	0.803	0.444-1.451	.47[f]	N/A
Age	N/A	N/A	0.010	0.021	1.010	0.969-1.054	.62[f]	N/A
HRQOL	N/A	N/A	–3.048	1.760	0.047	0.002-1.494	.08[f]	N/A
Comorbidity	N/A	N/A	0.029	0.193	1.030	0.705-1.504	.88[f]	N/A
NACT[g]	N/A	N/A	–0.036	0.061	0.965	0.857-1.086	.55[f]	N/A
Acute inpatient								
All M (SD)	0.73 (0.98)	0.60 (0.82)	N/A	N/A	N/A	N/A	N/A	N/A
Dependent variable: total, n	54 (n=35)	45 (n=32)	N/A	N/A	N/A	N/A	N/A	N/A
Independent variables:	N/A	N/A	N/A	N/A	N/A	N/A	.72[d]	.722
Group[e]	N/A	N/A	–0.182	0.2636	0.834	0.497-1.398	.49[f]	N/A
Age	N/A	N/A	0.006	0.0187	1.006	0.969-1.043	.76[f]	N/A
HRQOL	N/A	N/A	–1.291	1.634	0.275	0.011-6.768	.43[f]	N/A
Comorbidity	N/A	N/A	0.055	0.1702	1.057	0.757-1.476	.74[f]	N/A
NACT[g]	N/A	N/A	–0.068	0.0568	0.934	0.836-1.044	.23[f]	N/A
Fever[b] M (SD)	0.39 (0.70)	0.32 (0.55)	N/A	N/A	N/A	N/A	N/A	N/A
Dependent variable: Total, n	29 (n=22)	24 (n=21)	N/A	N/A	N/A	N/A	N/A	N/A
Independent variables:	N/A	N/A	N/A	N/A	N/A	N/A	.94[d]	.838
Group[e]	N/A	N/A	–0.20	0.32	0.815	0.433-1.535	.53[f]	N/A
Age	N/A	N/A	0.01	0.02	1.01	0.965-1.057	.66[f]	N/A
HRQOL	N/A	N/A	–1.04	1.94	0.353	0.008-15.75	.59[f]	N/A
Comorbidity	N/A	N/A	–0.13	0.23	0.876	0.557-1.378	.57[f]	N/A
NACT[g]	N/A	N/A	–0.01	0.07	0.988	0.868-1.124	.86[f]	N/A
Gastroenteritis[b] M (SD)	0.08 (0.36)	0.16 (0.40)	N/A	N/A	N/A	N/A	N/A	N/A

	Intervention group (n=74)	Control group (n=75)	B	SE	Standardized coefficient [Exp(B)]	95% CI	*P* value	χ^2/df[a]
Dependent variable: total, n	6 (n=4)	12 (n=11)	N/A	N/A	N/A	N/A	N/A	N/A
Independent variables:	N/A	N/A	N/A	N/A	N/A	N/A	.77[d]	1.186
Group[e]	N/A	N/A	0.66	0.53	1.931	0.686-5.440	.21[f]	N/A
Age	N/A	N/A	0.02	0.04	1.019	0.944-1.100	.63[f]	N/A
HRQOL	N/A	N/A	1.83	3.35	6.245	0.009-4422.840	.58[f]	N/A
Comorbidity	N/A	N/A	–0.31	0.41	0.737	0.328-1.657	.46[f]	N/A
NACT[g]	N/A	N/A	0.004	0.100	1.004	0.826-1.220	.97[f]	N/A
Anemia[b] M (SD)	0.05 (0.23)	0.05 (0.23)	N/A	N/A	N/A	N/A	N/A	N/A
Dependent variable: total, n	4 (n=4)	4 (n=4)	N/A	N/A	N/A	N/A	N/A	N/A
Independent variables:	N/A	N/A	N/A	N/A	N/A	N/A	.73[d]	.961
Group[e]	N/A	N/A	-0.10	0.74	0.901	0.210-3.863	.89[f]	N/A
Age	N/A	N/A	0.10	0.07	1.101	0.963-1.258	.16[f]	N/A
HRQOL	N/A	N/A	1.85	4.74	6.346	0.001-68734.141	.70[f]	N/A
Comorbidity	N/A	N/A	-0.94	0.73	0.389	0.093-1.633	.20[f]	N/A
NACT[g]	N/A	N/A	0.03	0.14	0.029	0.780-1.357	.89[f]	N/A
Urinary tract infection[b] M (SD)	0.04 (0.20)	0.05 (0.03)	N/A	N/A	N/A	N/A	N/A	N/A
Dependent variable: total, n	3 (n=3)	4 (n=3)	N/A	N/A	N/A	N/A	N/A	N/A
Independent variables:	N/A	N/A	N/A	N/A	N/A	N/A	.31[d]	1.268
Group[e]	N/A	N/A	–0.09	0.84	0.911	0.177-4.702	.91[f]	N/A
Age	N/A	N/A	0.03	0.05	1.030	0.926-1.146	.58[f]	N/A
HRQOL	N/A	N/A	4.01	6.39	55.11	0.000-15250488	.53[f]	N/A
Comorbidity	N/A	N/A	0.29	0.43	1.34	0.571-3.144	.50[f]	N/A
NACT[g]	N/A	N/A	0.07	0.14	1.08	0.818-1.426	.59[f]	N/A

[a]Pearson χ2 value divided by degrees of freedom (goodness of fit of the model).

[b]Nr of acute visits when a patient received the diagnose.

[c]N/A: not applicable.

[d]*P* value Omnibus test General Linear Model Negative Binomial Regression.

[e]Intervention/Control; Reference category=Intervention

[f]*P* value for the independent variable in the General Linear Model Negative Binomial Regression.

[g]Neoadjuvant chemotherapy treatments

Negative binomial multivariate regression analysis revealed that the independent variable group (intervention/control) did not significantly affect the predicted log odds of patients' health care costs (*P*=.949). For acute outpatient health care costs, the analysis showed that age, health-related quality of life at baseline, and comorbidities significantly predicted costs. Older age (*P*=.002) and better health-related quality of life (*P*<.001) were associated with lower acute outpatient health care costs. By contrast, a higher number of comorbidities was associated with increased acute health care costs (*P*=.02; Table 3).

Table 3. Breast cancer trial multivariate regression analysis of predictors for health care costs.

	Intervention group (n=74)	Control group (n=75)	Unstandardized coefficient	SE	Standardized coefficient [Exp(B)]	95% CI	*P* value	χ^2/df[a]
Care costs (€[b])								
Outpatient								
All costs Mean (SD)	27,571 (6392) n=74	26,348 (5800) n=75	N/A[c]	N/A	N/A	N/A	N/A	N/A
Dependent variable: Total costs	2,040,225	1,976,097	N/A	N/A	N/A	N/A	.95[d]	0.042
Independent variables:								
Group[e]	N/A	N/A	–0.038	0.166	0.963	0.695-1.334	.82[f]	N/A
Age	N/A	N/A	–0.001	0.012	0.999	0.976-1.023	.94[f]	N/A
HRQOL	N/A	N/A	–0.78	1.019	0.458	0.062-3.375	.44[f]	N/A
Comorbidity	N/A	N/A	–0.012	0.115	0.988	0.789-1.239	.92[f]	N/A
NACT[g]	N/A	N/A	0.014	0.033	1.014	0.951-1.081	.67[f]	N/A
Acute costs Mean (SD)	554 (597) n=51	562 (606) n=52	N/A	N/A	N/A	N/A	N/A	N/A
Dependent variable: Acute costs	40,998	42,183	N/A	N/A	N/A	N/A	<.001[d]	1.170[a]
Independent variables:								
Group[e]	N/A	N/A	0.043	0.166	1.044	0.754-1.446	.80[f]	N/A
Age	N/A	N/A	–0.036	0.012	0.964	0.942-0.987	.002[f]	N/A
HRQOL	N/A	N/A	–4.091	1.108	0.017	0.002-0.147	<.001[f]	N/A
Comorbidity	N/A	N/A	0.247	0.110	1.280	1.033-1.588	.02[f]	N/A
NACT[g]	N/A	N/A	–0.062	0.032	0.940	0.882-1.001	.05[f]	N/A
Inpatient								
All costs Mean (SD)	9206 (5254) n=69	9093 (5460) n=73	N/A	N/A	N/A	N/A	N/A	N/A
Dependent variable: Total costs	681,221	681,949	N/A	N/A	N/A	N/A	.84[d]	0.341[a]
Independent variables:								
Group[e]	N/A	N/A	0.008	0.165	1.008	0.73-1.393	.96[f]	N/A
Age	N/A	N/A	–0.004	0.012	0.996	0.973-1.018	.70[f]	N/A
HRQOL	N/A	N/A	–1.041	1.044	0.353	0.046-2.733	.32[f]	N/A
Comorbidity	N/A	N/A	–0.011	0.111	0.989	0.796-1.229	.92[f]	N/A
NACT[g]	N/A	N/A	–0.01	0.033	0.990	0.928-1.056	.77[f]	N/A

[a]€1=US $1.03.

[b]N/A: not applicable.

[c]*P* value Omnibus test General Linear Model Multivariate Regression.

[d]Pearson χ2 value divided by degrees of freedom (goodness of fit of the model).

[e]Intervention/Control; Reference category=Intervention

[f]*P* value for the independent variables in the General Linear Model Multivariate Regression Model.

[g]Neoadjuvant chemotherapy treatments

P-RCT

Health Outcome

The mean EQ-5DP before treatment was 0.88 in the intervention group and 0.89 in the control group. After treatment, the mean EQ-5DP was 0.87 in the intervention group and 0.88 in the control group (*P*=.51). The mean difference in EQ-5DP from before to after treatment was not statistically significant between the intervention and control groups (*P*=.94). The most prominent differences in change were observed in the dimensions of Pain/Discomfort and Anxiety/Depression (Multimedia Appendix 5).

Cost-Utility Analysis

The intervention group patients scored a mean QALY loss of –0.0008 (SD 0.0006), while the corresponding figure for the control group was –0.0009 (SD 0.0006; *P*<.001). The mean total cost of the Interaktor intervention per patient was €43 (SD €0.2). The mean total cost, including the intervention and all health care, was €3419 (SD €739) for the intervention group and €3537 (SD €689) for the control group. The ICERa was –€1,092,136 (SD €35,155,229; 95% CI –€3,274,774 to –€1,090,502). The mean total costs for the intervention and acute health care were €1219 (SD €593) in the intervention group and €802 (SD €281) in the control group. The ICERb was €745,987 (SD €16,006,924; 95% CI –€247,317 to €1,739,292). Lastly, when health care costs were excluded from the analysis, the ICERc was €13,118 (SD €1,314,743; 95% CI –€68,468 to €94,704; Table 4).

Table 4. Prostate cancer trial cost-utility analysis.

	Intervention group (n=75) Mean (SD)	Control group (n=75) Mean (SD)	*P* value[a]	*t* test (*df*)	95% CI
Health utility					
QALYs[b,c]	–0.0008 (0.0006)	–0.0009 (0.0006)	<.0001	6.419 (1998)	0.0001 to 0.0002
Incremental QALYs[b]	0.0002 (0.0008)	N/A[d]	N/A	N/A	N/A
Costs (€[e])					
Intervention costs	43 (0.2)	N/A	N/A	N/A	N/A
All health care costs[f] (€)					
Outpatient	2077 (1386)	2488 (2403)	N/A	N/A	N/A
Inpatient	1321 (5460)	1049 (4240)	N/A	N/A	N/A
Total	3419 (739)	3537 (689)	.0002	–4 (1988)	–183 to –57
Incremental[b]	–120 (1034)	N/A	N/A	N/A	N/A
Acute health care costs (€)					
Outpatient	121 (247)	126 (258)	N/A	N/A	N/A
Inpatient	1054 (5132)	684 (2335)	N/A	N/A	N/A
Total	1219 (593)	802 (281)	<.0001	20 (1426)	376 to 458
Incremental[b]	417 (659)	N/A	N/A	N/A	N/A
ICERa[b]	–1,092,136 (35,155,229)	N/A	N/A	N/A	–3,274,774 to 1,090,502
ICERb[b]	745,987 (16,006,924)	N/A	N/A	N/A	–247,317 to 1,739,292
ICERc[b]	13,118 (1,314,743)	N/A	N/A	N/A	–68,468 to 94,704

[a]Independent unpaired samples Student t test (2-tailed).

[b]Based on bootstrap.

[c]QALY: quality-adjusted life year.

[d]N/A: not applicable.

[e]€1=US $1.03.

[f]Excluding radiotherapy costs.

gICER: incremental cost-effectiveness ratio.

Figure 3 presents a cost-effectiveness plane showing the bootstrapped values of the intervention group's joint incremental costs and incremental QALYs compared with the control group for ICERa and ICERb. Figure 4 displays the cost-effectiveness plane with the corresponding values for ICERc.

Figure 3. Prostate cancer cost-effectiveness plane ICERa and ICERb. ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year.

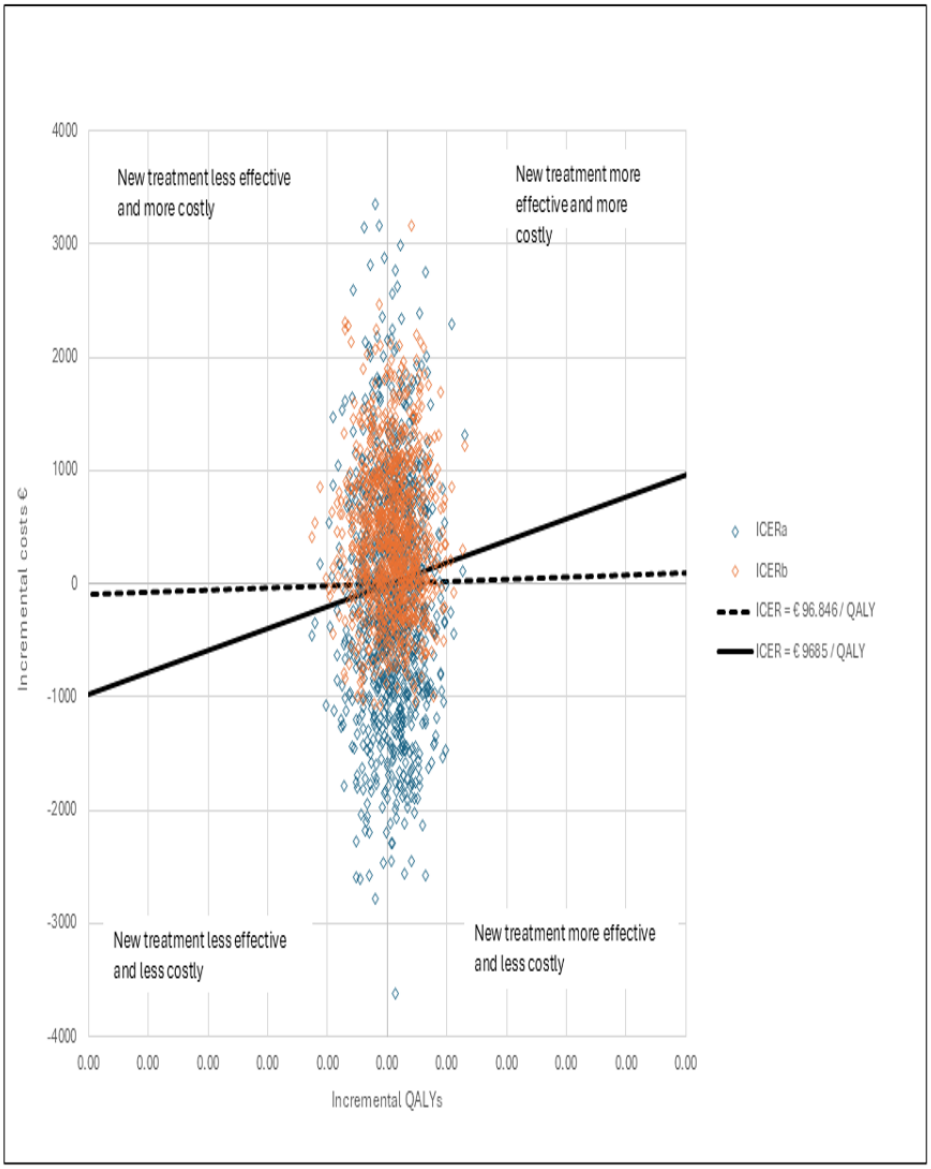
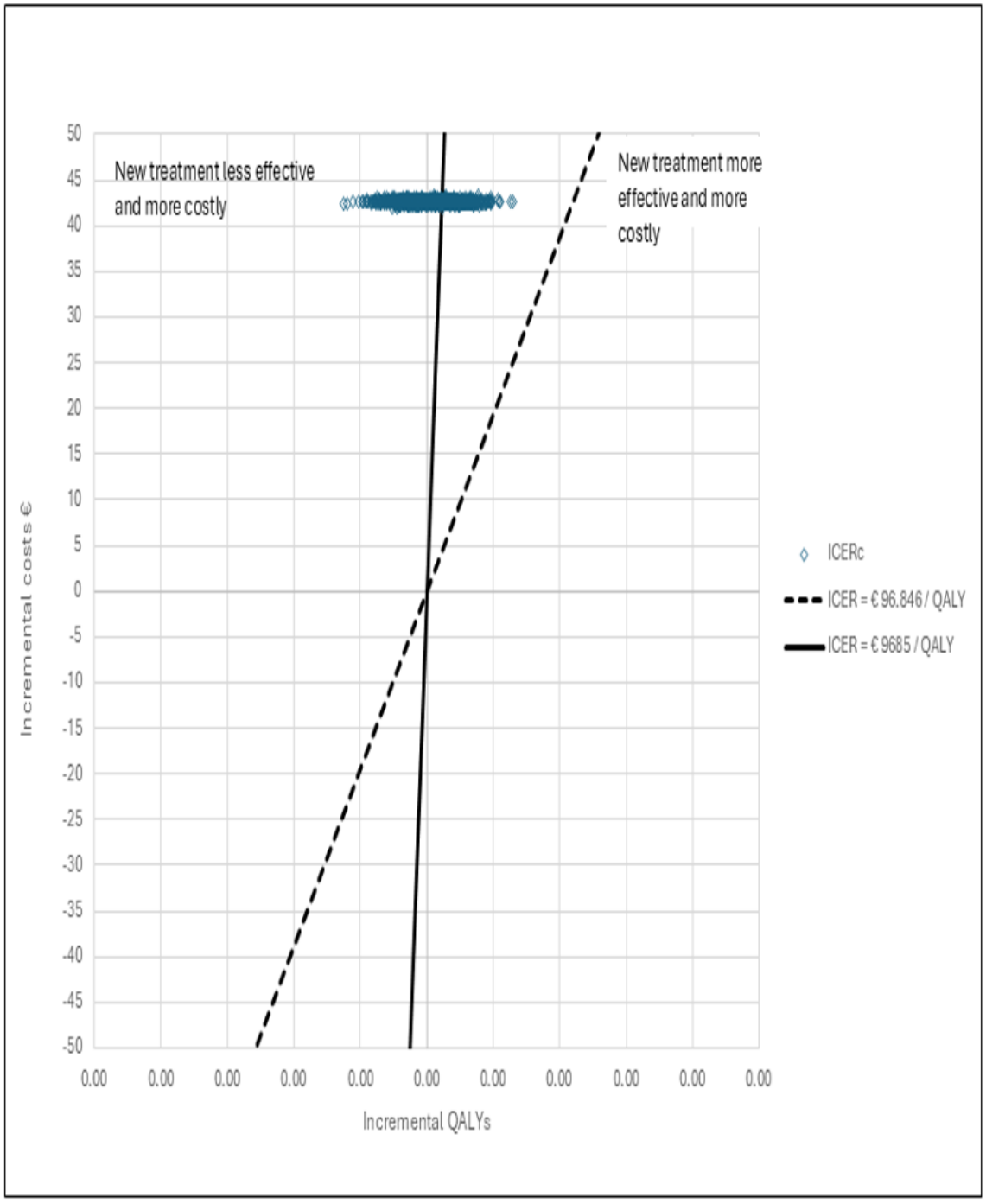


Figure 4. Prostate cancer cost-effectiveness plane ICERc. ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year.



Exploration of Health Care Utilization and Health Care Costs

The mean outpatient cost for patients in the intervention group was €2077 (SD €1386), while in the control group, it was €2488 (SD €2403); €121 out of €2077 (5.83%) outpatient cost in the intervention group was for acute care, compared with €126 out of €2488 (5.06%) in the control group. In both groups, 25 patients (33%) had an acute outpatient care visit.

Acute outpatient care for urological problems was required by 5 of 75 (7%) patients (7 visits) in the intervention group and 6 of 75 (8%) patients (14 visits) in the control group (Table 5). Regarding acute outpatient visits for urological problems that resulted in unplanned admissions from outpatient to inpatient care, this occurred for 1 patient in the intervention group (1 admission) and 2 patients in the control group (3 admissions).

Table 5. Prostate cancer trial multivariate regression analysis of predictors for health care visits.

	Intervention group (n=75)	Control group (n=75)	B	SE	Standardized coefficient [Exp(B)]	95% CI	*P* value	χ^2/df[a]
Variables								
Outpatient care visits								
Dependent variable: total visits, n	739	851	N/A[b]	N/A	N/A	N/A	.008[c]	.848
Independent variables:								
Group[d]	N/A	N/A	0.190	0.172	1.210	0.864-1.694	.27[e]	N/A
Age	N/A	N/A	0.019	0.016	1.019	0.989-1.051	.22[f]	N/A
HRQOL	N/A	N/A	–3.439	1.135	0.032	0.003-0.297	.002[f]	N/A
Comorbidity	N/A	N/A	0.007	0.072	1.007	0.875-1.159	.92[f]	N/A
Dependent variable: patients with acute visit, n	25	25					.88[f]	1.033
Independent variables:								
Group[g]	N/A	N/A	–0.017	0.350	0.983	0.495-1.953	.96[h]	N/A
Age	N/A	N/A	0.017	0.030	1.017	0.959-1.079	.56[h]	N/A
HRQOL	N/A	N/A	0.078	2.370	1.081	0.010-112.574	.97[h]	N/A
Comorbidity	N/A	N/A	0.078	0.150	1.081	0.807-1.450	.60[h]	N/A
Dependent variable: patients with acute visit for urological problems, n	5	6	N/A	N/A	N/A	N/A	.23[f]	1.048
Independent variables:								
Group[g]	N/A	N/A	–0.127	0.620	0.881	0.261-2.968	.84[h]	N/A
Age	N/A	N/A	–0.022	0.057	0.978	0.875-1.093	.69[h]	N/A
HRQOL	N/A	N/A	8.368	3.532	4304.883	4.239-4,371,830	.02[h]	N/A
Comorbidity	N/A	N/A	0.142	0.284	1.152	0.661-2.008	.62[h]	N/A
Dependent variable: patients with an unplanned admission, n	7	6	N/A	N/A	N/A	N/A	.43[f]	1.038
Independent variables:								
Group[g]	N/A	N/A	0.080	0.593	1.083	0.339-3.465	.89[h]	N/A
Age	N/A	N/A	-0.069	0.060	0.933	0.830-1.049	.25[h]	N/A
HRQOL	N/A	N/A	5.134	3.481	169.738	0.185-155,848.181	.14[h]	N/A
Comorbidity	N/A	N/A	0.030	0.252	1.03	0.628-1.69	.90[h]	N/A
Inpatient care visits								
Dependent variable: patients with a visit, n	11	10					.12[f]	1.068
Independent variables:								
Group[g]	N/A	N/A	0.027	0.487	1.027	0.396-2.665	.96[h]	N/A
Age	N/A	N/A	–0.065	0.049	0.937	0.852-1.030	.18[h]	N/A

		Intervention group (n=75)	Control group (n=75)	B	SE	Standardized coefficient [Exp(B)]	95% CI	*P* value	χ^2/df[a]
	HRQOL	N/A	N/A	6.017	2.918	410.398	1.349-124,898.645	.04[h]	N/A
	Comorbidity	N/A	N/A	–0.064	0.198	0.938	0.637-1.382	.75[h]	N/A

[a]Pearson χ2 value divided by degrees of freedom (goodness of fit of the model).

[b]N/A: not applicable.

[c]*P* value Omnibus test General Linear Model Multivariate Regression.

[d]Intervention/Control; Reference category=Intervention

[e]*P* value for the independent variable in the General Linear Model Multivariate Regression.

[f]*P* value Omnibus test Binary Logistic Regression Model

[g]Intervention/Control; reference category=Control

[h]*P* value for the independent variable in the Binary Logistic Regression Model.

The mean inpatient cost was €1321 (SD €5460) in the intervention group, compared with €1049 (SD €4240) in the control group. In the intervention group, €1054 out of €1321 (79.79%) inpatient care costs were attributed to acute care, while the corresponding figure in the control group was €684 out of €1049 (65.20%). The most common diagnoses during acute inpatient care episodes in the intervention group were R060 dyspnea (4 episodes) and I214 acute subendocardial infarction (3 episodes). In the control group, the most common diagnoses were N390 urinary tract infection (3 episodes) and anemia D630/D649 (2 episodes).

Negative binomial multivariate regression analysis revealed that higher HRQOL before treatment was associated with a decrease in the number of outpatient care visits (*P*=.002). Multivariate binary logistic regression indicated that higher HRQOL before treatment was negatively associated with both having an acute outpatient care visit for urological problems (*P*=.02) and having an inpatient care episode (*P*=.04). Therefore, patients with better HRQOL before treatment were less likely to have an acute outpatient visit for urological problems and less likely to experience an inpatient care episode (Table 5).

Discussion

Principal Findings

In both RCTs, patients in the intervention groups experienced fewer QALYs lost compared with those in the control groups. In the B-RCT, the quality of life of patients in the intervention group decreased significantly less during treatment compared with the control group; however, this was not observed in the P-RCT. The reduction in QALYs lost was achieved with a low intervention cost per patient. Unfortunately, while ICERs are particularly valuable for guiding decision makers, the ICERs from this study are somewhat difficult to interpret due to uncertainty, as illustrated in the cost-effectiveness planes. The variability in the ICERs arose from differences in patients' health care costs. When health care costs were excluded from the analysis, the RCTs showed ICERs slightly above the NBHW threshold for a low cost per QALY but still well below what the NBHW considers a high cost per QALY. These findings are encouraging and support the conceptual foundation of ePRO [48].

Comparison With Prior Work

Although the P-RCT showed a decrease in QALYs lost, the difference was relatively small. This finding is not uncommon. Snoswell et al [49] reviewed 25 cost-utility studies of telehealth interventions that reported costs from the health system perspective and changes in HRQOL. About one-third of these studies demonstrated cost savings and changes in effect, but most QALY improvements were marginal (range 0.0006-0.12). The authors concluded that this may be partly due to HRQOL instruments being neither sensitive nor appropriate for detecting the effects of changes in health service delivery. Demonstrating substantial cost savings from ePRO during curative cancer treatment may be challenging, given the relatively short time frame and the high variability in health care utilization, which necessitate large sample sizes. However, because productivity loss due to morbidity and mortality represents the most significant societal cost of cancer [50], further research should explore whether life quality improvements enable patients to continue working during treatment or return to work earlier and to a greater extent after treatment. Such a cost-effectiveness analysis could reveal societal cost savings.

There are a few studies available to compare with our findings, as CUAs of ePRO interventions remain limited [27]. We did not identify any studies evaluating the cost-effectiveness of ePRO during first-line curative cancer treatment. However, some studies have been conducted in the context of follow-up and advanced cancer care. For instance, Nixon et al [51] and Lizée et al [28] reported relatively low ICERs for ePRO in cancer survivors, whereas Van der Hout et al [52] found a small positive effect on HRQOL but no significant differences in direct or indirect medical costs among cancer survivors.

Evidence suggests that ePRO can reduce health care utilization [27]. For example, ePRO has been shown to positively impact outcomes such as emergency room visits, hospitalizations, and readmissions [27]. However, not all studies demonstrate these effects. Barbera et al [53], in a study conducted during adjuvant chemotherapy for breast cancer, did not observe a reduction in hospitalizations or readmissions. Similarly, Wheelock et al [54]

investigated the impact of ePRO during follow-up care after breast cancer treatment and found no reduction in health care resource use, including oncology-related appointments, physician visits, or medical tests. Lizée et al [28] observed a higher number of follow-up clinic visits in the ePRO intervention group compared with the control group. However, the intervention group also experienced longer overall survival, allowing more time for follow-up visits.

The study highlights the heterogeneity of the cancer population and the variation in health care use not only between but also within patient populations. For instance, nearly 104 of 149 (69.8%) patients treated for breast cancer made acute outpatient visits, whereas only about one-third (50/150) of patients with prostate cancer required such visits. This may be linked to the continuous health care contact that patients undergoing external RT maintain. In the context of older patients receiving cancer treatment, Nipp et al [55] demonstrated that age moderated the positive effects of ePRO on both ER visits and survival in patients with advanced cancer.

By analyzing ICD codes documented during health care visits to regional health care organizations, this study revealed that 5 of 75 (7%) intervention group patients and 6 of 75 (8%) control group patients undergoing treatment for prostate cancer had an acute outpatient visit for urological problems. Similarly, 4 of 74 (5%) intervention group patients and 11 of 75 (15%) control group patients treated for breast cancer had an acute outpatient visit for gastrointestinal symptoms. However, health care visits to general practitioners' clinics and health centers lacked ICD codes, meaning these figures may not fully capture the patients' health care utilization for those symptoms.

In this study, intervention group patients treated for breast cancer had more acute outpatient visits for fever/neutropenia, although the difference was not statistically significant. A similar increase in neutropenic events was observed by Absolom et al [10], who evaluated an intervention for patients undergoing chemotherapy, which aligns with national recommendations for managing chemotherapy patients presenting with this symptom [56]. In this study, the number of acute inpatient care episodes for fever/neutropenia was similar between the intervention and control groups.

Previous research has shown that health care utilization and costs during cancer treatment are complex [45]. This study examined whether age, comorbidity, and health status significantly influenced patients' health care utilization. As expected, the results suggest that health status has some impact on health care consumption. In the B-RCT, higher HRQOL before treatment was associated with reduced acute outpatient costs. In the P-RCT, higher HRQOL before treatment was associated with a lower likelihood of having an acute outpatient visit or inpatient care episode. Considering the results showing a smaller decline in quality of life among intervention group patients, these findings suggest that patients using the app receive timely and appropriate care, leading to more effective and prompt management of symptoms and adverse events associated with cancer treatment. This interpretation aligns with previously reported positive effects on health-related quality of life and symptom burden [11,19,20].

Limitations

The study presents unique and highly relevant findings for modern outpatient-based, personalized cancer care. A key strength is its randomized design, although some limitations should be noted. First, the ICERs must be interpreted with caution due to the uncertainty illustrated in the cost-effectiveness planes. Regarding costs, additional expenses for nurses handling alerts were not included, as staff interviews indicated that no increase in working hours was necessary (unpublished data). A similar assumption was made in the study by Nixon et al [51]. A recent study of a similar intervention also concluded that the intervention did not increase hospital clinicians' workload [10]. Finally, the Swedish valuation system assigns higher values to most conditions than the British system, presumably because it is based on patients' valuations of their conditions rather than, as in the British case, the public's valuations of hypothetical conditions [57]. Accordingly, the results regarding intervention effectiveness may have differed.

In the P-RCT, the dropout rate was notably high, potentially reducing the statistical power to detect significant differences. The reasons for nonresponse to outcome questionnaires remain unknown. Importantly, all patients used the app daily as instructed, with an adherence rate of 80% [6]. Although debated, imputation aims to accurately estimate the overall data distribution [58]. It is suggested that imputing missing values exceeding 10% increases the risk of bias [34,58]. By contrast, the use of ITT analysis is highly recommended in RCTs [34]. Although there is no specific threshold for missing values in health economic studies, it is emphasized that patterns of missing data should be reported [32]. The sample in our study is too small to analyze patterns, but the ITT principle presumably assumes missing data are random, though other mechanisms may also have contributed [59]. Based on the study design and sample size, a simple imputation method was applied in this study [57].

Health care costs were missing at random due to an administrative error, and values for no more than 2 patients per group were imputed. The risk of overestimating costs due to right-skewed data is therefore small. Given that the data on the EORTC dimension scores are approximately normally distributed, the imputation method appears to be accurate. Nevertheless, further studies are needed.

Conclusions

At a low weekly cost, the intervention reduced QALYs lost. The cost-effectiveness of the intervention, as defined by the ICER in relation to the Swedish NBHW, varied depending on the costs considered. For patients with breast cancer, the intervention was cost-effective when nonacute health care costs were excluded, whereas for patients with prostate cancer, cost-effectiveness was achieved when all health care costs were included. This suggests that the intervention has the potential to achieve cost-effectiveness. However, larger studies are needed, as there was considerable uncertainty regarding the ICERs due to significant variations in patients' health care costs.

Patients in the intervention group with breast cancer had more acute health care visits for neutropenia/fever, whereas more

patients in the control group were hospitalized for gastrointestinal symptoms. Only a few patients with prostate cancer were hospitalized for urological problems. These findings highlight the previously demonstrated positive effects on patients' symptom burden and suggest that the intervention may facilitate timelier and more effective symptom management. Future studies should assess cost-effectiveness from a societal perspective.

Acknowledgments

We thank Dr. Tracey A. Young and Dr. Clara Mukuria for kindly providing the original mapping algorithm syntax. We also extend our gratitude to the Region Stockholm Health Care Administration and the Confederation of Regional Cancer Centres for their supportive assistance with registry data. The study was supported by the Swedish Cancer Foundation (grant 160298), the Swedish Research Council (grant 521-2014-2723), the Kamprad Family Foundation for Entrepreneurship, Research, and Charity (grant 20150015), the Swedish Research Council for Health, Working Life and Welfare (grant 2014-04713), and Karolinska Institutet (internal grant).

Data Availability

The data sets generated and analyzed during this study are not publicly available as they contain information that could compromise the privacy and consent of the research participants. However, the transformed data are available upon reasonable request from the authors.

Disclaimer

The authors confirm that no generative artificial intelligence technology was used in the creation of the text, figures, or other informational content of this manuscript.

Conflicts of Interest

AL-E, KS, and MF were involved in developing the intervention. The others have no conflicts of interest to declare.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) flowchart for the B-RCT. B-RCT: neoadjuvant chemotherapy for breast cancer-randomized controlled trial.
[PDF File (Adobe PDF File), 97 KB - cancer_v11i1e53539_app1.pdf]

Multimedia Appendix 2
Cancer treatment B-RCT and P-RCT. B-RCT: neoadjuvant chemotherapy for breast cancer-randomized controlled trial; P-RCT: radiotherapy for prostate cancer-randomized controlled trial.
[DOCX File , 18 KB - cancer_v11i1e53539_app2.docx]

Multimedia Appendix 3
CONSORT (Consolidated Standards of Reporting Trials) flowchart P-RCT. P-RCT: radiotherapy for prostate cancer-randomized controlled trial.
[PDF File (Adobe PDF File), 96 KB - cancer_v11i1e53539_app3.pdf]

Multimedia Appendix 4
P-RCT complete cases analysis. P-RCT: radiotherapy for prostate cancer-randomized controlled trial.
[DOCX File , 20 KB - cancer_v11i1e53539_app4.docx]

Multimedia Appendix 5
Sociodemographic and clinical characteristics.
[DOCX File , 29 KB - cancer_v11i1e53539_app5.docx]

Multimedia Appendix 6
The SIMKOST (simulerad kostnad/simulated cost) variable.
[DOCX File , 23 KB - cancer_v11i1e53539_app6.docx]

Multimedia Appendix 7
ICD codes categorized. ICD: International Statistical Classification of Diseases and Related Health Problems.

[PDF File (Adobe PDF File), 41 KB - cancer_v11i1e53539_app7.pdf]

Multimedia Appendix 8
EQ-5DP dimensions.
[DOCX File , 20 KB - cancer_v11i1e53539_app8.docx]

Multimedia Appendix 9
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1074 KB - cancer_v11i1e53539_app9.pdf]

References

1. Soerjomataram I, Bray F. Planning for tomorrow: global cancer incidence and the role of prevention 2020-2070. Nat Rev Clin Oncol 2021 Oct 02;18(10):663-672 [FREE Full text] [doi: 10.1038/s41571-021-00514-z] [Medline: 34079102]
2. Aprile G, Pisa FE, Follador A, Foltran L, De Pauli F, Mazzer M, et al. Unplanned presentations of cancer outpatients: a retrospective cohort study. Support Care Cancer 2013 Feb;21(2):397-404 [FREE Full text] [doi: 10.1007/s00520-012-1524-6] [Medline: 22722887]
3. Lash R, Bell J, Reed S, Poghosyan H, Rodgers J, Kim K, et al. A systematic review of emergency department use among cancer patients. Cancer Nurs 2017;40(2):135-144 [FREE Full text] [doi: 10.1097/ncc.0000000000000360]
4. Basch E, Deal A, Kris M, Scher H, Hudis C, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. JCO 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/jco.2015.63.0830]
5. Trojan A, Leuthold N, Thomssen C, Rody A, Winder T, Jakob A, et al. The effect of collaborative reviews of electronic patient-reported outcomes on the congruence of patient- and clinician-reported toxicity in cancer patients receiving systemic therapy: prospective, multicenter, observational clinical trial. J Med Internet Res 2021 Aug 05;23(8):e29271 [FREE Full text] [doi: 10.2196/29271] [Medline: 34383675]
6. Crafoord M, Fjell M, Sundberg K, Nilsson M, Langius-Eklöf A. Engagement in an interactive app for symptom self-management during treatment in patients with breast or prostate cancer: mixed methods study. J Med Internet Res 2020 Aug 10;22(8):e17058 [FREE Full text] [doi: 10.2196/17058] [Medline: 32663140]
7. Langius-Eklöf A, Christiansen M, Lindström V, Blomberg K, Hälleberg Nyman M, Wengström Y, et al. Adherence to report and patient perception of an interactive app for managing symptoms during radiotherapy for prostate cancer: descriptive study of logged and interview data. JMIR Cancer 2017 Oct 31;3(2):e18 [FREE Full text] [doi: 10.2196/cancer.7599] [Medline: 29089290]
8. Richards R, Kinnersley P, Brain K, McCutchan G, Staffurth J, Wood F. Use of mobile devices to help cancer patients meet their information needs in non-inpatient settings: systematic review. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e10026 [FREE Full text] [doi: 10.2196/10026] [Medline: 30552082]
9. Judson T, Bennett A, Rogak L, Sit L, Barz A, Kris M, et al. Feasibility of long-term patient self-reporting of toxicities from home via the internet during routine chemotherapy. JCO 2013 Jul 10;31(20):2580-2585 [FREE Full text] [doi: 10.1200/jco.2012.47.6804]
10. Absolom K, Warrington L, Hudson E, Hewison J, Morris C, Holch P, et al. Phase III randomized controlled trial of eRAPID: eHealth intervention during chemotherapy. JCO 2021 Mar 01;39(7):734-747 [FREE Full text] [doi: 10.1200/jco.20.02015]
11. Fjell M, Langius-Eklöf A, Nilsson M, Wengström Y, Sundberg K. Reduced symptom burden with the support of an interactive app during neoadjuvant chemotherapy for breast cancer - a randomized controlled trial. Breast 2020 Jun;51:85-93 [FREE Full text] [doi: 10.1016/j.breast.2020.03.004] [Medline: 32247251]
12. Fridriksdottir N, Gunnarsdottir S, Zoëga S, Ingadottir B, Hafsteinsdottir EJG. Effects of web-based interventions on cancer patients' symptoms: review of randomized trials. Support Care Cancer 2018 Feb;26(2):337-351 [FREE Full text] [doi: 10.1007/s00520-017-3882-6] [Medline: 28921391]
13. Maguire R, McCann L, Kotronoulas G, Kearney N, Ream E, Armes J, et al. Real time remote symptom monitoring during chemotherapy for cancer: European multicentre randomised controlled trial (eSMART). BMJ 2021 Jul 21;374:n1647 [FREE Full text] [doi: 10.1136/bmj.n1647] [Medline: 34289996]
14. Osborn J, Ajakaiye A, Cooksley T, Subbe CP. Do mHealth applications improve clinical outcomes of patients with cancer? A critical appraisal of the peer-reviewed literature. Support Care Cancer 2020 Mar;28(3):1469-1479 [FREE Full text] [doi: 10.1007/s00520-019-04945-4] [Medline: 31273501]
15. Xu A, Wang Y, Wu X. Effectiveness of e-health based self-management to improve cancer-related fatigue, self-efficacy and quality of life in cancer patients: systematic review and meta-analysis. J Adv Nurs 2019 Dec;75(12):3434-3447 [FREE Full text] [doi: 10.1111/jan.14197] [Medline: 31566769]
16. Denis F, Basch E, Septans A, Bennouna J, Urban T, Dueck AC, et al. Two-year survival comparing web-based symptom monitoring vs routine surveillance following treatment for lung cancer. JAMA 2019 Jan 22;321(3):306-307 [FREE Full text] [doi: 10.1001/jama.2018.18085] [Medline: 30667494]

17. Handley N, Schuchter LM, Bekelman JE. Best Practices for Reducing Unplanned Acute Care for Patients With Cancer. J Oncol Pract 2018 May;14(5):306-313 [FREE Full text] [doi: 10.1200/JOP.17.00081] [Medline: 29664697]
18. van Egdom LSE, Oemrawsingh A, Verweij LM, Lingsma HF, Koppert LB, Verhoef C, et al. Implementing patient-reported outcome measures in clinical breast cancer care: a systematic review. Value Health 2019 Oct;22(10):1197-1226 [FREE Full text] [doi: 10.1016/j.jval.2019.04.1927] [Medline: 31563263]
19. Sundberg K, Wengström Y, Blomberg K, Hälleberg-Nyman M, Frank C, Langius-Eklöf A. Early detection and management of symptoms using an interactive smartphone application (Interaktor) during radiotherapy for prostate cancer. Support Care Cancer 2017 Jul;25(7):2195-2204 [FREE Full text] [doi: 10.1007/s00520-017-3625-8] [Medline: 28236145]
20. Gustavell T, Sundberg K, Segersvärd R, Wengström Y, Langius-Eklöf A. Decreased symptom burden following surgery due to support from an interactive app for symptom management for patients with pancreatic and periampullary cancer. Acta Oncologica 2019 Jul 09;58(9):1307-1314 [FREE Full text] [doi: 10.1080/0284186x.2019.1633473]
21. Granja C, Janssen W, Johansen M. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
22. Sullivan R, Peppercorn J, Sikora K, Zalcberg J, Meropol NJ, Amir E, et al. Delivering affordable cancer care in high-income countries. Lancet Oncol 2011 Sep;12(10):933-980. [doi: 10.1016/S1470-2045(11)70141-3] [Medline: 21958503]
23. Rising KL, Ward MM, Goldwater JC, Bhagianadh D, Hollander JE. Framework to Advance Oncology-Related Telehealth. JCO Clinical Cancer Informatics 2018 Dec(2):1-11 [FREE Full text] [doi: 10.1200/cci.17.00156]
24. The National Institute for HealthCare Excellence (NICE). NICE Glossary: cost-effectiveness analysis. NICE Glossary.: NICE; 2022. URL: https://www.nice.org.uk/Glossary?letter=C [accessed 2025-01-21]
25. The National Institute for HealthCare Excellence (NICE). NICE Glossary: Quality-adjusted life year. NICE. 2022. URL: https://www.nice.org.uk/glossary?letter [accessed 2025-01-21]
26. Swedish National Board of Health and Welfare. National Guidelines for adult dental care 2011. Swedish National Board of Health and Welfare. Stockholm, Sweden; 2011. URL: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/nationella-riktlinjer/2021-9-7549-tandvard-halsoekonomiskt-underlag.pdf [accessed 2025-01-21]
27. Lizán L, Pérez-Carbonell L, Comellas M. Additional value of patient-reported symptom monitoring in cancer care: a systematic review of the literature. Cancers (Basel) 2021 Sep 15;13(18):4615 [FREE Full text] [doi: 10.3390/cancers13184615] [Medline: 34572842]
28. Lizée T, Basch E, Trémolières P, Voog E, Domont J, Peyraga G, et al. Cost-effectiveness of web-based patient-reported outcome surveillance in patients with lung cancer. J Thorac Oncol 2019 Jun;14(6):1012-1020 [FREE Full text] [doi: 10.1016/j.jtho.2019.02.005] [Medline: 30776447]
29. Velikova G, Absolom K, Hewison J, Holch P, Warrington L, Avery K, et al. Electronic self-reporting of adverse events for patients undergoing cancer treatment: the eRAPID research programme including two RCTs. In: Electronic Self-Reporting of Adverse Events for Patients Undergoing Cancer Treatment: The eRAPID Research Programme Including Two RCTs. Southampton, UK: NIHR Journals Library; Feb 2022.
30. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]
31. Langius-Eklöf A, Crafoord MT, Christiansen M, Fjell M, Sundberg K. Effects of an interactive mHealth innovation for early detection of patient-reported symptom distress with focus on participatory care: protocol for a study based on prospective, randomised, controlled trials in patients with prostate and breast cancer. BMC Cancer 2017 Jul 04;17(1):466 [FREE Full text] [doi: 10.1186/s12885-017-3450-y] [Medline: 28676102]
32. Mukherjee K, Gunsoy N, Kristy R, Cappelleri J, Roydhouse J, Stephenson J, et al. Handling missing data in health economics and outcomes research (HEOR): a systematic review and practical recommendations. Pharmacoeconomics 2023 Dec;41(12):1589-1601 [FREE Full text] [doi: 10.1007/s40273-023-01297-0] [Medline: 37490207]
33. VAL databaserna Datalager för uppföljning av vårdhändelser i SLL. SLL. 2014. URL: http://www.gups.sll.se/val/Valhandbok.pdf [accessed 2025-01-21]
34. Molero-Calafell J, Burón A, Castells X, Porta M. Intention to treat and per protocol analyses: differences and similarities. J Clin Epidemiol 2024 Sep;173:111457 [FREE Full text] [doi: 10.1016/j.jclinepi.2024.111457] [Medline: 38977160]
35. Statistics Sweden (SCB). Prisomräknaren. Statistics Sweden (SCB). URL: https://www.scb.se/hitta-statistik/sverige-i-siffror/prisomraknaren/ [accessed 2025-01-23]
36. The R. Exchange rates. The Riksbank. 2022. URL: https://www.riksbank.se/en-gb/statistics/search-interest--exchange-rates/ [accessed 2025-01-21]
37. Briggs A, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Econ 1997;6(4):327-340. [doi: 10.1002/(sici)1099-1050(199707)6:4<327::aid-hec282>3.0.co;2-w] [Medline: 9285227]
38. Young T, Mukuria C, Rowen D, Brazier J, Longworth L. Mapping functions in health-related quality of life. Med Decis Making 2015 May 21;35(7):912-926 [FREE Full text] [doi: 10.1177/0272989x15587497]

39. Longworth L, Yang Y, Young T, Mulhern B, Hernández Alava M, Mukuria C, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. Health Technol Assess 2014 Feb;18(9):1-224 [FREE Full text] [doi: 10.3310/hta18090] [Medline: 24524660]
40. Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol: results from a UK general population survey. University of York. York: Centre for Health Economics, University of York; 1995. URL: https://www.york.ac.uk/che/pdf/DP138.pdf [accessed 2025-01-21]
41. Burström K, Sun S, Gerdtham UG, Henriksson M, Johannesson M, Levin LÅ, et al. Swedish experience-based value sets for EQ-5D health states. Qual Life Res 2014 Mar;23(2):431-442 [FREE Full text] [doi: 10.1007/s11136-013-0496-4] [Medline: 23975375]
42. O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. Med Care 1994 Feb;32(2):150-163 [FREE Full text] [doi: 10.1097/00005650-199402000-00006] [Medline: 8302107]
43. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288 [FREE Full text] [doi: 10.1177/1049732305276687] [Medline: 16204405]
44. Abu Bakar NS, Ab Hamid J, Mohd Nor Sham MSJ, Sham MN, Jailani AS. Count data models for outpatient health services utilisation. BMC Med Res Methodol 2022 Oct 05;22(1):261 [FREE Full text] [doi: 10.1186/s12874-022-01733-3] [Medline: 36199028]
45. Prince RM, Powis M, Zer A, Atenafu EG, Krzyzanowska MK. Hospitalisations and emergency department visits in cancer patients receiving systemic therapy: systematic review and meta-analysis. Eur J Cancer Care (Engl) 2019 Jan;28(1):e12909 [FREE Full text] [doi: 10.1111/ecc.12909] [Medline: 30238542]
46. Henry L, Sherise D, Jun L, Rhonda G, Kacy D, David C, et al. Avoidable and unavoidable ER utilization by cancer patients on systemic therapy. Oncology Issues 2021 Feb 01;36(1):42-51 [FREE Full text] [doi: 10.1080/10463356.2020.1849903]
47. Cygańska M, Kludacz-Alessandri M, Pyke C. Healthcare costs and health status: insights from the SHARE survey. Int J Environ Res Public Health 2023 Jan 12;20(2):1418 [FREE Full text] [doi: 10.3390/ijerph20021418] [Medline: 36674169]
48. Di Maio M, Gallo C, Leighl N, Piccirillo M, Daniele G, Nuzzo F, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. JCO 2015 Mar 10;33(8):910-915 [FREE Full text] [doi: 10.1200/jco.2014.57.9334]
49. Snoswell C, Taylor M, Comans T, Smith A, Gray L, Caffery L. Determining if telehealth can reduce health system costs: scoping review. J Med Internet Res 2020 Oct 19;22(10):e17298 [FREE Full text] [doi: 10.2196/17298] [Medline: 33074157]
50. Hofmarcher T, Lindgren P, Wilking N, Jönsson B. The cost of cancer in Europe 2018. Eur J Cancer 2020 Apr;129:41-49 [FREE Full text] [doi: 10.1016/j.ejca.2020.01.011] [Medline: 32120274]
51. Nixon N, Spackman E, Clement F, Verma S, Manns B. Cost-effectiveness of symptom monitoring with patient-reported outcomes during routine cancer treatment. Journal of Cancer Policy 2018 Mar;15:32-36 [FREE Full text] [doi: 10.1016/j.jcpo.2017.12.001]
52. van der Hout A, Jansen F, van Uden-Kraan CF, Coupé VM, Holtmaat K, Nieuwenhuijzen GA, et al. Cost-utility of an eHealth application 'Oncokompas' that supports cancer survivors in self-management: results of a randomised controlled trial. J Cancer Surviv 2021 Feb 12;15(1):77-86 [FREE Full text] [doi: 10.1007/s11764-020-00912-9] [Medline: 32656739]
53. Barbera L, Sutradhar R, Howell D, Sussman J, Seow H, Dudgeon D, et al. Does routine symptom screening with ESAS decrease ED visits in breast cancer patients undergoing adjuvant chemotherapy? Support Care Cancer 2015 Oct 26;23(10):3025-3032. [doi: 10.1007/s00520-015-2671-3] [Medline: 25711657]
54. Wheelock AE, Bock MA, Martin EL, Hwang J, Ernest ML, Rugo HS, et al. SIS.NET: a randomized controlled trial evaluating a web-based system for symptom management after treatment of breast cancer. Cancer 2015 Mar 15;121(6):893-899 [FREE Full text] [doi: 10.1002/cncr.29088] [Medline: 25469673]
55. Nipp R, Horick NK, Deal AM, Rogak LJ, Fuh C, Greer JA, et al. Differential effects of an electronic symptom monitoring intervention based on the age of patients with advanced cancer. Ann Oncol 2020 Jan;31(1):123-130 [FREE Full text] [doi: 10.1016/j.annonc.2019.09.003] [Medline: 31912785]
56. Regional Cancer Centres in Sweden. Akut onkologi nationellt vårdprogram Version 1.2. RCC Kunskapsbanken. 2023. URL: https://kunskapsbanken.cancercentrum.se/diagnoser/akut-onkologi/vardprogram/neutropen-febersepsis/ [accessed 2025-01-21]
57. Drummond MF, Schulper MJ, Claxton J, Stoddart G, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. In: Methods for the economic evaluation of health care programmes. Oxford: Oxford university press; 2015.
58. Lodder P. To Impute or not Impute, That's the Question. In G. 2014 Jan. URL: http://www.paultwin.com/wp-content/uploads/Lodder_1140873_Paper_Imputation.pdf [accessed 2025-01-25]
59. Dai S. Handling missing responses in psychometrics: methods and software. Psych 2021 Nov 19;3(4):673-693 [FREE Full text] [doi: 10.3390/psych3040043]

Abbreviations

B-RCT: neoadjuvant chemotherapy for breast cancer

CONSORT: Consolidated Standards of Reporting Trials
CUA: cost-utility analysis
ePRO: electronic patient-reported outcome
HRQOL: health-related quality of life
ICD: International Statistical Classification of Diseases and Related Health Problems
ICER: incremental cost-effectiveness ratio
ITT: intention-to-treat
NACT: neoadjuvant chemotherapy
NBHW: National Board of Health and Welfare
P-RCT: radiotherapy for prostate cancer
QALY: quality-adjusted life year
RT: radiotherapy
SEK: Swedish kronor

Edited by N Cahill; submitted 11.10.23; peer-reviewed by J Femke, J Zhu; comments to author 10.04.24; revised version received 23.05.24; accepted 07.01.25; published 11.03.25.

Please cite as:
Crafoord MT, Ekstrand J, Sundberg K, Nilsson MI, Fjell M, Langius-Eklöf A
Mobile Electronic Patient-Reported Outcomes and Interactive Support During Breast and Prostate Cancer Treatment: Health Economic Evaluation From Two Randomized Controlled Trials
JMIR Cancer 2025;11:e53539
URL: https://cancer.jmir.org/2025/1/e53539
doi:10.2196/53539
PMID:

Analyzing Geospatial and Socioeconomic Disparities in Breast Cancer Screening Among Populations in the United States: Machine Learning Approach

Soheil Hashtarkhani[1], PhD; Yiwang Zhou[2], PhD; Fekede Asefa Kumsa[1], PhD; Shelley White-Means[3], PhD; David L Schwartz[4], MD; Arash Shaban-Nejad[1], MPH, PhD

[1]Center for Biomedical Informatics, Department of Pediatrics, College of Medicine, University of Tennessee Health Science Center, 50 N. Dunlap St, Memphis, TN, United States

[2]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, United States

[3]College of Graduate Health Sciences, University of Tennessee Health Science Center, Memphis, TN, United States

[4]Department of Radiation Oncology, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, United States

Corresponding Author:
Arash Shaban-Nejad, MPH, PhD
Center for Biomedical Informatics, Department of Pediatrics, College of Medicine, University of Tennessee Health Science Center, 50 N. Dunlap St, Memphis, TN, United States

Abstract

Background: Breast cancer screening plays a pivotal role in early detection and subsequent effective management of the disease, impacting patient outcomes and survival rates.

Objective: This study aims to assess breast cancer screening rates nationwide in the United States and investigate the impact of social determinants of health on these screening rates.

Methods: Data on mammography screening at the census tract level for 2018 and 2020 were collected from the Behavioral Risk Factor Surveillance System. We developed a large-scale dataset of social determinants of health, comprising 13 variables for 72,337 census tracts. Spatial analysis employing Getis-Ord Gi statistics was used to identify clusters of high and low breast cancer screening rates. To evaluate the influence of these social determinants, we implemented a random forest model, with the aim of comparing its performance to linear regression and support vector machine models. The models were evaluated using R^2 and root mean squared error metrics. Shapley Additive Explanations values were subsequently used to assess the significance of variables and direction of their influence.

Results: Geospatial analysis revealed elevated screening rates in the eastern and northern United States, while central and midwestern regions exhibited lower rates. The random forest model demonstrated superior performance, with an R^2=64.53 and root mean squared error of 2.06, compared to linear regression and support vector machine models. Shapley Additive Explanations values indicated that the percentage of the Black population, the number of mammography facilities within a 10-mile radius, and the percentage of the population with at least a bachelor's degree were the most influential variables, all positively associated with mammography screening rates.

Conclusions: These findings underscore the significance of social determinants and the accessibility of mammography services in explaining the variability of breast cancer screening rates in the United States, emphasizing the need for targeted policy interventions in areas with relatively lower screening rates.

(JMIR Cancer 2025;11:e59882) doi:10.2196/59882

KEYWORDS

mammography; breast neoplasms; social determinants of health; geographic information systems; machine learning

Introduction

In the United States, breast cancer ranks as the second most prevalent form of cancer among women, surpassed only by skin cancer [1]. Annually, approximately 240,000 cases of breast cancer are diagnosed in women, and tragically, approximately 42,000 women succumb to this disease each year in the United States. This makes breast cancer the second leading cause of cancer-related mortality among women in the country, following lung cancer [2]. Screening for breast cancer serves as a crucial secondary prevention measure, aimed at identifying the disease at an early stage, prior to clinical manifestation. Early detection of breast cancer enables the implementation of less intensive treatment strategies, contributing to improved survival rates.

Mammography-based screening detects lesions before they achieved clinical visibility [3]. Evidence shows that high-quality routine screening programs have led to a 25% to 31% reduction in breast cancer–related mortality among women aged 50 to 69 years [4].

The US Preventive Services Task Force recommends mammography every two years for women aged 40 to 74 years [5]. Despite these recommendations, current research indicates disparities in mammography screening across different parameters, including variations among women residing in different regions and belonging to different races, with varying levels of median household incomes, health insurance statuses, and access to mammography services [6]. The 2022 Cancer Trends Progress Report revealed that 76% of women aged 50 - 74 years underwent mammogram testing, with rates varying from 74% among Hispanic women to 82% among non-Hispanic Black women [7]. Additionally, 64% of women with less than a high school education, 67.5% of women with incomes below 200% of the federal poverty level, and 75% of those who were Medicare beneficiaries underwent a mammogram test [7]. The Healthy People 2030 [8] has set a target to increase the proportion of breast cancer screenings to 80% [9].

Geospatial and machine learning models have proven effective in identifying the impact of social, natural, and built environments on health outcomes [10-12]. This study seeks to explain the geographical disparities in breast cancer screening across the United States and to explore the area-level socioeconomic factors associated with the rates of breast cancer screening. By examining these disparities, we seek to provide insights that can guide targeted interventions and policies aimed at improving equitable access to breast cancer screening services.

Methods

This cross-sectional study investigates the spatial and socioeconomic factors influencing mammography screening rates among women aged 50 to 74 years in the United States. The methodology section outlines the steps, including data collection, variable selection, descriptive analysis, spatial analysis, machine learning model implementation, and model performance evaluation.

Data Collection

Dependent Variables

The data for mammography screening rates in this study were sourced from the Centers for Disease Control and Prevention's (CDC) PLACES Project for the years 2018 and 2020, which used responses collected through the Behavioral Risk Factors Surveillance System (BRFSS) survey [13]. This survey specifically targeted female respondents aged 50-74 years, categorizing them as women who reported having undergone mammogram screenings and those who did not (excluding unknowns and refusals). Our data extraction process comprised two main stages. First, we extracted age-adjusted mammography screening rates at the county level for spatial analysis, facilitating the visualization of patterns across the entire country. Second, we obtained the crude rates (raw percentages) of mammography screening at the census tract level, a small geographic unit used by the US Census Bureau for collecting and analyzing statistical data, explanatory analysis, and prediction model development by machine learning methods.

Independent Variables

Based on a preliminary literature review, we selected independent variables for the study from various sources. We incorporated socioeconomic data from the CDC, the 2013 - 2017 American Community Survey, the United States Department of Agriculture, and the Health Resources and Services Administration. The analyzed variables encompass a range of factors, including urban-rural location, population density, the rate of older women (aged 55 to 74 years), poverty rate, ethnicity (Black and Hispanic), educational attainment, uninsured rate, median home value, social vulnerability index, and primary care shortage area.

To assess accessibility, we used data from the US Food and Drug Administration's Mammography Facility Database, which included geocoding the locations of 8706 mammography centers. The geodesic distance from each census tract to the nearest facility and the number of facilities within a 10-mile radius were calculated. Table 1 provides a comprehensive overview of the dependent and independent variables used in this study, including their names, sources, and definitions.

Table . Dependent and independent variables used in this study.

Variable name	Source	Unit	Definition
Dependent variables			
Mammography rate (2018)	CDC[a]	Percent	Crude percent of mammography use among women aged 50 - 74 years in 2018
Mammography rate (2020)	CDC	Percent	Crude percent of mammography use among women aged 50 - 74 years in 2020
Independent variables			
Urban-rural location	USDA[b]	Binary	Urban or rural tract as of 2019
Population density	2013 - 2017 ACS[c]	Per square mile	Number of people per square mile
Number of women aged ≥55 years	2013 - 2017 ACS	Percent	Estimated percent of the female population aged 55 or above
Poverty rate	Census ACS data	Percent	Estimated percent of all people that are living in poverty
Without health insurance	2013 - 2017 ACS	Percent	Estimated percent of the population without health insurance coverage
Higher education rate	2013 - 2017 ACS	Percent	Estimated percent of the population ≥25 years, with a bachelor's, graduate, or professional degree
Black population	2013 - 2017 ACS	Percent	Percent of the population that is Black or African American, by single census classification
Hispanic population	2013 - 2017 ACS	Percent	Percent of the population identified as Hispanic or Latino
Home value	2013 - 2017 ACS	Dollar	Estimated median value of an owner-occupied housing unit
Social vulnerability index	CDC	Index	Social vulnerability level as of 2020
Primary care shortage	HRSA[d]	Binary	Primary care health professional shortage area status as of 2020
Distance to nearest mammography facility	Calculated	Mile	Distance from the center of the census tract to the nearest accredited mammography facility
Number of mammography facilities	Calculated	Number	Number of mammography facilities within the 10-mile catchment of the census tract

[a]CDC: Center for Disease Control and Prevention.

[b]USDA: United States Department of Agriculture.

[c]ACS: American Community Survey.

[d]HRSA: Health Resources and Services Administration.

Analysis

Preprocessing

The primary objective of preprocessing was to handle missing values of both dependent and independent variables within the dataset. Due to the complexity of accounting for both spatial and temporal correlations in imputing breast cancer screening rates, we opted to exclude any census tracts that lacked mammography screening data in the BRFSS dataset for the years 2018 and 2020. Missing independent variables were imputed using the mean values for numerical data and the mode for binary data from the 20 closest neighboring records.

Thematic Mapping and Spatial Clustering

The age-adjusted rates for breast cancer screening were integrated into a shapefile of ArcGIS containing 3143 counties across all 50 states and the District of Columbia. Subsequently, the data was visualized using the natural break method [14] to enhance clarity. Using the Getis-Ord Gi statistic [15], we identified hotspots indicating areas with either high or low mammography screening rates. This spatial analysis allowed us to discern localized patterns and trends of breast cancer screening behavior.

Machine Learning Analysis

While constructing the predictive model, the response variable was the mean value of mammography screening rates in 2018 and 2020 for each census tract. The dataset was randomly split into two parts: 75% was used for training the model, and the remaining 25% was reserved for testing. This division allowed us to develop the model using the training data and then assess its predictive performance on the unseen testing data.

In this study, an ensemble learning algorithm known as random forest (RF) was employed to model the relationship between geospatial factors and breast cancer screening rates. Ensemble learning combines multiple models to improve the overall prediction accuracy and robustness, which is the rationale for choosing RF [16].

To enhance the efficacy of the RF model, we conducted a systematic hyperparameter search, where a predefined grid of values for the number of trees and the number of variables sampled at each split were explored to identify the optimal configuration [17]. We defined a grid of values for the number of trees and the number of variables sampled at each split.

We utilized the 5-fold cross-validation to evaluate the RF model's performance across different combinations of hyperparameters. In a 5-fold cross-validation, the dataset is split into 5 subsets, with each subset serving as the validation set once, while the other 4 subsets are used for training. This process helps in assessing the model's generalization ability. The model's performance was fine-tuned by selecting the combination of hyperparameters that minimized the root mean squared error (RMSE), a metric indicating the average difference between observed and predicted values. The RMSE is critical as it directly relates to the model's prediction accuracy, with lower values indicating better performance [18].

To benchmark the performance of the RF model, we also implemented the linear regression (LR) and support vector machine (SVM) models. The LR provides a straightforward baseline, while SVM is known for its effectiveness in high-dimensional spaces. The inclusion of these three algorithms was motivated by their complementary strengths in handling different data characteristics, allowing for a comprehensive comparison of predictive accuracy.

The models were implemented using the Scikit-learn package in Python, a widely used library for machine learning that provides efficient tools for model training, evaluation, and hyperparameter optimization [19].

Following the training process, predictions of breast cancer screening rates were made on a separate testing set. Model accuracy was evaluated using metrics such as R^2 and RMSE. R^2 represents the proportion of variance in the dependent variable explained by the model, serving as an indicator of goodness-of-fit. RMSE, as previously mentioned, measures the average difference between predicted and observed values, providing insight into the model's prediction error [18].

To interpret the model's predictions, we calculated Shapley Additive Explanations (SHAP) values for each feature. SHAP values provide a detailed understanding of how each feature contributes to the model's predictions [20]. By examining the mean SHAP values, the most influential variables in predicting breast cancer screening rates were identified. For variables with average SHAP values exceeding 0.3, scatterplots were created to explore the direction and magnitude of their effects on screening rates [21].

Ethical Considerations

The Institutional Review Board at the University of Tennessee Health Science Center determined that this study (24 - 10240-NHSR) qualifies for Not Human Subjects Research status as it does not involve human subjects as defined by 45 CFR 46.102. The data used in this study were obtained from the publicly available BRFSS dataset provided by the CDC. All study data were aggregated at the census tract level, and no individual-level data were accessed or analyzed, ensuring participant anonymity and compliance with ethical standards.

Results

Summary Statistics About Data

Of the 72,337 census tracts nationwide, 49,118 were eligible for inclusion in our analysis, as they had available mammography screening data. The mean mammography screening rate within these census tracts was 77% (SD 3.62) in 2018 and 76.51% (SD 3.71) in 2020. Table 2 provides a detailed overview of summary statistics for all variables considered in our analysis, encompassing the 49,118 included census tracts.

Table . Summary statistics for all dependent and independent variables for 49,118 census tracts included in the analysis.

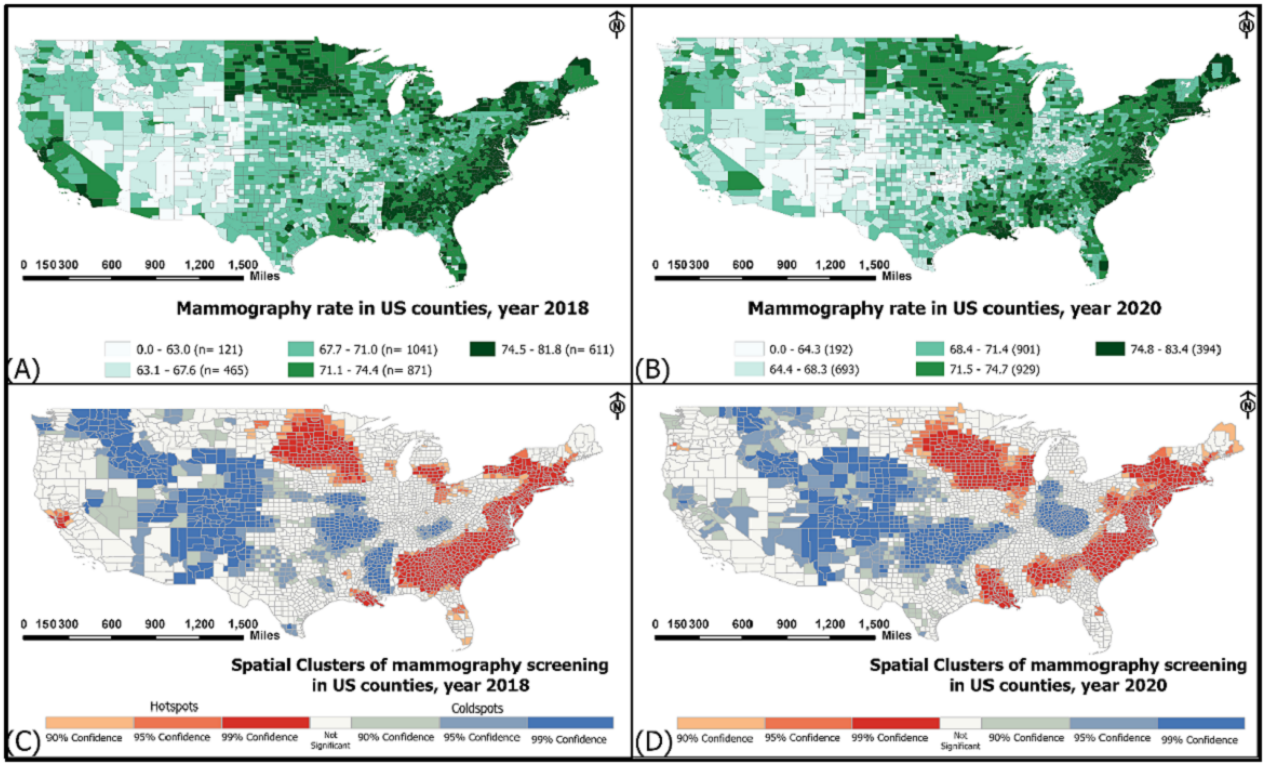
Variables	Missing values, n	Census tracts (N=49,118)
Mammography rate (2018) (%), mean (SD)	0	77 (3.6)
Mammography rate (2020) (%), mean (SD)	0	76.5 (3.7)
Location n (%)		
Rural	0	12,284 (25)
Urban	0	36,834 (75)
Population density (per square mile), mean (SD)	0	5,547.38 (13,334.53)
Women aged ≥55 years (%), mean (SD)	0	7.9 (4.0)
Poverty rate (%), mean (SD)	49	16.3 (12.5)
Without health insurance (%), mean (SD)	35	11.47 (7.83)
Higher education rate (%), mean (SD)	4	28.4 (18.6)
Black population (%), mean (SD)	1	15.2 (23.6)
Hispanic population (%), mean (SD)	1	12.7 (18.5)
Home value (US $), mean (SD)	727	203,834 (170,438)
Social vulnerability index, mean (SD)	82	0.59 (0.28)
Primary care shortage, n (%)		
Yes	0	28,031 (57.1)
No	0	21,087 (42.9)
Distance to nearest mammography (miles), mean (SD)	0	1.8 (3.2)
Number of mammography facilities, mean (SD)	0	18.1 (28.6)

Thematic Mapping and Spatial Clustering

Figures 1A and B illustrate the distribution of breast cancer screening rates across the 3143 US counties for the years 2018 and 2020, respectively. Regions in the eastern and northern parts of the country exhibited higher rates of breast cancer screening (>71%), while counties in the central, midwestern, and southern areas displayed comparatively lower rates (<63%). While these visual representations provide valuable insights, further confidence in the findings is derived from statistical and spatial analyses.

Figures 1C and D present the outcomes of Getis-Ord Gi statistics for the clustering of breast cancer screening rates across the United States in 2018 and 2020, respectively. The red areas (hotspots) on these maps represent spatial clusters characterized by high mammography rates, indicating that the screening rates and their neighboring values significantly surpass those in other regions. Conversely, the blue areas denote coldspots, representing spatial clusters with lower screening rates. The similarity in patterns between the two time points underscores the reliability of the observations and strengthens the robustness of the identified spatial clusters. The map also reveals certain disparities. For instance, counties along the western borders, such as California, experienced a decline in mammography rates from 2018 to 2020. Similarly, regions in Indiana, Texas, and Arkansas saw decreased rates of breast cancer screening during this period. Conversely, parts of Illinois and Louisiana showed reported mammography rates from 2018 to 2020.

Figure 1. Age-adjusted rates of breast cancer screening in US counties for (A) 2018 and (B) 2020. Spatial clusters in (C) 2018 and (D) 2020.



Machine Learning Analysis

Evaluation of the final RF model, along with the LR and SVM models, based on R^2 and RMSE of the testing dataset is presented in Figure 2. The results indicate that the RF model, with an optimal number of trees set to 500 and the number of nodes (m) set to 4, outperforms both LR and SVM. Specifically, the RF model achieved a higher R^2 value and a lower RMSE, indicating its superior ability to capture and predict the underlying patterns in the data. This performance underscores the suitability of the RF model for this analysis.

Figure 3 depicts the relative importance of each factor, as determined by SHAP values, at the census tract level in predicting the rate of breast cancer screening across the United States. The mean of SHAP values provides a measure of the overall contribution of each variable to the model's predictions. As evident from Figure 3, the proportion of the Black population is the most important factor, followed by the number of mammography facilities within a 10-mile distance and the higher education rate. For the subsequent analysis, we refined our focus to variables with SHAP values exceeding 0.3 (the top 6 variables) to assess the direction and magnitude of influence that each variable exerts on the prediction of breast cancer screening rates as the variables vary in value.

While assessing variable importance using mean SHAP values offers crucial insights into the most influential factors in predicting breast cancer screening rates, it does not elucidate the direction of their effects on the outcome variable across different variable values. To address this, we generated scatterplots of individual SHAP values for the selected six variables to examine the detailed changes in SHAP values across varying values of these variables. Figure 4 shows that higher proportions of the Black population, higher education levels, an increased number of mammography facilities, and a higher median home value exhibit positive associations with breast cancer screening rates. Conversely, a higher proportion of Hispanic ethnicity and a lack of health insurance demonstrate negative impacts on the screening rates.

Figure 2. Comparison of the performance of random forest, linear regression, and SVM models in predicting breast cancer screening rates. RMSE: root mean squared error; SVM: support vector machine.

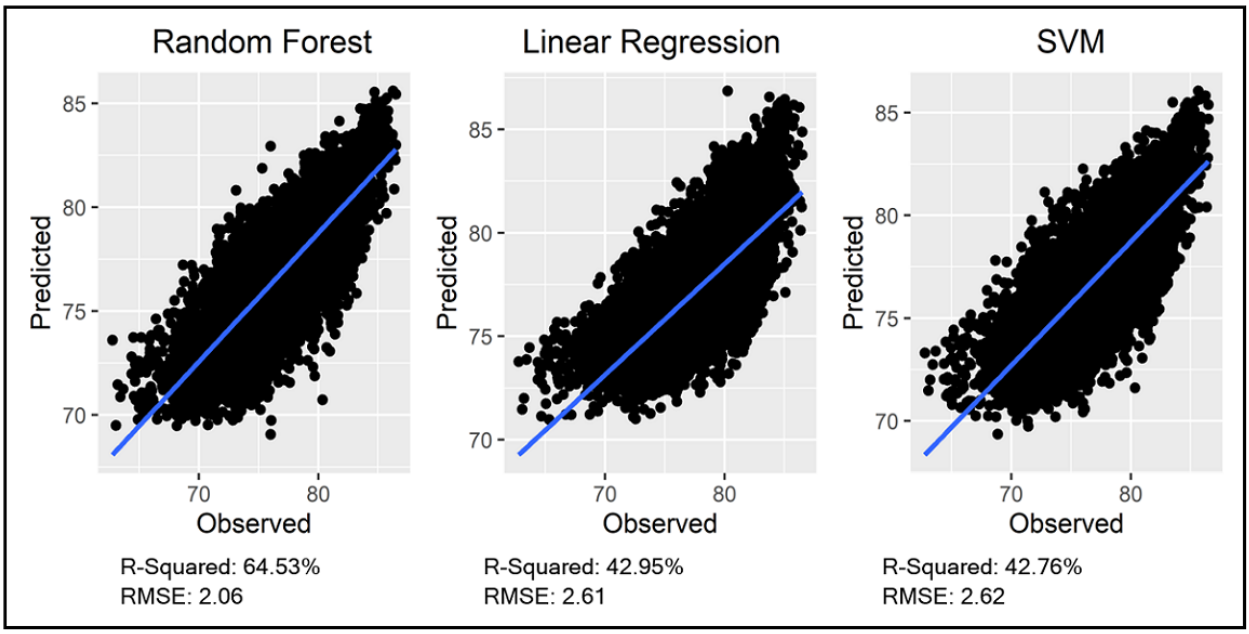


Figure 3. SHAP values of each census tract–level factor in predicting the rate of breast cancer screening across the United States. SHAP: Shapley Additive Explanations.

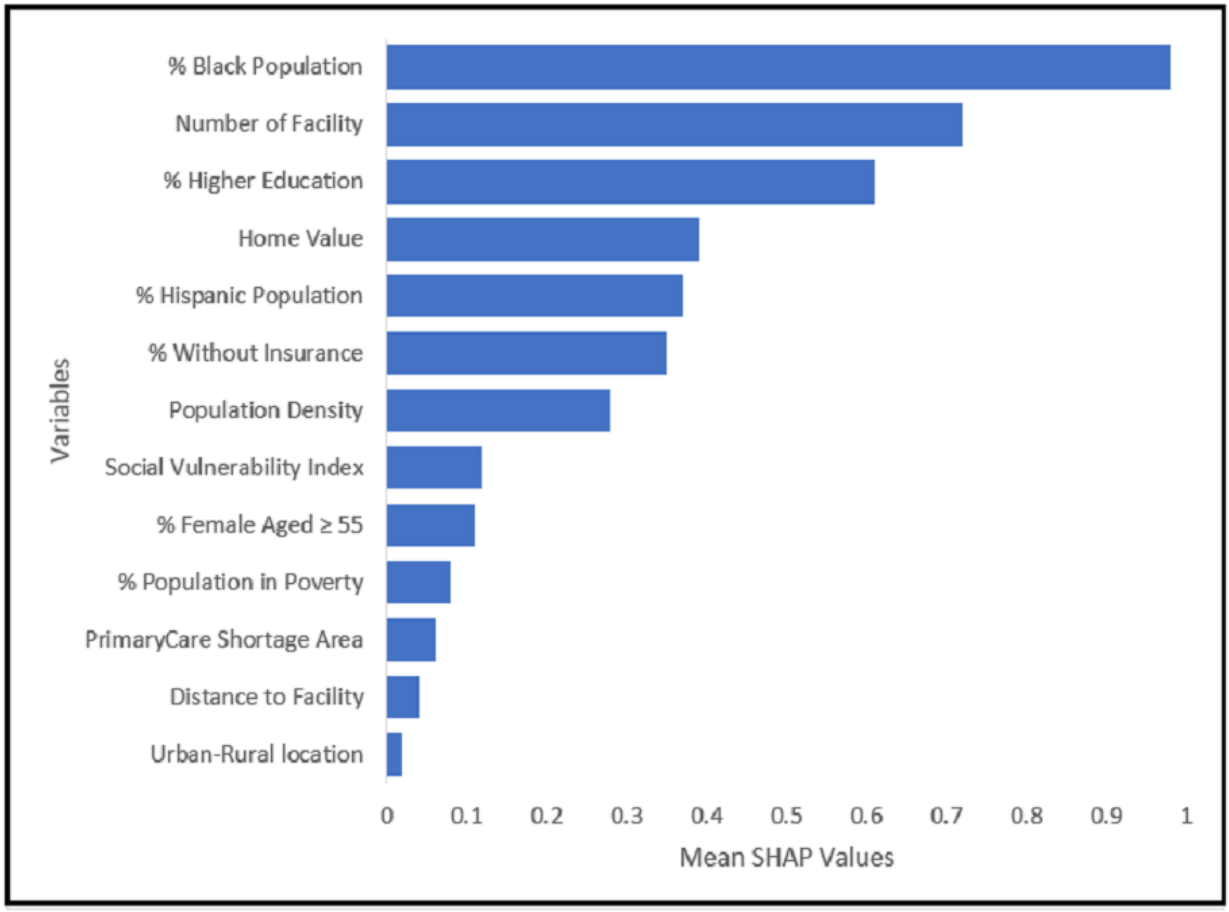
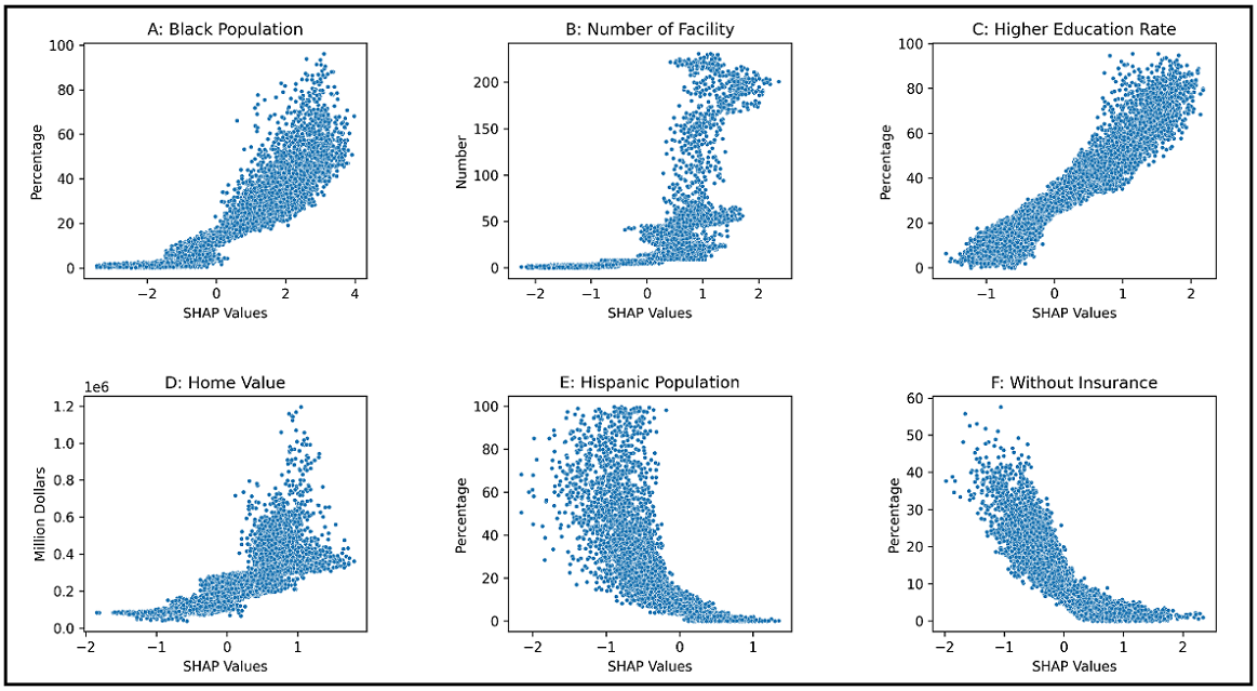


Figure 4. Direction of influence illustrated by individual SHAP values on the prediction of breast cancer screening rates for the selected six important variables. SHAP: Shapley additive explanations.



Discussion

Our research employed a combination of spatial analysis, statistical methods, and machine learning techniques to elucidate the disparities in breast cancer screening across the United States. Spatial patterns revealed clusters of low screening rates, particularly in central, midwestern, and southern regions, contrasted with hotspots of mammography rates, particularly evident along the east coast and in the northern parts of the United States. A predictive model for breast cancer screening rates was developed using the RF algorithm. Meanwhile, key influencing factors for predicting breast cancer screening rates were identified based on the mean SHAP values, including the proportion of the Black population, availability of mammography facilities, and higher education rates.

Spatial clustering identified through Getis-Ord Gi statistics reinforces the observed patterns and underscores their persistence across two distinct time points (2018 and 2020). The consistency of these spatial clusters suggests enduring factors influencing breast cancer screening behavior in specific areas, providing valuable information for policy makers and health care professionals seeking to implement targeted interventions.

It is crucial to acknowledge that the onset of the COVID-19 pandemic has significantly impacted various aspects of human life, including breast cancer screening [21]. The pandemic may have played a role in the decreased mean breast cancer screening rate of breast cancer from 77% in 2018 to 76.51% in 2020. The disruption of routine health care services and the challenges associated with social distancing could have affected mammography screening rates, particularly in urban areas with denser populations. However, it is important to consider that the BRFSS survey focuses on individuals who have undergone breast cancer screening in the last two years. Consequently, women who underwent mammography screening in the year prior to the COVID-19 pandemic could still respond affirmatively. The influence of COVID-19 on the average rate and pattern of breast cancer screening may vary significantly in 2022 and 2023, particularly in cities with higher disease prevalence. Additional investigations are warranted to understand the influence of COVID-19 on changes in breast cancer screening rates across the United States and globally.

Our machine learning analysis that uses an RF model contributes toward understanding the complex interplay of various factors influencing breast cancer screening rates. The RF model with optimal hyperparameters outperformed the LR and SVM models. The ability of RF to capture complex nonlinear relationships and interactions among influencing factors aligns with findings from other population-level studies, highlighting its superiority in predicting population health outcomes [22,23], which confirms our choice of RF as the primary model for our analysis.

While the RF model demonstrated superior performance in predicting breast cancer screening rates, there is a potential risk of overfitting, inherent to ensemble methods [24]. To mitigate this, we implemented cross-validation during the hyperparameter tuning process and evaluated the model's performance on a separate testing dataset to ensure that the RF model maintained its predictive accuracy on unseen data.

It is particularly noteworthy that a higher proportion of the Black population within a census tract was positively associated with increased mammography screening rates. This finding aligns with a 2022 cancer trends progress report, which revealed that 82% of Black women underwent mammography screening, while the screening rate was as low as 74% among other ethnic groups [1]. Possible explanations for this positive association

may include the effectiveness of targeted public health interventions and community-based outreach programs specifically designed to increase awareness and accessibility of breast cancer screening in these communities. Additionally, it may reflect a growing awareness and proactive behavior regarding breast cancer prevention among Black women, possibly influenced by public health campaigns and community support networks. Some studies also suggested that when access to health care is equitable, racial and ethnic minorities who are often more aware of their heightened risk, may be more likely to use preventive services like mammography [25]. However, despite the relatively higher mammography screening rates in areas with a larger Black population, it is crucial to underscore that Black women are 40% more likely to die from breast cancer compared to White women [26]. This disparity could be attributed to delays in diagnosis and treatment, particularly when a breast tissue abnormality is identified by mammography [27,28]

Our findings highlighted the importance of the number of available mammography facilities within a 10-mile radius, despite the relatively low SHAP value assigned to the distance to the nearest facility. A plausible explanation is that proximity to a facility may not always be a decisive factor, as various considerations such as affordability and type of insurance can significantly impact facility selection. Moreover, our research revealed that the education rate plays a pivotal role in determining breast cancer screening rates. This finding aligns with prior studies indicating that American women with lower educational attainment are less likely to undergo screening [29]. Educational attainment is closely linked to health literacy [27]; women with lower health literacy have a reduced likelihood of accessing health services, including breast cancer screening [28]. Moreover, women with lower educational attainment might face limited employment opportunities and a lack of jobs that offer access to employee health insurance, leading to a lower likelihood of consulting physicians who recommend mammography.

The variables of home value, rate of the Hispanic population, and rate of the uninsured population exhibited relatively similar and high SHAP values. Areas with higher home values and lower uninsured populations tend to have fewer financial barriers to accessing preventive services. According to existing literature, Hispanic women exhibit lower rates of breast cancer and mortality compared to non-Hispanic Black women and non-Hispanic White women [30]. This disparity could explain the lower screening rate in census tracts with higher proportions of Hispanic population.

Variables analyzed in this study are based on estimates from the CDC PLACES project, which uses a multilevel regression and poststratification approach. This method combines individual-level BRFSS data with demographic data from the US Census to produce reliable estimates at small geographic levels, including census tracts. The multilevel regression and poststratification method has been validated against direct survey data, ensuring that the aggregated rates at the census tract level are both stable and accurate for our analysis [31].

Our study has several limitations. The use of cross-sectional data restricts our capacity to establish causality, underscoring the importance of future research examining temporal changes in breast cancer screening rates. Moreover, as is inherent in all self-reported sample surveys, the BRFSS data may be susceptible to systematic errors stemming from noncoverage, nonresponse, or measurement bias. It is imperative to note that our study was conducted at an aggregate level; therefore, prudence is advised when extrapolating individual-level conclusions. The ecological fallacy, a key concern in population studies, underscores the necessity of avoiding assumptions about individual behaviors based solely on group-level observations.

This study provides a comprehensive analysis of breast cancer screening disparities in the United States, combining spatial, statistical, and machine learning approaches. The spatial patterns and influential factors identified in this study offer valuable insights for policy makers, health care professionals, and researchers striving to implement targeted interventions to reduce breast cancer screening disparities and improve overall public health outcomes. Ongoing research and targeted interventions are vital for achieving equitable access to breast cancer screening services and ultimately reducing the impact of this significant health issue.

Acknowledgments

We would like to thank Brianna M White, research coordinator at the Center for Biomedical Informatics, University of Tennessee Health Science Center, for facilitating and assisting in obtaining the institutional review board exemption letter. This study was partially supported by a grant from the Tennessee Department of Health and grant 5 R01MD018766-02 from the National Institutes of Health/National Institute on Minority Health and Health Disparities.

ChatGPT 3.5 was used for grammar checking and language editing of the final manuscript. The core content, analysis, results, and scientific contributions were developed independently by the authors without the use of generative artificial intelligence tools. All content was carefully reviewed and verified by the authors to ensure accuracy.

Data Availability

The data used in this study were sourced from the Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System. These publicly available datasets can be accessed through the Centers for Disease Control and Prevention website, subject to their data use agreement.

Authors' Contributions

Conceptualization: SH, ASN, DLS
Data curation: SH, YZ, SWM
Funding acquisition: ASN
Methodology: SH, ASN, DLS
Writing – original draft: SH, FAK
Writing – review and editing: SH, ASN, DLS, YZ, SWM, FAK
Spatial and machine learning analyses: SH, YZ
Supervision: ASN

Conflicts of Interest

None declared.

References

1. Cancer prevention & early detection facts & figures 2023-2024. American Cancer Society. 2023. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-prevention-and-early-detection-facts-and-figures/2023-cped-files/2023-cancer-prevention-and-early-detection.pdf [accessed 2024-12-26]
2. Breast cancer basics. Centers for Disease Control and Prevention. 2023. URL: https://www.cdc.gov/breast-cancer/about/index.html [accessed 2024-12-26]
3. Liu H, Xu C, Chen Q, Guo Z. The impact of aggregated census tract-based attributes on the zone-level mode choice difference between ridesourcing and taxi services: a case study in New York City. In: CICTP 2022: Intelligent, Green, and Connected Transportation: American Society of Civil Engineers; 2022:2991-3001. [doi: 10.1061/9780784484265.281]
4. Mitchell EP. U.S. Preventive Services Task Force final recommendation statement, evidence summary, and modeling studies on screening for lung cancer. J Natl Med Assoc 2021 Jun;113(3):239-240. [doi: 10.1016/j.jnma.2021.05.012] [Medline: 34274036]
5. Woloshin S, Jørgensen KJ, Hwang S, Welch HG. The new USPSTF mammography recommendations - a dissenting view. N Engl J Med 2023 Sep 21;389(12):1061-1064. [doi: 10.1056/NEJMp2307229] [Medline: 37721382]
6. Kim E, Moy L, Gao Y, Hartwell CA, Babb JS, Heller SL. City patterns of screening mammography uptake and disparity across the United States. Radiology 2019 Oct;293(1):151-157. [doi: 10.1148/radiol.2019190647] [Medline: 31429681]
7. Miller JW, King JA, Trivers KF, et al. Mammography use and association with social determinants of health and health-related social needs among women — United States, 2022. MMWR Morb Mortal Wkly Rep 2024 Apr 18;73(15):351-357. [doi: 10.15585/mmwr.mm7315e1] [Medline: 38652735]
8. Healthy people 2030. US Department of Health and Human Services. 2024. URL: https://health.gov/healthypeople [accessed 2024-08-28]
9. Cancer trends progress report - 2011/2012 update. National Cancer Institute. 2014 Aug 22. URL: https://progressreport.cancer.gov/sites/default/files/archive/report2011.pdf [accessed 2024-12-26]
10. Shin EK, Mahajan R, Akbilgic O, Shaban-Nejad A. Sociomarkers and biomarkers: predictive modeling in identifying pediatric asthma patients at risk of hospital revisits. NPJ Digit Med 2018;1:50. [doi: 10.1038/s41746-018-0056-y] [Medline: 31304329]
11. Dong W, Kim U, Rose J, et al. Geographic variation and risk factor association of early versus late onset colorectal cancer. Cancers (Basel) 2023 Feb 4;15(4):1006. [doi: 10.3390/cancers15041006] [Medline: 36831350]
12. Dong W, Bensken WP, Kim U, Rose J, Berger NA, Koroukian SM. Phenotype discovery and geographic disparities of late-stage breast cancer diagnosis across U.S. counties: a machine learning approach. Cancer Epidemiol Biomarkers Prev 2022 Jan;31(1):66-76. [doi: 10.1158/1055-9965.EPI-21-0838] [Medline: 34697059]
13. Behavioral Risk Factor Surveillance System. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/brfss/ [accessed 2024-12-26]
14. Data classification methods. ESRI ArcGIS Resource Center. 2018. URL: https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/data-classification-methods.htm [accessed 2024-01-07]
15. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. Geogr Anal 1992 Jul;24(3):189-206. [doi: 10.1111/j.1538-4632.1992.tb00261.x]
16. Schonlau M, Zou RY. The random forest algorithm for statistical learning. Stata J 2020 Mar;20(1):3-29. [doi: 10.1177/1536867X20909688]
17. Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing 2020 Nov;415:295-316. [doi: 10.1016/j.neucom.2020.07.061] [Medline: 34025035]
18. Willmott CJ. On the validation of model. Phys Geogr 1981 Jul;2(2):184-194. [doi: 10.1080/02723646.1981.10642213]
19. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12 [FREE Full text]

20. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); Dec 4-9, 2017; Long Beach, CA, USA URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf [accessed 2024-01-07]
21. Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720. [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
22. Ghafouri-Kesbi F, Rahimi-Mianji G, Honarvar M, Nejati-Javaremi A. Predictive ability of random forests, boosting, support vector machines and genomic best linear unbiased prediction in different scenarios of genomic evaluation. Anim Prod Sci 2017;57(2):229. [doi: 10.1071/AN15538]
23. Zhou Y, Ma M, Shi K, Peng Z. Estimating and interpreting fine-scale gridded population using random forest regression and multisource data. ISPRS Int J Geoinf 2020;9(6):369. [doi: 10.3390/ijgi9060369]
24. Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci 2004;44(1):1-12. [doi: 10.1021/ci0342472] [Medline: 14741005]
25. Goel MS, Wee CC, McCarthy EP, Davis RB, Ngo-Metzger Q, Phillips RS. Racial and ethnic disparities in cancer screening: the importance of foreign birth as a barrier to care. J Gen Intern Med 2003 Dec;18(12):1028-1035. [doi: 10.1111/j.1525-1497.2003.20807.x] [Medline: 14687262]
26. Giaquinto AN, Miller KD, Tossas KY, Winn RA, Jemal A, Siegel RL. Cancer statistics for African American/Black people 2022. CA Cancer J Clin 2022 May;72(3):202-229. [doi: 10.3322/caac.21718] [Medline: 35143040]
27. Jansen T, Rademakers J, Waverijn G, Verheij R, Osborne R, Heijmans M. The role of health literacy in explaining the association between educational attainment and the use of out-of-hours primary care services in chronically ill people: a survey study. BMC Health Serv Res 2018 May 31;18(1):394. [doi: 10.1186/s12913-018-3197-4] [Medline: 29855365]
28. McCaffery KJ, Holmes-Rovner M, Smith SK, et al. Addressing health literacy in patient decision aids. BMC Med Inform Decis Mak 2013;13:1-14. [doi: 10.1186/1472-6947-13-S2-S10] [Medline: 24624970]
29. Koç H, O'Donnell O, Van Ourti T. What explains education disparities in screening mammography in the United States? A comparison with the Netherlands. Int J Environ Res Public Health 2018 Sep 8;15(9):1961. [doi: 10.3390/ijerph15091961] [Medline: 30205539]
30. Bazzi T, Al-Husseini M, Saravolatz L, Kafri Z. Trends in breast cancer incidence and mortality in the United States from 2004-2018: a Surveillance, Epidemiology, and End Results (SEER)-based study. Cureus 2023 Apr;15(4):e37982. [doi: 10.7759/cureus.37982] [Medline: 37223193]
31. PLACES: local data for better health. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/places/methodology/ [accessed 2024-09-02]

Abbreviations

BRFSS: Behavioral Risk Factor Surveillance System
CDC: Centers for Disease Control and Prevention
LR: linear regression
RF: random forest
RMSE: root mean squared error
SHAP: Shapley Additive Explanations
SVM: support vector machine

Edited by N Cahill; submitted 25.04.24; peer-reviewed by U Kim, Z Salod; revised version received 15.09.24; accepted 27.11.24; published 16.01.25.

Please cite as:
Hashtarkhani S, Zhou Y, Kumsa FA, White-Means S, Schwartz DL, Shaban-Nejad A
Analyzing Geospatial and Socioeconomic Disparities in Breast Cancer Screening Among Populations in the United States: Machine Learning Approach
JMIR Cancer 2025;11:e59882
URL: https://cancer.jmir.org/2025/1/e59882
doi:10.2196/59882

Exploring the Social Media Discussion of Breast Cancer Treatment Choices: Quantitative Natural Language Processing Study

Daphna Y Spiegel[1], MD, MS; Isabel D Friesner[2,3,4], BA; William Zhang[2,3,4], BA; Travis Zack[2,4,5], MD, PhD; Gianna Yan[2,3,4]; Julia Willcox[1], BS; Nicolas Prionas[2], MD, PhD; Lisa Singer[2], MD, PhD; Catherine Park[2], MD; Julian C Hong[2,3,4], MD, MS

[1]Department of Radiation Oncology, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Ave, Boston, MA, United States

[2]Department of Radiation Oncology, University of California, San Francisco, San Francisco, CA, United States

[3]Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA, United States

[4]UCSF-UC Berkeley Joint Program in Computational Precision Health, San Francisco, CA, United States

[5]Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Daphna Y Spiegel, MD, MS
Department of Radiation Oncology, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Ave, Boston, MA, United States

Abstract

Background: Early-stage breast cancer has the complex challenge of carrying a favorable prognosis with multiple treatment options, including breast-conserving surgery (BCS) or mastectomy. Social media is increasingly used as a source of information and as a decision tool for patients, and awareness of these conversations is important for patient counseling.

Objective: The goal of this study was to compare sentiments and associated emotions in social media discussions surrounding BCS and mastectomy using natural language processing (NLP).

Methods: Reddit posts and comments from the Reddit subreddit r/breastcancer and associated metadata were collected using pushshift.io. Overall, 105,231 paragraphs across 59,416 posts and comments from 2011 to 2021 were collected and analyzed. Paragraphs were processed through the Apache Clinical Text Analysis Knowledge Extraction System and identified as discussing BCS or mastectomy based on physician-defined Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) concepts. Paragraphs were analyzed with a VADER (Valence Aware Dictionary for Sentiment Reasoning) compound sentiment score (ranging from −1 to 1, corresponding to negativity or positivity) and GoEmotions scores (0 - 1) corresponding to the intensity of 27 different emotions and neutrality.

Results: Of the 105,231 paragraphs, there were 7306 (6.94% of those analyzed) paragraphs mentioning BCS and mastectomy (2729 and 5476, respectively). Discussion of both increased over time, with BCS outpacing mastectomy. The median sentiment score for all discussions analyzed in aggregate became more positive over time. In specific analyses by topic, positive sentiments for discussions with mastectomy mentions increased over time; however, discussions with BCS-specific mentions did not show a similar trend and remained overall neutral. Compared to BCS, conversations about mastectomy tended to have more positive sentiments. The most commonly identified emotions included neutrality, gratitude, caring, approval, and optimism. Anger, annoyance, disappointment, disgust, and joy increased for BCS over time.

Conclusions: Patients are increasingly participating in breast cancer therapy discussions with a web-based community. While discussions surrounding mastectomy became increasingly positive, BCS discussions did not show the same trend. This mirrors national clinical trends in the United States, with the increasing use of mastectomy over BCS in early-stage breast cancer. Recognizing sentiments and emotions surrounding the decision-making process can facilitate patient-centric and emotionally sensitive treatment recommendations.

(JMIR Cancer 2025;11:e52886) doi:10.2196/52886

KEYWORDS

breast cancer; social media; patient decision-making; natural language processing; breast conservation; mastectomy

Introduction

Early-stage breast cancer has the complex challenge of carrying a favorable prognosis with multiple treatment options, including breast-conserving surgery (BCS) or mastectomy. Treatment decisions are therefore driven by patient preferences, making information gathering and decision analysis critical. Multiple randomized trials have shown that locoregional recurrence and survival rates are similar with breast-conserving therapy (BCT) or mastectomy, with recent data even suggesting improved survival with BCT [1,2]. Nonetheless, trends indicate that women with early-stage, nonhereditary breast cancer are increasingly choosing mastectomy [3].

Many factors contribute to patient decision-making for cancer therapy, including the growing influence of social media. Several previous studies have investigated the use of online forums and social media by patients with breast cancer [4-8]. As many as 77% of patients with breast cancer cite the internet as their primary information source [9]. Additionally, patients who are frequent users of online communication and social media tools experience increased decision-satisfaction [10]. Additionally, large language models, including ChatGPT, are being increasingly used by patients for medical decision making. As social media data, including that from Reddit, are used to train these models, these discussions are relevant to the information that patients receive [11,12].

There are limited data characterizing social media conversations surrounding the decision regarding BCS or mastectomy, which is important to understand to gain insights into national trends in the United States and inform the counseling process. We applied sentiment and emotion analyses with natural language processing (NLP) approaches to a popular breast cancer online community to compare the sentiments and associated emotions around conversations of BCS and mastectomy.

Methods

Data Source

The moderated Reddit subreddit r/breastcancer was created on December 3, 2011, and is self-described as “a support and information group for people who have been diagnosed with breast cancer and for their caregivers and loved ones.” As of January 2023, it had 13,700 subscribed members. Out of all internet users, 8% of men and 4% of women used Reddit; of those, 11% were aged 18 - 29 years, 7% were aged 30 - 39 years, 2% were aged 50 - 64 years, and 2% were aged 65+ years [13]. We selected Reddit for social and technical reasons, as it is anonymous, public, open, and interaction-centric. Reddit text is also frequently used to train NLP algorithms, reducing concerns about model applicability.

All posts, comments, and metadata from r/breastcancer from 2011 to 2021 were collected using pushshift.io [14]. Pushshift.io is a public social media archiving platform with real-time Reddit data for social media research. As of March 20, 2024, it has 908 citations according to Google Scholar. Data from Pushshift.io were accessed on February 4, 2022.

Data Preprocessing and Topic Identification

These posts and comments were separated into paragraphs based on line breaks to separate topics for analysis. We applied the Apache cTAKES (Clinical Text Analysis Knowledge Extraction System) v. 4.0.0 default clinical pipeline to identify mentions of BCS- or mastectomy-related terms mapped to the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT). cTAKES is an NLP software with multiple parts designed to process clinical free text. It had a sentence boundary detector, tokenizer, normalizer, part-of-speech tagger, shallow parser, and a named entity recognition annotator with negation. SNOMED keywords related to BCS or mastectomy were identified by a physician author (DYS). Examples of SNOMED words or phrases used in this analysis include *simple mastectomy*, *bilateral mastectomy*, and *modified radical mastectomy* as mastectomy keywords and *lumpectomy*, *excision biopsy*, and *segmental mastectomy* as BCS keywords. Analyses of paragraphs were based on references to BCS or mastectomy, nonexclusively. Paragraphs containing mentions of both were attributed to both treatments in the analysis.

Sentiment Analysis

Paragraphs (as a whole) were analyzed using VADER (Valence Aware Dictionary for Sentiment Reasoning) to generate compound sentiment scores from −1 to 1 (negative to positive). VADER is a popular sentiment analysis model trained on social media text, with performance comparable to more complex approaches and advantages of computational efficiency, explainability, and domain agnosticism or generalizability [12].

Emotion Classification

GoEmotions is the largest human-annotated dataset of fine-grained emotions, with 58,000 Reddit comments labeled for 27 emotions and neutrality [11]. We applied a publicly available BERT (Bidirectional Encoder Representations from Transformers) model from Google Research [15]. GoEmotions generates a score from 0 to 1 based on the intensity of the detected emotion. A paragraph was considered to express an emotion if its score was >0.5, as used in intensity annotations in the original GoEmotions benchmark studies [11].

Statistics were aggregated longitudinally for summary statistics per year. Years 2011 to 2014 had less than 100 posts discussing BCS or mastectomy; years 2011 - 2017 were pooled due to limited sample size. Sentiment scores were compared between BCS and mastectomy using a 2-tailed Student *t* test. Emotions across treatments were visualized using radar charts.

Ethical Considerations

The study data used in this analysis are anonymous, public, and open source. Therefore, there is minimal risk to performing these analyses. This study was approved by the University of San Francisco institutional review board, where the data collection and analyses were performed (IRB #21 - 35353).

Results

A total of 59,416 posts with 105,231 paragraphs on r/breastcancer, which were posted by 5845 users, were analyzed. There were 2729 mentions of BCS and 5476 mentions of

mastectomy, nonexclusively. Post volume increased over time (5282 in 2011 - 2017 to 44,235 in 2021). Words per paragraph had a slight increase over time, from a median of 28 (IQR 11 - 53) words per line in 2011 - 2017 to 26 (IQR 11 - 49) in 2018, 28 (IQR 12 - 50) in 2019, 30 (IQR 14 - 55) in 2020, and 30 (IQR 14 - 54) in 2021. The median number of comments per user was 2 (IQR 1 - 7).

Discussion of both BCS and mastectomy increased over time, but BCS outpaced mastectomy, with an increasing ratio of BCS to mastectomy mentions (0.312 in 2011 - 2017 to 0.583 in 2021). The median (IQR) sentiment score for all discussions became more positive annually: 0 (IQR –0.361 to 0.624) in 2011 - 2017 to 0.288 (–0.223 to 0.701) in 2021. Positive sentiments for mastectomy generally increased over time: median of 0 (IQR –0.599 to 0.726) in 2011 - 2017 to a median of 0.178 (IQR –0.511 to 0.73) in 2021. Similarly, the proportion of positive mastectomy-related discussions increased annually from 48.1% (151/314) in 2011 - 2017 to 53.3% (1107/2076) in 2021. Discussion of BCS remained neutral—median 0 (IQR –0.494 to 0.523) to median 0.039 (IQR –0.511 to 0.642)—and the proportion of positive BCS discussions did not show a similar trend year-to-year: 43.9% (43/98) in 2011-2017 to 52.3% (139/266) in 2019, and stable in 2021 to 50.7% (614/1211). Compared to BCS, conversations about mastectomy were more positive (P=.02), driven primarily by differences in 2021—median 0.178 (IQR –0.511 to 0.73) for mastectomy vs median 0.039 (IQR –0.511 to 0.642) for BCS, with P=.049.

The most common emotions across r/breastcancer were neutrality, gratitude, caring, approval, and optimism. The most common emotions for both BCS and mastectomy were similar: neutrality (BCS: 1001/2729, 36.68%; mastectomy: 1973/5476, 36.03%), caring (BCS: 242/2729, 8.87%; mastectomy: 547/5476, 9.99%), approval (BCS: 267/2729, 9.78%; mastectomy: 492/5476, 8.98%), realization (BCS: 284/2729, 10.41%; mastectomy: 542/5476, 9.9%), and curiosity (BCS: 237/2729, 8.68%; mastectomy: 459/5476, 8.38%) (Figure 1). Six emotions increased over time for all posts: approval (2011 - 2017: 354/5282, 6.7%; 2021: 4176/44,235, 9.44%), amusement (2011 - 2017: 34/5282, 0.64%; 2021: 677/44,235, 1.53%), desire (2011 - 2017: 37/5282, 0.7%; 2021: 552/44,235, 1.25%), disappointment (2011 - 2017: 64/5282, 1.21%; 2021: 787/44,235, 1.78%), excitement (2011 - 2017: 23/2729, 0.44%; 2021: 477/44,235, 1.08%), and realization (2011 - 2017: 259/5282, 4.9%; 2021: 2902/44,235, 6.56%). Fear (2011 - 2017: 203/5282, 3.84%; 2021: 1223/44,235, 2.76%) and neutrality (2011 - 2017: 2144/5282, 40.59%; 2021: 14,514/44,235, 32.81%) decreased over time (Table 1).

Figure 1. Radar chart of average emotion score across breast-conserving surgery (BCS) and mastectomy posts, respectively.

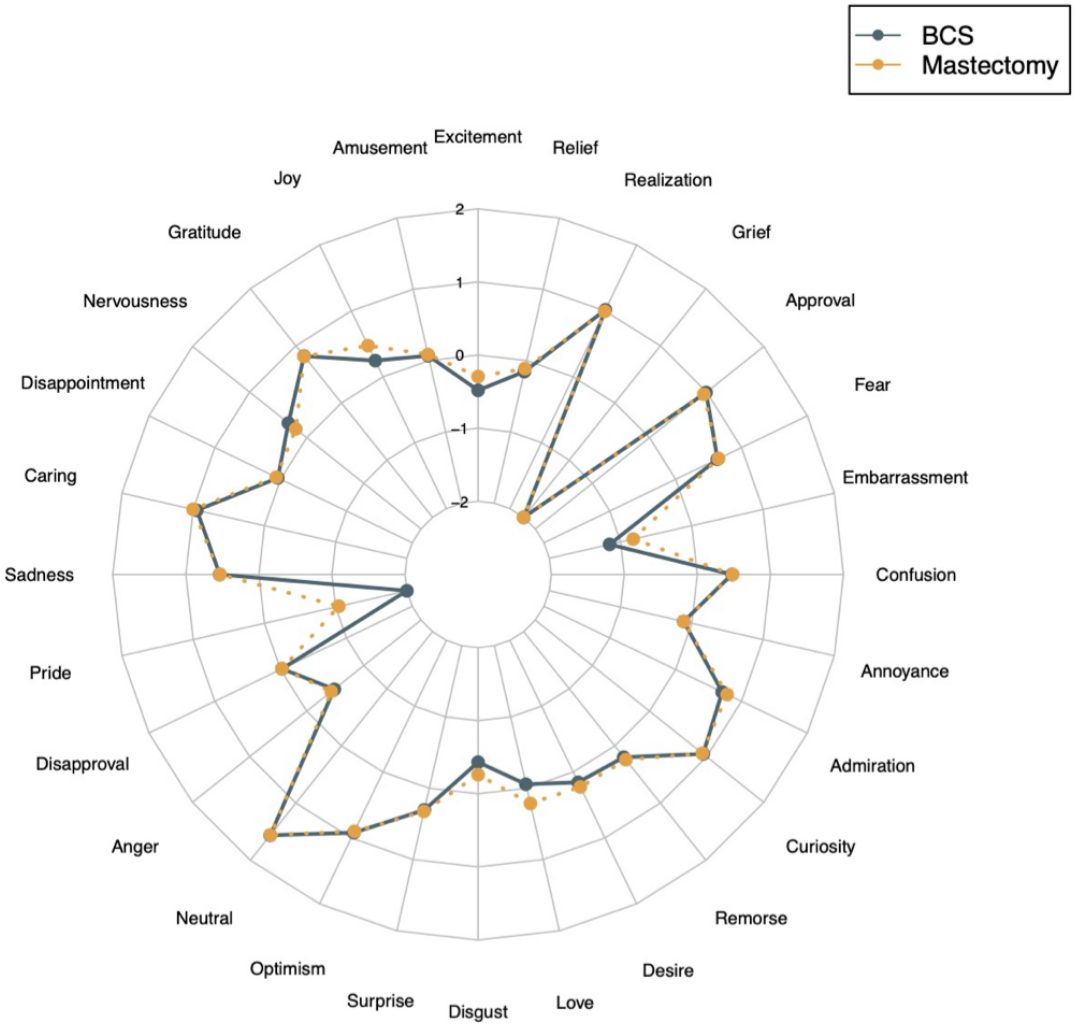


Table . Comparison of emotion trends.[a]

Emotion		All messages, n/N (%)	Breast conservation, n/N (%)	Mastectomy, n/N (%)
Amusement				
	Overall	1379/105,231 (1.31)	32/2729 (1.17)	67/5476 (1.22)
	2011 - 2017	34/5282 (0.64)	0/98 (0)	2/314 (0.64)
	2018	59/8144 (0.72)	3/174 (1.72)	3/410 (0.73)
	2019	109/11,943 (0.91)	1/266 (0.38)	10/673 (1.49)
	2020	500/35,627 (1.4)	16/980 (1.63)	32/2003 (1.6)
	2021	677/44,235 (1.53)	12/1211 (0.99)	20/2076 (0.96)
Anger				
	Overall	762/105,231 (0.72)	9/2729 (0.33)	20/5476 (0.37)
	2011 - 2017	48/5282 (0.91)	0/98 (0)	1/314 (0.32)
	2018	65/8144 (0.8)	0/174 (0)	2/410 (0.49)
	2019	68/11,943 (0.57)	0/266 (0)	1/673 (0.15)
	2020	245/35,627 (0.69)	4/980 (0.41)	5/2003 (0.25)
	2021	336/44,235 (0.76)	5/1211 (0.41)	11/2076 (0.53)
Annoyance				
	Overall	1536/105,231 (1.46)	21/2729 (0.77)	42/5476 (0.77)
	2011 - 2017	66/5282 (1.25)	0/98 (0)	3/314 (0.96)
	2018	105/8144 (1.29)	0/174 (0)	1/410 (0.24)
	2019	177/11,943 (1.48)	1/266 (0.38)	7/673 (1.04)
	2020	487/35,627 (1.37)	5/980 (0.51)	12/2003 (0.6)
	2021	701/44,235 (1.58)	15/1211 (1.24)	19/2076 (0.92)
Approval				
	Overall	9515/105,231 (9.04)	267/2729 (9.78)	492/5476 (8.98)
	2011 - 2017	354/5282 (6.7)	9/98 (9.18)	19/314 (6.05)
	2018	647/8144 (7.94)	15/174 (8.62)	41/410 (10)
	2019	1010/11,943 (8.46)	23/266 (8.65)	60/673 (8.92)
	2020	3328/35,627 (9.34)	979/980 (9.9)	184/2003 (9.19)
	2021	4176/44,235 (9.44)	123/1211 (10.16)	188/2076 (9.06)
Caring				
	Overall	11,640/105,231 (11.06)	242/2729 (8.87)	547/5476 (9.99)
	2011 - 2017	562/5282 (10.64)	9/98 (9.18)	39/314 (12.42)
	2018	1062/8144 (13.04)	23/174 (13.22)	61/410 (14.88)
	2019	1353/11,943 (11.33)	21/266 (7.89)	50/673 (7.43)
	2020	3958/35,627 (11.11)	97/980 (9.9)	199/2003 (9.94)
	2021	4705/44,235 (10.64)	92/1211 (7.6)	198/2076 (9.54)
Curiosity				
	Overall	6973/105,231 (6.63)	237/2729 (8.68)	459/5476 (8.38)
	2011 - 2017	429/5282 (8.12)	11/98 (11.22)	35/314 (11.15)
	2018	567/8144 (6.93)	14/174 (8.05)	28/410 (6.83)
	2019	724/11,943 (6.06)	24/266 (9.02)	44/673 (6.54)
	2020	2263/35,627 (6.35)	84/980 (8.57)	167/2003 (8.34)

Emotion		All messages, n/N (%)	Breast conservation, n/N (%)	Mastectomy, n/N (%)
	2021	2993/44,235 (6.77)	104/1211 (8.59)	185/2076 (8.91)
Desire				
	Overall	1139/105,231 (1.08)	39/2729 (1.43)	92/5476 (1.68)
	2011 - 2017	37/5282 (0.7)	1/98 (1.02)	3/314 (0.96)
	2018	72/8144 (0.88)	3/174 (1.72)	6/410 (1.46)
	2019	107/11,943 (0.9)	3/266 (1.13)	8/673 (1.19)
	2020	371/35,627 (1.04)	17/980 (1.73)	38/2003 (1.9)
	2021	552/44,235 (1.25)	15/1211 (1.24)	37/2076 (1.78)
Disappointment				
	Overall	1696/105,231 (1.61)	30/2729 (1.1)	64/5476 (1.17)
	2011 - 2017	64/5282 (1.21)	0/98 (0)	3/314 (0.96)
	2018	106/8144 (1.3)	1/174 (0.57)	3/410 (0.73)
	2019	174/11,943 (1.46)	2/266 (0.75)	7/673 (1.04)
	2020	565/35,627 (1.59)	9/980 (0.92)	24/2003 (1.2)
	2021	787/44,235 (1.78)	18/1211 (1.49)	27/2076 (1.3)
Disgust				
	Overall	630/105,231 (0.6)	10/2729 (0.37)	30/5476 (0.55)
	2011 - 2017	20/5282 (0.38)	0/98 (0)	0/314 (0)
	2018	35/8144 (0.43)	0/174 (0)	5/410 (1.22)
	2019	81/11,943 (0.68)	1/266 (0.38)	7/673 (1.04)
	2020	193/35,627 (0.54)	4/980 (0.41)	9/2003 (0.45)
	2021	301/44,235 (0.68)	5/1211 (0.41)	9/2076 (0.43)
Embarrassment				
	Overall	171/105,231 (0.16)	2/2729 (0.07)	8/5476 (0.15)
	2011 - 2017	9/5282 (0.17)	0/98 (0)	0/314 (0)
	2018	3/8144 (0.04)	0/174 (0)	1/410 (0.24)
	2019	19/11,943 (0.16)	0/266 (0)	0/673 (0)
	2020	55/35,627 (0.15)	0/980 (0)	1/2003 (0.05)
	2021	85/44,235 (0.19)	2/1211 (0.17)	6/2076 (0.29)
Excitement				
	Overall	933/105,231 (0.89)	9/2729 (0.33)	28/5476 (0.51)
	2011 - 2017	23/5282 (0.44)	0/98 (0)	2/314 (0.64)
	2018	37/8144 (0.45)	0/174 (0)	2/410 (0.49)
	2019	91/11,943 (0.76)	0/266 (0)	4/673 (0.59)
	2020	305/35,627 (0.86)	5/980 (0.51)	10/2003 (0.5)
	2021	477/44,235 (1.08)	4/1211 (0.33)	10/2076 (0.48)
Fear				
	Overall	3223/105,231 (3.06)	118/2729 (4.32)	247/5476 (4.51)
	2011 - 2017	203/5282 (3.84)	2/98 (2.04)	18/314 (5.73)
	2018	283/8144 (3.47)	4/174 (2.3)	26/410 (6.34)
	2019	396/11,943 (3.32)	10/266 (3.76)	26/673 (3.86)
	2020	1118/35,627 (3.14)	50/980 (5.1)	95/2003 (4.74)

Emotion		All messages, n/N (%)	Breast conservation, n/N (%)	Mastectomy, n/N (%)
	2021	1223/44,235 (2.76)	52/1211 (4.29)	82/2076 (3.95)
Joy				
	Overall	2392/105,231 (2.27)	48/2729 (1.76)	162/5476 (2.96)
	2011 - 2017	97/5282 (1.84)	0/98 (0)	8/314 (2.55)
	2018	136/8144 (1.67)	2/174 (1.15)	14/410 (3.41)
	2019	234/11,943 (1.96)	4/266 (1.5)	20/673 (2.97)
	2020	786/35,627 (2.21)	18/980 (1.84)	60/2003 (3)
	2021	1139/44,235 (2.57)	24/1211 (1.98)	60/2076 (2.89)
Love				
	Overall	2095/105,231 (1.99)	24/2729 (0.88)	89/5476 (1.63)
	2011 - 2017	96/5282 (1.82)	0/98 (0)	9/314 (2.87)
	2018	113/8144 (1.39)	0/174 (0)	9/410 (2.2)
	2019	178/11,943 (1.49)	3/266 (1.13)	7/673 (1.04)
	2020	678/35,627 (1.9)	7/980 (0.71)	31/2003 (1.55)
	2021	1030/44,235 (2.33)	14/1211 (1.16)	33/2076 (1.59)
Nervousness				
	Overall	1218/105,231 (1.16)	57/2729 (2.09)	85/5476 (1.55)
	2011 - 2017	68/5282 (1.29)	1/98 (1.02)	5/314 (1.59)
	2018	109/8144 (1.34)	2/174 (1.15)	2/410 (0.49)
	2019	155/11,943 (1.3)	5/266 (1.88)	6/673 (0.89)
	2020	404/35,627 (1.13)	22/980 (2.24)	29/2003 (1.45)
	2021	482/44,235 (1.09)	27/1211 (2.23)	43/2076 (2.07)
Neutral				
	Overall	36,373/105,231 (34.56)	1001/2729 (36.68)	1973/5476 (36.03)
	2011 - 2017	2144/5282 (40.59)	39/98 (39.8)	116/314 (36.94)
	2018	3166/8144 (38.8)	66/174 (37.93)	148/410 (36.1)
	2019	4502/11,943 (37.7)	102/266 (38.35)	288/673 (42.79)
	2020	12,047/35,627 (33.81)	336/980 (34.29)	689/2003 (34.4)
	2021	14,514/44,235 (32.81)	458/1211 (37.82)	732/2076 (35.26)
Pride				
	Overall	91/105,231 (0.09)	0/2729 (0)	5/5476 (0.09)
	2011 - 2017	5/5282 (0.09)	0/98 (0)	0/314 (0)
	2018	4/8144 (0.05)	0/174 (0)	1/410 (0.24)
	2019	9/11,943 (0.08)	0/266 (0)	0/673 (0)
	2020	34/35,627 (0.1)	0/980 (0)	2/2003 (0.1)
	2021	39/44,235 (0.09)	0/1211 (0)	2/2076 (0.1)
Realization				
	Overall	6602/105,231 (6.27)	284/2729 (10.41)	542/5476 (9.9)
	2011 - 2017	259/5282 (4.9)	10/98 (10.2)	27/314 (8.6)
	2018	438/8144 (5.38)	19/174 (10.92)	39/410 (9.51)
	2019	744/11,943 (6.23)	29/266 (10.9)	70/673 (10.4)
	2020	2259/35,627 (6.34)	100/980 (10.2)	198/2003 (9.89)

Emotion	All messages, n/N (%)	Breast conservation, n/N (%)	Mastectomy, n/N (%)
2021	2902/44,235 (6.56)	126/1211 (10.4)	208/2076 (10.02)

[a]Comparison of emotion trends overall and over time across all paragraphs and separated by breast conservation or mastectomy, nonexclusively, in r/breastcancer.

Five emotions became increasingly prevalent for BCS, although they were rare: anger (2011 - 2017: 0/98, 0%; 2021: 5/1211, 0.41%), annoyance (2011 - 2017: 0/98, 0%; 2021: 15/1211, 1.24%), disappointment (2011 - 2017: 0/98, 0%; 2021: 18/1211, 1.49%), disgust (2011 - 2017: 0/98, 0%; 2021: 5/1211, 0.41%), and joy (2011 - 2017: 0/98, 0%; 2021: 24/1211, 1.98%). No emotions showed a consistent trend for mastectomy-related posts (Table 1). Additionally, after 2017, realization and nervousness were more common for BCS than mastectomy annually. Realization, approval, and caring were the most strongly expressed emotions across both BCS (top decile scores: 0.52, 0.48, and 0.43, respectively) and mastectomy (top decile scores: 0.49, 0.41, and 0.5), with breast conservation being more associated with optimism (top decile score: 0.33).

Discussion

Principal Findings

Building upon past research in sentiment analysis of online discussions about breast cancer [4,5,16], NLP identified differences in social media discussions across BCS and mastectomy, reflecting trends reported clinically. Compared to previous studies [4,5,16] conducting sentiment analyses of online forums discussing cancer, our work focused specifically on surgical management options for patients with breast cancer. Our distinct NLP approaches identified that discussions surrounding mastectomy became increasingly positive over time, corresponding with concordant emotions. These findings are consistent with multiple studies that have found a growing trend of patients with early-stage breast cancer choosing mastectomy over BCS [3,17]. While it is not feasible to determine the reason for the observed increase in positive sentiments for mastectomy mentions based on the data available in this study, it does indicate a parallel with real-world patient decision-making.

Discussion in this breast cancer–specific forum increased substantially over time, confirming that patients are increasingly using social media as a resource. BCS and mastectomy-related posting increased, emphasizing trends in content-specific information. In a recent survey study, patients reported that their cancer diagnosis prompted them to join social media platforms, and over 80% of respondents reported using social media to gather information online [9]. The predominance of objectivity (neutrality emotion) and informative (realization and curiosity) emotions supports these findings.

Evidence surrounding treatment choice for early-stage breast cancer suggests the decision to pursue mastectomy over BCS is often driven by fear of recurrence and secondary cancers [18]. Our application of NLP identifies this in the online setting, with BCS posts more likely to express nervousness. Negative emotions such as anger, annoyance, disappointment, and disgust also became increasingly prevalent over time in BCS posts.

This study is limited by confounders. While VADER and GoEmotions are specifically developed for social media text and based on complementary approaches, they also may reflect inaccuracies and biases based on limitations in their training. Moreover, sentiments cannot be explicitly attributed to the topics themselves, but rather to the paragraphs associated with specific treatments. Nevertheless, these paragraphs likely reflect related discussions around each of these treatments or some aspect of related care.

These findings provide unique insight into patient decision-making. Social media reflects real-time discussions in a natural setting with less filtered discussion of patient concerns and experiences. Recognizing the sentiments and emotions expressed surrounding the treatment, the decision-making process can help clinicians create patient-centric recommendations.

Conclusion

As social media becomes more pervasive, patients are increasingly discussing options for breast cancer therapy online. NLP can characterize these candid online patient discussions at scale and help clinicians identify barriers to treatment decisions and strengthen counseling for patients. Additional studies will be required to see if ongoing social media sentiment trends continue to track patient decisions.

Acknowledgments

This study was supported in part by the ASCO Conquer Cancer Career Development Award, the Radiation Oncology Institute, and the UCSF Computational Cancer Award. Funders had no role in study design, data collection and analysis, decision to publish, and preparation of the manuscript. JCH is supported by the ASTRO-PCF Career Development Award to End Prostate Cancer. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect those of the American Society of Clinical Oncology or Conquer Cancer.

Data Availability

All data used for this analysis were obtained through a publicly available application programming interface (PushShift's Reddit API) using Python.

Authors' Contributions

Conceptualization: DYS, IDF, WZ, TZ, GY, NP, LS, CP, JCH
Data curation: DYS, IDF, WZ, TZ, GY, JW, JCH
Formal analysis: DYS, IDF, WZ, TZ, GY, JW, JCH
Investigation: DYS, IDF, WZ, TZ, GY, JW, JCH
Methodology: DYS, IDF, WZ, TZ, GY, JCH
Project administration: DYS, JCH
Resources: JCH
Software: IDF, WZ, TZ, GY, JCH
Supervision: DYS, JCH
Validation: DYS, NP, LS, CP, JCH
Visualization: DYS, IDF, WZ, JW, JCH
Writing – original draft: DYS, IDF, WZ, JW, JCH
Writing – review & editing: DYS, IDF, WZ, TZ, GY, JW, NP, LS, CP, JCH

Conflicts of Interest

JCH has received research funding from Roche, unrelated to this work.

References

1. Early Breast Cancer Trialists' Collaborative Group, Darby S, McGale P, et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. Lancet 2011 Nov 12;378(9804):1707-1716. [doi: 10.1016/S0140-6736(11)61629-2] [Medline: 22019144]
2. de Boniface J, Szulkin R, Johansson ALV. Survival after breast conservation vs mastectomy adjusted for comorbidity and socioeconomic status: a Swedish national 6-year follow-up of 48 986 women. JAMA Surg 2021 Jul 1;156(7):628-637. [doi: 10.1001/jamasurg.2021.1438] [Medline: 33950173]
3. Chang JM, Kosiorek HE, Dueck AC, et al. Trends in mastectomy and reconstruction for breast cancer: a twelve year experience from a tertiary care center. Am J Surg 2016 Dec;212(6):1201-1210. [doi: 10.1016/j.amjsurg.2016.08.020] [Medline: 27866726]
4. Portier K, Greer GE, Rokach L, et al. Understanding topics and sentiment in an online cancer survivor community. J Natl Cancer Inst Monogr 2013 Dec;2013(47):195-198. [doi: 10.1093/jncimonographs/lgt025] [Medline: 24395991]
5. Cabling ML, Turner JW, Hurtado-de-Mendoza A, et al. Sentiment analysis of an online breast cancer support group: communicating about tamoxifen. Health Commun 2018 Sep;33(9):1158-1165. [doi: 10.1080/10410236.2017.1339370] [Medline: 28678549]
6. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. JMIR Med Inform 2017 Jul 31;5(3):e23. [doi: 10.2196/medinform.7779] [Medline: 28760725]
7. Allen CG, Roberts M, Andersen B, Khoury MJ. Communication about hereditary cancers on social media: a content analysis of tweets about hereditary breast and ovarian cancer and Lynch syndrome. J Canc Educ 2020 Feb;35(1):131-137. [doi: 10.1007/s13187-018-1451-4]
8. Pluta M. Online self-disclosure and social sharing of emotions of women with breast cancer using Instagram: qualitative conventional content analysis. Chronic Illn 2022 Dec;18(4):834-848. [doi: 10.1177/17423953211039778] [Medline: 34494897]
9. Braun LA, Zomorodbakhsch B, Keinki C, Huebner J. Information needs, communication and usage of social media by cancer patients and their relatives. J Cancer Res Clin Oncol 2019 Jul;145(7):1865-1875. [doi: 10.1007/s00432-019-02929-9] [Medline: 31123824]
10. Wallner LP, Martinez KA, Li Y, et al. Use of online communication by patients with newly diagnosed breast cancer during the treatment decision process. JAMA Oncol 2016 Dec 1;2(12):1654-1656. [doi: 10.1001/jamaoncol.2016.2070] [Medline: 27468161]
11. Demszky D, Movshovitz-Attias D, Ko J, Cowen A, Nemade G, Ravi S. GoEmotions: a dataset of fine-grained emotions. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Association for Computational Linguistics; 2020:4040-4054. [doi: 10.18653/v1/2020.acl-main.372]
12. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. Proc Int AAAI Conf Web Soc Media 2014;8(1):216-225. [doi: 10.1609/icwsm.v8i1.14550]

13. Duggan M, Smith A. 6% of online adults are Reddit users. CORE. 2013 Jul 7. URL: https://core.ac.uk/outputs/71362264?source=oai [accessed 2025-01-02]
14. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The Pushshift Reddit dataset. Proc Int AAAI Conf Web Soc Media 2020;14(1):830-839. [doi: 10.1609/icwsm.v14i1.7347]
15. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online on Oct 11, 2018 URL: https://arxiv.org/abs/1810.04805 [accessed 2025-01-13] [doi: 10.48550/arXiv.1810.04805]
16. Park SK, Park HA, Lee J. Understanding the public's emotions about cancer: analysis of social media data. Int J Environ Res Public Health 2020 Sep 30;17(19):7160. [doi: 10.3390/ijerph17197160] [Medline: 33007865]
17. Kummerow KL, Du L, Penson DF, Shyr Y, Hooks MA. Nationwide trends in mastectomy for early-stage breast cancer. JAMA Surg 2015 Jan;150(1):9-16. [doi: 10.1001/jamasurg.2014.2895] [Medline: 25408966]
18. Katz SJ, Lantz PM, Janz NK, et al. Patient involvement in surgery treatment decisions for breast cancer. J Clin Oncol 2005 Aug 20;23(24):5526-5533. [doi: 10.1200/JCO.2005.06.217] [Medline: 16110013]

Abbreviations

BCS: breast-conserving surgery
BCT: breast-conserving therapy
BERT: Bidirectional Encoder Representations from Transformers
cTAKES: Clinical Text Analysis Knowledge Extraction System
NLP: natural language processing
SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms
VADER: Valence Aware Dictionary for Sentiment Reasoning

Edited by N Cahill; submitted 18.09.23; peer-reviewed by A Sarker, D Carvalho, T Meier; revised version received 21.10.24; accepted 29.11.24; published 28.01.25.

Please cite as:
Spiegel DY, Friesner ID, Zhang W, Zack T, Yan G, Willcox J, Prionas N, Singer L, Park C, Hong JC
Exploring the Social Media Discussion of Breast Cancer Treatment Choices: Quantitative Natural Language Processing Study
JMIR Cancer 2025;11:e52886
URL: https://cancer.jmir.org/2025/1/e52886
doi:10.2196/52886

Breast Cancer Screening Participation and Internet Search Activity in a Japanese Population: Decade-Long Time-Series Study

Noriaki Takahashi[1,2], MPH, BS; Mutsuhiro Nakao[3], MD, MPH, PhD; Tomio Nakayama[2], MD, PhD; Tsutomu Yamazaki[1,4], MD, PhD

[1]Graduate School of Medicine, International University of Health and Welfare, Tokyo, Japan

[2]Division of Screening Assessment and Management, National Cancer Center Institute for Cancer Control, Tokyo, Japan

[3]Institute of Stress Management, Showa University, Tokyo, Japan

[4]Innovation and Research Support Center, International University of Health and Welfare, Tokyo, Japan

Corresponding Author:
Noriaki Takahashi, MPH, BS
Graduate School of Medicine, International University of Health and Welfare, Tokyo, Japan

Abstract

Background: Breast cancer is a major health concern in various countries. Routine mammography screening has been shown to reduce breast cancer mortality, and Japan has set national targets to improve screening participation and increase public attention. However, collecting nationwide data on public attention and activity is not easy. Google Trends can reveal changes in societal interest, yet there are no reports on the relationship between internet search volume and nationwide participation rates in Japan.

Objective: This study aims to reveal and discuss the relationship between public awareness and actual behavior in breast cancer screening by examining trends in internet search volume for the keyword “breast cancer screening” and participation rates over a decade-long period.

Methods: This time-series study evaluated the association between internet search volume and breast cancer screening participation behavior among women aged 60 - 69 years in Japan from 2009 to 2019. Relative search volume (RSV) data for the search term “breast cancer screening (nyuugan-kenshin)” were extracted from Google Trends as internet search volume. Breast cancer screening and further assessment participation rates were based on government municipal screening data. Joinpoint regression analyses were conducted with weighted BIC to evaluate the time trends. An ethics review was not required because all data were open.

Results: The RSV for “breast cancer screening (nyuugan-kenshin)” peaked in June 2017 (100) and showed clear spikes in June 2016 (94), September (69), and October (77) 2015. No RSVs above 60 were observed except around these three specific periods, and the average RSV for the entire period was 30.7 (SD 16.2). Two statistically significant joinpoints were detected, rising in December 2013 and falling in June 2017. Screening participation rates showed a temporary increase in 2015 in a slowly decreasing trend, and no joinpoints were detected. Further assessment participation rates showed a temporary spike in 2015 in the middle of an increasing trend, with a statistically significant point of slowing increase detected in 2015. Post hoc manual searches revealed that Japanese celebrities’ breast cancer diagnoses were announced on the relevant dates, and many Japanese media reports were found.

Conclusions: This study found a notable association between internet search activity and celebrity cancer media reports and a temporal association with screening participation in breast cancer screening in Japan. Celebrity cancer media reports triggered internet searches for cancer screening, but this did not lead to long-term changes in screening participation behavior. This finding suggests what information needs to be provided to citizens to encourage participation in screening.

(JMIR Cancer 2025;11:e64020) doi:10.2196/64020

KEYWORDS

breast cancer; cancer screening; internet use; mass media; public health surveillance; health belief model; mammography; awareness; Japanese; Google

Introduction

Breast cancer is a major health concern that affects large numbers of women in various countries. It is the most common cancer in women, with approximately 2.3 million new cases diagnosed and 680,000 deaths reported in 2020 [1,2]. The disease burden of breast cancer is also high in Japan. The age-adjusted incidence rate of breast cancer continues to increase every year, and it has been reported that breast cancer accounts for about 20% of all cancers in women in Japan [3,4]. As Japan’s

population ages, the burden of breast cancer is predicted to increase [5].

Several studies have indicated the efficacy of mammography screening for breast cancer in reducing the burden of breast cancer. Routine mammography screening has been shown to reduce breast cancer mortality by 25% - 31% [6,7]. The long-term effects of mammography have also been shown in a 30-year follow-up study [8]. In Japan, national policy recommends biennial mammography for breast cancer screening in women older than 40 years [9]. Therefore, screening is a practical approach to reducing the long-term burden of breast cancer, and its importance is increasing in Japan.

Cancer screening programs in Japan are divided into two main types: municipal screening and workplace screening. Municipalities conduct screenings, and information is collected by the Ministry of Health, Labor, and Welfare and made available to the public. This is the only data on cancer screening for all regions in Japan for which the government reports actual statistics annually. This report includes data on screening and further assessment participation. The participation rate is the proportion of the target population that receives the primary screening test, which in the case of breast cancer screening is mammography. The further assessment participation rate is the proportion of women who, after a positive screening test result, receive the following test to confirm the diagnosis: fine needle aspiration cytology or core needle biopsy in breast cancer screening. Even if a person participates in screening, the effectiveness of cancer screening will not be fully realized unless the screening-positive person receives a further assessment. The further assessment participation rate is an important indicator, as is the screening participation rate. Japan has set national targets to improve screening participation, further assessment participation, and increase public attention to the importance of cancer screening [10]. Understanding the public's attention and behavior around cancer screening is critical to improving screening participation rates. However, it is not easy to collect nationwide data on public attention and activity to assess the association with screening participation.

Internet search volume has recently become one of the most valuable tools for exploring human interests and behavior. Google Trends is a popular open web-based tool that quantifies changes in internet search volume for a given term based on actual Google search history [11,12]. Google Trends is used for academic research in fields as diverse as social science, economics, language, and medicine and can also reveal changes in societal interest in public health issues [13]. Google Trends initially focused on detecting infectious disease outbreaks, and past studies have reported early detection of influenza outbreaks [14].

Google Trends is now expected to be used in noncommunicable disease areas such as mental health and preventive behaviors and is a potential source of information for understanding the public's interest in and behaviors around cancer screening [15]. Malaysia reported a significant correlation between Google Trends search patterns and Pink Ribbon Month, a breast cancer awareness campaign [16]. Among several internet search engines, Google was also shown to have the best search validity (regarding whether a web page could be opened) for breast cancer screening information [17]. In contrast, a previous Japanese study analyzed trends in cervical cancer and reported no change in public interest during the cervical cancer awareness month [18].

Therefore, there is considerable interest in the relationship between internet search activity and cancer screening. In Japan, it would be valuable to determine the relationship between public attention to cancer screening and participation rates at the national level to understand public awareness and behavior. However, there are no reports on the relationship between changes in internet search volume and long-term trends in nationwide participation rates in Japan.

This study examined the relationship between public awareness and actual behavior in breast cancer screening at the national level. This study is the first report in Japan to reveal and discuss the background of the relationship between trends in internet search volume for the keyword "breast cancer screening" and participation rates over a decade-long period. As an example of the application of epidemiologic research using internet search volume, this approach could provide knowledge for promoting cancer screening and providing appropriate information.

Methods

Study Design

This time-series study uses internet search volume and national cancer screening statistics. Internet search volume targets those who conducted searches in Japanese using Google in Japan. Cancer screening data targets municipal screening in Japan. Both data are openly available on the web.

Data Sources (Internet Search Volume)

Google Trends is a data tool that publishes the volume of keyword searches worldwide in Google Search, an internet search engine, since 2004. This tool allows users to access the relative search volume (RSV) but not the absolute number of searches. RSV is calculated on a scale of 0 to 100 based on the volume with the most searches per unit of time in the defined region, period, category, and search term. For example, RSV=30 means 30% of the highest search volume observed within a given condition. RSVs can assess changes in interest in a particular term by showing the relative value of search volume trends over time.

The search term was "nyuugan-kenshin," which means "breast cancer screening" in Japanese. Monthly RSV data from the Google Trends platform were retrieved on September 17, 2023. Since the Japanese term "nyuugan-kenshin" is written as one continuous word without any spaces, we did not enclose it in quotes when using it in Google Trends. Because Google Trends does not provide a "Topic" option for the Japanese term "nyuugan-kenshin," we used the "Search Term" option instead. It was set to Japan as the target region and 2009 - 2019 as the target period. To ensure that all possible contexts in which the term might appear were captured, we set "All categories" as the "Category" and "Web Search" as the "Search Type."

If a significant trend increase was observed, a post hoc manual search was conducted using the "Related Keywords" feature of Google Trends to see if any socially essential media reports might be related to the increase.

Data Sources (Cancer Screening)

For cancer screening, this study included screening participation rates and further assessment participation rates in municipal screening for breast cancer in Japan from 2009 to 2019. Municipal screening does not include individuals who participate in workplace screening. Consequently, when calculating screening participation rates for ages 40 years and older using the population as the denominator, workplace screening participants are excluded from the numerator. This omission can lead to fluctuations in time-series data, for example, if there is a change over time in the proportion of working individuals. Furthermore, when participation rates differ by age group, they are also affected by changes in the age distribution over time. To eliminate this effect as much as possible and to improve the time-series analysis's validity, the calculation of participation rates was restricted to women aged 60 - 69, who are mainly retired. The number of participants in screening, positive cases in screening, and further assessment participants were obtained from the "Report on Regional Public Health Services and Health Promotion Services" by the Ministry of Health, Labor, and Welfare [19]. Population data were obtained from the Statistics Bureau of the Ministry of Internal Affairs and Communications [20]. Screening participation rates were calculated by dividing the number of screening participants by the population of women in the target age group. The further assessment participation rates were calculated by dividing the number of further assessment participants by the number of positive screening cases. The recommended interval for breast cancer screening in Japan is once every two years, and the original "screening participation rates" are calculated by considering the number of participants screened for two years. However, this calculation method equalizes two years of information and may not detect sensitive value changes. As this study aimed to detect changes over time rather than absolute assessments, "screening participation rates" were defined as calculated values per one-year period and used in the analyses.

Statistical Analysis

Joinpoint regression analyses were performed to assess RSV trends quantitatively. This analysis is an appropriate way to examine data over time and statistically detect points of change in gradient [21]. The software used was the Joinpoint Regression Program (version 5.0.2, Statistical Research and Applications Branch, National Cancer Institute) [22]. The statistical method used for joinpoint detection was weighted BIC, a standard method in Program version 5.0 and later. Weighted BIC is the most flexible and adaptable method for various situations in this software. Joinpoint regression analysis requires many computing resources, and the calculation time depends on the maximum number of detectable joinpoints set before the calculation. This analysis's maximum number of joinpoints was set to 3 due to calculation time. Changing the maximum number of joinpoints can alter the significance level for individual tests and potentially change the number of joinpoints in the optimal model [22]. When no joinpoints were detected in the initial analysis, we conducted additional analyses with the maximum number set to two and one. The statistical significance level for joinpoint detection was set to 0.05.

Ethical Considerations

This study was conducted per the principles of the Declaration of Helsinki. An ethics review was not required because all data used in this study were open. For this type of study, formal consent is not required.

Results

Internet Search Volume

Figure 1 shows the trend of RSVs for the search term "breast cancer screening (nyuugan-kenshin)" from 2009 to 2019. The RSV peaked in June 2017 (100) and showed clear spikes in June 2016 (94), September 2015 (69), and October 2015 (77). No RSVs above 60 were observed except around these three specific periods. The average RSV for the entire period was 30.7 (SD 16.2). Figure 2 shows the results of the joinpoint regression analysis for RSVs. Two statistically significant joinpoints were detected, rising in December 2013 and falling in June 2017.

Figure 1. Monthly “breast cancer screening (nyuugan-kenshin)” relative search volumes in Japan from 2009 to 2019, based on Google Trends. The black arrows show the timing of media reports on the celebrities’ breast cancer diagnoses or passing.

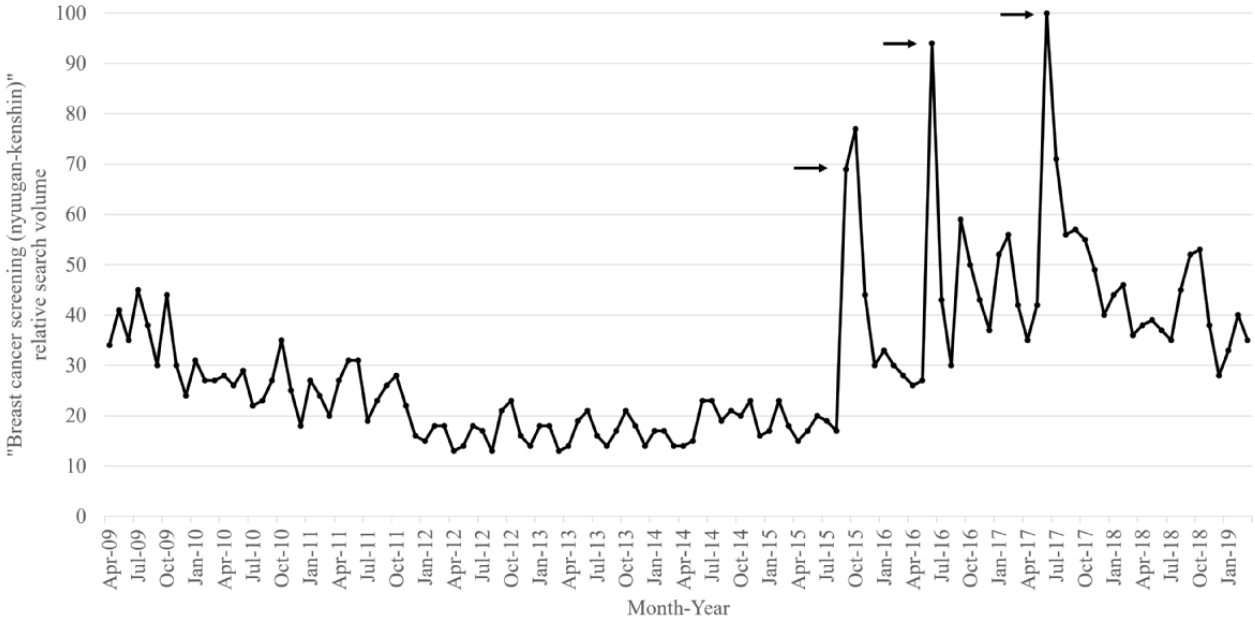
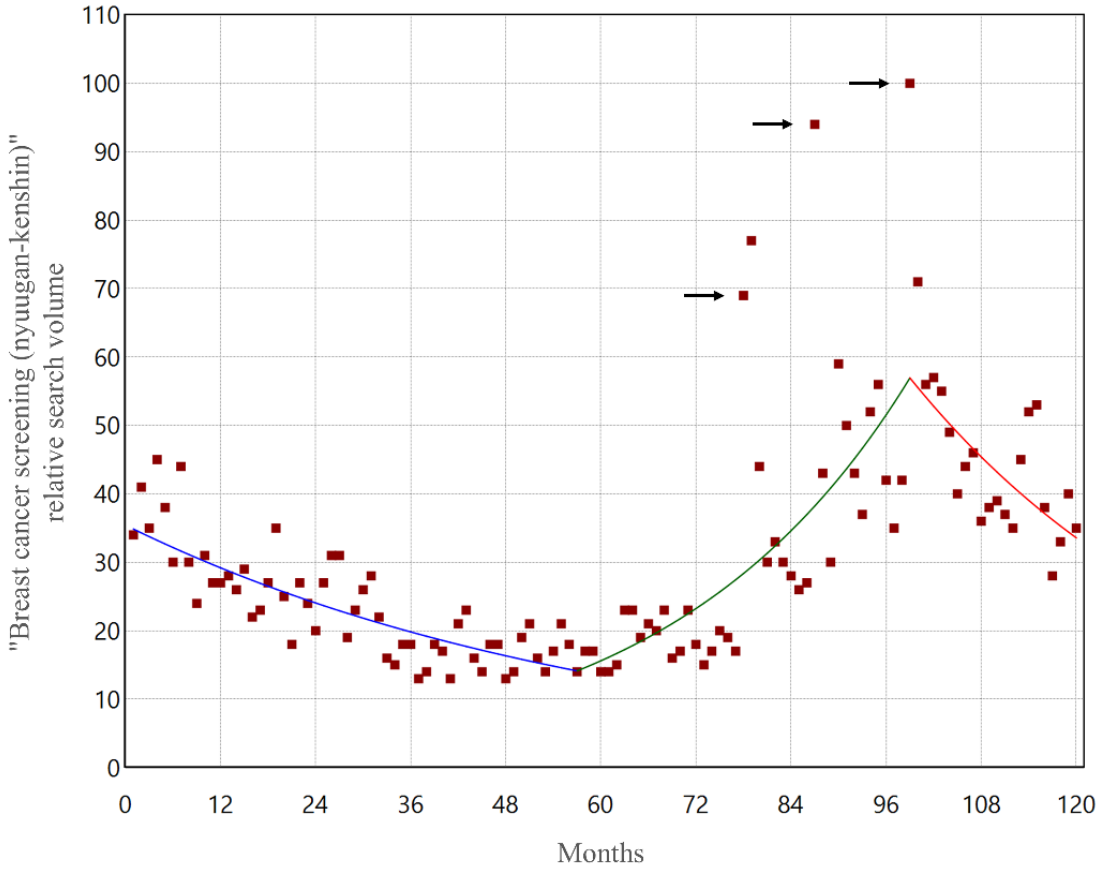


Figure 2. Joinpoint regression analysis of the monthly “breast cancer screening (nyuugan-kenshin)” relative search volumes in Japan from 2009 to 2019. Two significant joinpoints were detected (December 2013 and June 2017). The black arrows show the timing of media reports on the celebrities’ breast cancer diagnoses or passing.



Cancer Screening Participation

Figure 3 shows the trend of breast cancer screening participation rates from 2009 to 2019. Visual observation shows a temporary increase in 2015 in the slowly decreasing trend. Figure 4 shows the results of the joinpoint analysis for screening participation rates. No joinpoints were detected. Even in additional analyses with the maximum number set to two or one, no joinpoints were detected.

Figure 3. Annual breast cancer screening participation rates (mammography) among Japanese women aged 60 - 69 years from 2009 to 2019, based on municipal screening data. The rate is the proportion of screening participants in the target population.

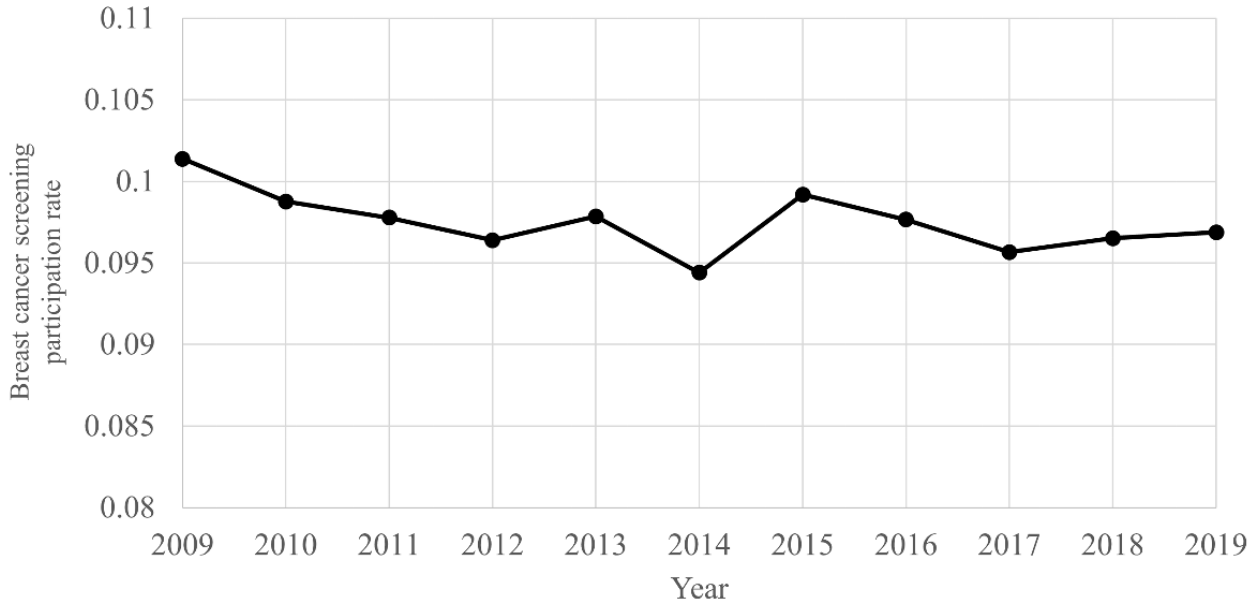


Figure 4. Joinpoint regression analysis of the annual breast cancer screening participation rates (mammography) among Japanese women aged 60 - 69 years from 2009 to 2019. No joinpoints were detected.

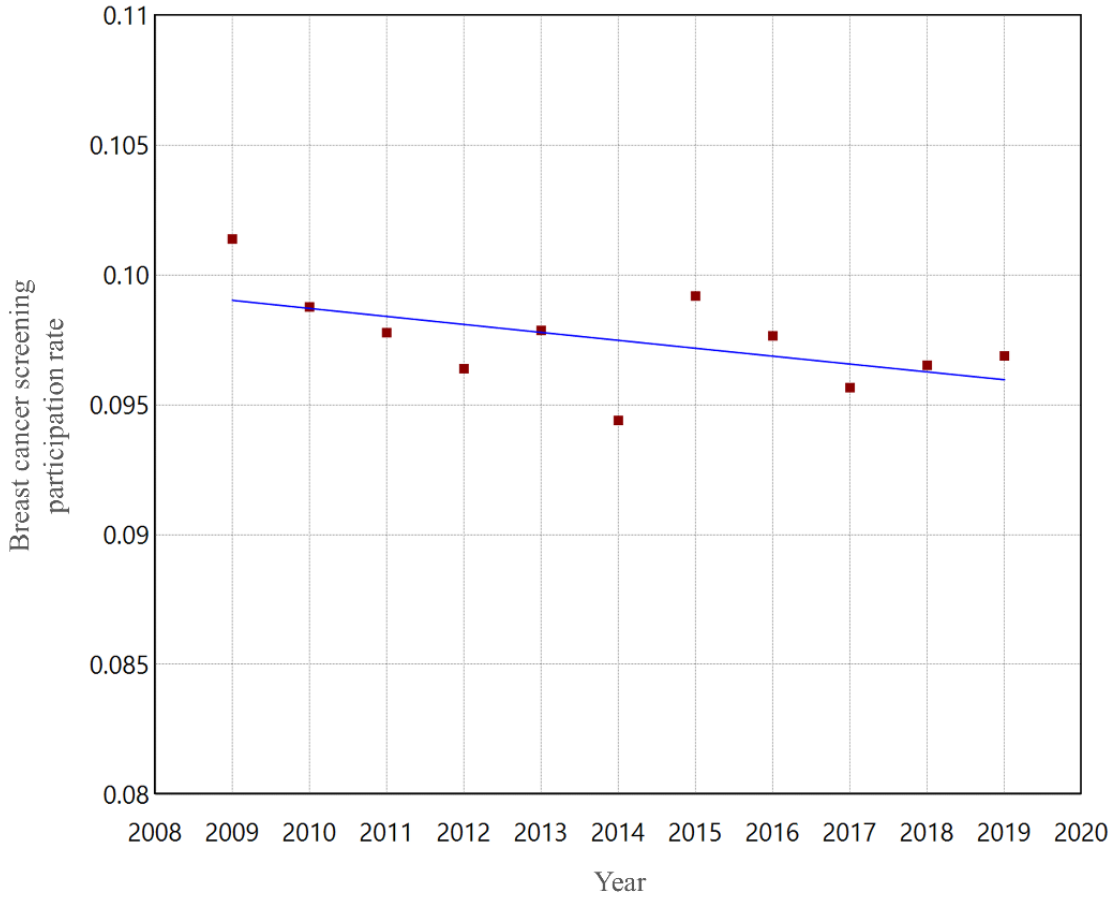


Figure 5 shows the trend of further assessment participation rates for breast cancer screening from 2009 to 2019. Visual observation shows a temporary spike in 2015 in the middle of an increasing trend. Figure 6 shows the result of the joinpoint analysis for further assessment participation rates. While the trend has been increasing for the entire period, a statistically significant point of slowing increase was detected in 2015. The year 2015 was the maximum for screening and further assessment participation rates, except for 2009 and 2019, the two ends of the period covered.

Figure 5. Annual breast cancer screening further assessment participation rates among Japanese women aged 60 - 69 years from 2009 to 2019, based on municipal screening data. The rate is the proportion of women who received fine needle aspiration cytology or core needle biopsy following a positive screening result.

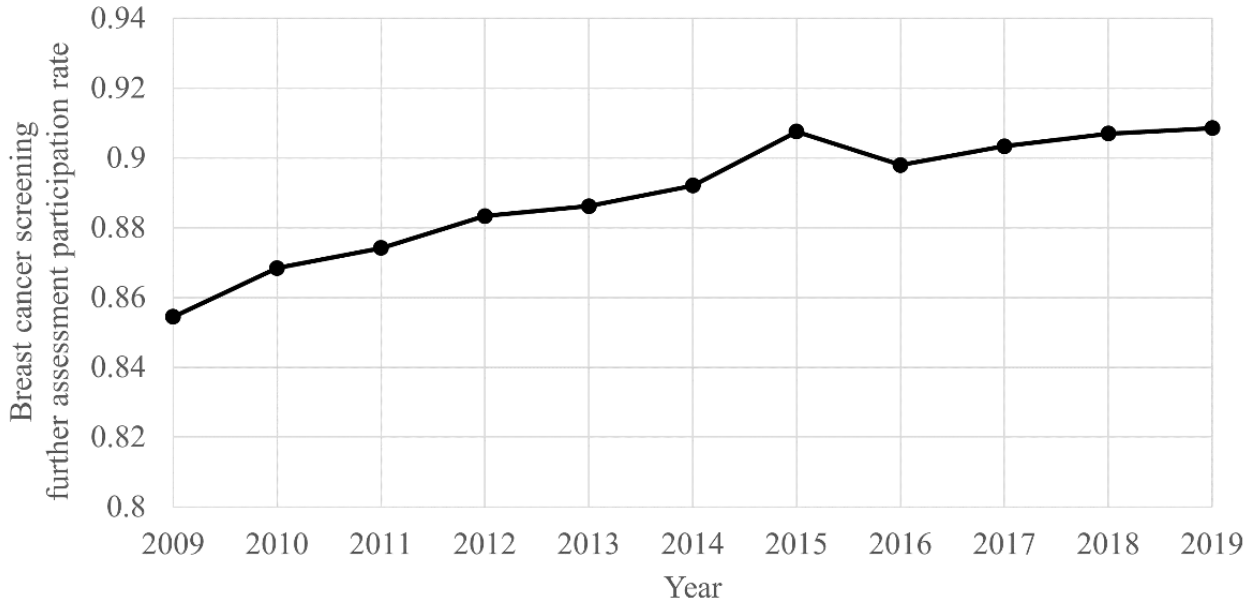
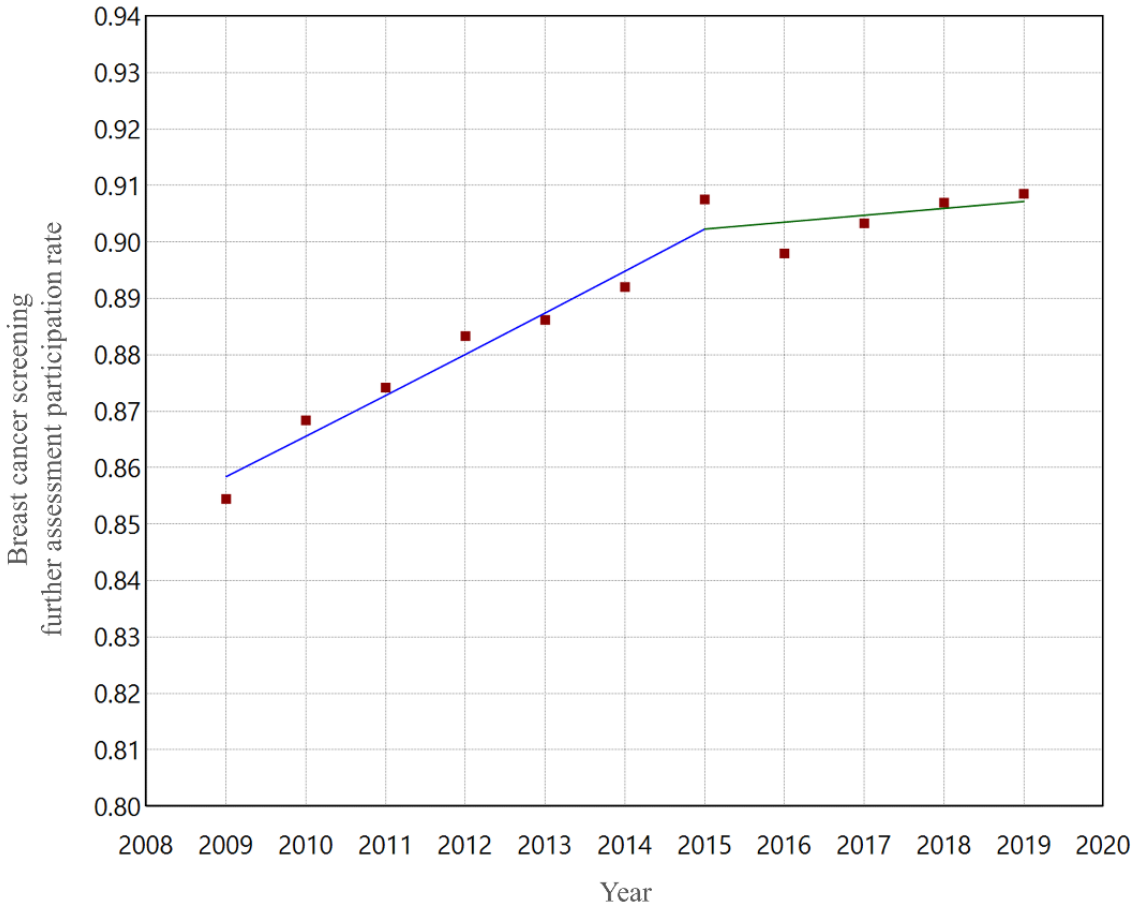


Figure 6. Joinpoint regression analysis of the annual breast cancer screening further assessment participation rates among Japanese women aged 60 - 69 years from 2009 to 2019. One significant joinpoint was detected in 2015.



Post Hoc Manual Search

When RSV spiked in 2015, 2016, and 2017, three periods were explored post hoc. A detailed review of the RSV data downloaded from Google Trends, restricting the period specified, revealed a notable spike on September 23 and 24, 2015, June 9, 2016, and June 23, 2017. The day's news and Google Trends "Related Keywords" were manually searched for these three periods. For the term "nyuugan-kenshin," only general keywords such as "examination," "cost," "mammography," "breastfeeding," and "symptoms" were suggested on each of the relevant dates. In contrast, as "Related Keywords" for "nyuugan (breast cancer)" on September 23 and 24, 2015, a Japanese celebrity, AH, was suggested as a nonmedical term. A Google search limited to the same period revealed that her breast cancer incidence was announced on September 23, 2015, and many Japanese media reports were found. In the same way, the Japanese celebrity MK's breast cancer incidence was announced on June 9, 2016, and her passing on June 23, 2017, confirmed many media reports.

Discussion

General Interpretation of the Results

We conducted a time-series study for 2009 - 2019 on Japanese internet search volume and breast cancer screening data for 60 - 69-year-olds. Internet search RSVs for "breast cancer screening" spiked notably in three specific periods in September 2015, June 2016, and June 2017. The joinpoint analysis for RSVs revealed two joinpoints in December 2013 and June 2017, showing an increase over the period, including the spikes mentioned above. The 10-year trend in internet searches for breast cancer screening was dynamically changing, down, up, and down. The joinpoint of screening data was not detected for screening participation rates but was detected in 2015 for further assessment. Upon examination of Figure 5, it might be reasonable to interpret this as a temporary increase in 2015 and a return to the original trend from 2016 onwards. Indeed, the screening and further assessment participation rates reached their maximum in 2015, except for 2009 and 2019, the two ends of the period covered. A post hoc search for the timing of the three RSV spikes from 2015 to 2017 was consistent with the dates of media reports of breast cancer incidences and the passing of Japanese celebrities.

It is worth noting that 2015 marked the timing of the first media reports on celebrities, the first notable increase in RSVs, and the short-term maximum in the screening and further assessment participation rates. In particular, the consistency of the three dates between celebrity media reports and search trends is evident. RSV provided by Google Trends is not an absolute number of searches but a relative measure of the maximum number of searches scaled to 100 within a defined period. Therefore, if there is even one moment of drastic increase in search, RSV for the rest of the period will be relatively low. The fact that the average volume for the entire period in this analysis was 30.7 (SD 16.2) highlights the magnitude of the three spikes. Many individuals became interested in breast cancer screening when AH and MK were featured in the media, which probably triggered internet search behavior. Additionally, screening and further assessment participation rates increased temporarily in 2015, suggesting some individuals may have engaged in screening participation behaviors. Celebrity media reports may have influenced individuals, leading to search and screening participation behavior. This contrasts with a previous Japanese study that showed no association between cervical cancer awareness months and RSVs for "cervical cancer" [18].

However, screening and further assessment participation rates showed only a temporary spike in 2015 and did not increase the long-term trend. RSVs also declined after the 2017 joinpoint. These findings provide insight into the mechanisms necessary for citizens to be concerned about, act on, and maintain their health. To interpret this study's results, referring to the findings of established behavioral models and previous studies would be appropriate.

Interpretation Based on the Health Belief Model

Participation behavior in cancer screening has been a critical subject of study in the Health Belief Model (HBM). HBM is a theoretical model in the behavioral sciences that aims to explain, predict, and promote individual health behaviors. This model was developed in the 1950s to understand the factors determining participation in immunization and screening [23-25]. HBM considers that individual health behaviors are determined by the interaction of six factors: "perceived susceptibility," "perceived severity," "perceived benefits," "perceived barriers," "self-efficacy," and "cues to action." The model is widely used to design education and promotion programs for health activities. In cancer screening, the model has been primarily used to improve participation in colorectal cancer screening and has been validated in several randomized controlled trials [26,27]. Recently, there has been much research on breast cancer screening [28-33].

It would be meaningful to interpret the results of this study based on HBM. In the media reports, AH was 48 years old, and MK was 32 when they were diagnosed. Consecutive media reports of breast cancer in young celebrities may have caused "perceived susceptibility" among the public. MK passed away about a year after her diagnosis was announced. The sad outcome of the celebrity, which had been worrying through media reports, would have caused "perceived severity" for the public. Indeed, the RSV peaked on June 23, 2017, when MK's passing was announced. MK's weblog about her fight against breast cancer has attracted attention in Japan and around the world, and she was named one of the BBC's 100 Women of the Year in 2016 [34,35]. Citizen exposure to a series of media reports may have fulfilled these elements in the HBM. Additionally, for those who were already aware of "susceptibility" and "severity," media reports on the cancer of celebrities may have become "cues to action" to help them take action.

There have been several reports on the impact of celebrity cancer media reports on the behavior of citizens [36,37]. In Australia, mammography screening appointments increased by 40% in the two weeks following media reports of singer Kylie Minogue's breast cancer diagnosis [38]. In the United States, Angelina Jolie's decision to share her experience with the increased risk of breast and ovarian cancer due to BRCA1 gene

mutations has improved public awareness of the disease and increased genetic testing and breast cancer screening. In particular, Angelina Jolie's influence was reported to be related to "perceived susceptibility" and "cues to action," which are elements of HBM [39,40].

These reports illustrate the appropriateness of interpreting the impact of personal cancer experiences and narratives on people's emotions and behaviors based on the HBM. The findings on these effects support the validity of the interpretation that the two celebrities' media reports were elements of "perceived susceptibility," "perceived susceptibility," and "cues to action" in the HBM.

Importance of Removing Barriers

Conversely, information that provides "perceived benefits" or "self-efficacy" for screening or removes "perceived barriers" is not directly included in the celebrity cancer media reports. In contrast to the case of Angelina Jolie, where there is a direct link between her actions and the benefits of preventive behavior, there is a gap in logic between media reports of celebrity breast cancer and the benefits of screening participation. For internet users, there are few barriers to search action. However, there are significant barriers to screening participation on an entirely different level than internet searches. To participate in screening, citizens must confirm the possible dates, times, locations, and costs, make an appointment, and go to mass screening sites or hospitals. Media reports and internet search activity showed a direct relationship, while screening participation behavior showed a limited response. This suggests that information from media reports and internet searches did not remove barriers to screening participation. Some citizens who participated in screenings triggered by the media reports may not have continued to behave because they were unaware of the benefits. This finding of limited participation versus notable search activity highlights the importance of removing "perceived barriers" in the HBM component. This study's post hoc manual search was limited to breast cancer information for 10 years. In today's Japan, where approximately half of the population will be diagnosed with some form of cancer in their lifetime, media reports on celebrities provide citizens with many opportunities to perceive the susceptibility and severity of cancer. Nonetheless, information to remove barriers does not occur unless someone intends it. The importance of "perceived barriers" in HBM elements has long been recognized [41]. A meta-analysis of 18 communication campaigns shows that "perceived benefits" and "perceived barriers" were consistently the most robust predictors [42]. This study supports the idea that removing barriers is an essential public action to encourage healthy behaviors.

Limitations and Strengths of This Study

There are several limitations to this study. First, there is a restriction due to the time-series analysis design. It is unknown whether specific individuals were exposed to media reports, performed search actions, or participated in screening because this is a comparison through time for the whole population. It is important to note that the results indicate only an association, and do not necessarily imply a causal relationship among celebrity news, search spikes, and screening uptake. Various real-world factors, such as concurrent public health campaigns or medical policies, could have influenced the keyword search volume, the screening participation rate, or both. Even if there is a match between exposure and outcome for an individual, it does not prove causation because confounding by unknown factors cannot be ruled out. Given this study's data sources and design, directly evaluating causality is complex and remains an issue for future research.

Second, there is a lack of data on workplace screening. Japan's cancer screening programs are divided into municipal screening and workplace screening. Due to incomplete legislation on workplace screening, data have not been collected and published for the entire country, and it was necessary to use only municipal screening data. In this study, to remove the effect of the lack of workplace screening data as much as possible, the age range for calculating participation rates was restricted to 60 - 69 years so that retirees would represent most of the population. Due to this restriction, the generalizability of the screening data is limited. In the future, once workplace screening data becomes available, it will be necessary to include those data in the analysis to more accurately evaluate trends in screening participation rates among the working population.

Third, the coverage of Google Trends data. Given that the RSV is based on Google search data, it does not reflect the interests of populations that do not use Google or internet search. The percentage of Japanese aged 60 - 69 years using the internet increased from 71.6% (60-64) and 58% (65-69) in 2009 to 90.5% (60-69) in 2019 [43]. This percentage and time change may have affected the results. Even if internet use was high enough in the age group 60s, RSV is an indicator that includes all ages and does not necessarily reflect search activity in the 60s. Furthermore, even if they use internet search, they may use a search engine other than Google. As of 2019, Google accounted for 92% of the global market share for internet search engines, 93% in Europe and 89% in North America, whereas in Japan, it was 75% [44]. While there is no doubt that Google holds the top market share in Japan, unlike Baidu in China or Yandex in Russia, its relatively lower share compared to Western regions could influence the validity of Google Trends data [45]. Further, an absolute assessment is impossible since RSV is a relative measure for a given period and search term.

Fourth, some of the methods and interpretations of this study were post hoc. In the analysis phase of this study, we found a marked increase in the RSV data for the keyword "breast cancer screening" in three specific periods. To explore background information, we performed a post hoc manual search and found that the media reports of the celebrity matched the RSV spikes. This manual search was not planned at the time the study was designed. Because these manual searches and discussions involve the arbitrariness of the researcher, careful attention should be paid to the validity of the interpretation of the results.

Furthermore, one possible reason that no joinpoints were detected for screening participation rates is that the limited number of data points may not have provided sufficient statistical power. To analyze one or more joinpoints, at least seven data points need to be observed [22]. Although this study had 11 data points for both screening and further assessment

participation rates, exceeding seven, the number of data points may still have been insufficient for detecting any joinpoints.

Despite these limitations, there are strengths to this study. Google Trends, an internet search volume, is a limited source of information that directly reflects changes in the preferences and interests of the entire population over time. Media reports related to changes in internet search volume will be revealed after data analysis. Therefore, the type of study that follows participants prospectively cannot discuss what this study did. Retrospective studies that question about past exposures may cause recall or information bias due to the validity of the questionnaire. Internet search volume has none of these biases and selection biases for study enrollment, so it directly reflects citizens' actual preferences. The screening data also have no self-reporting bias because they are actual values reported to the government by municipalities. This study design is conducive to exploring what influences public interest in cancer screening and leads to participation behavior.

Practical Considerations and Future Implications

We found that internet search volume for "breast cancer screening" was notably associated with media reports of the celebrity's cancer and was temporally associated with participation in screening. Although caution is needed in interpreting causal relationships, it is worth noting that the three periods in which the spike in internet search volume occurred match the media reports of the celebrities. It is reasonable to assume that media reports clearly impacted search activity. The results also showed barriers to screening participation and limitations to explaining behavior only by internet search volume. Established "model" and "effect" in behavioral medicine could describe these associations and limitations. Importantly, pragmatic data without educational intervention or questionnaire surveys supported these effects. This finding suggests that internet search volume is valuable for deciphering individuals' behavior. Internet search volume can also help verify the effectiveness of efficacy findings confirmed by exploratory methods, such as intervention trials, with real-world data. Although it is essential to evaluate various potential biases in internet search volume, with an understanding of its limitations, using it for epidemiologic studies will continue to be beneficial and may suggest improvements in public health policy and risk communication.

Conclusions

The public impact of celebrity cancer media reports found in this study will lead to the development of information and awareness methods to improve and sustain participation in cancer screening. A colorectal cancer awareness campaign conducted on television by Katie Couric, a well-known American television anchor, was associated with an increase in colonoscopy use [46]. The results of this study support the potential for such celebrity publicity for preventive health programs to be temporarily effective in Japan. If doing so, information on removing barriers should be included to maximize and sustain the effect.

Acknowledgments

We would like to thank Editage for English language editing. Additionally, we disclose that we used ChatGPT, a generative artificial intelligence, to assist in proofreading the English text of this manuscript.

Data Availability

This study used only open-source data. All original data are available from the cited sources. Any derivative data sets generated during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

NT and TN contributed to the conceptualization. Methodology was designed by NT, MN, and TY. Formal analysis, investigation, data curation, writing—original draft, and visualization were conducted by NT. Writing—review and editing was performed by MN, TN, and TY. Project administration was handled by MN, and supervision was provided by TN and TY.

Conflicts of Interest

None declared.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249. [doi: 10.3322/caac.21660] [Medline: 33538338]
2. World Health Organization. Breast cancer. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/breast-cancer [accessed 2024-06-10]
3. Japanese Ministry of Health, Labour and Welfare. Vital statistics in Japan. 2022. URL: https://www.mhlw.go.jp/english/database/db-hw/index.html [accessed 2024-06-10]
4. Center for Cancer Control and Information Services, National Cancer Center. Cancer statistics in Japan 2024. : Foundation for Promotion of Cancer Research; 2024. URL: https://ganjoho.jp/public/qa_links/report/statistics/2024_en.html [accessed 2024-06-10]

5. Nguyen PT, Saito E, Katanoda K. Long-term projections of cancer incidence and mortality in Japan and decomposition analysis of changes in cancer burden, 2020-2054: an empirical validation approach. Cancers (Basel) 2022 Dec 9;14(24):6076. [doi: 10.3390/cancers14246076] [Medline: 36551562]
6. Hamashima C, Ohta K, Kasahara Y, et al. A meta-analysis of mammographic screening with and without clinical breast examination. Cancer Sci 2015 Jul;106(7):812-818. [doi: 10.1111/cas.12693] [Medline: 25959787]
7. Nelson HD, Fu R, Cantor A, Pappas M, Daeges M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 U.S. Preventive Services Task Force recommendation. Ann Intern Med 2016 Feb 16;164(4):244-255. [doi: 10.7326/M15-0969] [Medline: 26756588]
8. Tabár L, Vitak B, Chen THH, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011 Sep;260(3):658-663. [doi: 10.1148/radiol.11110469] [Medline: 21712474]
9. Japanese Ministry of Health, Labour and Welfare. Cancer screening in municipalities [in Japanese]. 2024. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000059490.html [accessed 2024-06-10]
10. Japanese Ministry of Health, Labour and Welfare. Basic plan to promote cancer control programs [in Japanese]. 2023. URL: http://www.mhlw.go.jp/bunya/kenkou/gan_keikaku.html [accessed 2024-06-10]
11. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439. [doi: 10.2196/13439] [Medline: 31144671]
12. Google LLC. Google trends. 2024. URL: https://trends.google.com/trends [accessed 2024-06-10]
13. Nuti SV, Wayda B, Ranasinghe I, et al. The use of Google Trends in health care research: a systematic review. PLoS One 2014;9(10):e109583. [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
14. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nat New Biol 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
15. Schootman M, Toor A, Cavazos-Rehg P, et al. The utility of Google Trends data to examine interest in cancer screening. BMJ Open 2015 Jun 8;5(6):e006678. [doi: 10.1136/bmjopen-2014-006678] [Medline: 26056120]
16. Mohamad M, Kok HS. Using Google Trends data to study public interest in breast cancer screening in Malaysia. Asian Pac J Cancer Prev 2019 May 25;20(5):1427-1432. [doi: 10.31557/APJCP.2019.20.5.1427] [Medline: 31127903]
17. Wang L, Wang J, Wang M, Li Y, Liang Y, Xu D. Using Internet search engines to obtain medical information: a comparative study. J Med Internet Res 2012 May 16;14(3):e74. [doi: 10.2196/jmir.1943] [Medline: 22672889]
18. Hagiya H, Koyama T, Otsuka F. Impact of the cervical cancer awareness months on public interest in Japan: a Google Trends analysis, 2012-2021. Sci Rep 2022 Sep 13;12(1):15391. [doi: 10.1038/s41598-022-19798-x] [Medline: 36100649]
19. Japanese Ministry of Health, Labour and Welfare. Report on regional public health services and health promotion services 2019. 2022. URL: https://www.mhlw.go.jp/english/database/db-hss/rrphshps.html [accessed 2024-06-10]
20. Statistics Bureau of the Ministry of Internal Affairs and Communications. Population estimates. 2024. URL: https://www.stat.go.jp/english/data/jinsui/index.html [accessed 2024-06-10]
21. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med 2000 Feb;19(3). [doi: 10.1002/(SICI)1097-0258(20000215)19:3%3C335::AID-SIM336%3E3.0.CO;2-Z]
22. National Cancer Institute. Joinpoint trend analysis software. 2024. URL: https://surveillance.cancer.gov/joinpoint [accessed 2025-01-16]
23. Rosenstock IM. Why people use health services. Milbank Mem Fund Q 1966 Jul;44(3):94-127. [doi: 10.1111/j.1468-0009.2005.00425.x] [Medline: 5967464]
24. Becker MH, Maiman LA. Sociobehavioral determinants of compliance with health and medical care recommendations. Med Care 1975 Jan;13(1):10-24. [doi: 10.1097/00005650-197501000-00002] [Medline: 1089182]
25. Rosenstock IM. Historical origins of the health belief model. Health Educ Monogr 1974 Dec;2(4):328-335. [doi: 10.1177/109019817400200403]
26. Rawl SM, Menon U, Burness A, Breslau ES. Interventions to promote colorectal cancer screening: an integrative review. Nurs Outlook 2012;60(4):172-181. [doi: 10.1016/j.outlook.2011.11.003] [Medline: 22261002]
27. Rawl SM, Skinner CS, Perkins SM, et al. Computer-delivered tailored intervention improves colon cancer screening knowledge and health beliefs of African-Americans. Health Educ Res 2012 Oct;27(5):868-885. [doi: 10.1093/her/cys094] [Medline: 22926008]
28. Saei Ghare Naz M, Simbar M, Rashidi Fakari F, Ghasemi V. Effects of model-based interventions on breast cancer screening behavior of women: a systematic review. Asian Pac J Cancer Prev 2018 Aug 24;19(8):2031-2041. [doi: 10.22034/APJCP.2018.19.8.2031] [Medline: 30139040]
29. Sohl SJ, Moyer A. Tailored interventions to promote mammography screening: a meta-analytic review. Prev Med 2007 Oct;45(4):252-261. [doi: 10.1016/j.ypmed.2007.06.009] [Medline: 17643481]
30. Agide FD, Sadeghi R, Garmaroudi G, Tigabu BM. A systematic review of health promotion interventions to increase breast cancer screening uptake: from the last 12 years. Eur J Public Health 2018 Dec 1;28(6):1149-1155. [doi: 10.1093/eurpub/ckx231] [Medline: 29351597]
31. Jones CL, Jensen JD, Scherr CL, Brown NR, Christy K, Weaver J. The Health Belief Model as an explanatory framework in communication research: exploring parallel, serial, and moderated mediation. Health Commun 2015;30(6):566-576. [doi: 10.1080/10410236.2013.873363] [Medline: 25010519]

32. Ashing-Giwa K. Health behavior change models and their socio-cultural relevance for breast cancer screening in African American women. Women Health 1999;28(4):53-71. [doi: 10.1300/J013v28n04_04] [Medline: 10378345]
33. Champion VL. Revised susceptibility, benefits, and barriers scale for mammography screening. Res Nurs Health 1999 Aug;22(4):341-348. [doi: 10.1002/(SICI)1098-240X(199908)22:4%3C341::AID-NUR8%3E3.0.CO;2-P]
34. Sasaki Y, Ueno NT. The emerging impact of social media on cancer patient education in Japan. Oncologist 2018;23(9):e105–6. [doi: 10.1177/1090198174002004O3] [Medline: 29567825]
35. British Broadcasting Corporation. BBC 100 women 2016: who is on the list? 2019. URL: https://www.bbc.com/news/world-38012048 [accessed 2024-06-10]
36. Noar SM, Willoughby JF, Myrick JG, Brown J. Public figure announcements about cancer and opportunities for cancer communication: a review and research agenda. Health Commun 2014;29(5):445-461. [doi: 10.1080/10410236.2013.764781] [Medline: 23845155]
37. Hoffman SJ, Mansoor Y, Natt N, et al. Celebrities' impact on health-related knowledge, attitudes, behaviors, and status outcomes: protocol for a systematic review, meta-analysis, and meta-regression analysis. Syst Rev 2017 Jan 21;6(1):13. [doi: 10.1186/s13643-016-0395-1] [Medline: 28109320]
38. Chapman S, McLeod K, Wakefield M, Holding S. Impact of news of celebrity illness on breast cancer screening: Kylie Minogue's breast cancer diagnosis. Med J Aust 2005 Sep 5;183(5):247-250. [doi: 10.5694/j.1326-5377.2005.tb07029.x] [Medline: 16138798]
39. Borzekowski DLG, Guan Y, Smith KC, Erby LH, Roter DL. The Angelina effect: immediate reach, grasp, and impact of going public. Genet Med 2014 Jul;16(7):516-521. [doi: 10.1038/gim.2013.181] [Medline: 24357847]
40. Evans DG, Barwell J, Eccles DM, et al. The Angelina Jolie effect: how high celebrity profile can have a major impact on provision of cancer related services. Breast Cancer Res 2014 Sep 19;16(5):442. [doi: 10.1186/s13058-014-0442-6] [Medline: 25510853]
41. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
42. Carpenter CJ. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. Health Commun 2010 Dec;25(8):661-669. [doi: 10.1080/10410236.2010.521906] [Medline: 21153982]
43. Japanese Ministry of Internal Affairs and Communications. Information and communications statistics database. 2022. URL: https://www.soumu.go.jp/johotsusintokei/english [accessed 2024-06-10]
44. Statcounter. Statcounter global stats. 2024. URL: https://gs.statcounter.com [accessed 2025-01-16]
45. Wong MYZ, Gunasekeran DV, Nusinovici S, et al. Telehealth demand trends during the COVID-19 pandemic in the top 50 most affected countries: infodemiological evaluation. JMIR Public Health Surveill 2021 Feb 19;7(2):e24445. [doi: 10.2196/24445] [Medline: 33605883]
46. Cram P, Fendrick AM, Inadomi J, Cowen ME, Carpenter D, Vijan S. The impact of a celebrity promotional campaign on the use of colon cancer screening: the Katie Couric effect. Arch Intern Med 2003 Jul 14;163(13):1601-1605. [doi: 10.1001/archinte.163.13.1601] [Medline: 12860585]

Abbreviations

HBM: Health Belief Model
RSV: relative search volume

Edited by A Mavragani; submitted 06.07.24; peer-reviewed by P Stachteas, T Mitsunaga, Y Nishimura; revised version received 30.01.25; accepted 05.02.25; published 04.03.25.

Please cite as:
Takahashi N, Nakao M, Nakayama T, Yamazaki T
Breast Cancer Screening Participation and Internet Search Activity in a Japanese Population: Decade-Long Time-Series Study
JMIR Cancer 2025;11:e64020
URL: https://cancer.jmir.org/2025/1/e64020
doi:10.2196/64020

Assessing the Data Quality Dimensions of Partial and Complete Mastectomy Cohorts in the All of Us Research Program: Cross-Sectional Study

Matthew Spotnitz[1], MD, MPH; John Giannini[1], PhD; Yechiam Ostchega[1], PhD, RN; Stephanie L Goff[2], MD; Lakshmi Priya Anandan[3], MPH; Emily Clark[3], MPH; Tamara R Litwin[1], PhD, MPH; Lew Berman[1], PhD

[1]All of Us Research Program, National Institutes of Health, 6710B Rockledge Drive, Bethesda, MD, United States

[2]Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States

[3]Life Sciences Division, Leidos, Reston, VA, United States

Corresponding Author:
Matthew Spotnitz, MD, MPH
All of Us Research Program, National Institutes of Health, 6710B Rockledge Drive, Bethesda, MD, United States

Abstract

Background: Breast cancer is prevalent among females in the United States. Nonmetastatic disease is treated by partial or complete mastectomy procedures. However, the rates of those procedures vary across practices. Generating real-world evidence on breast cancer surgery could lead to improved and consistent practices. We investigated the quality of data from the *All of Us* Research Program, which is a precision medicine initiative that collected real-world electronic health care data from different sites in the United States both retrospectively and prospectively to participant enrollment.

Objective: The paper aims to determine whether *All of Us* data are fit for use in generating real-world evidence on mastectomy procedures.

Methods: Our mastectomy phenotype consisted of adult female participants who had CPT4 (Current Procedural Terminology 4), *ICD-9* (*International Classification of Diseases, Ninth Revision*) procedure, or SNOMED (Systematized Nomenclature of Medicine) codes for a partial or complete mastectomy procedure that mapped to Observational Medical Outcomes Partnership Common Data Model concepts. We evaluated the phenotype with a data quality dimensions (DQD) framework that consisted of 5 elements: conformance, completeness, concordance, plausibility, and temporality. Also, we applied a previously developed DQD checklist to evaluate concept selection, internal verification, and external validation for each dimension. We compared the DQD of our cohort to a control group of adult women who did not have a mastectomy procedure. Our subgroup analysis compared partial to complete mastectomy procedure phenotypes.

Results: There were 4175 female participants aged 18 years or older in the partial or complete mastectomy cohort, and 168,226 participants in the control cohort. The geospatial distribution of our cohort varied across states. For example, our cohort consisted of 835 (20%) participants from Massachusetts, but multiple other states contributed fewer than 20 participants. We compared the sociodemographic characteristics of the partial (n=2607) and complete (n=1568) mastectomy subgroups. Those groups differed in the distribution of age at procedure (P<.001), education (P=.02), and income (P=.03) levels, as per χ^2 analysis. A total of 367 (9.9%) participants in our cohort had overlapping CPT4 and SNOMED codes for a mastectomy, and 63 (1.5%) had overlapping *ICD-9* procedure and SNOMED codes. The prevalence of breast cancer–related concepts was higher in our cohort compared to the control group (P<.001). In both the partial and complete mastectomy subgroups, the correlations among concepts were consistent with the clinical management of breast cancer. The median time between biopsy and mastectomy was 5.5 (IQR 3.5-11.2) weeks. Although we did not have external benchmark comparisons, we were able to evaluate concept selection and internal verification for all domains.

Conclusions: Our data quality framework was implemented successfully on a mastectomy phenotype. Our systematic approach identified data missingness. Moreover, the framework allowed us to differentiate breast-conserving therapy and complete mastectomy subgroups in the *All of Us* data.

(JMIR Cancer 2025;11:e59298) doi:10.2196/59298

KEYWORDS

data quality; electronic health record; breast cancer; breast-conserving surgery; total mastectomy; modified radical mastectomy; public health informatics; cohort; assessment; women; United States; American; nonmetastatic disease; treatment; breast cancer surgery; real-world evidence; data; mastectomy; female; data quality framework; therapy

Introduction

Breast cancer is one of the most common forms of cancer in females worldwide and has a lifetime prevalence of 13%. The incidence in the United States is estimated to be greater than 297,000 women annually and increases with patient age [1,2]. In addition to patient age, breast cancer risk factors include BMI, early age of menarche, late age of menopause, family history or genetic risk, and environmental exposures [3].

Nonmetastatic breast cancer is treated surgically, and approximately 30% of patients have a complete mastectomy. An alternative to a complete mastectomy is breast-conserving therapy (BCT), which consists of breast-conserving surgery and radiation therapy [4]. In multiple randomized controlled trials, BCT has been shown to have similar long-term disease-free survival to a complete mastectomy [5-8].

A recent systematic review found that patients' choice of surgical treatment was multifaceted. Some factors that were associated with patients choosing a mastectomy over BCT were related to tumor characteristics and pathology. Others were sociodemographic or individual belief factors, such as body image, aversion to radiation, and physician preference [9]. In a prospective study of 180 patients, surgeons' preference was the strongest predictor of surgical treatment [10]. Accordingly, there is a need to compare a complete mastectomy to a partial mastectomy, the surgical component of BCT that encompasses lumpectomy, quadrantectomy, and other BCT-related surgical interventions.

We believe that a robust characterization of partial and complete mastectomy patients with data from the *All of Us* Research Program could generate valuable real-world evidence regarding breast cancer treatment and be used to provide evidence towards best practices for patients with the disease. The *All of Us* Research Program has electronic health records (EHRs) on more than 287,000 patients from 50 health care organizations within the United States. The program does targeted enrollment of groups that are underrepresented in biomedical research. Because the *All of Us* Research Program is one of the most comprehensive and diverse observational health care databases worldwide, those findings would represent real-world data associated with partial or complete mastectomy procedures [11].

Accordingly, to date, we are unaware of a study assessing the fitness for the use of *All of Us* and focusing on mastectomy as a treatment modality. Accordingly, the primary objective of this study is to determine whether the *All of Us* data are fit for an analysis of women who had a mastectomy.

Methods

Observational Medical Outcomes Partnership Common Data Model

The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) is the data standard used by the *All of Us* Research Program. The OMOP CDM consists of standardized concepts and relationships, allowing for harmonizing data from different sources. OMOP CDM concepts use codes from structured medical terminologies as the source (eg, CPT4 [Current Procedural Terminology 4], *ICD-9* [*International Classification of Diseases, Ninth Revision*], and LOINC [Logical Observation Identifiers Names and Codes]). The schema consists of standardized concept relationships across data tables [12-14].

We created partial and complete mastectomy phenotypes by manual selection. First, we selected CPT4, *ICD-9* procedure, and SNOMED (Systematized Nomenclature of Medicine) codes for those procedures manually. We chose CPT4, *ICD-9* procedure, and SNOMED source codes because these are the standards that are used by the OMOP CDM. Then, we searched ATHENA for the corresponding OMOP CDM concepts [15]. The partial and complete mastectomy OMOP CDM concept sets were the basis for our phenotype queries. Additionally, we restricted the phenotype to the earliest occurrence of a procedure and to female participants who were at least 18 years or older at the time of the procedure.

Primary Outcomes and Variables

Overview

We developed a data quality dimensions (DQD) framework and evaluation matrix that was adapted from Kahn et al [16]. The framework comprises 5 mutually exclusive and parsimonious dimensions that can be operationalized and applied to a mastectomy cohort as primary outcome variables: conformance, completeness, concordance, plausibility, and temporality. A prior study applied these dimensions to a ductal carcinoma in situ cohort data quality analysis [17].

DQD Framework

Concomitantly, we evaluated the application of the DQD framework to a mastectomy cohort. Each framework element was evaluated with respect to concept selection, internal verification, and external validation. The overarching principles of assessing the DQD include internal characteristics, described by Kahn et al [16] as verification; comparing external benchmarks as validation; and applying descriptive, inferential, and agreement statistics and data visualization. In practice, a researcher would decide whether the data associated with their constructed cohort meets their expectations for fitness of use based on the rating matrix [18]. For the DQD analysis, we selected OMOP CDM concepts related to risk factors and the medical management of breast cancer. Specifically, we included concepts that included but were not limited to breast cancer diagnoses, breast biopsies, screening and diagnostic breast imaging, endocrine therapy, anti–human epidermal growth factor receptor 2 (anti-HER2) therapy, tyrosine kinase inhibitors, chemotherapy, radiation therapy, laboratory measurements, and genetic risk factors [19,20]. Many of our codes had been validated in a prior ductal carcinoma in situ study [17]. Furthermore, a surgical oncologist (SLG) reviewed those codes, confirmed their appropriateness, and recommended additional codes for our analysis. Representative codes are shown in the supplemental appendix (Tables S1-S8 in Multimedia Appendix 1).

Sociodemographic Characterization and Geospatial Analysis

We characterized the geospatial distribution of the mastectomy cohort participants based on state address at the time of enrollment. Also, we characterized the mastectomy cohort, the partial and complete mastectomy subgroups, and control cohort according to sex at birth, race or ethnicity, age group, education, and income.

Analysis

All of Us participants enrolled between May 6, 2017, and July 1, 2022, provided consent to participate and had the option to authorize sharing of their EHRs. Upon enrollment, participants were required to fill out a basic self-reported survey, which includes information on sociodemographic characteristics, and could also consent to have additional data submitted to the program, including data from biospecimens, genomic sequences, and wearable data. All analyses presented in this paper used the *All of Us* Controlled Tier Dataset v7, released on April 20, 2023. The source data were formatted to be compatible with OMOP CDM (version 5.3.1) [21]. Additionally, the data curation team modified some of the drug table schemas for optimal use with *All of Us* data. Accordingly, we modified our queries to maximize the capture of drug exposure data. Missing data were not included in the analysis and we did not make statistical adjustments for missingness.

All programming and statistical analyses were performed in Python (version 3.7.12) and were implemented in a Jupyter Notebook (version 6.5.4). We used chi-squared statistics to test for independent association, Spearman coefficients to measure bivariate correlations, and data visualization to explore the application of the DQD. The level of significant differences was set at $P<.05$.

Ethical Considerations

The *All of Us* Research Program complies with multiple ethical considerations. First, it has an institutional review board (IRB) that reviews the protocol, informed consent, and other participant-facing materials for the *All of Us* Research Program. The IRB follows the regulations and guidance of the Office for Human Research Protections for all studies [22]. The *All of Us* IRB determined that the data that were used in this analysis were considered non–human subjects' research. Second, participants are provided with information on how the program operates, reasonable expectations, and participants' rights. Participants who agree to enroll sign consent forms [23]. Third, *All of Us* participants' data are removed of identifiers and coded to protect their privacy before they are made available to researchers. Reidentification or recontacting of participants is prohibited, and governance mechanisms ensure protection against reidentification or recontact of participants [24]. Fourth, *All of Us* participants who give blood, saliva, or urine samples receive a one-time compensation of US $25 [25]. Otherwise, no direct compensation is provided. Fifth, this paper is not focused on imaging, and thus we have not included any images in the supplemental material. In addition, we censor counts that are less than or equal to 20 to comply with program requirements for minimizing disclosure risk. Data and code used in this study are available as a featured workspace to registered researchers of the *All of Us* Researcher Workbench [26].

Results

Sample

In the *All of Us* database, 249,565 participants consented to participate in the study, were at least 18 years old, and selected assigned as female at birth in the *All of Us* "Basics" self-reported questionnaire. Of those, 172,401 (69%) signed an authorization to share clinical data and had at least one data record in a participating EHR. We created a cohort of 4175 (2.4%) patients with mastectomy procedures and a control cohort of 168,226 (97.6%) female participants who did not have a mastectomy. Out of the 4175 female participants who had mastectomy procedures, 316 (7.6%) had both partial and complete mastectomy procedures. The first occurrence of the procedure code was used for subgroup assignment.

We plotted the mastectomy (partial or complete) cohort's geospatial distribution to assess whether our cohort was distributed equally across the United States (Figure 1). A total of 835 (20%) participants of our cohort had medical records from Massachusetts, 656 (15.7%) from Arizona, 547 (13.1%) from Wisconsin, 468 (11.2%) from California, 386 (9.3%) from New York, 369 (8.8%) from Illinois, 245 (5.9%) from Florida, and 197 (5.9%) from Michigan. Many states had mastectomy cases for fewer than 20 participants in our cohort and were not reported due to disclosure risk guidelines.

Figure 1. Geospatial analysis of a mastectomy cohort (partial or complete). States with a white-gray fill color contributed no participants to the cohort. Data source: The *All of Us* research program.

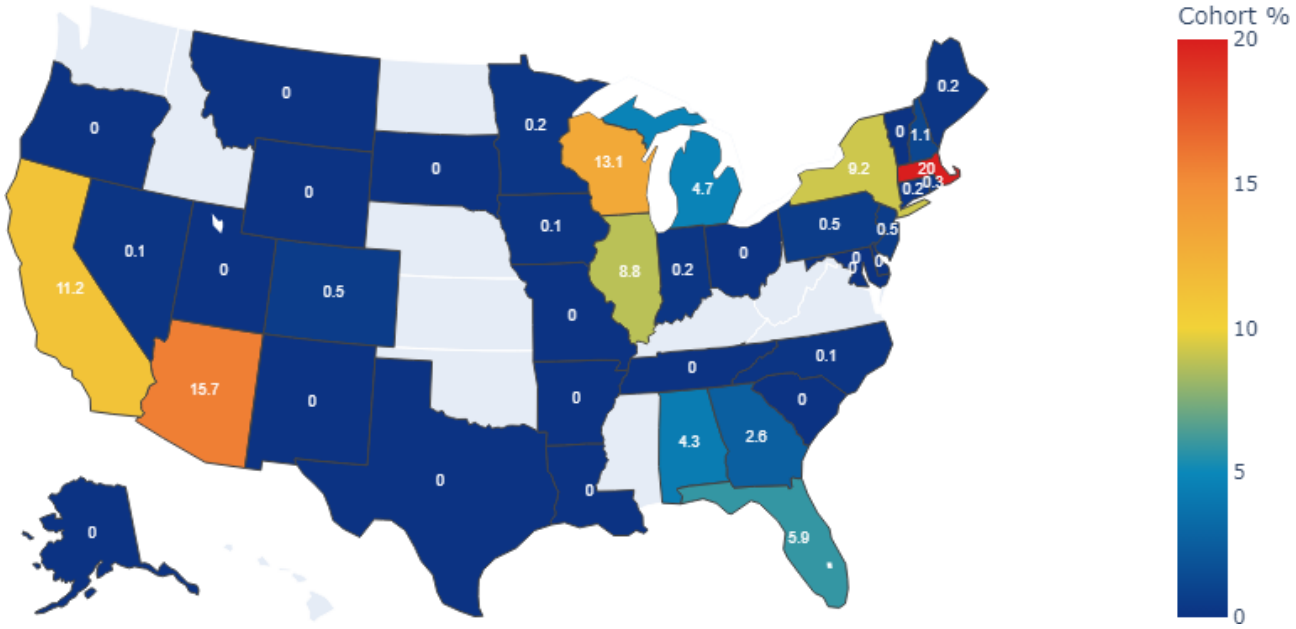


Table 1 provides a breakdown of the cohort by sex at birth, race or ethnicity, age group, education, and income by partial (n=2607) and complete (n=1568) mastectomy. Among the participants who underwent partial or complete mastectomy, the majority were white (66.9% and 70.4%, respectively), with procedures peaking between 40 and 79 years (89.5% and 82.6%, respectively), differ in achieving college or higher degree (52.2% and 58%, respectively), and household income greater or equal to 100k (27% and 32.3%, respectively). A similar sociodemographic comparison was performed for the mastectomy cohort and the control group (Table S8 in Multimedia Appendix 1).

Table . Sociodemographic characteristics of *All of Us* partial and complete mastectomy cohorts.

Demographic category	Partial mastectomy, n (%)	Complete mastectomy, n (%)	*P* value
Assigned sex at birth			—[a]
Female	2607 (100)	1568 (100)	
Race or ethnicity[b]			.14
Asian	80 (3.1)	55 (3.5)	
Black	380 (14.6)	180 (11.5)	
Hispanic	379 (14.5)	226 (14.4)	
Middle East and North Africa, Native Hawaiians, and Pacific Islanders	30 (1.2)	n≤20	
White	1744 (66.9)	1104 (70.4)	
Prefer not to answer, or skip	47 (1.8)	28 (1.8)	
None of these	27 (1.0)	n≤20	
Age at procedure (years)			<.001
18 - 39	187 (7.2)	258 (16.5)	
40 - 59	1140 (43.7)	851 (54.3)	
60 - 79	1193 (45.8)	444 (28.3)	
≥80 or <18	87 (3.3)	n≤20	
Education			.02
Never attended or grades 1 through 4 (primary)	21 (0.8)	n≤20	
Grades 5 through 8 (middle school)	47 (1.8)	23 (1.5)	
Grades 9 through 11 (some high school)	89 (3.4)	52 (3.3)	
Grade 12 or GED[c] (high school graduate)	346 (13.3)	187 (11.9)	
College 1 to 3 (some college, associate's degree, or technical school)	705 (27.0)	365 (23.3)	
College graduate	713 (27.4)	454 (29.0)	
Advanced degree (Master's, Doctorate, etc)	647 (24.8)	456 (29.0)	
Prefer not to answer, or skip	39 (1.5)	n≤20	
Annual household income (US $)			.03
Less than 10k	199 (7.6)	89 (5.7)	
10k-25k	243 (9.3)	149 (9.5)	
25k-35k	173 (6.6)	91 (5.8)	
35k-50k	202 (7.8)	110 (7.0)	
50k-75k	289 (11.1)	168 (10.7)	
75k-100k	270 (10.4)	162 (10.3)	
100k-150k	311 (11.9)	199 (12.7)	
150k-200k	149 (5.7)	120 (7.7)	
More than 200k	246 (9.4)	186 (11.9)	
Prefer not to answer	390 (15.0)	223 (14.2)	
Skip	135 (5.2)	71 (4.5)	

[a]Not applicable.

[b]More than one race or ethnicity category could have been selected.

[c]GED: General Educational Development.

Conformance

Data elements can be assessed according to standards. The *All of Us* Program uses SNOMED as a standard vocabulary. We created a butterfly plot to determine the overlap between CPT4, *ICD-9* procedure, and SNOMED procedure codes, as shown in Figure 2. Of the 4175 female participants in our cohort, 3376 (80.9%) had CPT4 codes only, 313 (7.5%) had both CPT4 and SNOMED codes, 176 (3.2%) had CPT4 and *ICD-9* procedure codes, and 63 (1.5%) had *ICD-9* procedure and SNOMED codes. A total of 54 (1.3%) female participants had overlapping CPT4, SNOMED, and *ICD-9* procedure codes (Figure 2). Thus, the overlap among standards was low.

To characterize the source data variance in the standards, we calculated the counts of the partial or complete mastectomy CPT4 codes in our cohort (Table 2). Of the 50 EHR-contributing *All of Us* sites, 24 reported mastectomy CPT4 codes. CPT4 code 19301 ("mastectomy, partial") was reported the most frequently by every site that contributed data to our cohort. The sets of distinct CPT4 codes that each site reported varied substantially, with the median site using 6 different CPT4 codes. We used data from within our cohort to verify conformance. However, we did not validate this dimension against an external benchmark because one was not available.

Figure 2. The butterfly plot of CPT4 (left), SNOMED (top right), and *ICD-9* (bottom right) mastectomy procedure codes. CPT4: Current Procedural Terminology 4; *ICD-9*: *International Classification of Diseases, Ninth Revision*; SNOMED: Systematized Nomenclature of Medicine. Data source: The *All of Us* research program.

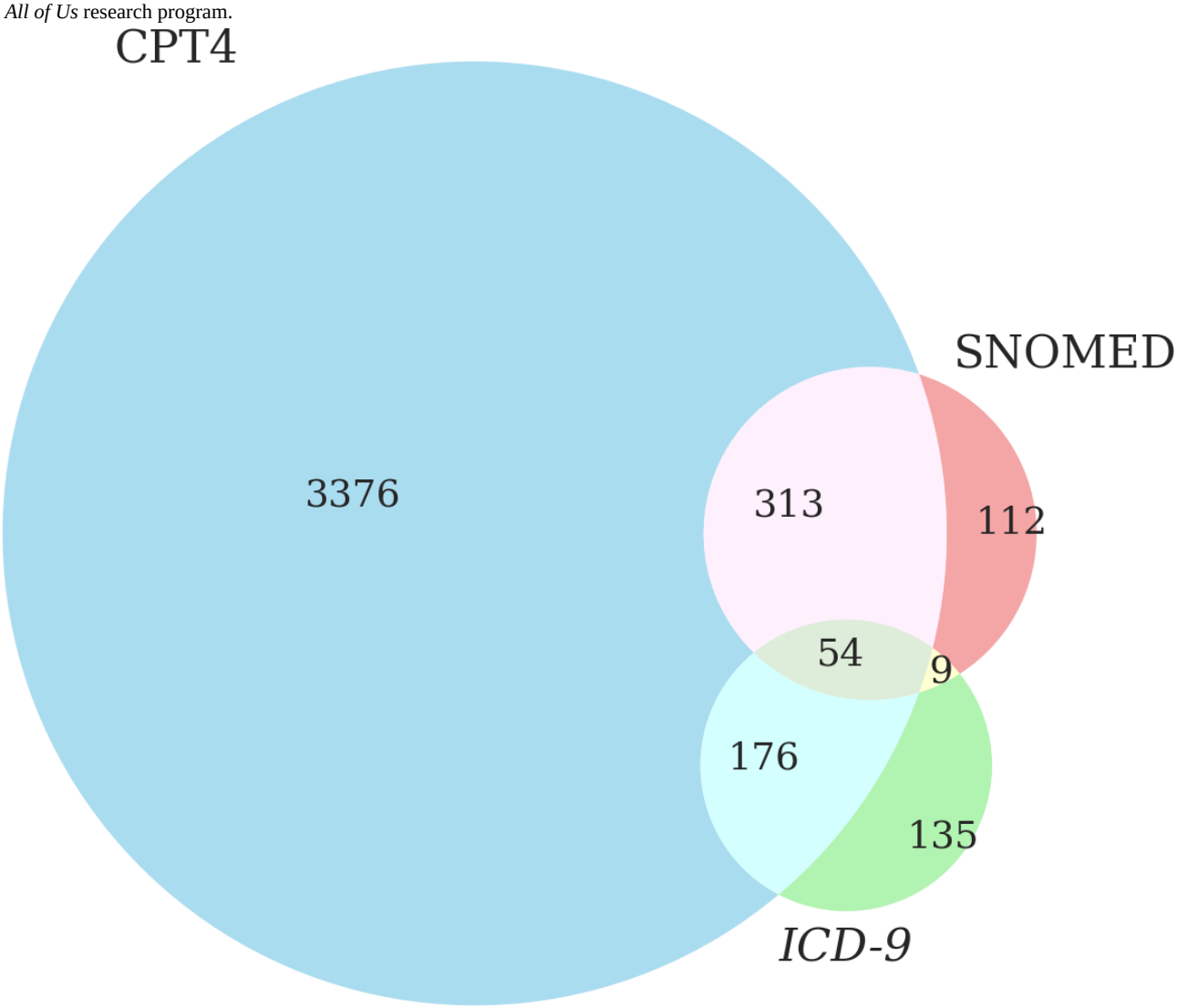


Table . Mastectomy Current Procedural Terminology 4 (CPT4) counts by code. Codes with ≤10 counts were omitted. Data source: The *All of Us* research program.

CPT4 code	Count
19301[a]	2366
19303[b]	1304
19307[c]	358
19302[d]	160
19160[e]	126
19180[f]	53
19162[g]	48
19304[h]	44
19240[i]	35

[a]CPT4 code 19301=mastectomy, partial (eg, lumpectomy, tylectomy, quadrantectomy, and segmentectomy).

[b]CPT4 code 19303=mastectomy, simple, complete.

[c]CPT4 code 19307=mastectomy, modified radical, including axillary lymph nodes, with or without pectoralis minor muscle, but excluding pectoralis major muscle CPT4.

[d]CPT4 code 19302=mastectomy, partial (eg, lumpectomy, tylectomy, quadrantectomy, segmentectomy); with axillary lymphadenectomy.

[e]CPT4 code 19160=mastectomy, partial.

[f]CPT4 code 19180=mastectomy, simple, complete.

[g]CPT4 code 19162=mastectomy, partial, with axillary lymphadenectomy.

[h]CPT4 code 19304=mastectomy, subcutaneous.

[i]CPT4 code 19240=mastectomy, modified radical, including axillary lymph nodes, with or without pectoralis minor muscle, but excluding pectoralis major muscle.

Completeness

We used concept prevalence within our cohort to evaluate data completeness. Table 3 shows the counts and percentages of female participants who did and did not have partial or complete mastectomy procedures, for which each specific clinical measure and intervention was present in the *All of Us* EHR at least once. The participants in our partial or complete mastectomy cohort had a higher prevalence of breast cancer associated OMOP CDM concepts compared to female control cohort participants who did not have a partial or complete mastectomy code. Specifically, comparing females who had a partial or complete mastectomy to females who had neither showed increased prevalence of diagnostic mammography (70.7% vs 13.5%), biopsy (61.2% vs 3.3%), or endocrine therapy (51% vs 1.8%), or chemotherapy (25.1% vs 4.3%). A χ^2 test indicated that partial or complete mastectomy procedures were associated with clinical measures and interventions (P<.001).

Table . Clinical measures and interventions for female participants who had a mastectomy and who did not have a mastectomy. Data source: The *All of Us* research program.

Clinical measure	Mastectomy cohort, n (%)	Nonmastectomy cohort, n (%)	*P* value
Procedures			<.001
Breast biopsy	2554 (61.2)	5625 (3.3)	
Diagnostic mammography	2951 (70.7)	22,731 (13.5)	
Radiation therapy	1656 (39.7)	1644 (1.0)	
Screening mammography	2143 (51.3)	45,071 (26.8)	
Surgery	4175 (100.0)	0 (0)	
Medications			<.001
Anti-HER2[a]	221 (5.3)	151 (0.1)	
CDK[b] 4/6 inhibitors	60 (1.4)	138 (0.1)	
Chemotherapy	1046 (25.1)	7152 (4.3)	
Endocrine therapy	2130 (51.0)	3109 (1.8)	
Goserelin	106 (2.5)	91 (0.1)	
Olaparib	≤20	41 (<0.1)	
Pembrolizumab	≤20	162 (0.1)	
Tyrosine kinase inhibitor	50 (1.2)	277 (0.2)	
Conditions			<.001
Breast cancer gene mutation	435 (10.4)	903 (0.5)	
Estrogen receptor status	235 (5.7)	180 (0.1)	

[a]anti-HER2: anti–human epidermal growth factor receptor 2.

[b]CDK: cyclin-dependent kinase.

Table 4 shows the counts and percentages for the partial and complete mastectomy subgroups. Each specific clinical measure and intervention was in the *All of Us* EHR at least once. The partial mastectomy subgroup, compared to the complete mastectomy subgroup, had a greater proportion of radiation therapy (49.4% vs 23.5%), endocrine therapy (54.7% vs 44.9%), screening mammography (58.8% vs 39%), and diagnostic mammography (77.5% vs 59.3%). By contrast, the complete mastectomy group when compared to the partial mastectomy subgroup had a greater proportion of breast cancer gene (BRCA) mutations (18%vs 5.8%). A χ^2 test indicated that partial and complete mastectomy subgroup categories were associated with clinical measures and interventions (*P*<.001).

Table . Clinical measures and interventions for female participants who had a partial mastectomy and who had a complete mastectomy. Data source: The *All of Us* research program.

Clinical measure	Partial mastectomy, n (%)	Complete mastectomy, n (%)	*P* value
Procedures			<.001
Breast biopsy	1728 (66.3)	826 (52.7)	
Diagnostic mammography	2021 (77.5)	930 (59.3)	
Radiation therapy	1288 (49.4)	368 (23.5)	
Screening mammography	1532 (58.8)	611 (39.0)	
Surgery	2607 (100.0)	1568 (100.0)	
Medications			<.001
Anti-HER2[a]	111 (4.3)	110 (7.0)	
CDK[b] 4/6 inhibitors	31 (1.2)	29 (1.8)	
Chemotherapy	574 (22.0)	472 (30.1)	
Endocrine therapy	1426 (54.7)	704 (44.9)	
Goserelin	51 (2.0)	55 (3.5)	
Olaparib	≤20	≤20	
Pembrolizumab	≤20	≤20	
Tyrosine kinase inhibitor	27 (1.0)	23 (1.5)	
Conditions			<.001
Breast cancer gene mutation	152 (5.8)	283 (18.0)	
Estrogen receptor status	162 (6.2)	73 (4.7)	

[a]anti-HER2: anti–human epidermal growth factor receptor 2.

[b]CDK: cyclin-dependent kinase.

To further characterize completeness, we used UpSet plots (Figures 3 and 4) to assess which combinations of clinical measurements and interventions were prevalent among participants in the partial and complete mastectomy subgroups. The plots show the counts of the concept sets on the left-hand side, and the counts of concept set combinations at the top. The makeup of the combinations is indicated by the dotted lines below. The most frequent combinations in the partial mastectomy subgroup are presented in Textbox 1.

Textbox 1. The most frequent combinations in the partial mastectomy subgroup.

- Combination 1: Surgery, diagnostic mammography, biopsy, screening mammography, endocrine therapy, and radiation therapy (298 cases)
- Combination 2: Surgery, diagnostic mammography, biopsy, screening mammography, and endocrine therapy (188 cases)
- Combination 3: Surgery, diagnostic mammography, biopsy, and screening mammography (174 cases)

Figure 3. Bar chart (top) and UpSet plot (bottom) of breast cancer–related diagnosis codes, procedures, medications, and genetic tests in female participants who had a partial mastectomy. Data source: The *All of Us* research program. anti-HER2: anti–human epidermal growth factor receptor 2; BRCA: breast cancer gene; CDK: cyclin-dependent kinase.

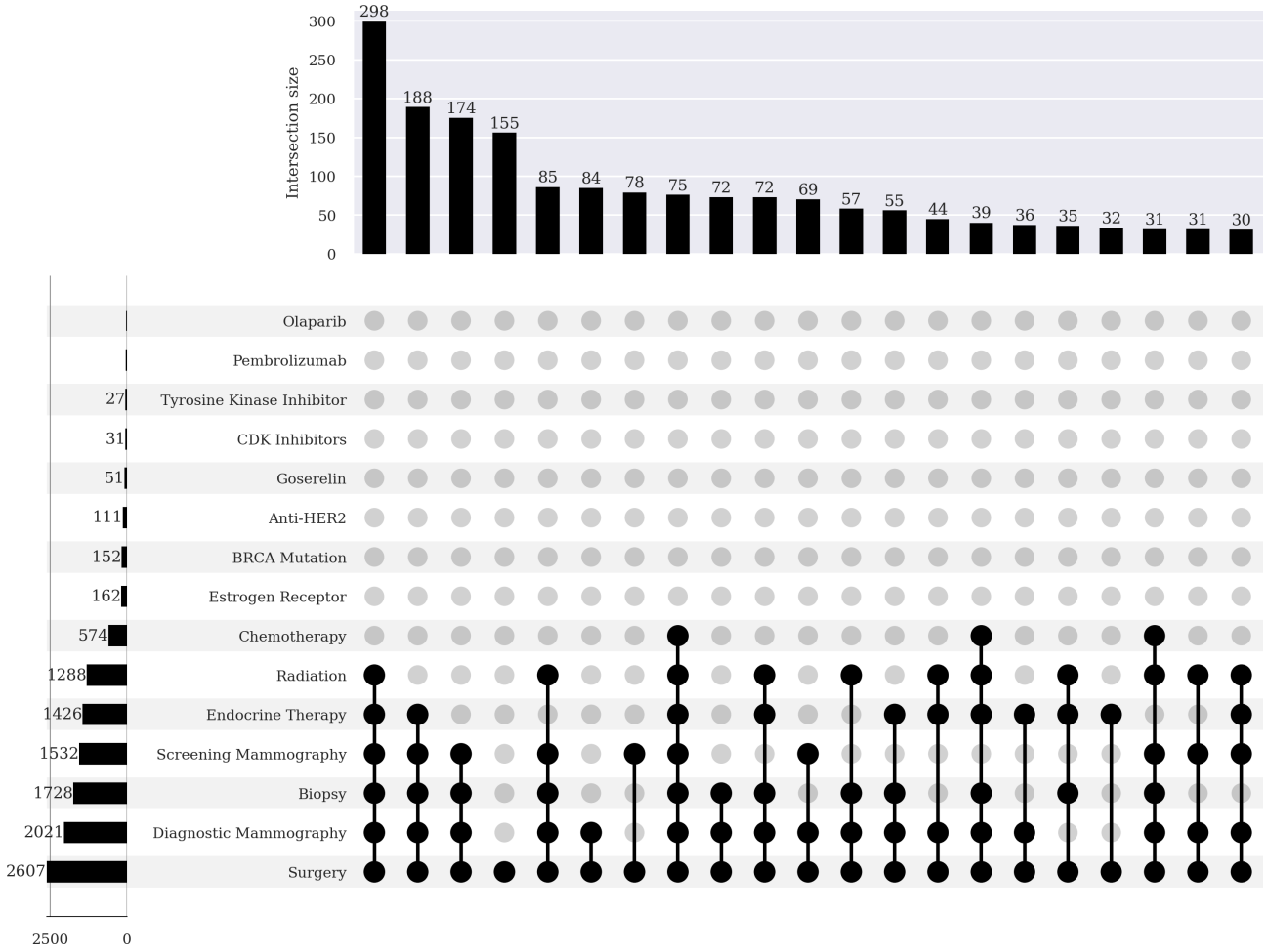
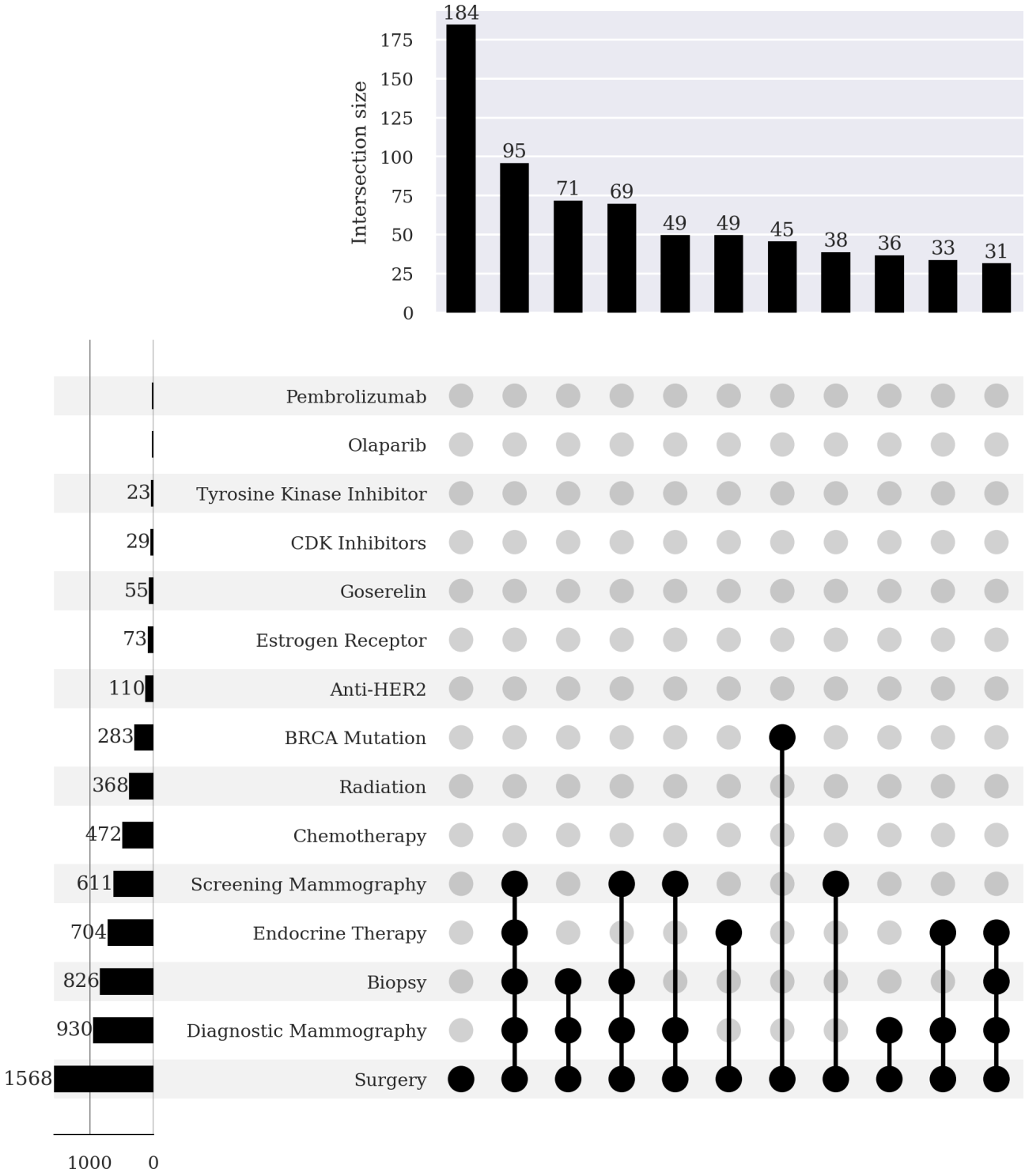


Figure 4. Bar chart (top) and UpSet plot (bottom) of breast cancer–related diagnosis codes, procedures, medications, and genetic tests in female participants who had a complete mastectomy. Data source: The *All of Us* research program. anti-HER2: anti–human epidermal growth factor receptor 2; BRCA: breast cancer gene; CDK: cyclin-dependent kinase.



We did not validate completeness because external benchmarks were not available.

The most frequent combinations in the complete mastectomy subgroup are presented in Textbox 2.

Textbox 2. The most frequent combinations in the complete mastectomy subgroup.

- Combination 1: Surgery (184 cases)
- Combination 2: Surgery, diagnostic mammography, biopsy, screening mammography, and endocrine therapy (95 cases)
- Combination 3: Surgery, diagnostic mammography, and biopsy (71 cases)

Concordance

We calculated the bivariate correlations between OMOP CDM concepts for clinical measures and interventions in the partial and complete mastectomy subgroups to measure concordance (Figures 5 and 6). The highest bivariate correlations for the partial mastectomy subgroup were between biopsy and diagnostic mammography (r=0.36) and chemotherapy and anti-HER2 therapy (r=0.36). We also calculated the bivariate correlations for the complete mastectomy subgroup; the highest bivariate correlations were between biopsy and diagnostic mammography (r=0.43), radiation therapy and chemotherapy (r=0.38), screening mammography and diagnostic mammography (r=0.37), and chemotherapy and anti-HER2 therapy (r=0.34).

Figure 5. Correlogram of medications, procedures, and genetic tests in the subgroup of partial mastectomy patients. Data source: The *All of Us* research program. anti-HER2: anti–human epidermal growth factor receptor 2; BRCA: breast cancer gene; CDK: cyclin-dependent kinase.

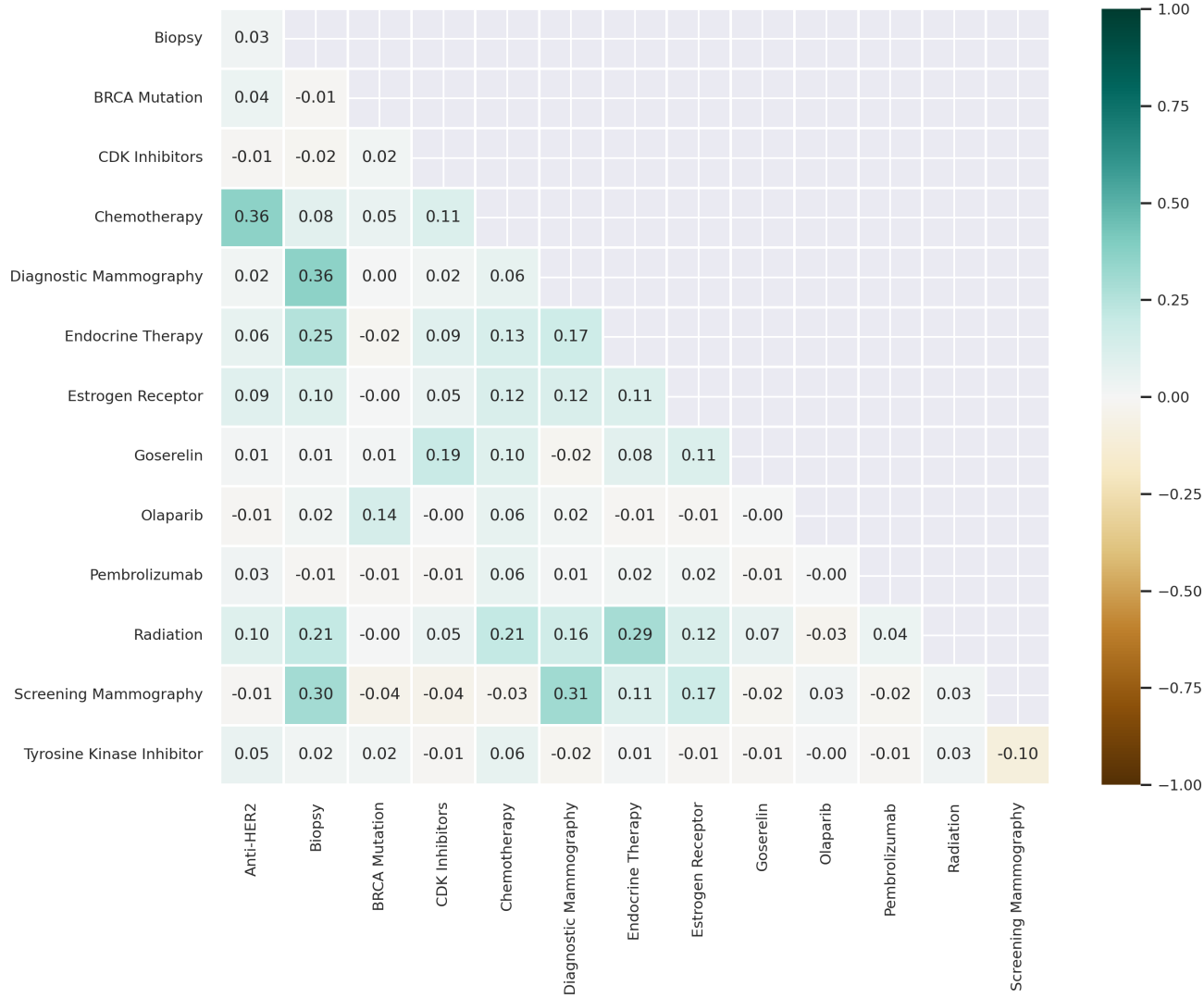
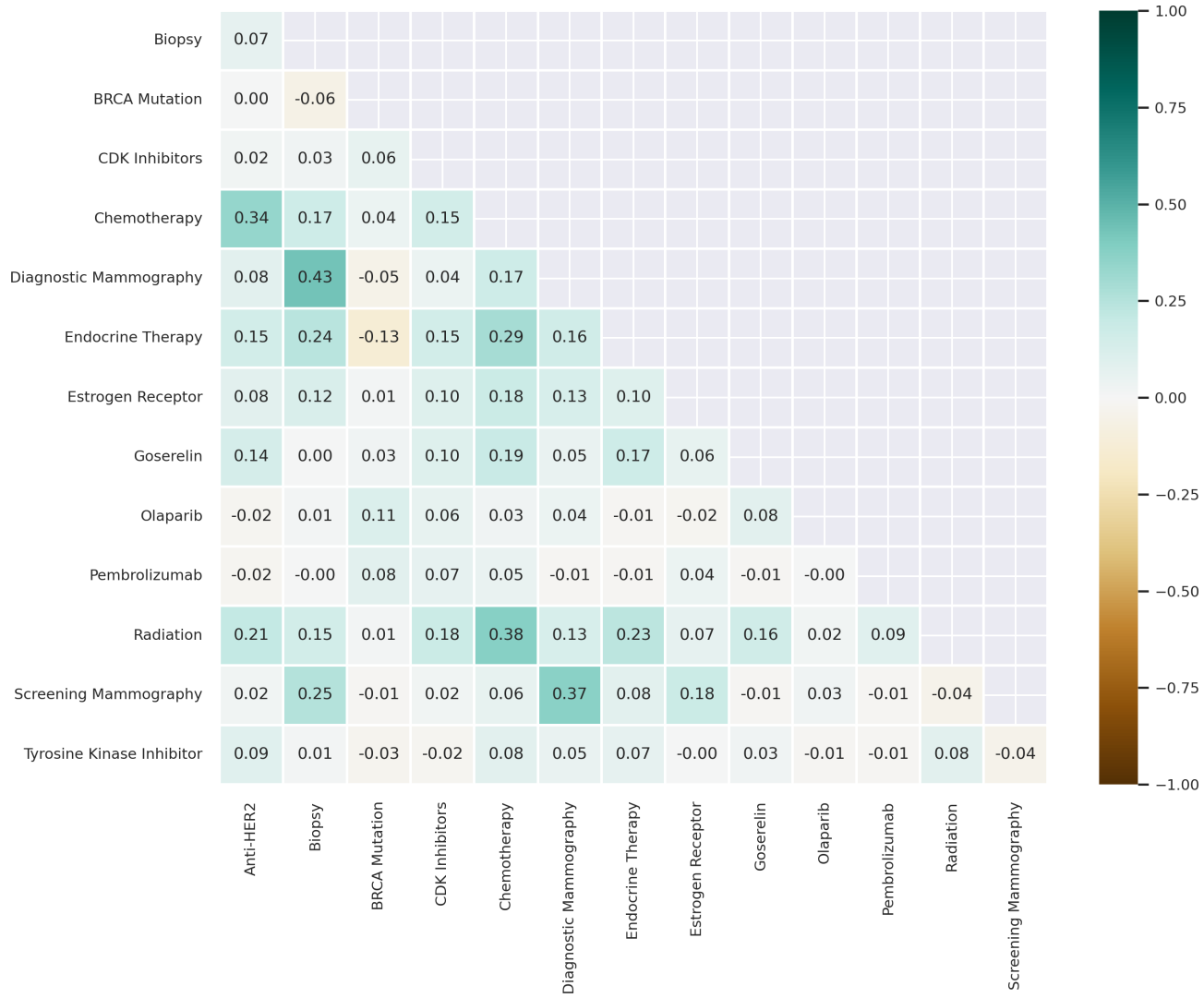


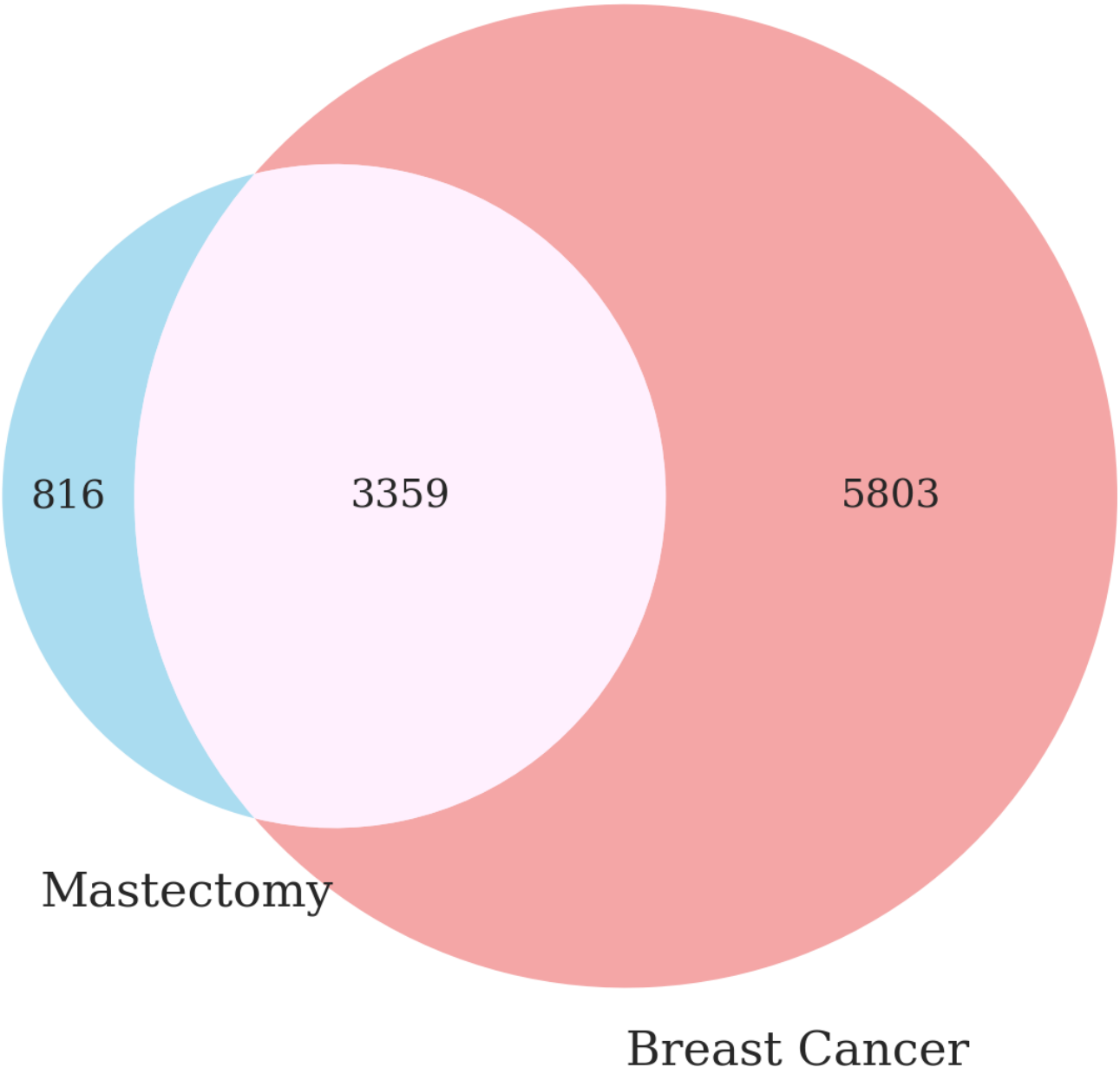
Figure 6. Correlogram of medications, procedures, and genetic tests in the subgroup of complete mastectomy patients. Data source: The *All of Us* research program. anti-HER2: anti–human epidermal growth factor receptor 2; BRCA: breast cancer gene; CDK: cyclin-dependent kinase.



The overlap of patients who had a mastectomy procedure and breast cancer diagnosis (eg, SNOMED 254837009 “Malignant Neoplasm of Breast”) is shown in Figure 7. Of the 816 (19.5%) of female participants who had a mastectomy code only, 277 (33.9%) had diagnosis codes for physical or radiographic findings (eg, breast lump, mammographic calcification of breast), premalignant disease, or benign disease within 1 year before the procedure.

We did not validate concordance because external benchmarks were not available.

Figure 7. The butterfly plot of the mastectomy procedure (left) and the breast cancer diagnosis (right) codes. Data source: The *All of Us* research program.



Plausibility

We assessed plausibility by characterizing distributions of clinical measurement and intervention concepts by age group. We stratified the analysis by partial and complete mastectomy procedures (Figures 8 and 9). We used the age at which a participant's surgical procedure was recorded in EHR rather than other internal characteristics. Our data support a clear association between age patterns and the rate of mastectomy surgery (see Table 1) and the literature [3].

Figure 8. Bar chart of clinical measures and interventions for female participants who had a partial mastectomy. Data source: The *All of Us* research program. anti-HER2: anti–human epidermal growth factor receptor 2; BRCA: breast cancer gene; CDK: cyclin-dependent kinase.

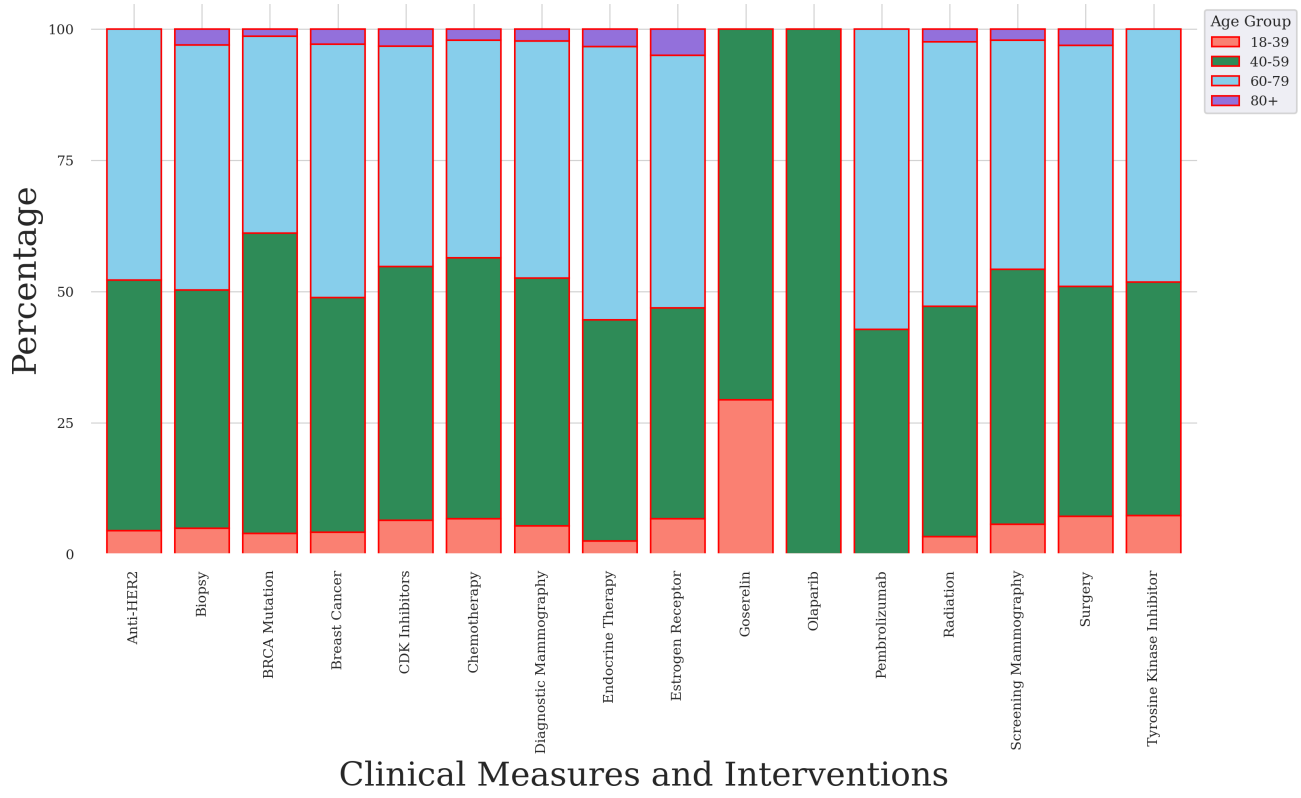
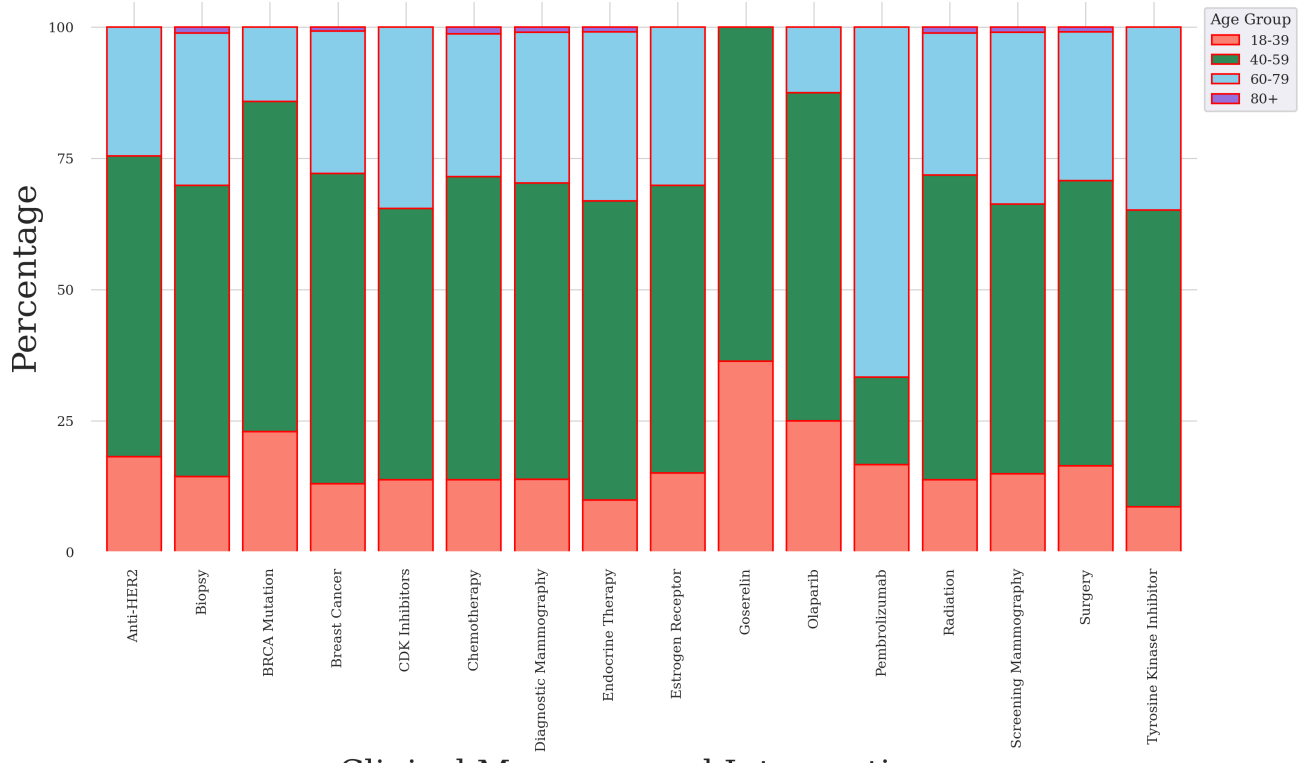


Figure 9. Bar chart of clinical measures and interventions for female participants who had a complete mastectomy. Data source: The *All of Us* research program. anti-HER2: anti–human epidermal growth factor receptor 2; BRCA: breast cancer gene; CDK: cyclin-dependent kinase.



For the partial mastectomy subgroup, clinical measures and interventions were most frequent in for adult female participants who were between 40 and 79 years of age (Figure 8). Specifically, BRCA mutation (57.2%) was most frequent in the 40 - to 59-year-old group. Biopsy (46.7%), radiation therapy (50.4%), surgery (45.9%), and endocrine therapy (52%) were most frequent in the 60 - to 79-year-old group. A χ^2 test indicated that age categories were associated with clinical measures and interventions (P<.001).

For the complete mastectomy subgroup, the frequencies of clinical measures and interventions were highest for adult female participants who were between 40 and 79 years of age (Figure 9). Specifically, BRCA mutation (62.9%), diagnostic mammography (56.5%), biopsy (55.4%), surgery (54.3%), endocrine therapy (56.9%), radiation therapy (58%), screening mammography (51.3%), and estrogen receptor status (54.8%) concepts were most frequent in the 40 - to 59-year-old age group. A χ^2 test indicated that age categories were associated with clinical measures and interventions (P<.001).

We did not validate plausibility because external benchmarks were not available.

Temporality

To assess temporality, we examined the time intervals between biopsy and mastectomy. A biopsy procedure was available for 2354 (56.4%) female participants in the partial or complete mastectomy cohort. There was a skewed time distribution from biopsy to surgery (right positive skew=9.9). Therefore, the median (5.5, IQR 3.5-11.2 weeks) better represents the distribution than the mean (18.4 weeks) for the time difference between biopsy and surgery.

We did not validate temporality because external benchmarks were not available.

Discussion

Principal Findings

The primary objective of this study is to determine whether the *All of Us* EHR data are fit for analyzing female participants who had a mastectomy. Indeed, this study provides valuable information to researchers on the quality of EHR data by operationalizing 5 DQD to the procedure-driven selection of the mastectomy cohort clinical measurements and interventions. We implemented concept selection and internal verification on all domains but were unable to validate them because external benchmarks were not available. Each domain provided unique information about data quality. In this study, our conformance analysis evaluated the overlap of procedure codes from different source vocabularies. The low overlap with SNOMED implies that there may be suboptimal linkage of procedure concepts with concepts from other domains because the standardized relationships may be underused. Furthermore, our method for evaluating conformance may be applicable to quantifying the amount of overlap between nonstandardized and standardized codes. The completeness DQD analysis can be used to identify disease-specific missingness in our data. The concordance analysis measures associations among concepts, which have implications for their relative missingness. The plausibility and temporality analyses are an effort to make the data quality issues transparent and comparable to existing clinical knowledge.

Despite the incompleteness of EHRs, breast cancer–related concepts were prevalent in our cohort. The correlations among those concepts were logical and consistent with the practice of treating breast cancer. For example, concepts for radiation therapy, which is an essential part of BCT, were more prevalent in the partial mastectomy subgroup. The completeness and correlations of our data allowed us to differentiate patients who had BCT from patients who had a complete mastectomy. Our cohort consisted of *All of Us* participants who had a mastectomy procedure at one of the participating sites. However, a greater number of participants may have had a mastectomy procedure at a site that was not part of our research network. Alternatively, diagnosis code-based phenotypes may have higher sensitivity and more false positives than procedure-based phenotypes.

This DQD paper is the first OMOP CDM study to evaluate the quality of partial or complete mastectomy procedure data with procedure-based phenotypes using *All of Us* EHR data. There are several distinct advantages to using a procedure-based phenotype over a diagnosis code-based phenotype. First, in the United States, procedure codes tend to be submitted by experts and can be subject to more rigorous quality checks than codes from other domains, which makes them more likely to be accurate. Second, a mastectomy is a disease-specific intervention for breast cancer. Therefore, a mastectomy phenotype should have a strong association with breast cancer. Third, procedure codes are well-defined and map to granular OMOP CDM concepts. Furthermore, the granularity of codes allows for differentiating partial from complete mastectomy procedures. Fourth, procedures are concrete events synchronizing a cohort to a point in the disease course. Synchronizing the cohort can be especially valuable for performing a treatment pathway analysis, a population-level estimation, or a patient-level prediction.

Comparisons to Prior Work

The relative proportions of the mastectomy cohort who had partial and complete mastectomy procedures were similar to the national averages [27]. However, we found that the frequencies of multiple concepts were lower than expected in our analysis. For example, 51% of our mastectomy cohort had endocrine therapy concepts, and only 5.6% had estrogen receptor status concepts.

Limitations

Our study had several limitations. First, the OMOP CDM breast cancer concepts had minimal information on the breast cancer stage, grade, pathology, laterality, and quadrant of a tumor. Consequently, adopting guidelines from other research networks, such as the National Comprehensive Cancer Network, was not feasible for our use because National Comprehensive Cancer Network guidelines are associated with specific tumor, node, and metastasis characteristics. Health Care Common Procedure Coding System and *International Classification of Diseases* procedure codes can help provide some information on mastectomy status; however, they are limited by their granularity and frequency in the dataset. Second, we wrote custom code to implement our phenotype and selected our concepts manually. Also, evaluating phenotypes with software packages such as CohortDiagnostics and Phevaluator is a possible future area of research. Third, our geospatial analysis was based on the participant's location at the time of enrollment. Some participants could have had surgical treatment in another state. Because our data does not identify the site, variation in practice patterns by institution or provider was unknown. These issues are potential sources of selection bias. Notwithstanding, we

recognize that institution and provider preferences can influence whether a patient undergoes a partial or complete mastectomy for breast cancer [9]. Future development with the *All of Us* Center for Linkage and Acquisition of Data may enable the effects of those preferences on patient procedure choice to be analyzed through the acquisition of health care claims data. Fourth, we restricted our analysis to female participants to reduce errors attributed to misclassification of participants' assigned sex at birth. A study that also includes males with breast cancer, who make up 1% of the breast cancer population, would be more generalizable [28]. Fifth, there was minimal data available for an external validation comparison.

Future Directions

Our study has shown that our data quality framework is systematic and comprehensive and can be implemented in a mastectomy use case. The results of our analysis could inform investigators about the feasibility of using *All of Us* data for follow-up studies. Furthermore, we encourage continued procedure-based phenotyping with our data. In summary, our methods can continue to assess data quality in the *All of Us* Research Program and they may lead to precision medicine studies applicable to diverse patient populations.

Conclusions

We successfully implemented a data quality framework to evaluate whether a mastectomy phenotype that uses *All of Us* data is fit for observational health care research. Our procedure-based phenotype overcame many EHR limitations. In a subgroup analysis, we achieved reasonable differentiation of BCT from complete mastectomy patients. We encourage the continued use of procedure-based phenotypes to evaluate data quality.

Acknowledgments

We gratefully acknowledge *All of Us* participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's (NIH) *All of Us* Research Program for making available the participant data examined in this study. This study used data from the *All of Us* Research Program's Controlled Tier Dataset V7, available to authorized users on the Researcher Workbench. The implementation of the program is supported by awards through the NIH Office of the Director. We did not use large language models to produce this manuscript. We did not use generative AI to produce any part of this work.

Data Availability

Data and code used in this study are available as a featured workspace to registered researchers of the *All of Us* Researcher Workbench [26].

Authors' Contributions

MS contributed to conceptualization, formal analysis, methodology, writing – original draft, and writing – review and editing; YO contributed to conceptualization, formal analysis, methodology, writing – original draft, writing – review and editing, and supervision; JG contributed to methodology, software, and visualization; SLG contributed to writing – review and editing, and validation; LPA contributed to methodology, software, visualization, writing – original draft, and writing – review and editing; EC contributed to project administration, formal analysis, writing - review and editing, and validation; TRL contributed to writing – review and editing; LB contributed to supervision, resources, methodology, conceptualization, formal analysis, writing – original draft, writing – review and editing, and validation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables containing representative codes, representative medications, sociodemographic characteristics.
[DOCX File, 51 KB - cancer_v11i1e59298_app1.docx]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424. [doi: 10.3322/caac.21492] [Medline: 30207593]
2. National Cancer Institute Surveillance, Epidemiology, and End Results Program cancer stat facts: female breast cancer. National Institutes of Health. URL: https://seer.cancer.gov/statfacts/html/breast.html [accessed 2024-12-30]
3. Singletary SE. Rating the risk factors for breast cancer. Ann Surg 2003 Apr;237(4):474-482. [doi: 10.1097/01.SLA.0000059969.64262.87] [Medline: 12677142]
4. McLaughlin SA. Surgical management of the breast: breast conservation therapy and mastectomy. Surg Clin North Am 2013 Apr;93(2):411-428. [doi: 10.1016/j.suc.2012.12.006] [Medline: 23464693]

5. Fisher B, Anderson S, Bryant J, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med 2002 Oct 17;347(16):1233-1241. [doi: 10.1056/NEJMoa022152] [Medline: 12393820]
6. Veronesi U, Saccozzi R, Del Vecchio M, et al. Comparing radical mastectomy with quadrantectomy, axillary dissection, and radiotherapy in patients with small cancers of the breast. N Engl J Med 1981 Jul 2;305(1):6-11. [doi: 10.1056/NEJM198107023050102] [Medline: 7015141]
7. Arriagada R, Lê MG, Rochard F, Contesso G. Conservative treatment versus mastectomy in early breast cancer: patterns of failure with 15 years of follow-up data. Institut Gustave-Roussy Breast Cancer Group. J Clin Oncol 1996 May;14(5):1558-1564. [doi: 10.1200/JCO.1996.14.5.1558] [Medline: 8622072]
8. Litière S, Werutsky G, Fentiman IS, et al. Breast conserving therapy versus mastectomy for stage I-II breast cancer: 20 year follow-up of the EORTC 10801 phase 3 randomised trial. Lancet Oncol 2012 Apr;13(4):412-419. [doi: 10.1016/S1470-2045(12)70042-6] [Medline: 22373563]
9. Gu J, Groot G, Boden C, Busch A, Holtslander L, Lim H. Review of factors influencing women's choice of mastectomy versus breast conserving therapy in early stage breast cancer: a systematic review. Clin Breast Cancer 2018 Aug;18(4):e539-e554. [doi: 10.1016/j.clbc.2017.12.013]
10. Molenaar S, Oort F, Sprangers M, et al. Predictors of patients' choices for breast-conserving therapy or mastectomy: a prospective study. Br J Cancer 2004 Jun 1;90(11):2123-2130. [doi: 10.1038/sj.bjc.6601835] [Medline: 15150557]
11. All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The "All of Us" Research Program. N Engl J Med 2019 Aug 15;381(7):668-676. [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
12. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578. [Medline: 26262116]
13. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc 2012;19(1):54-60. [doi: 10.1136/amiajnl-2011-000376] [Medline: 22037893]
14. Reich C, Ostropolets A, Ryan P, et al. OHDSI Standardized Vocabularies – a large-scale centralized reference ontology for international data harmonization. J Am Med Inform Assoc 2024 Feb 16;31(3):583-590. [doi: 10.1093/jamia/ocad247] [Medline: 38175665]
15. Observational Health Data Sciences and Informatics (OHDSI) Network, ATHENA Search Engine. URL: https://athena.ohdsi.org/search-terms/start [accessed 2024-12-30]
16. Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. EGEMS (Wash DC) 2016;4(1):1244. [doi: 10.13063/2327-9214.1244] [Medline: 27713905]
17. Berman L, Ostchega Y, Giannini J, et al. Application of a data quality framework to ductal carcinoma in situ using electronic health record data from the All of Us Research Program. JCO Clin Cancer Inform 2024 Aug;8:e2400052. [doi: 10.1200/CCI.24.00052] [Medline: 39178364]
18. Karr AF, Sanil AP, Banks DL. Data quality: a statistical perspective. Stat Methodol 2006 Apr;3(2):137-173. [doi: 10.1016/j.stamet.2005.08.005]
19. Waks AG, Winer EP. Breast cancer treatment: a review. J Am Med Assoc 2019 Jan 22;321(3):288-300. [doi: 10.1001/jama.2018.19323] [Medline: 30667505]
20. Association of Breast Surgery at Baso 2009. Surgical guidelines for the management of breast cancer. Eur J Surg Oncol 2009;35 Suppl 1:1-22. [doi: 10.1016/j.ejso.2009.01.008] [Medline: 19299100]
21. The basics survey. All of Us Research Program. URL: https://databrowser.researchallofus.org/survey/the-basics [accessed 2024-12-30]
22. All of Us Institutional Review Board (IRB). All of Us Research Program. URL: https://allofus.nih.gov/about/who-we-are/institutional-review-board-irb-of-all-of-us-research-program [accessed 2024-12-30]
23. The All of Us consent process. All of Us Research Program. URL: https://allofus.nih.gov/about/protocol/all-us-consent-process [accessed 2024-12-30]
24. Precision Medicine Initiative: privacy and trust principles. All of Us Research Program. URL: https://allofus.nih.gov/protecting-data-and-privacy/precision-medicine-initiative-privacy-and-trust-principles [accessed 2024-12-30]
25. What participants receive. All of Us Research Program. URL: https://www.joinallofus.org/what-participants-receive [accessed 2024-12-30]
26. Welcome to the All of Us Research Hub. All of Us Research Program. URL: https://www.researchallofus.org [accessed 2024-12-30]
27. Kummerow KL, Du L, Penson DF, Shyr Y, Hooks MA. Nationwide trends in mastectomy for early-stage breast cancer. JAMA Surg 2015 Jan;150(1):9-16. [doi: 10.1001/jamasurg.2014.2895] [Medline: 25408966]
28. Anderson WF, Devesa SS. In situ male breast carcinoma in the Surveillance, Epidemiology, and End Results database of the National Cancer Institute. Cancer 2005 Oct 15;104(8):1733-1741. [doi: 10.1002/cncr.21353] [Medline: 16138363]

Abbreviations

anti-HER2: anti–human epidermal growth factor receptor 2
BCT: breast-conserving therapy
BRCA: breast cancer gene
CPT4: Current Procedural Terminology 4
DQD: data quality dimensions
EHR: electronic health record
***ICD-9*:** *International Classification of Diseases, Ninth Revision*
IRB: institutional review board
LOINC: Logical Observation Identifiers Names and Codes
OMOP CDM : Observational Medical Outcomes Partnership Common Data Model
SNOMED: Systematized Nomenclature of Medicine

Edited by N Cahill; submitted 08.04.24; peer-reviewed by EA Voss, V Sekhon; revised version received 23.12.24; accepted 30.12.24; published 11.03.25.

Please cite as:
Spotnitz M, Giannini J, Ostchega Y, Goff SL, Anandan LP, Clark E, Litwin TR, Berman L
Assessing the Data Quality Dimensions of Partial and Complete Mastectomy Cohorts in the All of Us Research Program: Cross-Sectional Study
JMIR Cancer 2025;11:e59298
URL: https://cancer.jmir.org/2025/1/e59298
doi:10.2196/59298

Examining Demographic, Geographic, and Temporal Patterns of Melanoma Incidence in Texas From 2000 to 2018: Retrospective Study

Kehe Zhang[1,2*], MS; Madison M Taylor[3*], BSA; Jocelyn Hunyadi[1,2], MPH; Hung Q Doan[4], PhD, MD; Adewole S Adamson[5], MPP, MD; Paige Miller[6], MPH, PhD; Kelly C Nelson[4], MD; Cici Bauer[1,2], MS, PhD

[1]Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, E819, 1200 Pressler St., Houston, TX, United States

[2]Center for Spatial-Temporal Modeling for Applications in Population Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

[3]John P. and Kathrine G. McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, United States

[4]Department of Dermatology, Division of Internal Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

[5]Division of Dermatology, Dell Medical School, The University of Texas at Austin, Austin, TX, United States

[6]Cancer Epidemiology and Surveillance Branch, Texas Department of State Health Services, Austin, TX, United States

[*]these authors contributed equally

Corresponding Author:
Cici Bauer, MS, PhD
Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, E819, 1200 Pressler St., Houston, TX, United States

Abstract

Background: Melanoma currently ranks as the fifth leading cancer diagnosis and is projected to become the second most common cancer in the United States by 2040. Melanoma detected at earlier stages may be treated with less-risky and less-costly therapeutic options.

Objective: This study aims to analyze temporal and spatial trends in melanoma incidence by stage at diagnosis (overall, early, and late) in Texas from 2000 to 2018, focusing on demographic and geographic variations to identify high-risk populations and regions for targeted prevention efforts.

Methods: We used melanoma incidence data from all 254 Texas counties from the Texas Cancer Registry (TCR) from 2000 to 2018, aggregated by county and year. Among these, 250 counties reported melanoma cases during the period. Counties with no cases reported in a certain year were treated as having no cases. Melanoma cases were classified by SEER Summary Stage and stratified by the following four key covariates: age, sex, race and ethnicity, and stage at diagnosis. Incidence rates (IRs) were calculated per 100,000 population, and temporal trends were analyzed using joinpoint regression to determine average annual percentage changes (AAPCs) with 95% CIs for the whole time period (2000 - 2018), the most recent 10-year period (2009 - 2018), and the most recent 5-year period (2014 - 2018). Heat map visualizations were developed to assess temporal trends by patient age, year of diagnosis, stage at diagnosis, sex, and race and ethnicity. Spatial cluster analysis was conducted using Getis-Ord Gi* statistics to identify county-level geographic clusters of high and low melanoma incidence by stage at diagnosis.

Results: A total of 82,462 melanoma cases were recorded, of which 74.7% (n=61,588) were early stage, 11.3% (n=9,352) were late stage, and 14% (n=11,522) were of unknown stage. Most cases were identified as males and non-Hispanic White individuals. Melanoma IRs increased from 2000 to 2018, particularly among older adults (60+ years; AAPC range 1.20%-1.84%; all *P* values were <.001), males (AAPC 1.59%; *P*<.001), and non-Hispanic White individuals (AAPC of 3.24% for early stage and 2.38% for late stage; *P*<.001 for early stage and *P* = .03 for late state). Early-stage diagnoses increased while the rates of late-stage diagnoses remained stable for the overall population. The spatial analysis showed that urban areas had higher early-stage incidence rates (*P*=.06), whereas rural areas showed higher late-stage incidence rates (*P*=.05), indicating possible geographic-based differences in access to dermatologic care.

Conclusions: Melanoma incidence in Texas increased over the study time period, with the most-at-risk populations being non-Hispanic White individuals, males, and individuals aged 50 years and older. The stable rates of late-stage melanoma among racial and ethnic minority populations and rural populations highlight potential differences in access to diagnostic care. Future prevention efforts may benefit from increasing access to dermatologic care in areas with higher rates of late-stage melanoma at diagnosis.

(JMIR Cancer 2025;11:e67902) doi:10.2196/67902

KEYWORDS

melanoma incidence; melanoma screening; geographic disparity; geospatial analysis; joinpoint regression; demographic variation; temporal trend analysis; stage at diagnosis

Introduction

Melanoma currently ranks as the fifth leading cancer diagnosis overall and is projected to become the second most diagnosed cancer in the United States by 2040 [1]. Nationally, melanoma incidence rates (IRs) have shown distinct temporal trends across age, sex, race, and ethnicity. Since 2006, IRs have decreased for adolescents and young adults but increased for older adults, with an annual percent change (APC) of 2.5% for individuals older than 65 years between 2006 and 2015 [2,3]. From 2006 to 2021, IRs have increased for non-Hispanic White and Hispanic individuals (APC of 1.7% and 0.6%, respectively) but decreased for non-Hispanic Black individuals (APC of –1.2%) [3]. Non-Hispanic White males older than 50 years have maintained the highest incidence over the last 2 decades [3].

Previous studies have reported that the areal-level social determinants of health (SDoH) are associated with disparities in melanoma incidence, stage at diagnosis, and survival outcomes. For example, using spatial cluster analysis and a multinomial logistic regression model, a study in Florida found that patients with melanoma who live in census tracts with higher percentage of poverty are more likely to have a late-stage diagnosis [4]. Another study examining the national cancer database from 2011 to 2020 used a chi-square test and found that patients from urban areas are more likely to have an early-stage melanoma (P<.001) [5]. Similarly, a study in Texas using a multinomial logistic regression model found that patients from counties with persistent poverty (≥20% of residents at or below the federal poverty level for the past two decennial censuses) have higher incidence-based melanoma mortality [6]. Melanoma survival rates vary dramatically depending on stage at diagnosis, with nearly guaranteed 5-year survival for early-stage (localized) diagnoses [3]. However, racial and ethnic minority groups are more likely to be diagnosed with melanoma at advanced stages when compared with stage-matched White patients (chi-square tests with P<.001) [3,7,8]. Therefore, understanding how stage-specific melanoma incidence varies across time by patient demographics and geographic location can inform data-driven early detection efforts to improve melanoma morbidity and mortality.

Although Texas had lower all-stage melanoma IRs compared with the national average during 2017 and 2021 (14.9 vs 22.7 cases per 100,000 population), it reported the highest percentage of late-stage cutaneous melanoma diagnoses in the contiguous United States (18.2% in Texas vs 14.1% nationally) [9]. One strategy to shift melanoma detection from late to earlier stages is to increase screening via dermatologists. However, many regions of Texas lack access to dermatologists [10,11]. Alternatively, primary care providers (PCPs) may provide essential skin-cancer detection services [12], yet significant training barriers often preclude early skin cancer diagnosis by PCPs [13]. Geographically targeted education and telementoring efforts to support PCP melanoma diagnosis [14] could potentially enhance early melanoma detection, particularly in areas with a high late-stage melanoma IR. However, the spatial and temporal distribution of stage-based melanoma incidence in Texas has yet to be thoroughly explored to identify these critical locations.

In this study, we analyzed Texas Cancer Registry (TCR) melanoma cases from 2000 to 2018. Using novel data visualizations, we identified trends in melanoma incidence by year of diagnosis, patient demographics, and stage at diagnosis. We also investigated county-level geographic patterns of melanoma incidence across Texas over time. Understanding these trends can guide the development of risk-based interventions to improve melanoma outcomes at the population level.

Methods

Data

Melanoma incidence data from 2000 to 2018 were obtained from the TCR and aggregated by 254 Texas counties and year. Among these, 250 counties reported melanoma cases during the period. Counties with no cases reported in a certain year were treated as having no cases. Melanoma cases were categorized by stage at diagnosis using the Surveillance, Epidemiology, and End Results (SEER) summary stage system, which differs from the more clinically oriented National Comprehensive Cancer Network melanoma-specific staging guidelines [10]. While SEER summary stage data categories have changed over time, we organized TCR melanoma cases into three groups: early stage (SEER stages 0 [in situ], 1 [localized], and 2 [regional by direct extension only]), late stage (SEER stages 3 to 5 [regional] and stage 7 [distant]), and unknown (SEER stage 9: unknown, unstaged, and unspecified).

For demographic stratification, we considered 7 age groups (18 - 29, 30 - 39, 40 - 49, 50 - 59, 60 - 69, 70 - 79, and ≥80 years old), 2 sex groups (female and male), and 4 racial and ethnic groups (non-Hispanic White, non-Hispanic Black, Hispanic, and non-Hispanic Others). Annual county-level population estimates stratified by age, sex, and race and ethnicity were obtained from the National Institutes of Health (NIH) SEER county population data [15].

Patient county of residence at the time of diagnosis was classified as rural or urban using the 2013 US Department of Agriculture Rural-Urban Continuum Codes (RUCCs) [16]. RUCCs 1 - 3 were classified as urban and RUCCs 4 - 9 were classified as rural. Counties were also classified as either with or without persistent poverty using the US Economic Development Administration's 2021 persistent poverty data [17].

Incidence Calculation and Trend Analysis

Annual melanoma incidence-based rates were calculated by sex, age, and racial and ethnic groups as described in the "Data" section. Given the substantial variation in melanoma incidence across different age groups, with higher rates typically observed in older age groups, previous analyses often reported age-adjusted rates to allow for more comparable temporal trends across different populations. However, in this study, we adopted a novel data visualization approach using temporal heat maps, which effectively incorporate known confounders such as age and provide a clearer and more intuitive representation of trends [18]. The heat maps display the year of diagnosis (x-axis), age at diagnosis (y-axis), and calculated stage-specific incidence per 100,000 population as a blue (lower rates) to red (higher rates) color gradient. The heat maps used the Akima interpolation method [19] to generate a smoothed surface from observed data points, providing a visually coherent presentation of temporal trends in IRs.

Spatial Cluster Analysis

We calculated the annual melanoma incidence-based rate for each county in Texas (n=254) using county-specific population data. Spatial cluster analyses were then performed using Getis-Ord Gi* statistics. The Gi* statistic indicates the degree of spatial clustering: positive values indicate that a county and its neighboring areas have higher-than-average rates, while negative values suggest lower-than-average rates. Statistical significance was assessed via Monte Carlo simulation, comparing the observed Gi* values to a reference distribution generated from simulated spatial data. The results categorize counties' spatial clustering significance as follows: Very High (Gi* stat>0 and $P<.01$), High (Gi* stat>0 and $0.01 \leq P<.05$), Somewhat High (Gi* stat>0 and $0.05 \leq P<.10$), Insignificant ($P>.10$), and Low (Gi* stat<0 and $P<.10$). These categorizations allow for the identification of counties with significantly higher or lower melanoma IRs than would be expected by random chance and represent successive thresholds for interpreting spatial clusters without implying a strict ranking of significance. Chi-square tests were used to examine the relationship between spatial clustering categories and urban-rural or poverty status. All data analyses and visualizations were conducted in R (version 4.2.1; R Core Team) [20].

Joinpoint Trend Analysis

To assess temporal changes in melanoma IR, we performed a state-level joinpoint trend analysis to identify years when significant shifts in trends occured. We calculated the APC in IR using the weighted least-squares method, stratified by stage at diagnosis and among different demographic groups [21]. The APC represents the annual rate of change in IR over a specified period, assuming a constant percentage change each year. For instance, an APC of 2% would indicate that an IR of 100 per 100,000 would increase to 102 per 100,000 in the following year. We allowed for a maximum number of two joinpoints over the 19-year study period. Using the Joinpoint Trend Analysis Software (Joinpoint Regression Program; version 5.3.0.0) from SEER [22], we identified specific years with significant changes in the temporal trends and determined the final number of joinpoints using permutation tests. In addition, we derived a summary measure, the average annual percentage change (AAPC), over three fixed time periods: 2000 - 2018 (entire study period), 2009 - 2018 (most recent 10 y), and 2014 - 2018 (most recent 5 y), based on the joinpoint regression model for the full period from 2000 to 2018. For example, an AAPC of 2% for the 2010 - 2018 period would indicate that the IR increased by an average of 2% annually during these years. The 95% CIs for both APC and AAPC were derived using empirical quantile methods.

Ethical Considerations

The study was approved by the institutional review board (IRB) at the University of Texas Health Science Center at Houston (IRB: HSC-SPH-23 - 0483). Melanoma TCR data were obtained from the Texas Department of State Health Services (DSHS) via a formal data request that included an IRB application. Upon IRB approval from the DSHS, a waiver of informed consent was granted because study constitutes secondary research using existing data, involves no more than minimal risk to the participants, and therefore does not require additional consent.

To protect participants' privacy and confidentiality, all data were de-identified prior to analysis. The dataset included only non-identifiable variables such as year of diagnosis, county of residence at diagnosis, demographic characteristics (e.g., sex, race/ethnicity), and birth year. No names, contact information, or medical record numbers were included. The research team adhered to all DSHS data use agreements and institutional data security protocols. Access to the data was limited to approved study personnel and stored on encrypted, password-protected servers within secure institutional networks.

Results

Overview

From 2000 to 2018, the TCR reported 82,462 melanoma cases (Table 1). Among these, 61,588 (74.7%) were diagnosed at an early stage, 9352 (11.3%) at a late stage, and 11,522 (14%) had an unknown stage at diagnosis. The demographic subgroups with the most cases included individuals aged 60 - 69 years (18,959 cases, 23.0%), males (49,058 cases, 59.5%), and non-Hispanic White individuals (74,943 cases, 90.9%). The demographic distribution of melanoma cases by stage largely mirrored these overall trends. The racial and ethnic distribution showed that non-Hispanic White individuals dominated both early- and late-stage cases, although their proportion was slightly lower in late-stage cases.

Figure 1 presents temporal heat maps of stage-specific population-adjusted melanoma IRs (cases per 100,000 population) across demographic subgroups (sex, age, and race and ethnicity) from 2000 to 2018 in Texas. When considering all stages, melanoma IRs slightly increased over time to around 90 per 100,000 population. The highest increase was observed in the 80+ age group from 79 in 2000 to 104 in 2018 (AAPC 1.84%, 95% CI 1.27-2.40), while the 70 - 79 age group saw the highest increase from 2014 - 2018 (AAPC 2.75%, 95% CI 1.84-5.03). Conversely, IRs for the 18 - 29 age group declined from 6 in 2014 to 4.6 per 100,000 in 2018 (AAPC 3.05%, 95% CI –3.66 to –2.46). IRs for those aged 30 - 49 years remained

stable at approximately 10 throughout the study period (Figure 1and Multimedia Appendix 1).

Table . Summary of patient demographics by Surveillance, Epidemiology, and End Results (SEER) summary stage system groupings at diagnosis.

Variable	Early stage[a] (N=61,588), n (%)	Late stage[b] (N=9,352), n (%)	Unknown stage[c] (N=11,522), n (%)	All cases (N=82,462), n (%)
Age group (years)				
18 - 29	2,196 (3.6)	391 (4.2)	536 (4.7)	3,123 (3.8)
30 - 39	4,466 (7.3)	758 (8.1)	927 (8.0)	6,151 (7.5)
40 - 49	7,599 (12.3)	1,287 (13.8)	1,589 (13.8)	10,475 (12.7)
50 - 59	11,822 (19.2)	2,049 (21.9)	2,149 (18.7)	16,020 (19.4)
60 - 69	14,412 (23.4)	2,108 (22.5)	2,439 (21.2)	18,959 (23.0)
70 - 79	13,027 (21.2)	1,712 (18.3)	2,156 (18.7)	16,895 (20.5)
≥80	8,066 (13.1)	1,047 (11.2)	1,726 (15.0)	10,839 (13.1)
Sex				
Male	36,300 (58.9)	6,016 (64.3)	6,742 (58.5)	49,058 (59.5)
Female	25,288 (41.1)	3,335 (35.7)	4,780 (41.5)	33,403 (40.5)
Race and ethnicity				
Non-Hispanic White	56,250 (91.3)	8,247 (88.2)	10,446 (90.7)	74,943 (90.9)
Non-Hispanic Black	304 (0.5)	122 (1.3)	109 (0.9)	535 (0.6)
Hispanic	3,294 (5.3)	924 (9.9)	784 (6.8)	5,002 (6.1)
Non-Hispanic Others	340 (0.6)	56 (0.6)	63 (0.5)	459 (0.6)
Unknown	1,400 (2.3)	3 (0.0)	120 (1.0)	1,523 (1.8)

[a]Early stage: SEER stages 0 (in situ), 1 (localized), and 2 (regional by direct extension only).

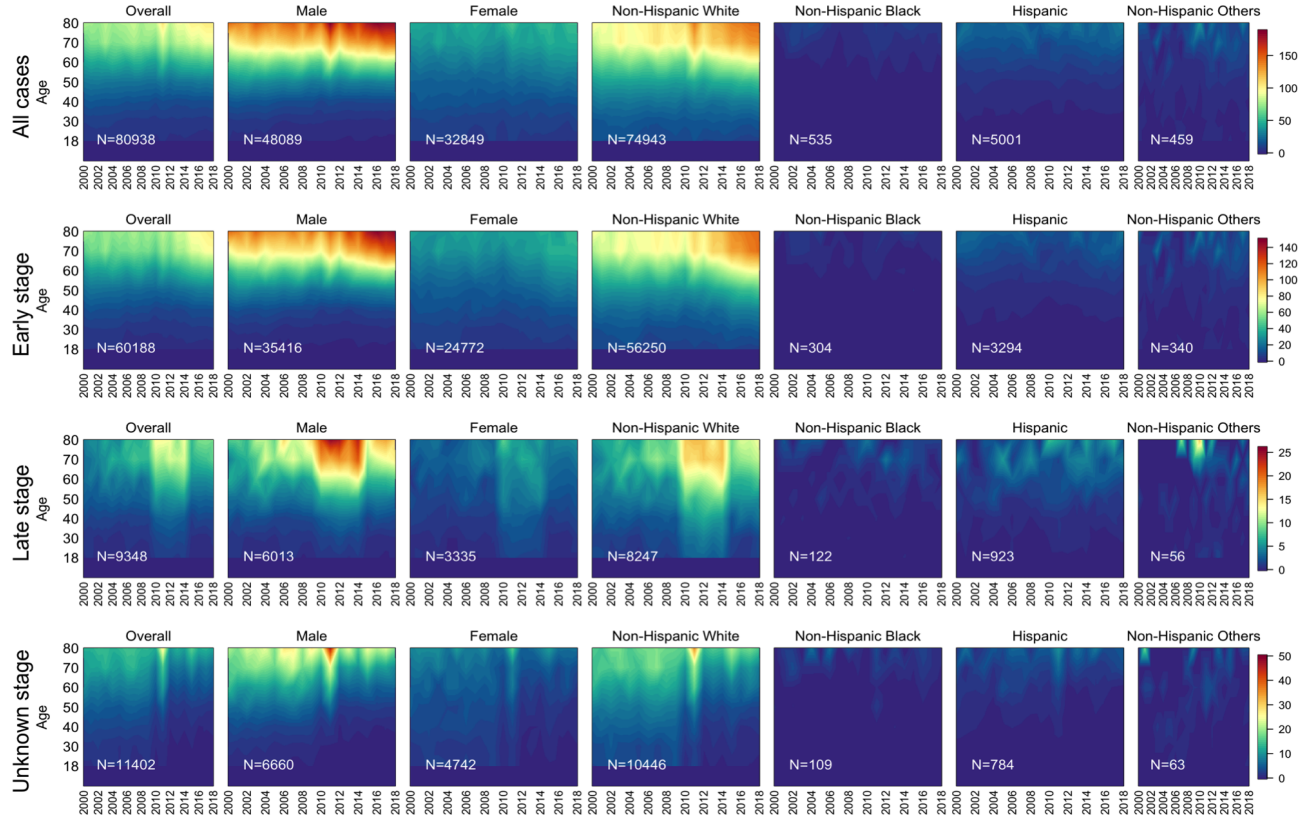
[b]Late stage: SEER stages 3 to 5 (regional) and stage 7 (distant).

[c]Unknown stage: SEER stage 9: unknown, unstaged, unspecified.

IRs for males appear much higher than females among 50+ years, reaching over 140 per 100,000 in the 70+ age group and over 180 per 100,000 in the 80+ age group by 2018 (Figure 1). In contrast, among younger age groups (18-49), females showed slightly higher rates than males over time. In addition, melanoma incidence increased at a more rapid rate for males from 2009 to 2018, with an AAPC of 3.48% (95% CI 2.49-5.19) compared with 2.47% (95% CI 1.89-3.26) for females (Multimedia Appendix 1).

Stratifying by race and ethnicity, non-Hispanic White individuals displayed the highest rates across all age groups, with noticeable increases over the study period for those aged 50 years and older. The older non-Hispanic White patients (70+ years) showed rates over 130 per 100,000 by 2018. For the 18 - 29 age group, non-Hispanic White patients had a slight decrease in IRs from 16.5 per 100,000 in 2000 to 11.4 per 100,000 in 2018. Other racial and ethnic groups, such as non-Hispanic Black and Hispanic individuals, maintained much lower IRs mostly below 30 per 100,000 across all age groups and years (Figure 1).

Figure 1. Temporal heat maps presenting melanoma incidence rates (per 100,000 population) by columns of overall population, sex, and racial and ethnicity groups. Each row panel shows a different stage at diagnosis: all cases, early stage, late stage, and unknown stage. Numbers in the lower left corner of each panel indicate the total number of melanoma cases in Texas from 2000 to 2018.



To facilitate comparison with national statistics, age-adjusted IRs (to the 2000 US standard population) are presented by stage at diagnosis, age, gender, and race and ethnicity in Multimedia Appendix 2.

The incidence of early-stage melanoma has increased more rapidly than late-stage melanoma incidence (Figure 1). Early-stage melanoma temporal trends mirrored those of overall cases, with the highest rates for non-Hispanic White males older than 70 years, reaching 114 cases per 100,000 by 2018. Other racial and ethnic groups displayed minimal variation over time, with rates remaining low throughout the study period (under 10 cases per 100,000 for non-Hispanic Black individuals and under 25 per 100,000 for Hispanic individuals). Late-stage melanoma IRs were lower than early-stage incidence across all demographics. The overall population showed rates below 15 cases per 100,000, with a slight increase over time in older males, peaking around 25 per 100,000 between 2011 - 2014 (refer to additional details in Discussion). Non-Hispanic White patients had the highest rates (maximum 16 per 100,000 in the 80+ age group), while non-Hispanic Black and Hispanic patients had consistently low late-stage IRs (mostly below 5 cases per 100,000). The IRs for unknown stage cases remained relatively stable over time across all demographic groups.

When investigating stage by race and ethnicity, early-stage diagnoses predominated across all races and ethnicities. However, Hispanic and non-Hispanic Black individuals had proportionately more late stage at diagnosis melanomas than non-Hispanic White patients (Multimedia Appendix 3). The AAPC for late-stage melanoma cases was similar for Hispanic (AAPC 2.38%, 95% CI 0.53-3.73) and non-Hispanic White patients (AAPC 2.66%, 95% CI –0.98 to 7.17) and higher for non-Hispanic Black patients (AAPC 5.79%, 95% CI 2.61-9.01; Multimedia Appendix 1).

Spatial Cluster Analysis

Maps showing high- and low-melanoma incidence spatial clusters are presented in Figure 2. When considering all melanoma cases, spatial clusters with significantly higher-than-average melanoma incidence (median IR 42 per 100,000, IQR 23-62) were primarily in northwestern Texas from 2000 to 2006, with a shift to central Texas between 2007 and 2015, and then to counties surrounding Dallas by 2018. Spatial clusters with lower-than-average melanoma incidence were clustered near southern and western Texas (median IR 11, IQR 0-24). The spatial patterns for early-stage and overall melanoma cases were similar with median IRs of 35 per 100,000 (IQR 19-52) in high-incidence spatial clusters, and 6 per 100,000 (IQR 0-17) in low-incidence spatial clusters. However, the spatial clusters for higher-than-average late-stage cases showed distinct patterns with localization to northwestern Texas from 2000 to 2007, a shift toward southeast Texas between 2010 and 2014, and a return to central-northern Texas in 2015.

Overlaying the 2018 spatial clusters with rural counties (hatched lines), we observed that clusters of higher-than-average late-stage melanoma incidence significantly overlap with rural areas (P=.05; Figure 3A). Similarly, clusters of higher-than-average early-stage melanoma incidence appear to overlap with urban areas (P=.06; Figure 3A). In contrast, clusters of lower-than-average incidence for overall, early-stage, and

late-stage melanoma appear to overlap with persistent poverty counties (all *P* values were <.001; Figure 3B).

Figure 2. Spatial cluster analysis of melanoma incidence rates (cases per 100,000 population) by melanoma stage at diagnosis (all cases, early stage, and late stage) and selected years from 2000 to 2018 using Gi* statistics. Classifications were defined as follows: very high (Gi*>0 and *P*<.01), high (Gi*>0 and 0.01 ≤ *P*<.05), somewhat high (Gi*>0 and 0.05 ≤ *P*<.10), insignificant (*P*>.10), and low (Gi*<0 and *P*<.10). Red-shaded areas represent high-incidence clusters, depicting clusters of counties with significantly higher incidence rates compared with the statewide average incidence rate. Blue-shaded areas highlight clusters of counties with significantly lower incidence rates compared with the state average.

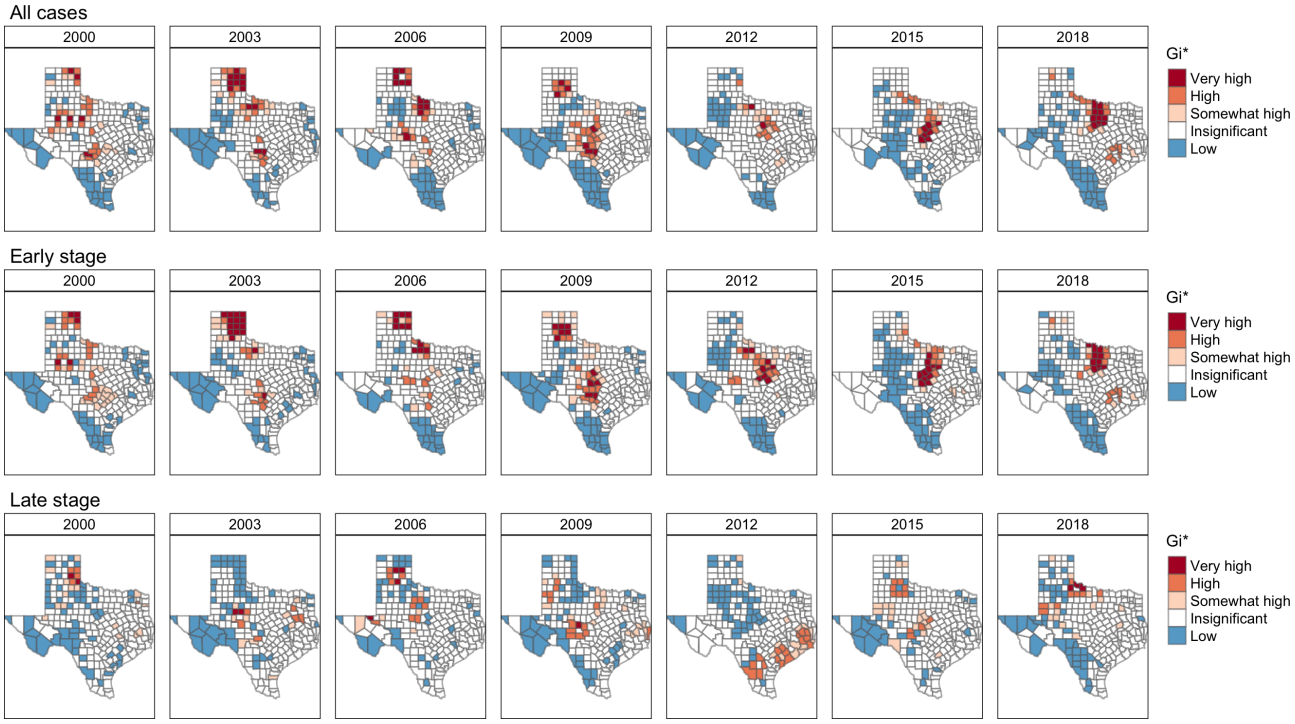
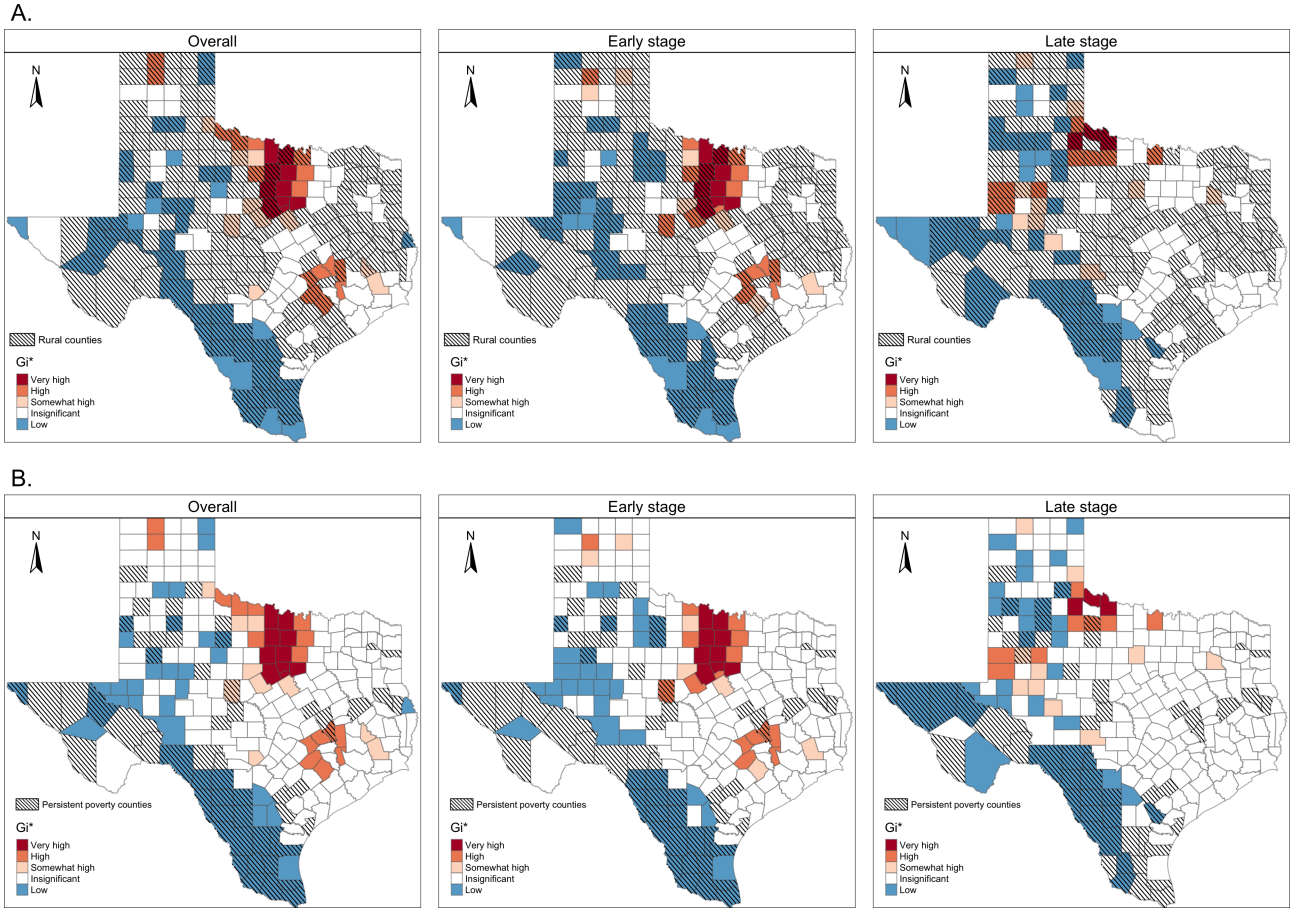


Figure 3. Spatial clusters of melanoma incidence rates overlaid with (A) rural counties and (B) persistent poverty counties in 2018. Classifications were based on Gi* statistics as follows: very high (Gi*>0 and P<.01), high (Gi*>0and 0.01 ≤ P<.05), somewhat high (Gi*>0 and 0.05 ≤ P<.10), insignificant (P>.10), and low (Gi*<0 and P<.10).



Discussion

Principal Findings

Texas has the highest proportion of late-stage melanoma cases relative to the total number of reported melanoma cases in the contiguous United States and is the second most populated state. Identifying the regions and patient populations that disproportionately bear the burden of late-stage melanoma at diagnosis is crucial for developing targeted early detection efforts. Our spatial clustering analysis of Texas revealed that high-incidence clusters of early-stage melanoma were primarily localized in urban, well-resourced areas, whereas high-incidence clusters of late-stage melanoma were concentrated in rural areas . This disparity may be partly explained by the lower density of dermatologists in rural areas [11,23]. Patients with melanoma living in these areas may experience delayed diagnosis [24], need to travel longer distances to receive surgical management, and have decreased melanoma-specific survival [25]. These findings, while observational, highlight the importance of addressing structural disparities in health care access. Therefore, any early detection intervention must be tailored to be feasible in rural, lower-resourced settings.

Melanoma IRs exhibit substantial variation by age, with older patients experiencing significantly higher rates compared with younger patients. Because the age distribution of the population often shifts over time, age-adjusted rates are commonly used to compare trends across different time periods and geographic regions. These adjustments typically use a fixed reference year for the age distribution, which may lack the robustness to fully capture ongoing demographic change. As the population continue to age and shift in age structure, a static reference year may obscure important trends and fail to accurately reflect the current risk landscape. In this study, we used a novel visualization approach using temporal heat maps that directly incorporate age as a variable, providing a more adaptive and precise method for identifying shifts in melanoma incidence trends over time. This visualization highlighted pronounced incidence variations, particularly among late-stage male and non-Hispanic White patients from 2010 to 2014. We identified two TCR data sources which may explain this variation: Texas Health Care Information Collection (THCIC) and eMaRC Plus (Centers for Disease Control and Prevention [CDC]). THCIC, established by the Texas legislature, collects data on health care activities in hospitals and health maintenance organizations [26]. During a pilot from 2010 to 2013, THCIC identified melanoma cases not otherwise reported to TCR, leading to the inclusion of 559 cases in the TCR database, most of which were categorized as unknown stage at diagnosis. The second data source, eMaRC Plus, a software developed by the CDC to receive and process Health Level Seven files from pathology laboratories, was used from 2010 to 2018. eMaRC Plus identified 8,786 cases, with 93.1% being early-stage and 6.8% unknown stage. The peaks in early-stage and unknown-stage

incidence observed in 2011 and onward may be attributed to these two data sources. However, the increase in late-stage incidence from 2010 to 2014 could not be fully explained by these data sources, suggesting that temporary reporting inconsistencies may warrant further investigation to fully understand their impact on the identified temporal trends.

Despite these data source differences, descriptive trends also revealed that older non-Hispanic White men comprise the majority of late-stage melanoma cases at diagnosis. This specific demographic provides a clear target cohort for refining early melanoma detection efforts. Given that older men generally have lower rates of skin self-awareness [27], promoting early detection through their family members or PCPs may offer greater opportunity. Our analysis of race and ethnicity revealed that, while most melanoma cases were diagnosed at early stages across all groups, Hispanic and non-Hispanic Black patients experienced proportionately more late stage at diagnosis melanomas than non-Hispanic White counterparts. Although the absolute number of melanoma cases is small compared with other cancers, these findings reinforce the concept that melanoma can impact individuals of all races and ethnicities.

Strengths and Limitations

This study has several strengths and limitations. A key strength is the innovative visualization approach, which enabled a comprehensive analysis of temporal trends, stratified by age, sex, race and ethnicity, and stage at diagnosis. The use of TCR data allowed for the inclusion of patient residential information at the time of diagnosis, as well as additional data captures from pilot studies that would otherwise be unavailable. Furthermore, the geospatial analysis provided a comprehensive examination of patterns across all cases and by specific stages, offering valuable insights that could guide future interventions and educational efforts. The study also presents several limitations, particularly the large proportion of cases with an unknown stage at diagnosis. More precise data on these cases would enable better stage classification and improve the stage-specific analyses. In addition, the TCR data has limitations in capturing SDoH, which, if included, could provide a deeper understanding of the health disparities associated with melanoma outcomes.

Conclusions

Our study provides valuable guidance for future early melanoma detection efforts. Such efforts must be feasible in rural, lower-resourced areas of the state and focus on patients at highest risk of late-stage melanoma at diagnosis. Multimodal approaches, which combine foundational dermatology training for interested PCPs [14], telementoring to support PCP incorporation of skin cancer detection examinations into practice [28], and efficient store-and-forward eConsults [29] to reduce dermatology access gaps, offer promising pathways to improve early melanoma detection in low-resource settings.

Acknowledgments

This project is funded by CPRIT RP230036 (MPI: CB and KCN).

Data Availability

The data used in this study were obtained from the Texas Cancer Registry via a formal data request, accompanied by an institutional review board application. For inquiries regarding data access, please contact the Texas Cancer Registry at CancerData@dshs.texas.gov.

Authors' Contributions

CB conceived and designed the analysis. KCN contributed to the acquisition of the data. KZ and CB contributed to data processing and curation and conducted the formal data analysis. KZ and MMT wrote the initial draft of the manuscript. CB and KCN supervised the study and acquired funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Average annual percent change (AAPC) in melanoma incidence rates and corresponding 95% CI, stratified by stage at diagnosis (overall, early, and late) and demographic factors (age, sex, and race and ethnicity), for the time intervals 2000-2018, 2009-2018, and 2014-2018. Results were based on Joinpoint regression analysis for the time interval of 2000-2018.
[DOCX File, 16 KB - cancer_v11i1e67902_app1.docx]

Multimedia Appendix 2
Age-adjusted melanoma incidence rates (cases per 100,000 population) by stage at diagnosis and demographics groups, standardized to the 2000 US population.
[PNG File, 718 KB - cancer_v11i1e67902_app2.png]

Multimedia Appendix 3

Melanoma incidence rates by race and ethnicity, and stage at diagnosis from 2000 to 2018.
[PNG File, 124 KB - cancer_v11i1e67902_app3.png]

References

1. Rahib L, Wehner MR, Matrisian LM, et al. Estimated projection of US cancer incidence and death to 2040. JAMA Netw Open 2021 Apr 1;4(4):e214708. [doi: 10.1001/jamanetworkopen.2021.4708] [Medline: 33825840]
2. Paulson KG, Gupta D, Kim TS, et al. Age-specific incidence of melanoma in the United States. JAMA Dermatol 2020 Jan 1;156(1):57-64. [doi: 10.1001/jamadermatol.2019.3353]
3. SEER*Explorer. Surveillance, Epidemiology, and End Results Program, National Cancer Institute. URL: https://seer.cancer.gov/statistics-network/explorer/application.html?sel=18&datatype=1&graphtype=2&compareBy=sex&chk_sex_3=3&chk_sex_2=2&rate_type=2&race=1&age_range=18&hdn_stage=101&advopt_precision=1&advopt_show_ci=on&hdn_view=0&advopt_show_apc=on&advopt_display=2 [accessed 2024-02-21]
4. Hu S, Sherman R, Arheart K, et al. Predictors of neighborhood risk for late-stage melanoma: addressing disparities through spatial analysis and area-based measures. J Invest Dermatol 2014 Apr;134(4):937-945. [doi: 10.1038/jid.2013.465] [Medline: 24335896]
5. Akers R, Coogan A, Kasper L, et al. Melanoma disparities in urban vs rural communities in the era of immunotherapy: an NCDB analysis. J Clin Oncol 2023;41:9578 [FREE Full text]
6. Madrigal K, Morris L, Zhang K, et al. Persistent poverty and incidence-based melanoma mortality in Texas. Cancer Causes Control 2024 Jun;35(6):973-979. [doi: 10.1007/s10552-023-01841-5] [Medline: 38421511]
7. Qian Y, Johannet P, Sawyers A, et al. The ongoing racial disparities in melanoma: an analysis of the surveillance, epidemiology, and end results database (1975-2016). J Am Acad Dermatol 2021 Jun;84(6):1585-1593. [doi: 10.1016/j.jaad.2020.08.097] [Medline: 32861710]
8. Ward-Peterson M, Acuña JM, Alkhalifah MK, et al. Association between race/ethnicity and survival of melanoma patients in the United States over 3 decades: a secondary analysis of SEER data. Medicine 2016 Apr;95(17):e3315. [doi: 10.1097/MD.0000000000003315] [Medline: 27124020]
9. Incidence rate report by state: melanoma of the skin (late stage^), 2017-2021, all races (includes hispanic), both sexes, all ages, sorted by percentlate. State Cancer Profiles. URL: https://statecancerprofiles.cancer.gov/incidencerates/index.php?stateFIPS=00&areatype=state&cancer=053&race=00&sex=0&age=001&stage=211&year=0&type=incd&sortVariableName=percentlate&sortOrder=desc#results [accessed 2024-08-08]
10. Glazer AM, Farberg AS, Winkelmann RR, et al. Analysis of trends in geographic distribution and density of US dermatologists. JAMA Dermatol 2017 Apr 1;153(4):322-325. [doi: 10.1001/jamadermatol.2016.5411] [Medline: 28146246]
11. Shah M, Burshtein J, Zakria D, Rigel D. Analysis of trends in US dermatologist density and geographic distribution. J Am Acad Dermatol 2024 Aug;91(2):338-341. [doi: 10.1016/j.jaad.2024.03.037] [Medline: 38574771]
12. Seiverling EV, Prentiss MA, Stevens K, et al. Impact of dermoscopy training for primary care practitioners on number needed to biopsy to detect melanoma. PRiMER 2023;7:4. [doi: 10.22454/PRiMER.2023.276659] [Medline: 36845847]
13. Najmi M, Brown AE, Harrington SR, et al. A systematic review and synthesis of qualitative and quantitative studies evaluating provider, patient, and health care system-related barriers to diagnostic skin cancer examinations. Arch Dermatol Res 2022 May;314:329-340. [doi: 10.1007/s00403-021-02224-z]
14. Nelson KC, Seiverling EV, Gonna N, et al. A pilot educational intervention to support primary care provider performance of skin cancer examinations. J Canc Educ 2023 Feb;38:364-369. [doi: 10.1007/s13187-021-02126-8] [Medline: 35013902]
15. U.S. county population data - 1969-2023. Surveillance, Epidemiology, and End Results Program, National Cancer Institute. 2023. URL: https://seer.cancer.gov/popdata/index.html [accessed 2023-03-08]
16. Rural-urban continuum codes. Economic Research Service, US Department of Agriculture. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx [accessed 2024-03-29]
17. FY2021 persistent poverty counties. US Economic Development Administration. 2021. URL: https://www.eda.gov/sites/default/files/2023-03/EDA_FY23_PPCs.xlsx [accessed 2025-04-29]
18. Jalal H, Buchanich JM, Roberts MS, et al. Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. Science 2018 Sep 21;361:eaau1184. [doi: 10.1126/science.aau1184] [Medline: 30237320]
19. Akima H, Gebhardt A, Petzold T, et al. Akima: interpolation of irregularly and regularly spaced data. The Comprehensive R Archive Network. 2022. URL: https://CRAN.R-project.org/package=akima [accessed 2025-04-11]
20. The R project for statistical computing. R Foundation for Statistical Computing. 2022. URL: https://www.R-project.org [accessed 2024-05-11]
21. Annual percent change (APC) and confidence interval. National Cancer Institute. URL: https://surveillance.cancer.gov/help/joinpoint/setting-parameters/method-and-parameters-tab/apc-aapc-tau-confidence-intervals/estimate-average-percent-change-apc-and-confidence-interval [accessed 2024-06-10]
22. Joinpoint trend analysis software. National Cancer Institute. 2023 May. URL: https://surveillance.cancer.gov/joinpoint/ [accessed 2025-04-23]
23. Feng H, Berk-Krauss J, Feng PW, et al. Comparison of dermatologist density between urban and rural counties in the United States. JAMA Dermatol 2018 Nov 1;154(11):1265-1271. [doi: 10.1001/jamadermatol.2018.3022] [Medline: 30193349]

24. Strömberg U, Peterson S, Holmberg E, et al. Cutaneous malignant melanoma show geographic and socioeconomic disparities in stage at diagnosis and excess mortality. Acta Oncol 2016 Aug;55(8):993-1000. [doi: 10.3109/0284186X.2016.1144934] [Medline: 26935355]
25. Godinich BM, Hensperger V, Guo W, et al. Barriers to malignant melanoma diagnosis in rural areas in the United States: a systematic review. JAAD Reviews 2024 Sep;1:29-41. [doi: 10.1016/j.jdrv.2024.06.001]
26. Texas Health Care Information Collection (THCIC). Texas Department of State Health Services. URL: https://www.dshs.texas.gov/center-health-statistics/texas-health-care-information-collection [accessed 2025-04-11]
27. Oliveria SA, Christos PJ, Halpern AC, et al. Patient knowledge, awareness, and delay in seeking medical attention for malignant melanoma. J Clin Epidemiol 1999 Nov;52(11):1111-1116. [doi: 10.1016/s0895-4356(99)00083-9] [Medline: 10527006]
28. Seiverling E, Ahrns H, Stevens K, et al. Dermoscopic lotus of learning: implementation and dissemination of a multimodal dermoscopy curriculum for primary care. J Med Educ Curric Dev 2021;8. [doi: 10.1177/2382120521989983] [Medline: 33598548]
29. Gellatly ZS, Lagha IB, Ternov NK, et al. The role of dermoscopy in provider-to-provider store-and-forward dermatology eConsults: a scoping review of the recent literature. Curr Derm Rep 2023 Dec;12:169-179. [doi: 10.1007/s13671-023-00407-7]

Abbreviations

AAPC: average annual percentage change
APC: annual percent change
CDC: Centers for Disease Control and Prevention
DSHS: Department of State Health Services
IR: incidence rate
IRB: institutional review board
NIH: National Institutes of Health
PCP: primary care provider
RUCC: Rural-Urban Continuum Code
SDoH: social determinants of health
SEER: Surveillance, Epidemiology, and End Results
TCR: Texas Cancer Registry
THCIC: Texas Health Care Information Collection

Edited by N Cahill; submitted 23.10.24; peer-reviewed by PJ Gruenewald, S Young; revised version received 26.02.25; accepted 26.02.25; published 02.05.25.

Please cite as:
Zhang K, Taylor MM, Hunyadi J, Doan HQ, Adamson AS, Miller P, Nelson KC, Bauer C
Examining Demographic, Geographic, and Temporal Patterns of Melanoma Incidence in Texas From 2000 to 2018: Retrospective Study
JMIR Cancer 2025;11:e67902
URL: https://cancer.jmir.org/2025/1/e67902
doi:10.2196/67902

Assessing Public Interest in Mammography, Computed Tomography Lung Cancer Screening, and Computed Tomography Colonography Screening Examinations Using Internet Search Data: Cross-Sectional Study

Zachary D Zippi[1], PharmD; Isabel O Cortopassi[2], MD; Rolf A Grage[2], MD; Elizabeth M Johnson[2], MD; Matthew R McCann[2], MD; Patricia J Mergo[2], MD; Sushil K Sonavane[2], MD; Justin T Stowell[2], MD; Brent P Little[2], MD

[1]Florida International University College of Medicine, Miami, FL, United States

[2]Mayo Clinic College of Medicine and Science, Mayo Clinic Florida, 4500 San Pablo Road, Jacksonville, FL, United States

Corresponding Author:
Brent P Little, MD
Mayo Clinic College of Medicine and Science, Mayo Clinic Florida, 4500 San Pablo Road, Jacksonville, FL, United States

Abstract

Background: The noninvasive imaging examinations of mammography (MG), low-dose computed tomography (CT) for lung cancer screening (LCS), and CT colonography (CTC) play important roles in screening for the most common cancer types. Internet search data can be used to gauge public interest in screening techniques, assess common screening-related questions and concerns, and formulate public awareness strategies.

Objective: This study aims to compare historical Google search volumes for MG, LCS, and CTC and to determine the most common search topics.

Methods: Google Trends data were used to quantify relative Google search frequencies for these imaging screening modalities over the last 2 decades. A commercial search engine tracking product (keywordtool.io) was used to assess the content of related Google queries over the year from May 1, 2022, to April 30, 2023, and 2 authors used an iterative process to agree upon a list of thematic categories for these queries. Queries with at least 10 monthly instances were independently assigned to the most appropriate category by the 2 authors, with disagreements resolved by consensus.

Results: The mean 20-year relative search volume for MG was approximately 10-fold higher than for LCS and 25-fold higher than for CTC. Search volumes for LCS have trended upward since 2011. The most common topics of MG-related searches included nearby screening locations (60,850/253,810, 24%) and inquiries about procedural discomfort (28,970/253,810, 11%). Most common LCS-related searches included CT-specific inquiries (5380/11,150, 48%) or general inquiries (1790/11,150, 16%), use of artificial intelligence or deep learning (1210/11,150, 11%), and eligibility criteria (1020/11,150, 9%). For CTC, the most common searches were CT-specific inquiries (1800/5590, 32%) or procedural details (1380/5590, 25%).

Conclusions: Over the past 2 decades, Google search volumes have been significantly higher for MG than for either LCS or CTC, although search volumes for LCS have trended upward since 2011. Knowledge of public interest and queries related to imaging-based screening techniques may help guide public awareness efforts.

(JMIR Cancer 2025;11:e53328) doi:10.2196/53328

KEYWORDS

lung cancer; lung cancer screening; breast cancer; mammography; colon cancer; CT colonography; Google search; internet; Google Trends; imaging-based; cancer screening; search data; noninvasive; cancer; CT; online; public awareness; big data; analytics; patient education; screening uptake

Introduction

Worldwide, an estimated 20 million new cancer diagnoses and 9.7 million deaths occurred in 2022. The 3 most common types of newly diagnosed malignancies were lung cancer (12.4%), female breast cancer (11.6%), and colorectal cancer (9.6%). Lung cancer and colorectal cancer were the most common causes of cancer-related mortality, with breast cancer in fourth place after liver cancer [1]. Noninvasive imaging, such as mammography (MG), low-dose computed tomography (CT) for lung cancer screening (LCS), and CT colonography (CTC), plays important roles in the early detection of the most common cancer types and has demonstrated efficacy in reducing cancer-related and all-cause mortality rates [2,3]. Encouraging results from large screening trials in several countries [4-6] have

driven global interest in imaging-based lung, breast, and colorectal screening [7-12] and have prompted efforts to initiate and optimize screening worldwide [13-15].

A 2013 Pew Research Center survey found that 72% of adults had pursued web-based health information over the past year, with 77% performing an initial search using an internet search engine [16]. Analysis of internet search volumes for topics related to these imaging examinations may identify opportunities for improved patient outreach and education. Google search trends have been shown to correlate with both epidemiological data and public interest [17-19]. However, few studies have looked at general search volumes for each screening method [20-22], and to our knowledge no publication has documented a detailed topic-level analysis comparing all 3 types of image-based screening.

Methods

Overview

Google Trends was used to assess long-term variation in worldwide relative search volumes for the terms "mammography," "lung cancer screening," and "CT colonography" for the period January 1, 2004 (the earliest date for which Google Trends data were retrievable) to April 1, 2023. Google was chosen, as it is the most frequently used internet search engine, consistently capturing more than 80% of the worldwide internet search market [23].

Keyword tool (keywordtool.io), a commercial search engine tracking product, was used to query average volumes of monthly Google searches in English worldwide for the period May 1, 2022, to April 30, 2023 [24]. Keyword tool uses the output of the Google autocomplete tool to extract the most common entries at the search prompt, which made it optimal to use for this study design [24]. All questions and question-like queries entered at the Google search prompt relating to the terms "mammography," "lung cancer screening," and "CT colonography" were extracted. Two of the authors (ZDZ and BPL) used an iterative process to agree upon a list of thematic categories for queries. Search results with at least 10 monthly searches were independently assigned to the most appropriate category by the 2 graders, with disagreements resolved by consensus.

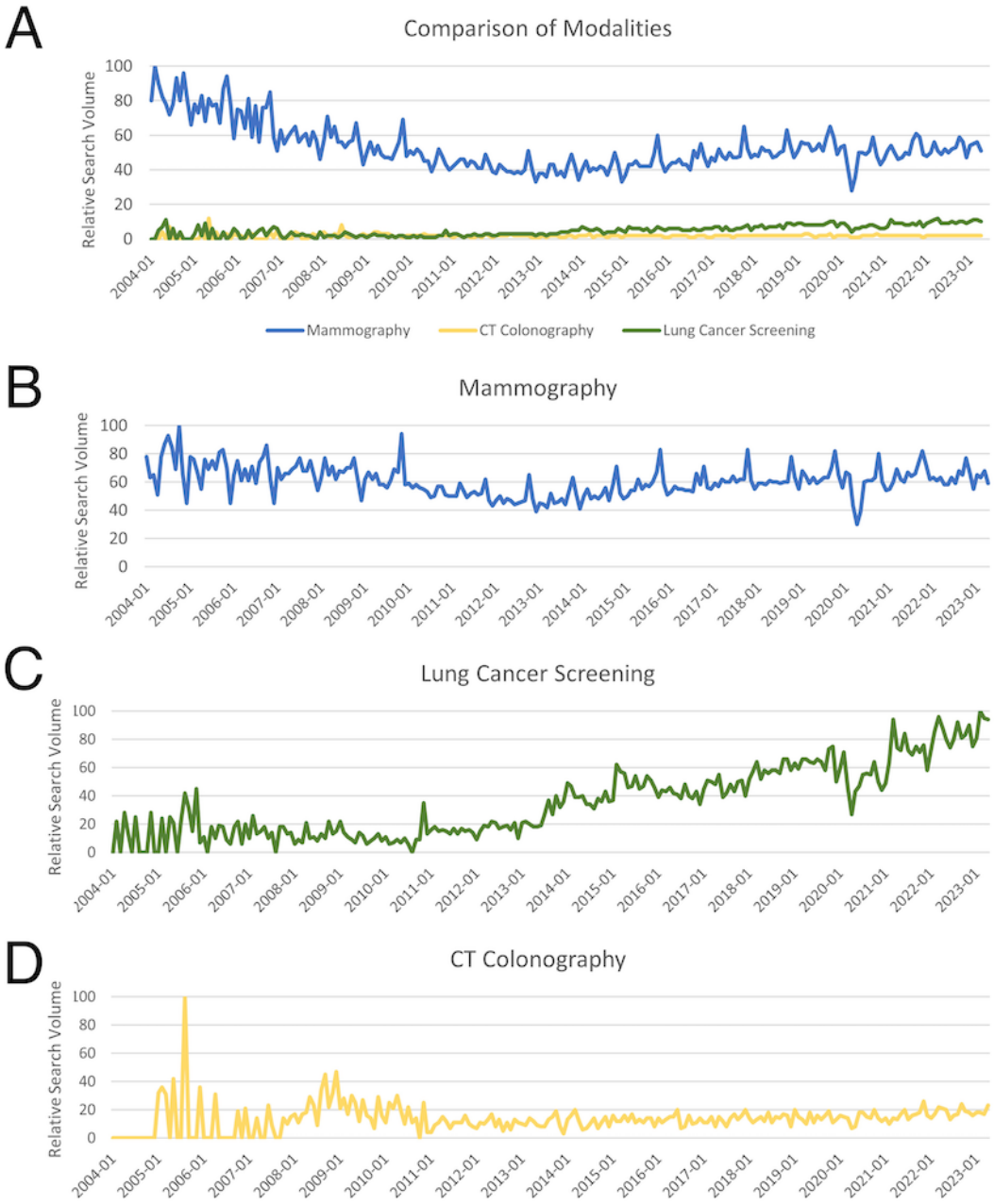
Data Analysis

Search volumes for each screening type were plotted as normalized values (relative search volumes comprising the format of Google Trends data output), with 100 representing the highest relative search volume and 1 representing the lowest, with 0 representing a search volume of zero or insufficient search volume to calculate. Average monthly search volumes for each category were grouped by screening modality and presented in tabular format as the number of searches and as a percentage of total searches for the corresponding modality.

Results

A plot of Google Trends data from 2004 to 2023 comparing MG, LCS, and CTC Google searches is shown in Figure 1. The median 20-year relative search volume was 50 (SD 13; IQR 43-57) for MG, 5 (SD 3; IQR 2-7) for LCS, and 2 (SD 1.2; IQR 1-2) for CTC. On average, the search frequency for LCS was only one-tenth of that for MG, and CTC search volume was only one–twenty-fifth of MG search volume. Search volumes for LCS steadily increased from 2011 to 2023, while searches for MG and CTC plateaued.

Figure 1. Relative worldwide monthly Google search volumes for imaging-based screening examinations. Graph A compares the relative average worldwide monthly search volumes for mammography, lung cancer screening, and CT colonography from January 1, 2004, to April 1, 2023. Graphs B, C, and D illustrate search volumes for the modalities taken independently. All data were obtained from Google Trends for worldwide searches in English. CT: computed tomography.



The relative search volumes for MG alone (Figure 1) across the period ranged from 30 to 100 (median 60; IQR 54-67); there was an overall decrease in relative search volume from the high in 2004 (100) to 2013 (39), followed by an overall increase from 2013 to 2019 (82), with the nadir in 2020 (30) corresponding to the onset of the COVID-19 pandemic. Relative search volumes for LCS alone ranged from 0 to 100; there was an overall decrease in relative search volumes from 2005 (90) to 2010 (0), and a subsequent uptrend after 2011, culminating in a high of 100 in 2019. Relative search volumes for CTC alone ranged from 0 to 100 with the highest volume in 2005, with a secondary peak in 2008 (47), and a plateau from 2010 to 2023 with an average of 14.

Topic analysis using Keyword tool data for the period from May 1, 2022, to April 30, 2023, showed that MG-related topics had the highest total monthly search volume (621,810 searches/month), followed by those related to LCS (23,250 searches/month), and CTC (17,690 searches/month). From these searches, a total of 751 queries classified as questions or question-like queries by the Keyword tool interface were extracted as follows: 442 for MG, 178 for LCS, and 131 for CTC. Of these, 332 had a search volume of >10 per month. The reviewers identified 39, 14, and 11 thematic categories for MG, LCS, and CTC, respectively. The most common categories of queries for MG were nearby screening locations (60,850/253,810, 24%), general inquiries (52,460/253,810, 21%), pain associated with screening (28,970/253,810, 11%), and eligibility criteria (ages) (16,160/253,810, 6%) (Table 1). For LCS, the most frequent categories of queries were CT-specific inquiries (5380/11,150, 48%), general inquiries (1790/11,150, 16%), artificial intelligence (AI) or deep learning use in lung screening (1210/11,150, 11%), screening eligibility criteria (ages or pack-years) (1020/11,150, 9%), and nearby screening locations (750/11,150, 7%) (Table 1). The most common categories of queries related to CTC were CT-specific inquiries (1800/5590, 32%), screening procedural details (1380/5590, 25%), performance compared with colonoscopy (870/5590, 16%), and screening preparation (such as colon preparation and sedation) (720/5590, 13%) (Table 1). Examples of the most common queries for each image-based screening are presented in Table 2.

Table . Average monthly search volumes for imaging-based screening examinations by search topic category. The table displays the average monthly Google search volumes for imaging-based screening examinations from May 1, 2022, to April 30, 2023. All data were obtained from Keyword tool (keywordtool.io) for worldwide searches performed in English. The top 10 thematic categories of search topics are listed for each modality.

Imaging-based cancer screening search query categories	Average monthly search volume, n (%)
MG[a] (n=253,810)	
Nearest screening locations	60,850 (24)
General inquiries	52,460 (21)
Pain associated with screening	28,970 (11)
Cancer imaging characteristics	28,890 (11)
Screening eligibility criteria	16,160 (6)
Screening procedural details	11,810 (5)
Comparison of screening modalities (MG vs MRI[b])	9120 (4)
General screening inquiries	8180 (3)
Opportunities for no-cost screening	6810 (3)
Breast tomosynthesis	6770 (3)
Other categories (combined)	23,790 (9)
LCS[c] (n=11,150)	
CT[d]-specific inquiries	5380 (48)
General inquiries	1790 (16)
Artificial intelligence and deep learning in LCS	1210 (11)
Screening eligibility criteria	1020 (9)
Nearest screening locations	750 (7)
Opportunities for no-cost screening	420 (4)
Screening procedural details	230 (2)
Insurance coverage	150 (1)
LCS trials	90 (0.8)
Imaging accuracy	40 (0.4)
Other categories (combined)	70 (0.6)
CTC[e] (n=5590)	
CT-specific inquiries	1800 (32)
Screening procedural details	1380 (25)
Imaging modality comparison (colonoscopy vs CTC)	870 (16)
Prescreening procedural preparation	720 (13)
Coding (*ICD*[f]/CPT[g])	230 (4)
Nearest screening locations	230 (4)
Pain associated with screening	220 (4)
Opportunities for no-cost screening	60 (1)
Diagnostic capabilities	60 (1)
Insurance coverage	10 (0.2)
Other categories (combined)	10 (0.2)

[a]MG: mammography.

[b]MRI: magnetic resonance imaging.

[c]LCS: lung cancer screening.

[d]CT: computed tomography.

[e]CTC: computed tomography colonography.

[f]*ICD*: *International Classification of Diseases*.

[g]CPT: current procedural terminology.

Table . The table shows examples of the 5 most common inquiries for each imaging-based cancer screening examination from May 1, 2022, to April 30, 2023. All data were obtained from Keyword tool (keywordtool.io) for worldwide searches performed in English.

	Example #1	Example #2	Example #3
Mammography			
Nearest screening locations	"mammography near me"	"mammogram without referral near me"	"mammogram near me now"
General inquires	"what is the mammography"	"why mammography is important"	"mammography versus mammogram"
Pain associated with screening	"is mammography painful"	"do mammograms hurt"	"what does a mammogram feel like"
Cancer imaging characteristics	"mammogramof breast cancer"	"mammography of breast cancer images"	"mammography of fibroadenoma"
Screening eligibility criteria	"mammography at what age"	"when should mammograms be done"	"mammography before 40"
Lung cancer screening			
CT[a]-specific inquiries	"lung cancer screening with low dose ct"	"low dose lung cancer screening"	"lung cancer detection using ct scan images"
General inquires	"lung cancer screening tests"	"what is a lung cancer screening"	"lung cancer screening for smokers"
Artificial intelligence and deep learning in LCS[b]	"lung cancer detection using image processing"	"lung cancer detection using deep learning"	"lung cancer detection using machine learning"
Screening eligibility criteria	"lung cancer screening ages"	"criteria for lung cancer screening"	"who should be screened for lung cancer"
Nearest screening locations	"lung cancer screening near me"	"mobile lung cancer screening near me"	"private lung cancer screening near me"
CT colonography			
CT-specific inquiries	"what is a ct colonography"	"ct colonography with contrast"	"ct colonography without contrast"
Screening procedural details	"ct colonography procedure"	"does ct colonography use contrast"	"how is ct colonography performed"
Imaging modality comparison (colonoscopy versus CTC[c])	"ct colonography versus colonoscopy cost"	"ct colonography versus colonoscopy sensitivity"	"is ct colonography as good as colonoscopy"
Prescreening procedural preparation	"what is the prep for a ct colonography"	"what is the prep for a ct colonography nhs"	—[d]
Coding (*ICD*[e]/CPT[f])	"ct colonography cpt code"	"ct colonography cpt"	—

[a]CT: computed tomography.

[b]LCS: lung cancer screening.

[c]CTC: computed tomography colonography.

[d]Not available.

[e]*ICD*: *International Classification of Diseases*.

[f]CPT: current procedural terminology.

Discussion

Principal Findings

Overview

Our study revealed much higher Google search volumes for MG topics over the past 2 decades than for LCS and CTC. The average search frequency for LCS was only approximately one-tenth of that for MG, and CTC search volume was only one–twenty-fifth of MG search volume. LCS average search volumes increased from 2011 to 2023, while searches for MG and CTC plateaued. MG-related topics had the highest total monthly search volume, followed by those related to LCS and CTC. Frequently searched topics varied across modalities and included nearby screening locations, procedural details or associated pain, and eligibility criteria.

MG consistently exhibited higher levels of search interest compared with LCS or CTC, possibly reflecting the longer history of MG and various longstanding initiatives focused on breast screening and women's health [25,26]. Despite the higher

absolute mortality for lung cancer than for breast cancer globally, search volumes for topics related to LCS have remained much lower than those for MG, although LCS search volumes have experienced an uptrend since 2011, when the National Lung Screening Trial was published [3]. MG was first established in 1913 and fully adopted in the 1970s [25], in contrast to low-dose CT for LCS, which gained widespread recognition in 2011. The early peak of interest in LCS in the 2004‐2006 period coincides with attention generated by early CT LCS programs, followed by a decline in interest until 2011 [27,28]. The relative search volumes for CTC have been persistently low over the last 2 decades, especially compared with MG (2%). While CTC is recognized by the national medical societies and formal guidelines of several countries as a viable screening option [29-31], other organizations have raised concerns regarding the strength of evidence supporting CTC or cost-effectiveness compared with colonoscopy [32,33]. The low search volumes might reflect a lack of awareness or desire for this screening option among the public.

Our topic analysis of recent volumes for queries related to image-based screening techniques showed both similarities and differences across imaging modalities. For example, in the case of MG, a substantial portion of searches (60,850/253,810, 24%) involved the nearest screening center, a topic with substantial although lower percentages of search volumes for LCS (750/11,150, 7%) and CTC (230/5590, 4%), highlighting a possible target of increased or more effective publicity. A relatively high volume of searches related to procedural aspects of CTC (1380/5590, 25%) may suggest a relative unfamiliarity with details of this specific imaging modality. A common topic of searches was potential procedural discomfort or pain in MG (28,970/253,810, 11%), with a lower percentage for CTC (220/5590, 1%), suggesting an opportunity for providers to better address procedural comfort that might otherwise decrease screening uptake. Cost or insurance coverage for screening was in the top 10 most-searched topics for all 3 screening modalities but comprised a minority of searches in terms of percentage; although cost and coverage pose concerns for some individuals, it is notable that several other topics, such as the nearest screening locations, procedural pain, and eligibility criteria, had on average 2- to 3-fold higher search volumes. AI and deep learning were common search topics for LCS (1210/11,150, 11%) but were not common queries for MG or CTC; this is somewhat surprising, as AI in the form of computer-aided detection is commonly used in MG.

Comparison to Prior Work

Previous research has examined internet search trends for cancer screening examinations. Snyder et al [20] examined Google search volume trends for cancer screening terms during the first stages of the COVID-19 pandemic, finding a temporary decline in searches for terms related to MG, LCS, colonoscopy, and pap smear. Rosenkrantz and Prabhu [22] performed a Google Trends analysis of the relative frequency of Google searches for MG, LCS, CTC, and prostate magnetic resonance imaging from 2004 to 2014, finding a slight progressive decline over the decade in searches for MG, a decrease from 2004 to 2010 in searches for LCS, followed by a persistent increase beginning in 2011, and an overall decade-long decline in searches for CTC; these findings are consistent with our analysis, although we show that searches for LCS continued to rise from 2014 to 2023 and that searches for MG and CTC have continued to plateau. Our analysis also has the advantage of comparing both the relative and absolute volumes of searches across modalities (MG, LCS, and CTC) instead of relying solely on relative search volumes, showing that searches related to MG greatly exceeded those related to LCS and CTC.

A variety of other methods have been used to assess public interest in and knowledge of cancer screening, including interviews, focus group discussions, questionnaires, and news coverage analysis [10,34,35]. Raju et al's [34] survey of LCS-eligible individuals who chose not to participate in screening found that 19% of individuals had concerns about the distance to the screening site, and 14% of individuals had concerns about insurance coverage, recalling some of the frequent search topics for LCS in our study. In a Google News analysis of news coverage of MG from 2006 to 2015 [21], the most frequently covered topics included screening MG controversies (29%), new breast imaging technology (23%), imaging of dense breasts (11%), and public screening initiatives (11%), topics that were not among the most common MG searches in our study.

Strengths and Limitations

To our knowledge, our study is the first to use internet search engine data to gauge both general and topic-level public interest across imaging-based screening modalities by providing both relative search frequencies and estimates of absolute search volumes. We examined 2 decades of Google search volumes across 3 key imaging-based cancer screening examinations and documented the most common themes of recent related search queries.

Our study had several limitations. First, there are limitations inherent in any method of search volume estimation. Google Trends reports search data in relative terms and may report a value of 0 when search frequencies are low. For topic analysis, we captured data representing a snapshot of queries over a year, while queries may change over time. Search volumes were also estimated through a proprietary tool that uses Google autocomplete as a proxy for search volumes, as actual absolute search volumes are not available from Google Trends. However, our goal was not to measure exact search volumes but rather to analyze and compare trends across modalities and to discern the topics of the most frequent searches. Second, the analysis of thematic categories is intrinsically subjective; we attempted to mitigate this by using 2 observers, with final decisions rendered by consensus. Third, we analyzed worldwide Google searches in English and did not perform a country or region-specific analysis; while such comparative analyses are potentially of interest because of differences in screening guidelines and behavior across countries, our main objectives were to provide a general and comprehensive analysis of search interest across modalities and to determine the main topics of search queries. Future studies may incorporate comparisons of search queries across countries. Finally, internet search volumes may not directly translate to individual concerns, real-world behaviors, or screening uptake; however, individuals frequently

turn to the internet as a source of medical information [16], and internet searches have often been used as a proxy of public interest in a wide variety of health-related topics [17-20,22].

Conclusions

In conclusion, MG has generated consistently higher Google search volumes over the past 2 decades than LCS and CTC, but search interest in LCS has been on an upward trend since 2011. Frequently searched topics varied across modalities and included nearby screening locations, procedural details or associated pain, and eligibility criteria. These search trends might inform the development of communication strategies related to screening and aid in addressing frequently asked questions from the public.

Acknowledgments

Generative artificial intelligence was not used for any portion of the study or for any portion of manuscript writing.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the proprietary nature of one of the search analysis tools used but are available from the corresponding author on reasonable request.

Authors' Contributions

ZDZ and BPL contributed to data curation, formal analysis, and methodology. ZDZ, IOC, RAG, EMJ, MRM, PJM, SKS, JTS, and BPL were involved in conceptualization, drafting the original manuscript, and reviewing and editing the final version.

Conflicts of Interest

None declared.

References

1. Global cancer burden growing, amidst mounting need for services. World Health Organization. URL: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services [accessed 2024-01-07]
2. Myers ER, Moorman P, Gierisch JM, et al. Benefits and harms of breast cancer screening: a systematic review. JAMA 2015 Oct 20;314(15):1615-1634. [doi: 10.1001/jama.2015.13183] [Medline: 26501537]
3. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011 Aug 4;365(5):395-409. [doi: 10.1056/NEJMoa1102873] [Medline: 21714641]
4. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. N Engl J Med 2020 Feb 6;382(6):503-513. [doi: 10.1056/NEJMoa1911793] [Medline: 31995683]
5. Nyström L, Bjurstam N, Jonsson H, Zackrisson S, Frisell J. Reduced breast cancer mortality after 20+ years of follow-up in the Swedish randomized controlled mammography trials in Malmö, Stockholm, and Göteborg. J Med Screen 2017 Mar;24(1):34-42. [doi: 10.1177/0969141316648987] [Medline: 27306511]
6. Field JK, Vulkan D, Davies MPA, et al. Lung cancer mortality reduction by LDCT screening: UKLS randomised trial results and international meta-analysis. Lancet Reg Health Eur 2021 Nov;10:100179. [doi: 10.1016/j.lanepe.2021.100179] [Medline: 34806061]
7. McInnerney D, Simmonds I, Hancock N, et al. Yorkshire Lung Screening Trial (YLST) pathway navigation study: a protocol for a nested randomised controlled trial to evaluate the effect of a pathway navigation intervention on lung cancer screening uptake. BMJ Open 2024 Jul 9;14(7):e084577. [doi: 10.1136/bmjopen-2024-084577] [Medline: 38986555]
8. Tao W, Yu X, Shao J, Li R, Li W. Telemedicine-enhanced lung cancer screening using mobile computed tomography unit with remote artificial intelligence assistance in underserved communities: initial results of a population cohort study in Western China. Telemed J E Health 2024 Jun;30(6):e1695-e1704. [doi: 10.1089/tmj.2023.0648] [Medline: 38436233]
9. de-Torres JP, Alcaide AB, Campo A, et al. Lung cancer screening in people with COPD: the Pamplona-IELCAP experience. Arch Bronconeumol 2024 Feb;60(2):95-100. [doi: 10.1016/j.arbres.2023.12.012]
10. Ren W, Chen M, Qiao Y, Zhao F. Global guidelines for breast cancer screening: a systematic review. Breast 2022 Aug;64:85-99. [doi: 10.1016/j.breast.2022.04.003] [Medline: 35636342]
11. Caruso D, Polici M, Bellini D, Laghi A. ESR essentials: imaging in colorectal cancer-practice recommendations by ESGAR. Eur Radiol 2024 Sep;34(9):5903-5910. [doi: 10.1007/s00330-024-10645-3] [Medline: 38418627]
12. Smyth AE, Healy CF, Aherne EA, MacMathuna P, Morrin MM, Fenlon HM. National survey of CT colonography practice in Ireland. Ir Med J 2016 Jun 10;109(6):419. [Medline: 27814436]
13. Lam S, Bai C, Baldwin DR, et al. Current and future perspectives on computed tomography screening for lung cancer: a roadmap from 2023 to 2027 from the international association for the study of lung cancer. J Thorac Oncol 2024 Jan;19(1):36-51. [doi: 10.1016/j.jtho.2023.07.019] [Medline: 37487906]

14. Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. Nat Rev Clin Oncol 2023 Sep;20(9):624-639. [doi: 10.1038/s41571-023-00798-3] [Medline: 37479810]
15. Tavakoli B, Feizi A, Zamani-Alavijeh F, Shahnazi H. Factors influencing breast cancer screening practices among women worldwide: a systematic review of observational and qualitative studies. BMC Womens Health 2024 Apr 27;24(1):268. [doi: 10.1186/s12905-024-03096-x] [Medline: 38678224]
16. Health online 2013. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013 [accessed 2023-07-05]
17. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206. [doi: 10.2196/16206] [Medline: 32310818]
18. Watad A, Watad S, Mahroum N, et al. Forecasting the West Nile virus in the United States: an extensive novel data streams-based time series analysis and structural equation modeling of related digital searching behavior. JMIR Public Health Surveill 2019 Feb 28;5(1):e9176. [doi: 10.2196/publichealth.9176] [Medline: 30601755]
19. Gianfredi V, Bragazzi NL, Mahamid M, et al. Monitoring public interest toward pertussis outbreaks: an extensive Google Trends-based analysis. Public Health 2018 Dec;165:9-15. [doi: 10.1016/j.puhe.2018.09.001] [Medline: 30342281]
20. Snyder A, Jang S, Nazari IS, et al. Google search volume trends for cancer screening terms during the COVID-19 pandemic. J Med Screen 2021 Jun;28(2):210-212. [doi: 10.1177/0969141321999426] [Medline: 33663240]
21. Young Lin LL, Rosenkrantz AB. The U.S. online news coverage of mammography based on a Google News search. Acad Radiol 2017 Dec;24(12):1612-1615. [doi: 10.1016/j.acra.2017.05.011] [Medline: 28693758]
22. Rosenkrantz AB, Prabhu V. Public interest in imaging-based cancer screening examinations in the United States: analysis using a web-based search tool. AJR Am J Roentgenol 2016 Jan;206(1):113-118. [doi: 10.2214/AJR.15.14840] [Medline: 26700342]
23. Bianchi T. Market share of leading desktop search engines worldwide from january 2015 to january 2025. Statista. URL: https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines [accessed 2024-04-22]
24. Find great keywords using google autocomplete. Keyword Tool. URL: https://keywordtool.io [accessed 2023-05-05]
25. Gold RH, Bassett LW, Widoff BE. Highlights from the history of mammography. Radiographics 1990 Nov;10(6):1111-1131. [doi: 10.1148/radiographics.10.6.2259767] [Medline: 2259767]
26. O'Mahony M, Comber H, Fitzgerald T, et al. Interventions for raising breast cancer awareness in women. Cochrane Database Syst Rev 2017 Feb 10;2(2):CD011396. [doi: 10.1002/14651858.CD011396.pub2] [Medline: 28185268]
27. Nawa T, Nakagawa T, Kusano S, Kawasaki Y, Sugawara Y, Nakata H. Lung cancer screening using low-dose spiral CT: results of baseline and 1-year follow-up studies. Chest 2002 Jul;122(1):15-20. [doi: 10.1378/chest.122.1.15] [Medline: 12114333]
28. Wilson DO, Weissfeld JL, Fuhrman CR, et al. The Pittsburgh Lung Screening Study (PLuSS): outcomes within 3 years of a first computed tomography scan. Am J Respir Crit Care Med 2008 Nov 1;178(9):956-961. [doi: 10.1164/rccm.200802-336OC] [Medline: 18635890]
29. Wolf AMD, Fontham ETH, Church TR, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin 2018 Jul;68(4):250-281. [doi: 10.3322/caac.21457] [Medline: 29846947]
30. Spada C, Hassan C, Bellini D, et al. Imaging alternatives to colonoscopy: CT colonography and colon capsule. European Society of Gastrointestinal Endoscopy (ESGE) and European Society of Gastrointestinal and Abdominal Radiology (ESGAR) guideline - update 2020. Eur Radiol 2021 May;31(5):2967-2982. [doi: 10.1007/s00330-020-07413-4] [Medline: 33104846]
31. Moreno C, Kim DH, Bartel TB, et al. ACR appropriateness criteria ® colorectal cancer screening. J Am Coll Radiol 2018 May;15(5):S56-S68. [doi: 10.1016/j.jacr.2018.03.014]
32. Lin JS, Perdue LA, Henrikson NB, Bean SI, Blasi PR. Screening for colorectal cancer: updated evidence report and systematic review for the US preventive services task force. JAMA 2021 May 18;325(19):1978-1998. [doi: 10.1001/jama.2021.4417] [Medline: 34003220]
33. Kriza C, Emmert M, Wahlster P, Niederländer C, Kolominsky-Rabas P. An international review of the main cost-effectiveness drivers of virtual colonography versus conventional colonoscopy for colorectal cancer screening: is the tide changing due to adherence? Eur J Radiol 2013 Nov;82(11):e629-e636. [doi: 10.1016/j.ejrad.2013.07.019] [Medline: 23938237]
34. Raju S, Khawaja A, Han X, Wang X, Mazzone PJ. Lung cancer screening: characteristics of nonparticipants and potential screening barriers. Clin Lung Cancer 2020 Sep;21(5):e329-e336. [doi: 10.1016/j.cllc.2019.11.016] [Medline: 32371161]
35. Pal A, Taneja N, Malhotra N, et al. Knowledge, attitude, and practice towards breast cancer and its screening among women in India: a systematic review. J Cancer Res Ther 2021;17(6):1314-1321. [doi: 10.4103/jcrt.JCRT_922_20] [Medline: 34916359]

Abbreviations:

AI: artificial intelligence
CT: computed tomography
CTC: computed tomography colonography

LCS: lung cancer screening
MG: mammography

Edited by N Cahill; submitted 03.10.23; peer-reviewed by S Lampridis, VM Tiryaki; revised version received 23.01.25; accepted 30.01.25; published 11.03.25.

Please cite as:
Zippi ZD, Cortopassi IO, Grage RA, Johnson EM, McCann MR, Mergo PJ, Sonavane SK, Stowell JT, Little BP
Assessing Public Interest in Mammography, Computed Tomography Lung Cancer Screening, and Computed Tomography Colonography Screening Examinations Using Internet Search Data: Cross-Sectional Study
JMIR Cancer 2025;11:e53328
URL: https://cancer.jmir.org/2025/1/e53328
doi:10.2196/53328

Spatiotemporal Correlation Analysis for the Incidence of Esophageal and Gastric Cancer From 2010 to 2019: Ecological Study

Zixuan Cui[1], BMed; Chen Suo[1,2,3*], PhD; Yidan Zhao[1], BMed; Shuo Wang[1], BMed; Ming Zhao[1,2], BMed; Ruilin Chen[1], BMed; Linyao Lu[3,4], BMed; Tiejun Zhang[1,2,3,5], PhD; Xingdong Chen[3,4,5,6*], PhD

[1]Department of Epidemiology, School of Public Health, Fudan University, Dongan Road 130, Shanghai, China

[2]Shanghai Institute of Infectious Disease and Biosecurity, Shanghai, China

[3]Fudan University Taizhou Institute of Health Sciences, Jiangsu, China

[4]State Key Laboratory of Genetic Engineering, Human Phenome Institute, Zhangjiang Fudan International Innovation Center, Shanghai, China

[5]Yiwu Research Institute of Fudan University, Zhejiang, China

[6]National Clinical Research Center for Aging and Medicine, Huashan Hospital, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Chen Suo, PhD
Department of Epidemiology, School of Public Health, Fudan University, Dongan Road 130, Shanghai, China

Abstract

Background: Esophageal and gastric cancer were among the top 10 most common cancers worldwide. In addition, sex-specific differences were observed in the incidence. Due to their anatomic proximity, the 2 cancers have both different but also shared risk factors and epidemiological features. Exploring the potential correlated incidence pattern of them, holds significant importance in providing clues in the etiology and preventive strategies.

Objective: This study aims to explore the spatiotemporal correlation between the incidence patterns of esophageal and gastric cancer in 204 countries and territories from 2010 to 2019 so that prevention and control strategies can be more effective.

Methods: The data of esophageal and gastric cancer were sourced from the Global Burden of Disease (GBD). Spatial autocorrelation analysis using Moran I in ArcGIS 10.8 (Esri) was performed to determine spatial clustering of each cancer incidence. We classified different risk areas based on the risk ratio (RR) of the 2 cancers in various countries to the global, and the correlation between their RR was evaluated using Pearson correlation coefficient. Temporal trends were quantified by calculating the average annual percent change (AAPC), and the correlation between the temporal trends of both cancers was evaluated using Pearson correlation coefficients.

Results: In 2019, among 204 countries and territories, the age-standardized incidence rates (ASIR) of esophageal cancer ranged from 0.91 (95% CI 0.65-1.58) to 24.53 (95% CI 18.74-32.51), and the ASIR of gastric cancer ranged from 3.28 (95% CI 2.67-3.91) to 43.70 (95% CI 34.29-55.10). Malawi was identified as the highest risk for esophageal cancer (male RR=3.27; female RR=5.19) and low risk for gastric cancer (male RR=0.21; female RR=0.23) in both sexes. Spatial autocorrelation analysis revealed significant spatial clustering of the incidence for both cancers (Moran I>0.20 and P<.001). A positive correlation between the risk of esophageal and gastric cancer was observed in males (r=0.25, P<.001). The ASIR of both cancers showed a decreasing trend globally. The ASIR for esophageal and gastric cancer showed an AAPC of −1.43 (95% CI −1.58 to −1.27) and −1.76 (95% CI −2.08 to −1.43) in males, and −1.93 (95% CI −2.11 to −1.75) and −1.79 (95% CI −2.13 to −1.46) in females. In addition, a positive correlation between the temporal trends in ASIR for both cancers was observed at the global level across sexes (male r=0.98; female r=0.98).

Conclusions: Our study shows that there was a significant spatial clustering of the incidence for esophageal and gastric cancer and a positive correlation between the risk of both cancers across countries was observed in males. In addition, a codescending incidence trend between both cancers was observed at the global level.

(JMIR Cancer 2025;11:e66655) doi:10.2196/66655

KEYWORDS

spatiotemporal analysis; spatiotemporal correlation; esophageal cancer; gastric cancer; cancer; global burden of disease; GBD; average annual percentage change; incidence; epidemiology

Introduction

Globally, cancer is a leading cause of mortality, with its significance continually rising [1-4]. It was estimated that in 2019, cancer accounted for 250 million disability-adjusted life years [3]. Concurrently, according to the Global Cancer Statistics 2020, esophageal and gastric cancer were among the top 10 most common cancers worldwide [5]. In addition, sex-specific differences were observed in the incidence [5]. Due to their anatomic proximity, esophageal and gastric cancer have both different but also a number of shared risk factors and epidemiological features [6-8]. Several published articles have demonstrated that frequent consumption of hot beverages and poor oral hygiene are risk factors for esophageal cancer [9,10], while a high-sodium diet and *Helicobacter pylori* infection are risk factors for gastric cancer [11,12]. Smoking and heavy alcohol consumption are associated with the incidence of both esophageal and gastric cancer [13,14]. There are also some pathological injuries involving both esophagus and stomach, such as gastroesophageal reflux disease and Barrett esophagus, which is related to reflux [15,16]. Furthermore, the high-incidence regions for esophageal cancer are primarily distributed in East Asia, Southeastern Africa, and Northern Europe, while the high-incidence regions for gastric cancer are mainly distributed in East Asia and Eastern Europe [5]. The spatial distribution of these 2 cancers exhibits certain overlapping patterns. However, there has been no global-level spatiotemporal correlation analysis simultaneously examining the incidence of both esophageal and gastric cancer [17-22]. Therefore, systematic exploring the spatiotemporal correlation in the incidence of esophageal and gastric cancer holds significant importance.

Through the spatiotemporal correlation analysis of global incidence of esophageal and gastric cancer, we aim to provide insightful information that would contribute to the prevention and control of these cancers worldwide and the rational allocation of global public health resources.

Methods

Data Sources

The Global Burden of Diseases, Injuries, and Risk Factors Study 2019 (GBD 2019), coordinated by the Institute for Health Metrics and Evaluation (IHME) offers comprehensive and comparable data on the epidemiological burden of esophageal and gastric cancer, which includes age-standardized incidence rates (ASIR) recorded annually spanning from 1990 to 2019 [4]. The data of esophageal and gastric cancer were sourced from the Global Health Data Exchange (GHDx) query tool [23]. The GHDx query tool in the GBD 2019 database includes data from 204 countries and territories by sex from 1990 to 2019. Furthermore, according to the geographic location, the world was divided into 21 regions such as East Asia in GBD 2019 database. The quality and integrity of data can be significantly influenced by advancements in medical records and data collection technologies. Therefore, we extracted data on the ASIR (per 100,000 person-years) with 95% uncertainty interval (UI) of esophageal and gastric cancer from 2010 to 2019 due to the higher quality and integrity of recent data in our study. The ASIR was calculated by the direct method. Standardization was crucial in this study as it eliminates the bias when comparing rates. The 95% UI is a range of values that reflects the certainty of an estimate. The sociodemographic index (SDI) is a composite indicator developed by GBD researchers to characterize the developmental status of countries and territories, and it is closely associated with health outcomes. It is the geometric mean of 0 to 1 indices of total fertility rate under the age of 25, mean education for those ages 15 and older, and lag distributed income per capita [24]. The IHME provides publicly available SDI Reference Quintiles and the SDI values for all locations from 1950 to 2020 [24]. Based on the provided SDI Reference Quintiles, all locations can be classified into 5 categories, including low SDI (SDI≤0.46), low-middle SDI (0.46<SDI≤0.61), middle SDI (0.61<SDI≤0.69), high-middle SDI (0.69<SDI≤0.81), and high SDI (SDI>0.81).

Statistical Analysis

Spatial Distribution and Correlation of Esophageal and Gastric Cancer

In the analysis of spatial distribution and correlation of the ASIR of esophageal and gastric cancer, we used ArcGIS 10.8 (Esri) to calculate Global Moran I [25,26], which was used to evaluate whether the incidence of each cancer exhibited spatial clustering. Global Moran I measure spatial autocorrelation based on both feature locations and feature values simultaneously, where the feature here refers to different countries and territories. The tool calculates the Moran I value and both a zI-score and *P* value to evaluate the significance of that Index. The Moran I statistic for spatial autocorrelation was given as Multimedia Appendix 1.

Moran I values range between −1 and +1. For the Global Moran I statistic, the null hypothesis states that the attribute being analyzed is randomly distributed among the features in the study area. When the *P* value is statistically significant, we may reject the null hypothesis, and a positive value indicates the spatial distribution of high values and low values in the dataset is more spatially clustered than would be expected if underlying spatial processes were random; a negative value indicates the spatial distribution of high values and low values in the dataset is more spatially dispersed than would be expected if underlying spatial processes were random. A dispersed spatial pattern often reflects some type of competitive process—a feature with a high value repels other features with high values, and it is similar in a feature with a low value. When the *P* value is not statistically significant, we cannot reject the null hypothesis, indicating a random spatial distribution. In spatial autocorrelation analysis, the conceptualization of spatial relationships adheres to the Inverse_Distance rule.

In addition, by comparing the ASIR levels of esophageal and gastric cancer in each country and territory against those of the global ASIR levels, we obtained the corresponding risk ratio (RR) to quantify the level of relative risk. Countries and territories with an RR value≤0.50 were classified as low-risk area, while an 0.50<RR value<2.00 and RR value≥2.00 were classified as medium-risk area and high-risk area, respectively.

We then used the Pearson correlation coefficient to evaluate the correlation between the RR of esophageal and gastric cancer in both sex groups. We also evaluated the relationship between SDI quintiles and the distribution of different risk areas.

Temporal Trend and Correlation of Esophageal and Gastric Cancer

In the analysis of temporal trends and correlation of the ASIR of esophageal and gastric cancer, we used the National Cancer Institute Joinpoint Regression Program (version 5.0.2) to calculate the average annual percent change (AAPC) of the ASIR for esophageal and gastric cancer across 204 countries and territories during 2010 - 2019. The CI was set at 95%. AAPC is a statistical indicator used to describe the average annual change rate of the ASIR over a certain period. It provides a comprehensive assessment of the overall trend of ASIR over the entire study period by weighting the annual percentage change over multiple time periods. If both the estimation of AAPC and its lower boundary of 95% CI were >0, the ASIR was considered to be in an increasing trend; if both the estimation of AAPC and its upper boundary of 95% CI were <0, the ASIR was considered to be in a decreasing trend. Otherwise, the ASIR was considered to be stable over time. In the correlation analysis, the Pearson correlation coefficient was used to evaluate the correlation between the temporal trends in the ASIR of esophageal and gastric cancer. A Pearson correlation coefficient >0 with a *P* value less than the specified significance level indicated a significant positive correlation, while a coefficient <0 with a *P* value less than the specified significance level indicated a significant negative correlation. In addition, we assessed the correlation between the temporal trends in the ASIR of both cancers across SDI quintiles.

Given the significant sex differences in the incidence of esophageal and gastric cancer [27,28] and the lifestyle disparities, all analyses were stratified by sex. ArcGIS (version 10.8) and Joinpoint Regression Program (version 5.0.2) were used for spatial autocorrelation analysis and temporal trend analysis of ASIR respectively. All data analyses were conducted in software R (version 4.3.2; R Foundation for Statistical Computing) and R Studio (Posit). A *P* value of less than .05 was considered statistically significant.

Ethical Considerations

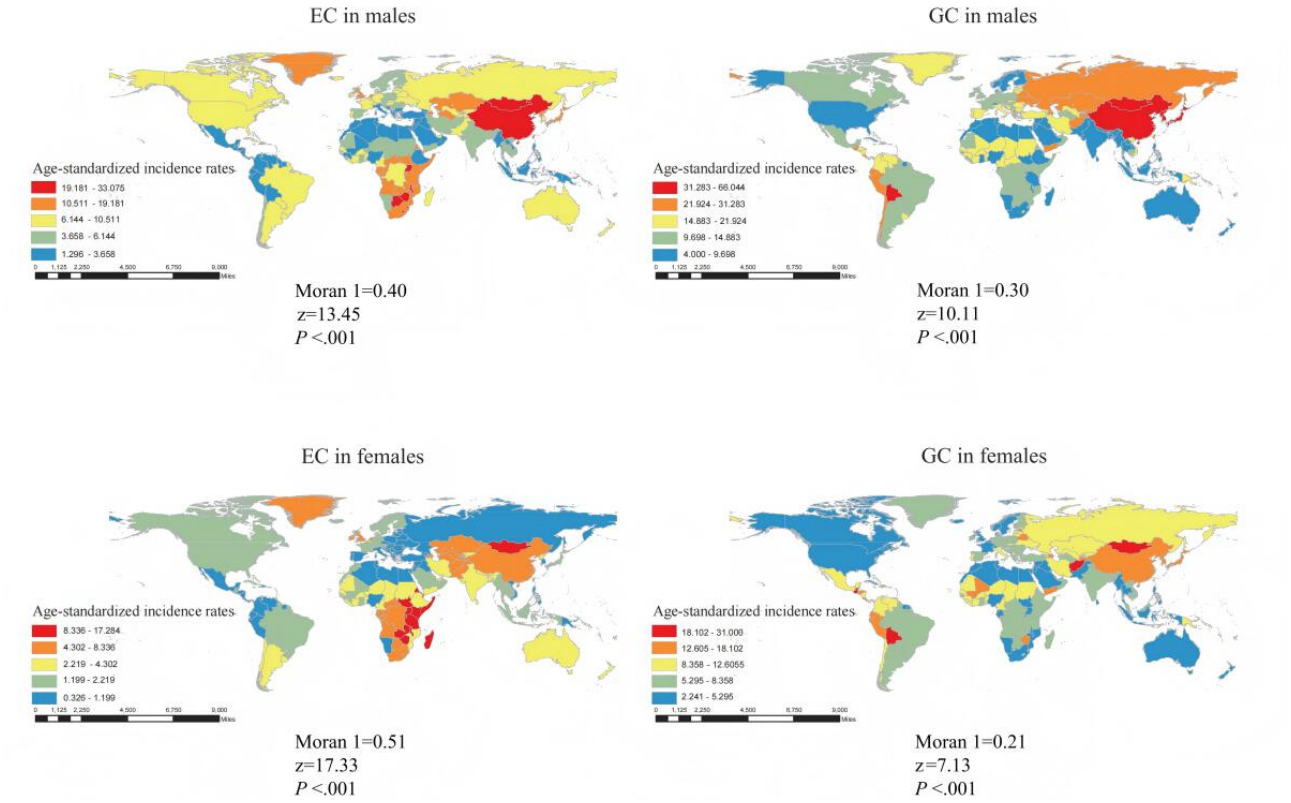
This study did not involve human participants and animals. Ethics approval was not applicable for this study, as this study used existing good quality data that were aggregated at the population level. Data available for download on IHME websites are publicly available and can be used, shared, modified or built upon by noncommercial users in accordance with the IHME Free-of-Charge Non-Commercial User Agreement [29].

Results

Spatial Distribution and Correlation Between the Esophageal and Gastric Cancer Incidence in 2019

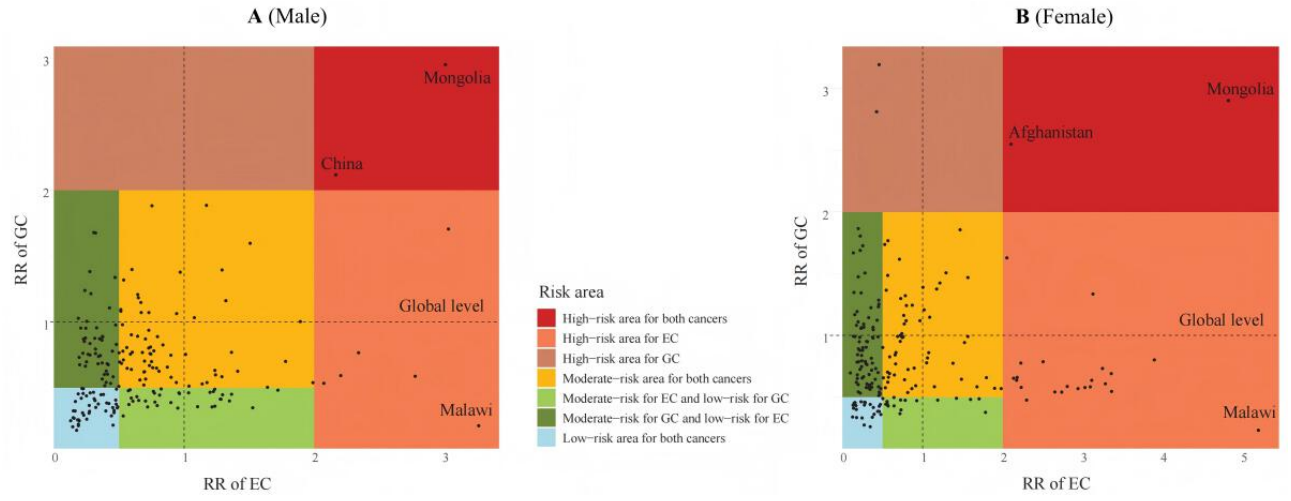
The study included data on the incidence of esophageal and gastric cancer from 204 countries and territories and 21 geographic regions. In 2019, the global ASIR of esophageal and gastric cancer were 6.51 (95% CI 5.69-7.25) and 15.59 (95% CI 14.11-17.15) per 100,000 person-years, respectively, with considerable variation observed across regions and sex groups. Among the 204 countries and territories provided by the GBD dataset, the ASIR of esophageal cancer ranged from a minimum of 0.91 (95% CI 0.65-1.58) in Nigeria to a maximum of 24.53 (95% CI 18.74-32.51) in Malawi, while the ASIR of gastric cancer ranged from a minimum of 3.28 (95% CI 2.67-3.91) in Malawi to a maximum of 43.70 (95% CI 34.29-55.10) in Mongolia. Overall, the ASIR of esophageal and gastric cancer was significantly higher in males than in females, and within the sex group, the ASIR of gastric cancer was significantly higher than that of esophageal cancer at the global level (Figure 1). In 2019, Malawi had the highest ASIR of esophageal cancer in both sexes (male: 33.08, 95% CI 24.44-44.43; female: 17.28, 95% CI 12.64-23.50), and Mongolia was ranked the top 5 for both esophageal (male: 30.48, 95% CI 22.58-37.90; female: 16.03, 95% CI 7.50-21.37) and gastric cancer (male: 66.04, 95% CI 51.50-82.68; female: 28.18, 95% CI 21.53-36.56) in both sexes (Table S1 in Multimedia Appendix 2). In a broader scale, among males, the ASIR of esophageal and gastric cancer in East Asia, Central Asia and high-income Asia Pacific all ranked within the top 6 among 21 geographic regions. Among males, East Asia had the highest ASIR for both esophageal (21.70, 95% CI 16.37-26.61) and gastric cancer (46.67, 95% CI 37.63-56.82), while High-income Asia Pacific ranked fifth for esophageal cancer (10.68, 95% CI 8.87-12.89) and second for gastric cancer (41.91, 95% CI 35.58-49.40), and Central Asia ranked sixth (9.51, 95% CI 8.37-11.53) and fifth (24.08, 95% CI 21.82-26.57), respectively. Similarly, in females, the regions with the top 6 ASIR for esophageal and gastric cancer were East Asia and Central Asia. East Asia ranked second for esophageal cancer (6.67, 95% CI 4.39-8.38) and third for gastric cancer (15.65, 95% CI 12.80-19.05), while Central Asia ranked fifth (4.70, 95% CI 4.19-5.29) and sixth (10.76, 95% CI 9.81-11.87), respectively (Figures S1 and S2 in Multimedia Appendix 2). The spatial autocorrelation analysis also showed that there was a significant clustering phenomenon in the spatial distribution of the incidence of esophageal and gastric cancer in both sex groups (all Moran I>0.20 and *P* value <.001; Figure 1).

Figure 1. ASIR of esophageal cancer (EC) and gastric cancer (GC) in 204 countries and territories in 2019, by sex. EC: esophageal cancer; GC: gastric cancer; ASIR: age-standardized incidence rate.



The 2019 risk area classification results were shown in Figure 2. In 2019, China and Mongolia belonged to the high-risk area for both esophageal and gastric cancer in males (Figure 2), while Afghanistan and Mongolia belonged to the high-risk area for both cancers in the females (Figure 2). Mongolia was identified as an overlapped high-risk area for both esophageal cancer and gastric cancer in males and females. In addition, the number of countries and territories belonging to the high-risk area for esophageal cancer alone was significantly higher in females than in males, and the number of countries and territories belonging to the low-risk area for both esophageal and gastric cancer was much lower in females than in males. Furthermore, no country or territory in males belonged to the high-risk area for gastric cancer alone, while in females, there were 2 countries (Guatemala and Bolivia). It is noteworthy that Malawi had the highest RR value for esophageal cancer (male RR=3.27; female RR=5.19) and a low RR value for gastric cancer (male RR=0.21; female RR=0.23) in both males and females. We further conducted Pearson correlation analysis for the RR values of esophageal and gastric cancer, and found that there was a positive correlation between the RR values of esophageal and gastric cancer in males (r=0.25; P<.001), but no significant correlation was found in females (r=0.02; P=.77).

Figure 2. Risk area classification map of esophageal cancer (EC) and GC (gastric cancer) for 204 countries and territories in 2019, by sex. A. Risk area classification map of EC and GC for 204 countries and territories in males. B. Risk area classification map of EC and GC for 204 countries and territories in females. EC: esophageal cancer; GC: gastric cancer; ASIR: age-standardized incidence rates; RR: risk ratio; obtained by comparing the ASIR levels of esophageal and gastric cancer in each country and territory against those of the global ASIR levels.



Temporal Trend and Correlation Between Esophageal and Gastric Cancer Incidence From 2010 to 2019

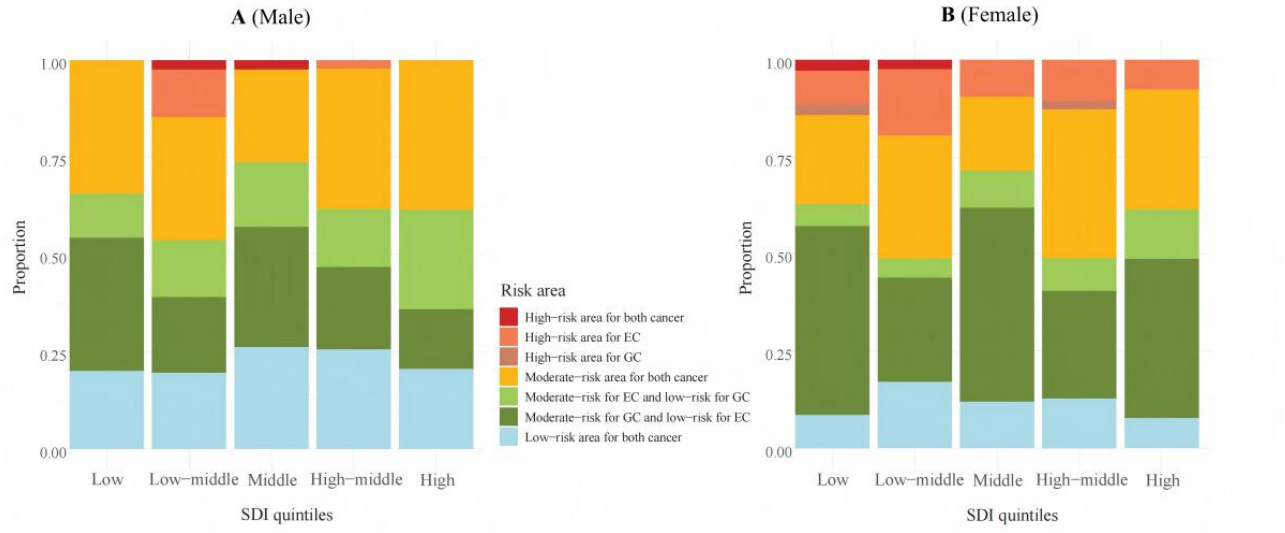
From 2010 to 2019, at the global level, the ASIR of esophageal cancer ranged from a peak of 7.50 (95% CI 6.10-8.02) in 2010 to a lowest point of 6.44 (95% CI 5.79-7.00) in 2017, while the ASIR of gastric cancer ranged from a peak of 18.25 (95% CI 17.14-19.22) in 2010 to a lowest point of 15.59 (95% CI 14.11-17.15) in 2019. And the ASIR of esophageal and gastric cancer decreased significantly in both sex groups globally. The AAPC for esophageal and gastric cancer in males was −1.43 (95% CI −1.58 to −1.27) and −1.76 (95% CI −2.08 to −1.43), respectively, while in females, the AAPC for esophageal and gastric cancer was −1.93 (95% CI −2.11 to −1.75) and −1.79 (95% CI −2.13 to −1.46), respectively. We selected all the high-risk countries and territories which had the top 5 ASIR of esophageal or gastric cancer in 2019 in both sex groups (14 countries and territories were finally presented in the table due to partial overlap; Table S2 in Multimedia Appendix 3). In males, the ASIR for esophageal and gastric cancer both showed a decreasing trend in 10 countries and territories, while in females, this trend was observed in 11 countries and territories. Furthermore, in all countries and territories except Bolivia, there was a strong positive correlation between the temporal trends in the ASIR of esophageal and gastric cancer (Pearson correlation coefficient>0.5). In addition, it is noteworthy that despite Cabo Verde being among the top 5 countries for ASIR of esophageal and gastric cancer in males, the ASIR for these cancers continued to show a significant increasing trend in males, with an increasing trend also observed in females.

We also performed Pearson correlation analyses for the temporal trends between esophageal and gastric cancer ASIR at the global level and across all 204 countries and territories. At the global level, the Pearson correlation coefficient (r) for the temporal trends in the ASIR for both cancers were 0.98 in males and 0.98 in females. The results also showed that a positive correlation between temporal trends in the ASIR of the 2 cancers was observed in most countries and territories, regardless of sex (Figure 3 in Multimedia Appendix 3). A normality test on the distribution of Pearson correlation coefficients across 204 countries and territories for both males and females indicated that the distributions were not normal. Wilcoxon rank-sum test results showed that there were differences (P<.001) in the distribution of Pearson correlation coefficient data between males and females in these 204 countries and territories, and the median Pearson correlation coefficient was higher for females than for males.

The Correlation Between SDI and Both Risk Areas and Temporal Trend

Based on the SDI Reference Quintiles publicly provided by the IHME and the SDI values for all locations since 1950 [24], we collected the SDI for 204 countries and territories in 2019, and classified these countries and territories into different SDI quintiles according to the SDI Reference Quintiles. The composition of risk areas for different SDI quintiles was shown in Figure 3. As shown in Figure 3, the high-risk area for both cancers all distributed in the middle and below SDI quintiles. In males, high-risk areas were distributed only in the low-middle, middle, and high-middle SDI quintiles (Figure 3), while high-risk areas were distributed across all SDI quintiles in females (Figure 3). The chi-square test results show that there was no difference in the composition ratio of risk areas among various SDI quintiles in both sex groups.

Figure 3. Stacked bar plot of sociodemographic index (SDI) quintiles and different risk areas, by sex. (A) The stacked bar plot of SDI quintiles and different risk areas in males. (B) The stacked bar plot of SDI quintiles and different risk areas in females. EC: esophageal cancer; GC: gastric cancer; SDI: sociodemographic index.



As shown in Table 1, the ASIR of esophageal and gastric cancer both showed a downward trend in all SDI quintiles from 2010 to 2019, with all 95% CI excluding 0. In both sexes, the average annual percentage decrease in ASIR for esophageal and gastric cancer in the middle SDI quintile was both substantial. In addition, the decrease in ASIR for esophageal cancer in this quintile over the past decade was significantly more substantial than in other SDI quintiles. In the middle SDI quintile, the average annual percentage decrease in ASIR for esophageal cancer over the decade was more substantial than that for gastric cancer across both sexes. Conversely, in other SDI quintiles except for females in the low-middle SDI, the decrease in ASIR

for esophageal cancer was less substantial than that for gastric cancer. Pearson correlation analysis indicated that the correlation between the temporal trends in the ASIR for esophageal and gastric cancer was high in all SDI quintiles (Table 1 and Figure S4 in Multimedia Appendix 4). Since the distributions of Pearson correlation coefficients in each SDI quintile was nonnormal in both sex groups after normality test, the Kruskal-Wallis test was used to assess the differences in correlation coefficients among these SDI quintiles. The Kruskal-Wallis test indicated no significant differences in the correlation between the temporal trends in the ASIR of esophageal and gastric cancer across various SDI quintiles in both sex groups.

Table . Average annual percent change (AAPC) in age-standardized incidence rates (ASIR) of esophageal cancer (EC) and gastric cancer (GC) among different sociodemographic index (SDI) quintiles, 2010 - 2019 (AAPC, %).

Location	Male (95% CI)	*r* [a]	Female (95% CI)	*r* [a]
Low SDI		0.85		0.94
EC	−0.27[b] (−0.42 to −0.12)		−0.39[b] (−0.46 to −0.33)	
GC	−1.37[b] (−1.53 to −1.21)		−0.74[b] (−0.84 to −0.65)	
Low−middle SDI		0.88		0.95
EC	−0.21[b] (−0.34 to −0.09)		−0.43[b] (−0.74 to −0.12)	
GC	−1.19[b] (−1.41 to −0.98)		−0.67[b] (−0.82 to −0.52)	
Middle SDI		0.87		0.94
EC	−2.56[b] (−2.77 to −2.35)		−3.79[b] (−4.04 to −3.53)	
GC	−1.90[b] (−2.13 to −1.67)		−2.26[b] (−2.64 to −1.88)	
High−middle SDI		0.76		0.94
EC	−0.95[b] (−1.28 to −0.62)		−1.58[b] (−1.84 to −1.31)	
GC	−1.66[b] (−1.84 to −1.48)		−2.13[b] (−2.28 to −1.97)	
High SDI		0.88		0.95
EC	−0.92[b] (−1.07 to −0.76)		−0.79[b] (−0.88 to −0.69)	
GC	−2.46[b] (−2.68 to −2.23)		−1.94[b] (−2.31 to −1.56)	

[a]Median Pearson correlation coefficients between temporal trends in the ASIR of esophageal and gastric cancer for countries and territories in each SDI quintile from 2010 to 2019.

[b]Indicates that 95% CI did not include 0.

Discussion

Principal Results

This study is the first to simultaneously report on the global temporal trends and spatial distribution of the ASIR for esophageal and gastric cancer, as well as their correlation based on GBD 2019 data. Our results show that there was a significant spatial clustering of the incidence for both cancers and a positive correlation between the risk of esophageal and gastric cancer across countries was observed in males. In addition, a codescending incidence trend between both cancers was observed in most countries and territories.

Due to the significant differences in the incidence of esophageal and gastric cancer between males and females [27,28], and the disparities in lifestyle, sex stratification was conducted in all analysis. In consistent with the previous studies, the risk of both esophageal and gastric cancer was high in Asia (especially East Asia and Central Asia) [5,30,31]. In addition, among Asian countries, the incidence of esophageal and gastric cancer was particularly serious in Mongolia (in 2019, the ASIR of both esophageal and gastric cancer ranked in the top 5 among both sexes). For Mongolia, the high incidence of esophageal cancer may be associated with high levels of fluoride in drinking water or drinking hot milk tea [32]. In addition, local cooking and heating mainly rely on coal and wood, resulting in high levels of fine particulate matter (diameter<2.5 μm), may also be one of the main reasons for the high incidence of esophageal cancer [33]. The spatial clustering phenomenon and the variation of geographical distribution in ASIR of esophageal and gastric cancer may be due to different lifestyles and environmental factors caused by differences in geographical and socioeconomic factors, as well as different histological subtypes of both cancers [34]. Compared with gastric cancer, there were more countries and territories with extremely high ASIR for esophageal cancer relative to the global level, which may be the reason why there were more countries and territories classified as the high-risk area for esophageal cancer than gastric cancer in this study. Although females were generally considered to have a lower risk of esophageal and gastric cancer than males, there were more countries and territories where females were at extremely high risk of esophageal and gastric cancer relative to the global

level. Those countries and territories, where the risk of esophageal and gastric cancer was much higher than the global level should be taken seriously. In addition, it is noteworthy that Malawi exhibited the highest risk for esophageal cancer alongside a very low risk for gastric cancer in both sex groups. The previous studies found that African countries with a higher incidence of esophageal cancer tend to have a lower estimated supply of selenium in their diets [35]. Therefore, the generally low selenium intake of the population in Malawi mainly caused by the reduced soil-to-crop selenium transfers in the local typical low pH soils may be one of the reasons for this phenomenon [35-37]. Since Malawi is one of the major tobacco producers in Africa, smokers have easier access to self-rolled tobacco (without filters), which will lead to smokers of this form of tobacco receive a higher dose of the carcinogenic products within the tobacco in comparison to store-bought cigarettes, thereby increasing the risk of developing esophageal cancer [38]. In addition, it has been reported that mycotoxins such as fumonisin B-1 stored in grain are fairly common in maize samples from Malawi, and although a direct causal relationship has not been established, this may be one of the reasons for the high incidence of esophageal cancer there [38]. The main source of energy in Malawian households is wood burning, which produces incomplete combustion products such as polycyclic aromatic hydrocarbons and may also be responsible for the high incidence of esophageal cancer [39]. In Pearson correlation analysis on the RR values for esophageal and gastric cancer among males and females, a positive correlation between the risk of incidence for esophageal and gastric cancer was observed in males, whereas no such correlation was found in females. Taking into account the possibility that the results may be influenced by a greater number of extreme values in females, we conducted a correlation analysis after removing some outliers. Nevertheless, no statistically significant correlation was found. One possible explanation is that smoking, as a primary common risk factor for both esophageal and gastric cancer, occurs at a significantly higher prevalence in males than in females (about 4 - 5 times higher) [40].

From 2010 to 2019, there was a global trend of decline in the ASIR for both esophageal and gastric cancer among males and females. The decline in the incidence of both cancers may be attributed to the improvements in the socioeconomic level and population health awareness in recent years [41-43]. It is worth mentioning that there were still some countries and territories, even some high-risk countries and territories including Cabo Verde, exhibited an increasing trend in the ASIR of esophageal and gastric cancer from 2010 to 2019. Although the significant increase of cancer registers compared with the past may be one of the reasons, it is necessary to actively control the incidence of esophageal and gastric cancer for those countries and territories showing an increasing trend. The reason for the stronger positive correlation between the temporal trends of esophageal and gastric cancer in females compared to that in males is unclear. However, it is certain that a significant positive correlation exists, which may be related to the shared risk factors due to their anatomic proximity.

When exploring the correlation between SDI and risk areas classified by RR values of esophageal and gastric cancer, no statistically significant difference was found in the composition ratio of risk areas across various SDI quintiles. This may be related to the low number of countries and territories belonging to the high-risk area for both cancers and gastric cancer alone. Consistent with the global trend, all SDI quintiles exhibited a decreasing trend in the ASIR of esophageal and gastric cancer, which indicated that, overall, the incidence of esophageal and gastric cancer had been improving in regions with different SDI quintiles from 2010 to 2019. Compared with other quintiles, the ASIR of esophageal and gastric cancer in middle SDI quintile both showed a substantial decreasing trend, which may be related to the high ASIR of both esophageal and gastric cancer in middle SDI quintile. Although the ASIR of gastric cancer was significantly higher than that of esophageal cancer, in contrast to other SDI quintiles, the decreasing trend in ASIR for esophageal cancer over the decade was more substantial than that for gastric cancer in middle SDI quintile. This opposite phenomenon, observed in the middle SDI quintile, calls for more rational public policies to strengthen the control of gastric cancer related risk factors. The correlation between the temporal trends in the incidence of esophageal and gastric cancer was strong among different SDI quintiles, and no statistically significant difference of this correlation was found across various SDI quintiles.

Strengths and Limitations

Our study has numerous strengths, including the ability to visually display the risk of both esophageal and gastric cancer in various countries and territories through risk area classification. This aids in developing prevention strategies tailored to shared or distinct risk factors of the 2 digestive cancers, while also considering local characteristics. Our study also has some limitations. First, the data we analyzed is sourced from the GBD 2019, the quality of some data, especially that from low-income or low SDI countries, is difficult to be guaranteed. Second, esophageal cancer includes 2 histological subtypes—esophageal squamous cell carcinoma and esophageal adenocarcinoma [44,45], while gastric cancer includes cardia gastric cancer and noncardia gastric cancer subtypes [46,47]. There are analogous and distinct etiologies with modifiable risk factors between these subtypes, as well as epidemiological characteristics [48-51]. However, due to the limitations of the GBD database, our study did not differentiate the subtypes of esophageal cancer and gastric cancer. Third, the results of Moran I are highly dependent on the choice of the conceptualization of spatial relationships. For global analysis, selecting an appropriate conceptualization is challenging, especially when regions are separated by natural barriers such as oceans. Contiguity measure may fail to capture spatial relationships across such barriers, while distance measure requires careful consideration of distance thresholds. In addition, Moran I is typically calculated for a single time point, which does not account for dynamic changes in spatial patterns over time. Furthermore, more detailed work is required to identify the primary modifiable risk factors in high-risk areas for each cancer, which will aid in more precisely reducing the global incidence of esophageal and gastric cancer.

Conclusions

This spatiotemporal correlation study simultaneously investigates esophageal cancer and gastric cancer, which share many risk factors. The results shows that there was a significant spatial clustering of the incidence for esophageal and gastric cancer and a positive correlation between the risk of both cancers across countries was observed in males. In addition, a codescending incidence trend between both cancers was observed at the global level. Despite the overall declining trend in the incidence rates of both esophageal and gastric cancer, they still pose a heavy disease burden worldwide. Analyzing the correlations in the global distribution and temporal trends of the incidence of esophageal and gastric cancer can help to gain a deeper understanding of the homogeneity and heterogeneity in the incidence pattern of these 2 cancers to optimize the allocation of global public health resources.

Acknowledgments

We sincerely appreciate the works by the Global Burden of Disease study 2019 collaborators and thank all the researchers and collaborators who were involved in this study for their efforts and contributions. We acknowledge financial supports from the National Key Research and Development Program of China (grant number: 2023YFC2508001), National Natural Science Foundation of China (grant numbers: 82073637, 82122060, and 82473700), Shanghai Municipal Science and Technology Major Project (grant numbers: ZD2021CY001 and 2023SHZDZX02).

Data Availability

The datasets analyzed during this study are available in the GBD repository [23].

Authors' Contributions

ZC and CS developed the study design and conducted data collection. ZC analyzed and interpreted the data. ZC drafted the article; CS, XC, and TZ critically revised it critically for important intellectual content. All authors gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Formulas for calculating the Moran I.
[DOCX File, 16 KB - cancer_v11i1e66655_app1.docx]

Multimedia Appendix 2
The age-standardized incidence rates (ASIR) of both cancers in 2019.
[DOCX File, 4506 KB - cancer_v11i1e66655_app2.docx]

Multimedia Appendix 3
Average annual percent change (AAPC) in age-standardized incidence rates (ASIR) of both cancers and the correlation of temporal trends.
[DOCX File, 117 KB - cancer_v11i1e66655_app3.docx]

Multimedia Appendix 4
Correlation of temporal trends in different sociodemocratic index (SDI) quintiles.
[DOCX File, 2590 KB - cancer_v11i1e66655_app4.docx]

References

1. Foreman KJ, Marquez N, Dolgert A, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016-40 for 195 countries and territories. Lancet 2018 Nov 10;392(10159):2052-2090. [doi: 10.1016/S0140-6736(18)31694-5] [Medline: 30340847]
2. Bray F, Laversanne M, Weiderpass E, Soerjomataram I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. Cancer 2021 Aug 15;127(16):3029-3030. [doi: 10.1002/cncr.33587] [Medline: 34086348]
3. Global Burden of Disease 2019 Cancer Collaboration, Kocarnik JM, Compton K, et al. Cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life years for 29 cancer groups from 2010 to 2019: A systematic analysis for the Global Burden of Disease Study 2019. JAMA Oncol 2022 Mar 1;8(3):420-444. [doi: 10.1001/jamaoncol.2021.6987] [Medline: 34967848]

4. Tran KB, Lang JJ, Compton K, et al. The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet 2022 Aug;400(10352):563-591. [doi: 10.1016/S0140-6736(22)01438-6]
5. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021 May;71(3):209-249. [doi: 10.3322/caac.21660] [Medline: 33538338]
6. Arnold M, Ferlay J, van Berge Henegouwen MI, Soerjomataram I. Global burden of oesophageal and gastric cancer by histology and subsite in 2018. Gut 2020 Sep;69(9):1564-1571. [doi: 10.1136/gutjnl-2020-321600] [Medline: 32606208]
7. Navarro Silvera SA, Mayne ST, Gammon MD, et al. Diet and lifestyle factors and risk of subtypes of esophageal and gastric cancers: classification tree analysis. Ann Epidemiol 2014 Jan;24(1):50-57. [doi: 10.1016/j.annepidem.2013.10.009] [Medline: 24239095]
8. Wang SM, Katki HA, Graubard BI, et al. Population attributable risks of subtypes of esophageal and gastric cancers in the United States. Am J Gastroenterol 2021 Sep 1;116(9):1844-1852. [doi: 10.14309/ajg.0000000000001355] [Medline: 34240714]
9. Yang CS, Chen XL. Research on esophageal cancer: With personal perspectives from studies in China and Kenya. Int J Cancer 2021 Jul 15;149(2):264-276. [doi: 10.1002/ijc.33421] [Medline: 33270917]
10. Sheikh M, Poustchi H, Pourshams A, et al. Individual and combined effects of environmental risk factors for esophageal cancer based on results from the Golestan cohort study. Gastroenterology 2019 Apr;156(5):1416-1427. [doi: 10.1053/j.gastro.2018.12.024] [Medline: 30611753]
11. Usui Y, Taniyama Y, Endo M, et al. Helicobacter pylori, homologous-recombination genes, and gastric cancer. N Engl J Med 2023 Mar 30;388(13):1181-1190. [doi: 10.1056/NEJMoa2211807]
12. Tsugane S, Sasazuki S. Diet and the risk of gastric cancer: review of epidemiological evidence. Gastric Cancer 2007;10(2):75-83. [doi: 10.1007/s10120-007-0420-0] [Medline: 17577615]
13. Wu Y, Li Y, Giovannucci E. Potential impact of time trend of lifestyle risk factors on burden of major gastrointestinal cancers in China. Gastroenterology 2021 Dec;161(6):1830-1841. [doi: 10.1053/j.gastro.2021.08.006] [Medline: 34389341]
14. Okada E, Ukawa S, Nakamura K, et al. Demographic and lifestyle factors and survival among patients with esophageal and gastric cancer: The Biobank Japan Project. J Epidemiol 2017 Mar;27(3):S29-S35. [doi: 10.1016/j.je.2016.12.002]
15. Spechler SJ. Barrett esophagus and risk of esophageal cancer: a clinical review. JAMA 2013 Aug 14;310(6):627-636. [doi: 10.1001/jama.2013.226450] [Medline: 23942681]
16. Shaheen N, Ransohoff DF. Gastroesophageal reflux, barrett esophagus, and esophageal cancer: scientific review. JAMA 2002 Apr 17;287(15):1972-1981. [doi: 10.1001/jama.287.15.1972] [Medline: 11960540]
17. Li P, Jing J, Guo W, et al. The associations of air pollution and socioeconomic factors with esophageal cancer in China based on a spatiotemporal analysis. Environ Res 2021 May;196:110415. [doi: 10.1016/j.envres.2020.110415] [Medline: 33159927]
18. Rao Z, Xie X, Tang X, et al. The spatiotemporal correlation of PM2.5 concentration on esophageal cancer hospitalization rate in Fujian province of China. Environ Sci Pollut Res Int 2022 Sep;29(44):67325-67335. [doi: 10.1007/s11356-022-20587-2] [Medline: 35524092]
19. Mancini S, Ravaioli A, Giuliani O, et al. Gastric cancer incidence in the Romagna Region of Italy: A spatial and temporal analysis. Dig Liver Dis 2015 Dec;47(12):1076-1081. [doi: 10.1016/j.dld.2015.09.001] [Medline: 26476476]
20. Papoila AL, Riebler A, Amaral-Turkman A, et al. Stomach cancer incidence in Southern Portugal 1998-2006: a spatio-temporal analysis. Biom J 2014 May;56(3):403-415. [doi: 10.1002/bimj.201200264] [Medline: 24596314]
21. Sun X, Zhao D, Liu Y, et al. The long-term spatial-temporal trends and burden of esophageal cancer in one high-risk area: A population-registered study in Feicheng, China. PLoS ONE 2017;12(3):e0173211. [doi: 10.1371/journal.pone.0173211]
22. Ghasemi-Kebria F, Semnani S, Fazel A, et al. Esophageal and gastric cancer incidence trends in Golestan, Iran: An age-period-cohort analysis 2004 to 2018. Int J Cancer 2023 Jul 1;153(1):73-82. [doi: 10.1002/ijc.34518] [Medline: 36943026]
23. Global Burden of Disease Study 2019 (GBD 2019) Results. Institute for Health Metrics and Evaluation (IHME). 2020. URL: https://gbd2019.healthdata.org/gbd-results/ [accessed 2024-01-25]
24. Global Burden of Disease Study 2019 (GBD 2019). Socio-Demographic Index (SDI) 1950–2019. 2020. URL: https://ghdx.healthdata.org/record/ihme-data/gbd-2019-socio-demographic-index-sdi-1950-2019 [accessed 2024-01-25]
25. Tan H, Liang L, Yin X, Li C, Liu F, Wu C. Spatiotemporal analysis of pertussis in Hunan Province, China, 2009-2019. BMJ Open 2022 Sep 8;12(9):e055581. [doi: 10.1136/bmjopen-2021-055581] [Medline: 36691220]
26. Xue M, Zhong J, Gao M, et al. Analysis of spatial-temporal dynamic distribution and related factors of tuberculosis in China from 2008 to 2018. Sci Rep 2023 Mar 27;13(1):4974. [doi: 10.1038/s41598-023-31430-0] [Medline: 36973322]
27. Jiang Q, Shu Y, Jiang Z, et al. Burdens of stomach and esophageal cancer from 1990 to 2019 and projection to 2030 in China: Findings from the 2019 Global Burden of Disease Study. J Glob Health 2024 Jan 5;14:04025. [doi: 10.7189/jogh.14.04025] [Medline: 38180951]
28. Wang S, Zheng R, Arnold M, et al. Global and national trends in the age‐specific sex ratio of esophageal cancer and gastric cancer by subtype. Int J Cancer 2022 Nov;151(9):1447-1461. [doi: 10.1002/ijc.34158]

29. IHME's Free-of-Charge Non-commercial User Agreement. URL: https://www.healthdata.org/Data-tools-practices/data-practices/ihme-free-charge-non-commercial-user-agreement [accessed 2024-12-17]
30. López MJ, Carbajal J, Alfaro AL, et al. Characteristics of gastric cancer around the world. Crit Rev Oncol Hematol 2023 Jan;181:103841. [doi: 10.1016/j.critrevonc.2022.103841] [Medline: 36240980]
31. Lin L, Li Z, Yan L, Liu Y, Yang H, Li H. Global, regional, and national cancer incidence and death for 29 cancer groups in 2019 and trends analysis of the global cancer burden, 1990-2019. J Hematol Oncol 2021 Nov 22;14(1):197. [doi: 10.1186/s13045-021-01213-z] [Medline: 34809683]
32. Kamangar F, Nasrollahzadeh D, Safiri S. The global, regional, and national burden of oesophageal cancer and its attributable risk factors in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol 2020 Jun;5(6):582-597. [doi: 10.1016/S2468-1253(20)30007-8] [Medline: 32246941]
33. Flowerday CE, Lundrigan P, Kitras C, Nguyen T, Hansen JC. Utilizing low-cost sensors to monitor indoor air quality in Mongolian Gers. Sensors (Basel) 2023 Sep 7;23(18):7721. [doi: 10.3390/s23187721] [Medline: 37765777]
34. Arnold M, Abnet CC, Neale RE, et al. Global burden of 5 major types of gastrointestinal cancer. Gastroenterology 2020 Jul;159(1):335-349. [doi: 10.1053/j.gastro.2020.02.068] [Medline: 32247694]
35. Jaskiewicz K, Marasas WF, Rossouw JE, Niekerk FE, Heine Tech EW. Selenium and other mineral elements in populations at risk for esophageal cancer. Cancer 1988 Dec 15;62(12):2635-2639. [doi: 10.1002/1097-0142(19881215)62:12<2635::aid-cncr2820621232>3.0.co;2-5] [Medline: 3191467]
36. Chasimpha SJD, Parkin DM, Masamba L, Dzamalala CP. Three-year cancer incidence in Blantyre, Malawi (2008-2010). Int J Cancer 2017 Aug 15;141(4):694-700. [doi: 10.1002/ijc.30777] [Medline: 28493322]
37. Hurst R, Siyame EWP, Young SD, et al. Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi. Sci Rep 2013;3:1425. [doi: 10.1038/srep01425] [Medline: 23478344]
38. Kaimila B, Mulima G, Kajombo C, et al. Tobacco and other risk factors for esophageal squamous cell carcinoma in Lilongwe Malawi: Results from the Lilongwe esophageal cancer case: Control study. PLOS Glob Public Health 2022;2(6):e0000135. [doi: 10.1371/journal.pgph.0000135] [Medline: 36962303]
39. Mlombe YB, Rosenberg NE, Wolf LL, et al. Environmental risk factors for oesophageal cancer in Malawi: A case-control study. Malawi Med J 2015 Sep;27(3):88-92. [doi: 10.4314/mmj.v27i3.3] [Medline: 26715952]
40. Reitsma MB, Fullman N, Ng M, et al. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: a systematic analysis from the Global Burden of Disease Study 2015. Lancet 2017 May;389(10082):1885-1906. [doi: 10.1016/S0140-6736(17)30819-X]
41. Garcia-Saiso S, Marti M, Pesce K, et al. Artificial intelligence as a potential catalyst to a more equitable cancer care. JMIR Cancer 2024 Aug 12;10:e57276. [doi: 10.2196/57276] [Medline: 39133537]
42. Ramadan M, Aboalola D, Aouabdi S, et al. Influence of Breast Cancer Awareness Month on public interest of breast cancer in high-income countries between 2012 and 2022: Google Trends analysis. JMIR Cancer 2024 Aug 12;10:e49197. [doi: 10.2196/49197] [Medline: 39133912]
43. Pan P, Yu C, Li T, et al. Evaluating the quality of cancer-related WeChat public accounts: Cross-sectional study. JMIR Cancer 2024 May 30;10:e52156. [doi: 10.2196/52156] [Medline: 38814688]
44. Morgan E, Soerjomataram I, Rumgay H, et al. The global landscape of esophageal squamous cell carcinoma and esophageal adenocarcinoma incidence and mortality in 2020 and projections to 2040: New estimates from GLOBOCAN 2020. Gastroenterology 2022 Sep;163(3):649-658. [doi: 10.1053/j.gastro.2022.05.054] [Medline: 35671803]
45. Abbas G, Krasna M. Overview of esophageal cancer. Ann Cardiothorac Surg 2017 Mar;6(2):131-136. [doi: 10.21037/acs.2017.03.03] [Medline: 28447001]
46. Yang L, Kartsonaki C, Yao P, et al. The relative and attributable risks of cardia and non-cardia gastric cancer associated with Helicobacter pylori infection in China: a case-cohort study. Lancet Public Health 2021 Dec;6(12):e888-e896. [doi: 10.1016/S2468-2667(21)00164-X] [Medline: 34838195]
47. Thrift AP, El-Serag HB. Burden of gastric cancer. Clin Gastroenterol Hepatol 2020 Mar;18(3):534-542. [doi: 10.1016/j.cgh.2019.07.045] [Medline: 31362118]
48. Li P, Jing J, Liu W, Wang J, Qi X, Zhang G. Spatiotemporal patterns of esophageal cancer burden attributable to behavioral, metabolic, and dietary risk factors from 1990 to 2019: Longitudinal observational study. JMIR Public Health Surveill 2023 Oct 6;9:e46051. [doi: 10.2196/46051] [Medline: 37801354]
49. Ho ALK, Smyth EC. A global perspective on oesophageal cancer: two diseases in one. Lancet Gastroenterol Hepatol 2020 Jun;5(6):521-522. [doi: 10.1016/S2468-1253(20)30047-9] [Medline: 32246940]
50. Han Y, Zhu X, Hu Y, et al. Electronic health record-based absolute risk prediction model for esophageal cancer in the Chinese population: Model development and external validation. JMIR Public Health Surveill 2023 Mar 15;9:e43725. [doi: 10.2196/43725] [Medline: 36781293]
51. van den Brandt PA. The impact of a healthy lifestyle on the risk of esophageal and gastric cancer subtypes. Eur J Epidemiol 2022 Sep;37(9):931-945. [doi: 10.1007/s10654-022-00899-w] [Medline: 35982188]

Abbreviations

AAPC: average annual percent change
ASIR: age-standardized incidence rates
GBD: Global Burden of Disease
GHDx: Global Health Data Exchange
IHME: Institute for Health Metrics and Evaluation
RR: risk ratio
SDI: socio-demographic index
UI: uncertainty interval

Edited by N Cahill; submitted 19.09.24; peer-reviewed by AR Sevilla, YL Cheong; revised version received 15.12.24; accepted 16.12.24; published 29.01.25.

Please cite as:
Cui Z, Suo C, Zhao Y, Wang S, Zhao M, Chen R, Lu L, Zhang T, Chen X
Spatiotemporal Correlation Analysis for the Incidence of Esophageal and Gastric Cancer From 2010 to 2019: Ecological Study
JMIR Cancer 2025;11:e66655
URL: https://cancer.jmir.org/2025/1/e66655
doi:10.2196/66655

Original Paper

Analyzing Online Search Trends for Kidney, Prostate, and Bladder Cancers in China: Infodemiology Study Using Baidu Search Data (2011-2023)

Shuangquan Lin[1], MD; Lingxing Duan[1], MD; Xiangda Xu[1], MD; Haichao Cao[1], MD; Xiongbing Lu[1], MD; Xi Wen[2], MPH; Shanzun Wei[1], MD, PhD

[1]Urology Department, The Second Affiliated Hospital of Nanchang University, Nanchang University, Nanchang, China

[2]Urology Laboratory, The Second Affiliated Hospital of Nanchang University, Nanchang University, Nanchang, China

Corresponding Author:
Shanzun Wei, MD, PhD
Urology Department
The Second Affiliated Hospital of Nanchang University
Nanchang University
1st Mingde Rd Donghu
Nanchang, 330000
China
Phone: 1 458 800 6725
Email: Zunny377@icloud.com

Abstract

Background: Cancers of the bladder, kidney, and prostate are the 3 major genitourinary cancers that significantly contribute to the global burden of disease (GBD) and continue to show increasing rates of morbidity and mortality worldwide. In mainland China, understanding the cancer burden on patients and their families is crucial; however, public awareness and concerns about these cancers, particularly from the patient's perspective, remain predominantly focused on financial costs. A more comprehensive exploration of their needs and concerns has yet to be fully addressed.

Objective: This study aims to analyze trends in online searches and user information–seeking behaviors related to bladder, kidney, and prostate cancers—encompassing descriptive terms (eg, "bladder cancer," "kidney cancer," "prostate cancer") as well as related synonyms and variations—on both national and regional scales. This study leverages data from mainland China's leading search engine to explore the implications of these search patterns for addressing user needs and improving health management.

Methods: The study analyzed Baidu Index search trends for bladder, kidney, and prostate cancers (from January 2011 to August 2023) at national and provincial levels. Search volume data were analyzed using the joinpoint regression model to calculate annual percentage changes (APCs) and average APCs (AAPCs), identifying shifts in public interest. User demand was assessed by categorizing the top 10 related terms weekly into 13 predefined topics, including diagnosis, treatment, and traditional Chinese medicine. Data visualization and statistical analyses were performed using Prism 9. Results revealed keyword trends, demographic distributions, and public information needs, offering insights into health communication and management strategies based on online information-seeking behavior.

Results: Three cancer topics were analyzed using 39 search keywords, yielding a total Baidu Search Index (BSI) of 43,643,453. From 2011 to 2015, the overall APC was 15.2% (P<.05), followed by –2.8% from 2015 to 2021, and 8.9% from 2021 to 2023, with an AAPC of 4.9%. Bladder, kidney, and prostate cancers exhibited AAPCs of 2.8%, 3.9%, and 6.8%, respectively (P<.05). The age distribution of individuals searching for these cancer topics varied across the topics. Geographically, searches for cancer were predominantly conducted by people from East China, who accounted for approximately 30% of each cancer search query. Regarding user demand, the total BSI for relevant user demand terms from August 2022 to August 2023 was 676,526,998 out of 2,570,697,380 (15.74%), representing only a limited total cancer-related search volume.

Conclusions: Online searches and inquiries related to genitourinary cancers are on the rise. The depth of users' information demands appears to be influenced by regional economic levels. Cancer treatment decision-making may often involve a family-centered approach. Insights from internet search data can help medical professionals better understand public interests and concerns, enabling them to provide more targeted and reliable health care services.

(JMIR Cancer 2025;11:e57414) doi:10.2196/57414

KEYWORDS

bladder cancer; kidney cancer; prostate cancer; Baidu Index; infodemiology; public interest; patients' concern

Introduction

Cancer poses a significant burden on global public health [1]. In urology, bladder, kidney, and prostate cancers are the 3 primary genitourinary cancers contributing to the global burden of disease (GBD), with persistently high morbidity and mortality rates [1,2]. According to a GBD report from 2019, the annual global incidence rates of bladder, kidney, and prostate cancers have risen by 154.78%, 123.34%, and 169.11%, respectively, over the past 2 decades, making them the most prominent cancers in the field of urology [3,4]. In mainland China, the incidence rates of bladder, kidney, and prostate cancers in 2019 had doubled compared with 1990 and are projected to triple by 2030 [2], with 192,390 cases of bladder cancer, 126,980 cases of kidney cancer, and 315,310 cases of prostate cancer. Therefore, precautionary measures are essential in addition to gathering information and assessing the disease burden using real-world data.

Infodemiology was defined as "The framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet." It has been shown to effectively highlight public health issues, assess the impact of health care policies, and uncover public concerns during global pandemics, as well as in chronic and contagious diseases, along with related public acceptance [5-9]. Examining underlying trends in user behavior and the specific demands associated with major genitourinary cancers could potentially provide insights into regional health information–seeking behaviors and population-level interests.

Cancer imposes a significant burden on patients and their families, typically measured in terms of financial costs and clinical outcomes [10,11]. However, there is limited understanding of its broader impacts, such as public awareness, emotional well-being, and social participation. In China, previous studies have examined prevalent noncancer urological issues and their online visibility using data from the Baidu Index [6,12,13], a tool that analyzes search behaviors to reflect public interest in health topics. While prior research has explored general cancer-related searches, these studies primarily focused on incidence correlations and population-level disparities across 28 cancer types [14], offering limited insights into specific user demands or temporal and geographic patterns.

This study aims to address this gap by focusing on 3 major genitourinary cancers: bladder cancer, kidney cancer, and prostate cancer. Using Baidu Index data, we analyzed internet search trends, user needs, and associated geospatial and temporal patterns. By identifying search behaviors and topics of interest, we seek to provide actionable insights into public health awareness and address unmet needs, potentially contributing to the improvement and guidance of health care strategies.

Methods

Keyword Selecting and Data Retrieving

Baidu (Baidu, Inc.), the leading search engine in mainland China, accounts for 92.1% of the search volume and 93.1% of user coverage [15]. Its analytics platform, Baidu Index, allows for tracking keyword popularity trends and analyzing related user demands [6,7,14]. Comparable to Google's platform on a global scale, Baidu has been validated as a reliable tool for studying online search trends and user behavior in infodemiology research within China [16,17].

This study primarily focuses on analyzing the temporal search trends of cancer-related terms associated with kidney, bladder, and prostate cancers. Based on defined criteria, these terms are characterized as compounds [6], combining morphemes denoting a urological organ with those indicating tumor-related concepts. The key morphemes identified were (1) "肾脏" or "肾" (the kidney), (2) "膀胱" (the bladder), (3) "前列腺" (the prostate), and (4) "肿瘤" (the tumor). The Baidu Index platform automatically matched these combinations with all available search keywords, including synonyms and complex derivatives. Measures were implemented to prevent duplication and omissions, following approaches detailed in previous studies [6,18]. Synonyms and complex derivatives were screened and selected to minimize ambiguity and bias arising from language habits, as previously described [6,13]. All available search keywords related to these 3 cancer types were categorized based on their connotations and are listed in Multimedia Appendix 1.

The Baidu Index platform consists of 3 key modules: (1) the search trend module, (2) the user demand module, and (3) the demographic portrait module. These modules enable the analysis of search demand from multiple perspectives, including popularity trends, topic-related concerns, and geodemographic features [6,13]. Search popularity is quantified using the Baidu Search Index (BSI), a key metric based on daily recorded search demand. With integrated data on location, gender, age, and other elements, trends and demographic profiles of the population can be visualized and retrieved [6,13]. Search trend data, available since 2011, were retrieved from the search trend module of the Baidu Index platform [14]. Data at both provincial and national scales were collected for the period from January 1, 2011, to July 31, 2023. The most recent data from the geodemographic and user demand modules were obtained from the user demand module on the Baidu Index official website [9,19].

Data Processing and Statistical Analysis

The Baidu Index is a data-sharing platform that leverages extensive user behavior data to measure search trends. By tracking the frequency of unique keyword searches and their weighting within Baidu's overall search volume, it provides a metric for keyword popularity. This study collected data on a daily, weekly, monthly, and yearly basis to capture a

comprehensive view of cancer-related search patterns. Sequential plotting of BSI data for each term was conducted to illustrate trends in public interest. Changes in trends over time were analyzed using the Joinpoint Regression Model (program version 4.7.0.0; Statistical Research and Applications Branch, National Cancer Institute). This model, well-suited for time-series analysis of large data sets, identifies statistically significant shifts in trends. The annual percentage change (APC) was calculated to summarize yearly trends within specified intervals, measuring year-over-year percentage changes. The average APC (AAPC) was used to evaluate trends over extended periods, providing a more stable estimate of the overall trend direction and rate of change [20,21]. For each topic—bladder cancer, kidney cancer, and prostate cancer—the public demand trend was illustrated through sequentially plotted BSI data. Intergroup differences were analyzed using the Student *t* test and Kruskal-Wallis test, as appropriate. A *P* value of <.05 was considered statistically significant.

In the user demand section, the top 10 most frequently mentioned words related to each search keyword were listed weekly and sorted by cancer type. This allowed for the identification and analysis of the most prominent and commonly discussed topics for each cancer type. In line with previous findings, we used a 13-topic system to categorize user demand–related terms, helping to clarify users' main concerns and implied intentions [13]. Aside from some random or off-topic terms, these categories were defined as follows: (1) complaint, (2) inquiry, (3) treatment and decision, (4) health issue, (5) diagnosis, (6) hospital and service, (7) symptom confirmation, (8) tests and examinations, (9) prognosis, (10) traditional Chinese medicine (TCM) complaint, (11) TCM inquiry, (12) TCM ingredient, and (13) TCM regimen.

All databases were constructed using Excel 2019 (Microsoft Corporation). The APC was calculated with the Joinpoint Regression Model, program version 4.7.0.0 (Statistical Research and Applications Branch, National Cancer Institute). Statistical analysis and figure creation were performed using Prism 9 for macOS (version 9.5.0 (525); GraphPad Software).

Ethical Considerations

We used publicly available, anonymized data that can be accessed without special permissions. As the data are aggregated and publicly accessible, IRB approval or exemption was not required.

Results

Available Trends Data in Urology Cancer Topics

We identified and confirmed 39 valid search keywords on the Baidu Index platform. These keywords are theme-based synonyms and moderate derivative terms that convey specific motives or demands. Among these, 13 keywords pertained specifically to bladder cancer, while 15 and 11 keywords were related to kidney cancer and prostate cancer, respectively. For theme categorization, 4 topics were identified: (1) complaint, with 9 keywords; (2) inquiry, with 23 keywords; (3) treatment, with 4 keywords; and (4) prognosis, with 3 keywords. All available search keywords related to these 3 urological cancers, along with their English equivalent translations, are listed in Multimedia Appendix 1.

The general search volume for all 3 urological cancers increased to a mean of 10,737.74 (SD 1026.29) from an initial mean of 5975.68 (SD 770.42). Specifically, the average daily search volume for bladder cancer rose to 3453.09 (SD 337.44) in 2023 from an initial average of 2275.72 (SD 302.17). For kidney cancer, the average daily search volume increased to 2976.78 (SD 319.64) from an initial average of 1706.84 (SD 262.95). Similarly, the search volume for prostate cancer grew to a mean of 4307.87 (SD 417.68) from 1993.12 (SD 297.99). According to the trend module, the total BSI for these top 3 urological cancers was 43,643,453. Specifically, the 13-year summed BSI ratio was 37.37% (15,972,271/43,643,453) for bladder cancer, 28.27% (12,079,106/43,643,453) for kidney cancer, and 34.36% (15,592,076/43,643,453) for prostate cancer (Figure 1). Regarding topic preferences for each cancer, the BSI ratio for complaint and inquiry was dominant, accounting for 90.26%, 96.10%, and 79.53% across all 4 topics for bladder cancer, kidney cancer, and prostate cancer, respectively (Figure 2).

To illustrate search trends over time since January 1, 2011, the daily request–based BSI for each cancer was analyzed both overall and by specific topics. The significance of these trends was evaluated using the APC model, as shown in Figures 1 and 2.

Based on the average annual BSI counts, a general growth in search requests for all 3 urological cancers was observed. The overall APC was 15.2% (*P*<.05) from 2011 to 2015, –2.8% from 2015 to 2021, and 8.9% from 2021 to 2023, resulting in an AAPC of 4.9%. For bladder cancer, the APC was 8.3% (*P*<.05) from 2011 to 2019, –11.7% from 2019 to 2021, and 7.4% from 2021 to 2023, with an AAPC of 2.8%. For kidney cancer, the APC was 8.0% (*P*<.05) from 2011 to 2019, –9.6% from 2019 to 2021, and 11.4% from 2021 to 2023, resulting in an AAPC of 3.9%. For prostate cancer, the APC was 17.7% (*P*<.05) from 2011 to 2015, –3.1% from 2015 to 2020, and 10.4% from 2021 to 2023, yielding an AAPC of 6.8% (*P*<.05).

Specifically within the bladder cancer theme, the APCs for the *complaint* topic were 8.0% (*P*<.05) from 2011 to 2021 and –5.1% from 2021 to 2023, with an AAPC of 1.2%. For the *inquiry* topic, the APCs were 20.8% (*P*<.05) from 2011 to 2014 and 1.2% from 2014 to 2023, with an AAPC of 4.8% (*P*<.05). For the *prognosis* topic, the APCs were –20.8% (*P*<.05) from 2011 to 2014, 15.6% from 2014 to 2018, and –6.9% from 2018 to 2023, resulting in an AAPC of –4.4%. For the *treatment* topic, the APCs were –5.4% from 2011 to 2016, 7.6% from 2016 to 2019, and –23.9% (*P*<.05) from 2019 to 2023, with an AAPC of –9.4%.

In the kidney cancer theme, the APCs for the *complaint* topic were 6.8% (*P*<.05) from 2011 to 2019 and –3.6% from 2019 to 2023, with an AAPC of 3.2% (*P*<.05). For the *inquiry* topic, the APCs were 14.0% (*P*<.05) from 2011 to 2017, –9.2% (*P*<.05) from 2017 to 2021, and 17.8% from 2021 to 2023, resulting in an AAPC of 6.2% (*P*<.05). For the *prognosis* topic, the APCs were –19.0% (*P*<.05) from 2011 to 2019 and 15.7% from 2019 to 2023, with an AAPC of –11.0% (*P*<.05). For the *treatment* topic, the APCs were 2.5% from 2011 to 2014,

–47.1% from 2014 to 2018, and 42.6% from 2019 to 2023, with an AAPC of –12.2% (*P*<.05). In the prostate cancer theme, the APCs for the *complaint* topic were 25.4% (*P*<.05) from 2011 to 2013, 3.1% from 2013 to 2017, and –1.4% from 2017 to 2023, resulting in an AAPC of 4.2%. For the *inquiry* topic, the APCs were 21.2% (*P*<.05) from 2011 to 2015, –4.6% from 2015 to 2020, and 11.8% from 2020 to 2023, with an AAPC of 11.8%. For the *prognosis* topic, the APCs were –6.2% (*P*<.05) from 2011 to 2018 and 2.5% from 2018 to 2023, with an AAPC of –3.0% (*P*<.05). For the *treatment* topic, the APCs were 41.6% (*P*<.05) from 2011 to 2015, –5.8% from 2015 to 2018, and 17.4% (*P*<.05) from 2018 to 2023, resulting in an AAPC of 18.2% (*P*<.05).

Figure 1. Online search trends in bladder, kidney, and prostate cancer topics since 2011. (A) Searching trend of each cancer topic; (B) Sum BSI proportion of each cancer topic. APC: annual percentage change; BSI: Baidu Search Index.

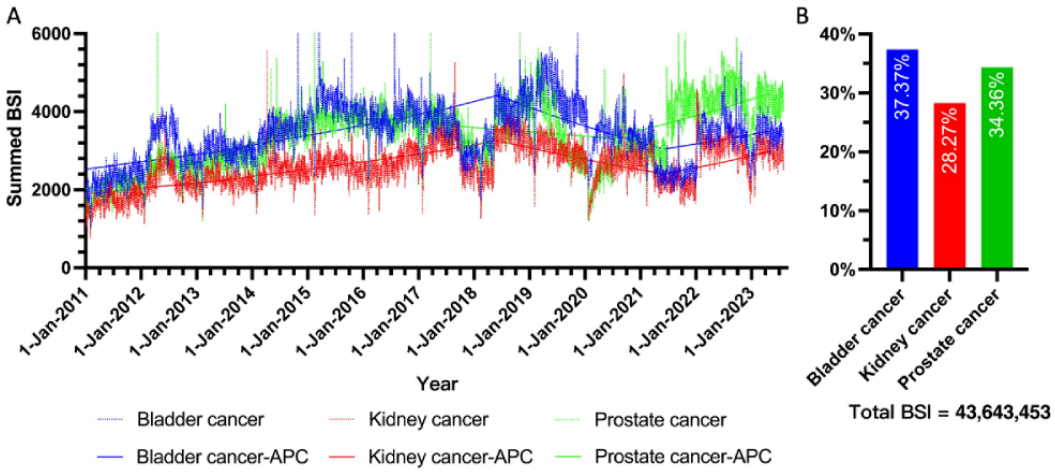
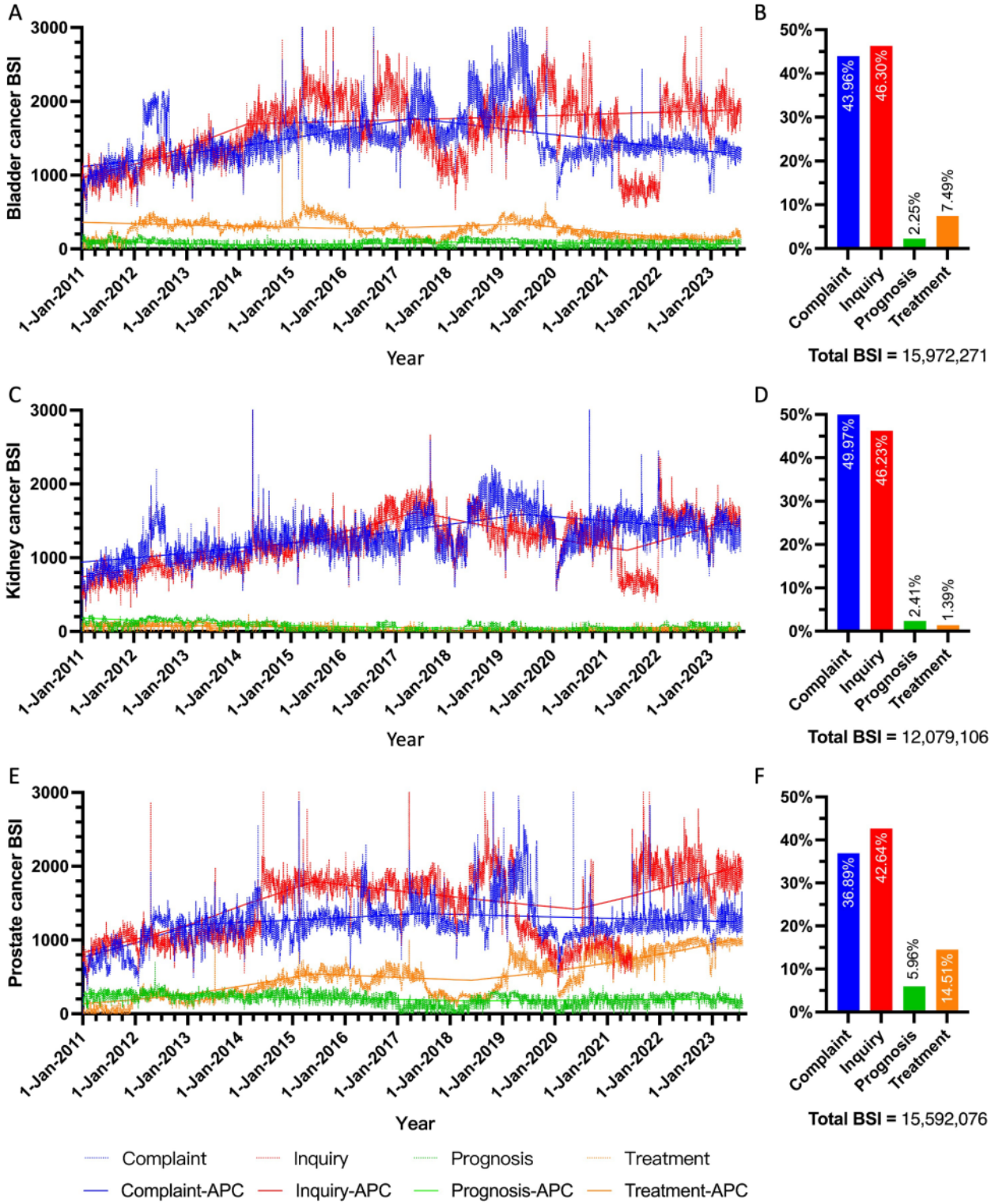


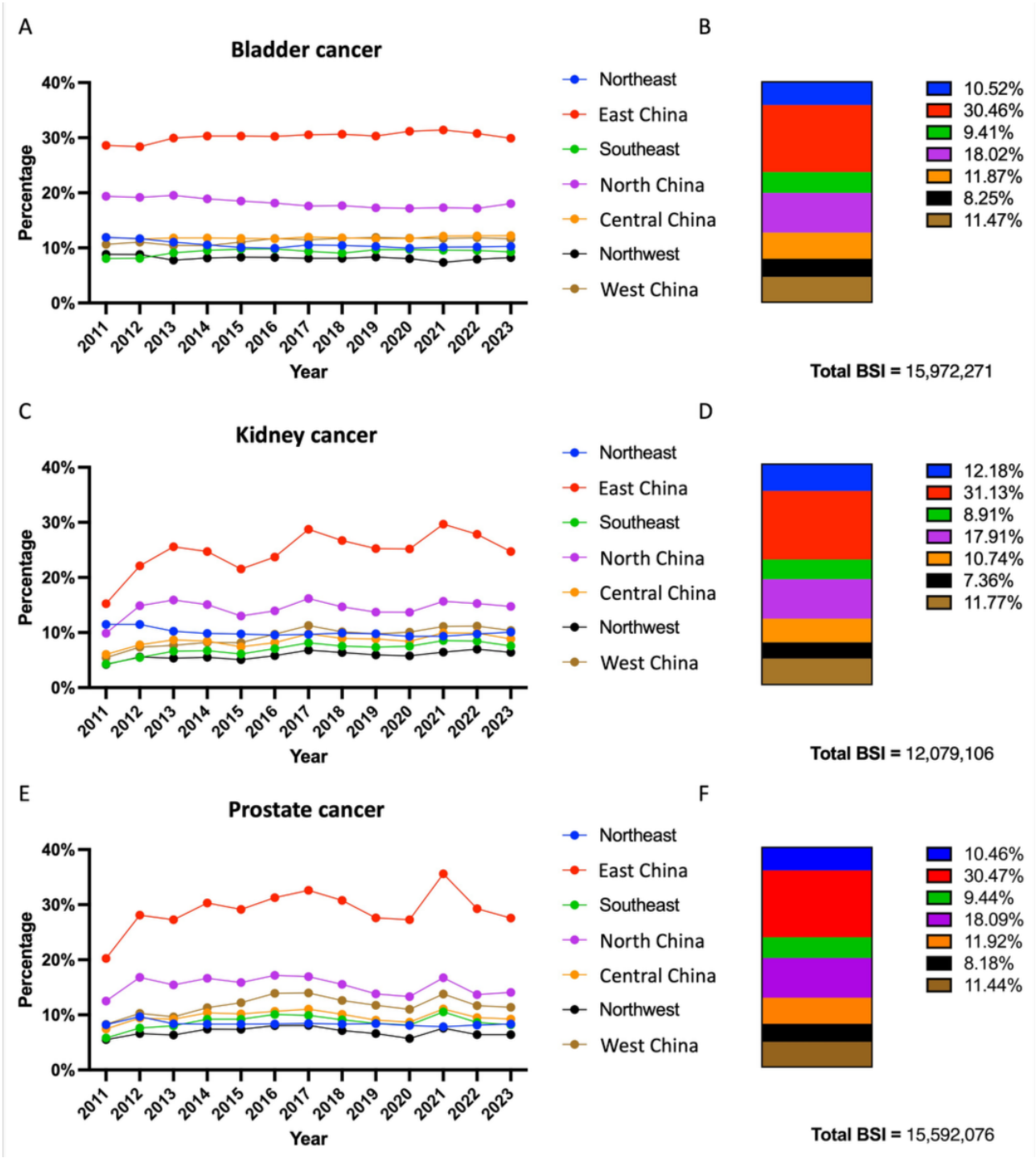
Figure 2. Online search trends for specific topics related to bladder, kidney, and prostate cancer since 2011. (A) Searching trend of specific topics in bladder cancer. (B) Sum BSI proportion of specific topics in bladder cancer. (C) Searching trend of specific topics in kidney cancer. (D) Sum BSI proportion of specific topics in kidney cancer. (E) Searching trend of specific topics in prostate cancer. (F) Sum BSI proportion of specific topics in prostate cancer. APC: annual percentage change; BSI: Baidu Search Index.



Geographic Differences

The geographic distribution of each cancer type was calculated based on provincial BSI data and categorized according to the 7 Chinese administrative divisions [6]. Figure 3 shows the 13-year regional BSI proportions for each cancer type with valid search records. Search requests were predominantly from East China, accounting for 30.46%, 31.13%, and 30.47% of bladder, kidney, and prostate cancer searches, respectively, followed by North China. Search demand from Central China, South China, and West China was comparable, each contributing around 11%. The Northeast and Northwest regions ranked the lowest, collectively accounting for approximately 8% of searches for each urological cancer.

Figure 3. Regional distribution of online search in bladder, kidney, and prostate cancer topics. (A) Annual BSI trend for each region in the topic of bladder cancer. (B) Total search rates for each region on the topic of bladder cancer. (C) Annual BSI trend for each region in the topic of kidney cancer. (D) Total search rates for each region on the topic of kidney cancer. (E) Annual BSI trend for each region in the topic of prostate cancer. (F) Total search rates for each region on the topic of prostate cancer. BSI: Baidu Search Index.



Demographic Differences

From the demographic distribution analysis, variations in age and gender distribution were observed across each cancer theme and specific topic. In the bladder cancer theme, search requests for each topic were primarily made by individuals aged 20-29 and 30-39 years. For kidney cancer, the topics of complaint, inquiry, and prognosis were predominantly searched by the 20-29-year age group, whereas the topic of treatment was mainly searched by individuals aged 40-49 years. In the prostate cancer theme, the topics of inquiry and treatment were primarily requested by the 30-39-year age group, while searches for prognosis were mainly made by those aged 40-49 years. Notably, no dominant age group was identified for searches in the complaint topic for prostate cancer (Table 1).

Table 1. Demographic differences in each cancer topic.

Theme and topic	≤19 years	20-29 years	30-39 years	40-49 years	≥50 years	Female	Male
Bladder cancer							
Complaint, n/N (%)	2447/40,701 (6.01)	12,411/40,701 (30.49)	14,187/40,701 (34.86)	6954/40,701 (17.09)	4702/40,701 (11.55)	22,332/40,701 (54.87)	18,369/40,701 (45.13)
Enquiry, n/N (%)	2149/58,805 (3.65)	10,237/58,805 (17.41)	18,226/58,805 (30.99)	15,246/58,805 (25.93)	12,947/58,805 (22.02)	34,729/58,805 (59.06)	24,076/58,805 (40.94)
Treatment, n/N (%)	45/2567 (1.75)	907/2567 (35.33)	730/2567 (28.44)	486/2567 (18.93)	398/2567 (15.50)	1195/2567 (46.55)	1371/2567 (53.41)
Prognosis, n/N (%)	256/6017 (4.25)	1298/6017 (21.57)	2217/6017 (36.85)	1245/6017 (20.69)	1001/6017 (16.64)	3342/6017 (55.54)	2675/6017 (44.46)
Kidney cancer							
Complaint, n/N (%)	2548/40,181 (6.34)	12,059/40,181 (30.01)	14,170/40,181 (35.27)	7312/40,181 (18.20)	4092/40,181 (10.18)	20,819/40,181 (51.81)	19,362/40,181 (48.19)
Enquiry, n/N (%)	1927/49,219 (3.92)	9211/49,219 (18.71)	15,897/49,219 (32.30)	12,657/49,219 (25.72)	9527/49,219 (19.36)	28,829/49,219 (58.57)	20,390/49,219 (41.43)
Treatment, n/N (%)	0/1328 (0.00)	221/1328 (16.64)	296/1328 (22.29)	738/1328 (55.57)	73/1328 (5.50)	664/1328 (50.00)	664/1328 (50.00)
Prognosis, n/N (%)	0/1913 (0.00)	201/1913 (10.51)	705/1913 (36.85)	453/1913 (23.68)	554/1913 (28.96)	1325/1913 (69.26)	588/1913 (30.74)
Prostate cancer							
Complaint, n/N (%)	2011/39,725 (5.06)	8413/39,725 (21.18)	10,388/39,725 (26.15)	8457/39,725 (21.29)	10,456/39,725 (26.32)	17,602/39,725 (44.31)	22,123/39,725 (55.69)
Enquiry, n/N (%)	3786/59,541 (6.36)	14,445/59,541 (24.26)	20,435/59,541 (34.32)	12,156/59,541 (20.42)	8719/59,541 (14.64)	27,446/59,541 (46.10)	32,095/59,541 (53.90)
Treatment, n/N (%)	116/4462 (2.60)	1007/4462 (22.57)	1344/4462 (30.12)	1063/4462 (23.82)	932/4462 (20.89)	2053/4462 (46.01)	2409/4462 (53.99)
Prognosis, n/N (%)	833/30,625 (2.72)	3479/30,625 (11.36)	8780/30,625 (28.67)	9460/30,625 (30.89)	8073/30,625 (26.36)	16,363/30,625 (53.43)	14,262/30,625 (46.57)

Keywords, Related Terms, and Search Frequency

During the data-providing period from August 15, 2022, to August 13, 2023, 27,065 out of 31,200 words were identified as in-topic, representing valid user demand. The total BSI for these relevant user demand terms was 676,526,998, accounting for only 676,526,998 of 2,570,697,380 (15.74%) total search requests. Detailed distributions of these relevant terms and their search frequencies are shown in Figure 4. The valid search ratio and demand distribution were also summarized overall (Figure 4) and specifically for each cancer theme (Figures 5 and 6). Additionally, the 3 most representative user-demand issues were identified and ranked based on frequency and search popularity (Tables 2 and 3).

Figure 4. erm categories related to all cancers (bladder, kidney, and prostate) in the Baidu Index user demand module (August 2022 to August 2023). (A) The most frequently appearing related words (word units) in Baidu Index searches related to bladder, kidney, and prostate cancers. (B) The most searched related words in Baidu Index inquiries related to bladder, kidney, and prostate cancers. BSI: Baidu Search Index; TCM: traditional Chinese medicine.

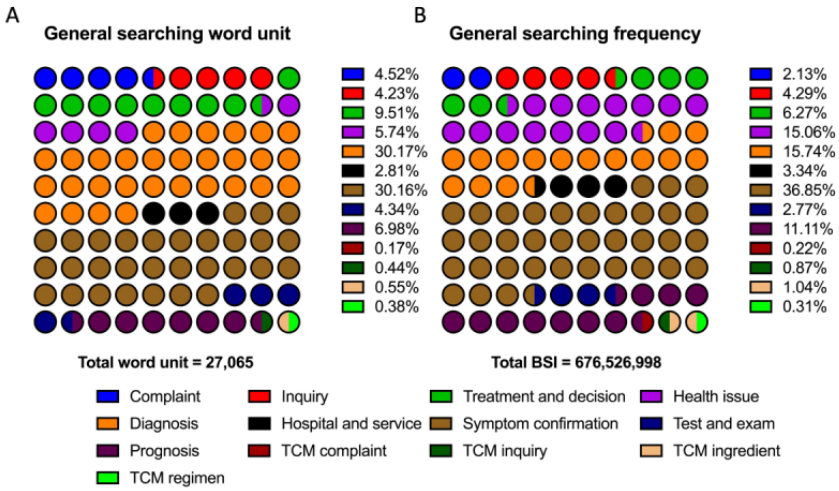


Figure 5. Term categories of the most frequently appearing related words for each cancer (August 2022 to August 2023). (A) Most frequently appearing related words (word units) in Baidu Index searches related to (A) bladder cancer, (B) kidney cancer, and (C) prostate cancer. TCM: traditional Chinese medicine.

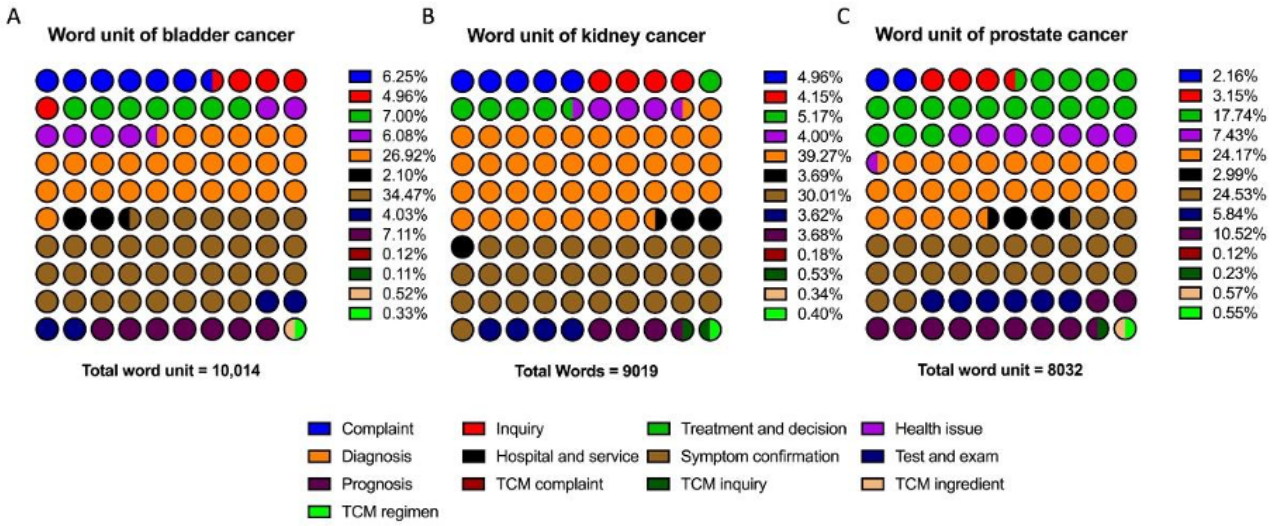


Figure 6. Term categories related to the most searched related words for each cancer (August 2022 to August 2023). Most searched related words (word units) in Baidu Index searches related to (A) bladder cancer, (B) kidney cancer, and (C) prostate cancer. BSI: Baidu Search Index; TCM: traditional Chinese medicine.

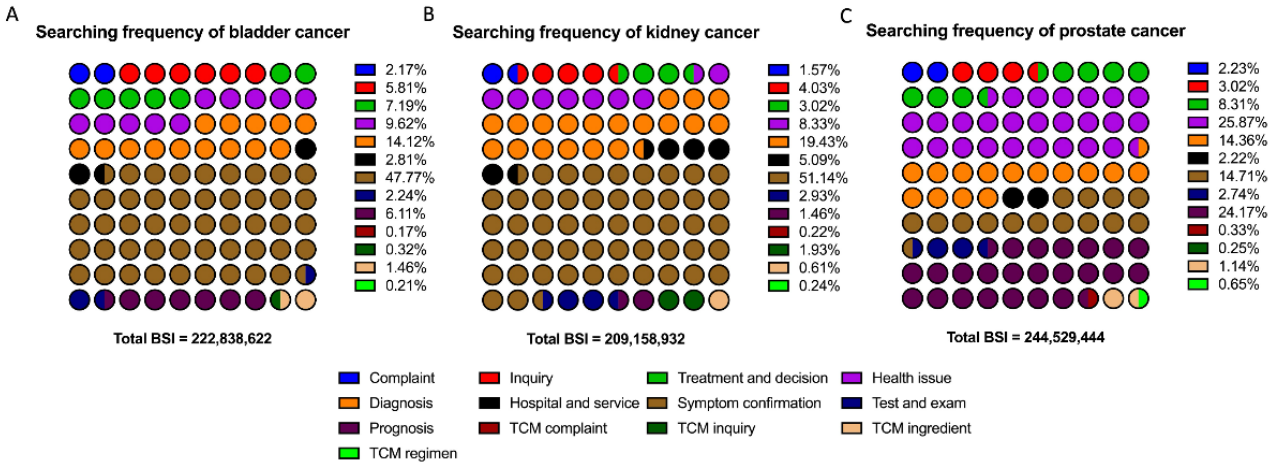


Table 2. The top 3 most frequently appearing related words (word units) searched in the Baidu Index for each type of cancer.

Category and term	Bladder cancer	Count, n	Kidney cancer	Count, n	Prostate cancer	Count, n
Complain						
Term 1	血尿(Hematuria)	96	血尿 (Hematuria)	51	尿潴留 (Urinary retention)	19
Term 2	尿血 (Urinate with blood)	79	无痛血尿 (Painless hematuria)	40	血精 (Hemospermia)	9
Term 3	无痛血尿 (Painless hematuria)	74	尿血 (Urinate with blood)	33	前列腺痛 (Prostatic pain)	7
Etiology and causes						
Term 1	小便时尿出血是怎么回事 (What is urinate with blood)	81	肾肿瘤分类 (Phenotypes in kidney cancer)	19	前列腺钙化是什么意思 (What is prostatic calcification)	34
Term 2	尿频尿急尿不尽是什么原因造成的 (What caused urinary frequency, urgency, and incomplete urinate)	26	肾积水是什么原因造成的怎么治疗 (What caused the hydronephrosis)	13	前列腺钙化灶是什么意思 (What does prostatic calcification mean)	26
Term 3	尿频繁是什么原因 (What caused frequently urination)	19	小便时尿出淡红色是怎么回事 (What caused reddish urination)	10	前列腺炎怎么引起的 (What caused prostatitis)	11
Treatment and pharmaceutical						
Term 1	膀胱癌治疗 (Treatment for bladder cancer)	115	肾癌的治疗 (Treatment for kidney cancer)	19	前列腺癌治疗 (Treatment for prostate cancer)	129
Term 2	膀胱炎怎么治疗 (Treatment for cystitis)	21	肾囊肿怎么治疗 (Treatment for renal cyst)	11	前列腺增生的最佳治疗方法 (Best treatment for prostate cancer)	96
Term 3	前列腺癌治疗 (Treatment for prostate cancer)	13	肾肿瘤切除 (Nephrectomy for kidney cancer)	10	前列腺癌的治疗 (Prostate cancer treatment)	74
Health care–related terms						
Term 1	膀胱 (Bladder)	169	肿瘤 (Tumor)	24	前列腺 (Prostate)	174
Term 2	四种癌已经不是癌了 (4 no longer diseases defined as cancer)	117	肾脏 (Kidney)	23	四种癌已经不是癌了 (4 no longer diseases defined as Cancer)	165
Term 3	芳香胺 (Aromatic amines)	28	肾癌饮食 (Diet for kidney)	17	前例腺 (“Prastate”)	7
Diagnosis						
Term 1	膀胱癌 (Bladder cancer)	499	肾癌 (Kidney cancer)	395	前列腺癌 (Prostate cancer)	422
Term 2	膀胱肿瘤 (Bladder tumor)	180	肾囊肿 (Renal cyst)	242	前列腺癌骨转移 (Prostate cancer with bone metastasis)	115
Term 3	膀胱炎 (Cystitis)	169	肾肿瘤 (Tumors in kidney)	230	前列腺炎 (Prostatitis)	108
Health care services and commodities						
Term 1	北京大学第一医院 (The first affiliate Hospital of PKU)	4	中国最好的肾病医院 (The best hospital for treating kidney diseases)	42	海外医疗 (Oversee medication service)	15
Term 2	泌尿外科哪个医院好 (What is the best urology hospital)	4	治疗肾癌最好的医院 (The best hospital for treating kidney cancers)	30	麻省总医院 (Massachusetts General Hospital)	14
Term 3	吴阶平 (Prof Jiepin Wu)	3	肿瘤医院全国排名一 (National NO.1 Oncology Hospital)	10	厚朴方舟 (Hopenoak.com)	7
Diagnosis confirmation						
Term 1	膀胱癌早期是什么症状 (Early symptoms of bladder cancer)	431	肾癌早期的五个表现 (Early symptoms of bladder cancer)	376	前列腺癌症状有哪些 (What are the symptoms of prostate cancer)	356

Category and term	Bladder cancer	Count, n	Kidney cancer	Count, n	Prostate cancer	Count, n
Term 2	膀胱炎是什么症状表现 (What are the symptoms of cystitis)	204	肾癌症状 (Symptoms of kidney cancer)	134	前列腺癌症状 (Symptoms of prostate cancer)	149
Term 3	膀胱癌症状 (Symptoms of bladder cancer)	202	肾衰竭的早期症状表现 (Early symptoms of kidney cancer)	123	前列腺癌的症状 (Prostate cancer symptoms)	226
Test and examination						
Term 1	膀胱镜 (Cystoscope)	140	肾钙化 (Renal calcification)	12	前列腺特异性抗原 (Prostate specific antigen)	98
Term 2	膀胱检查 (Bladder examination)	19	肾功能检查哪些项目 (Items of renal function test)	10	PSA (PSA)	72
Term 3	尿常规能检查出什么 (What can routine urinary test tell)	10	肾穿刺 (Renal puncturing)	10	前列腺炎一杯水自测 (Confirming prostatitis with a cup of water)	30
Prognosis						
Term 1	膀胱癌能活多久 (How long one can live with diagnosed bladder cancer)	210	肾癌晚期能活多久 (How long one can live with diagnosed late-stage kidney cancer)	51	前列腺癌能活多久 (How long one can live with diagnosed prostate cancer)	342
Term 2	膀胱癌晚期 (Late-stage bladder cancer)	125	晚期肾癌 (Late-stage kidney cancer)	17	前列腺炎有什么症状和危害性 (Symptoms and hazard of prostatitis)	213
Term 3	前列腺癌能活多久 (How long one can live with diagnosed prostate cancer)	67	肺癌晚期能活多久 (How long one can live with diagnosed late-stage lung cancer)	16	前列腺癌晚期 (Late-stage prostate cancer)	122
TCM[a] diagnosis						
Term 1	疾在腠理 (Disease sign on the skin)	1	肾虚 (Insufficiency in “Shen” essence)	3	湿热症疹状 (Rashes of the humid heat symptoms)	2
Term 2	肾精亏耗 (Depletion of “Shen” essence)	1	肾阳虚 (Insufficiency in “Shen” essence of “Yang”)	2	五心烦热 (Sphoria with feverish sensation in chest, palms, and soles)	1
Term 3	肾虚 (Insufficiency in “Shen” essence)	1	肾阴虚 (Insufficiency in “Shen” essence of “Yin”)	2	湿热疹 (Rashes of the humid heat)	1
TCM diagnosis confirmation						
Term 1	肾虚的表现症状有哪些 (What are the symptoms of insufficiency in “Shen” essence)	3	肾虚的表现症状有哪些 (What are the symptoms of insufficiency in “Shen” essence)	21	中医治疗前列腺 (TCM treatment of prostate gland)	2
Term 2	肾虚的症状 (Symptoms of insufficiency in “Shen” essence)	2	如何保养肾 (How to maintain kidney with “Shen” essence)	2	中医治疗癌症 (TCM treatment of cancer)	1
Term 3	拔罐的好处与功效 (Benefit and efficacy of cupping cup)	1	肾俞穴 (Acupoint of “Shen Yu”)	2	中医治疗肿瘤 (TCM treatment of tumor)	1
TCM regimen						
Term 1	五味子 (*Schisandra chinensis* Turcz. Baill.)	2	丝瓜子 (Loofah seed)	1	车前草 (Plantago asiatica L.)	2
Term 2	黑枸杞的作用与功效 (The efficacy of Lycium ruthenicum Murr)	2	丝瓜的功效与作用禁忌 (The efficacy and contraindication of loofah)	1	三七的副作用太大了 (The side effect of Panax pseudoginsenga)	1
Term 3	东革阿里的功效 (The efficacy of Tongkat Ali)	2	丝瓜蒂 (*Luffa cylindrica* （L.） Roem.)	1	东阿阿胶250克价格 (Price of donkey-hide gelatin)	1
TCM remedy and materials						

Category and term	Bladder cancer	Count, n	Kidney cancer	Count, n	Prostate cancer	Count, n
Term 1	银花泌炎灵片 (Tablet of “YinHuaMiYanLin”)	3	云南白药气雾剂的作用与功效 (The efficacy of “Yunnan Baiyao” spray)	1	金水宝胶囊的功效与主治 (The efficacy and indications of “JinshuiBao” caspule)	3
Term 2	仙鹤神针 (The miraculous needle of “Crane”)	2	加味二陈汤 (Potion of two old ingredient with extra additional)	1	抗肿瘤最强的中草药 (The strongest anti-tumor TCM herb)	2
Term 3	桂枝茯苓丸 (Guizhi Ling Pills)	2	华蟾素 (Cinobufagin)	1	乌头赤石脂丸 (Aconite red halloysite)	1

[a]TCM: traditional Chinese medicine.

Table 3. The top 3 most searched related words in the Baidu Index for each type of cancer.

Category and terms	Bladder cancer	Baidu Search Index	Renal cancer	Baidu Search Index	Prostate cancer	Baidu Search Index
Complain						
Term 1	尿血 (Urinate with blood)	11,24,546	尿隐血 (Occult hematuria)	5,68,694	白肺 (“White Lung”)	36,90,652
Term 2	血尿 (Hematuria)	9,34,444	血尿 (Hematuria)	4,91,164	尿潴留 (Urinary retention)	2,08,682
Term 3	小便尿完过一会又想尿 (Small moment of urinate urge after pee)	5,17,276	尿血 (Urinate blood)	4,82,462	阳痿 (Impotence)	1,46,838
Etiology and causes						
Term 1	尿频尿急尿不尽是什么原因造成的 (What caused urinary frequency, urgency, and incomplete urinate)	25,14,540	白肺是什么意思 (What does the “White Lung” mean)	7,60,196	前列腺钙化是什么意思 (What does prostate calcification mean)	11,11,802
Term 2	小便时尿出血是怎么回事 (Why is urinate with blood)	24,30,620	尿酸高是什么引起的原因 (What causes hyperuricemia)	5,51,308	前列腺钙化灶是什么意思 (What is prostatic calcification)	10,02,476
Term 3	尿频繁是什么原因 (What causes frequently urination)	11,30,904	尿液发红褐色怎么回事 (What causes reddish urine)	4,84,088	前列腺炎怎么引起的 (What causes prostatitis)	8,72,092
Treatment and pharmaceutical						
Term 1	甲流吃什么药效果最好 (What is the medication of Influenza A virus)	19,31,970	肾结石怎么排出来最快方法 (The fastest way of urinate out the kidney stone)	4,53,232	前列腺增生的最佳治疗方法 (The best way of treating BPH)	51,31,986
Term 2	布洛芬混悬液 (Ibuprofen Suspension)	18,84,472	靶向治疗是什么意思 (What does the targeted therapy mean)	2,30,322	前列腺增生的症状 (Symptoms of BPH)	18,04,768
Term 3	尿路感染10分钟解决方法 (Method of eradicate UTI within 10 minutes)	10,87,080	腰椎间盘突出最好的治疗方法 (Best treatment for lumber disc protrusion)	1,88,922	前列腺炎吃什么药效果好见效快 (What medication for prostatitis effect promptly)	11,47,908
Health care–related terms						
Term 1	中国知网 (CKNI.COM)	35,69,116	中国知网 (CKNI.COM)	72,65,752	前列腺 (Prostatitis)	1,46,50,248
Term 2	四种癌已经不是癌了 (4 no longer cancer defined cancers)	30,23,448	咸阳疫情最新消息 (The latest pandemic news in Xianyang City)	6,48,786	四种癌已经不是癌了 (4 no longer cancer defined cancers)	43,33,372
Term 3	膀胱 Bladder()	25,49,522	肿瘤 (Tumor)	4,80,378	中国知网 (CKNI.COM)	22,87,492
Diagnosis						
Term 1	膀胱癌 (Bladder caner)	59,88,140	肾囊肿 (Renal cyst)	91,41,308	前列腺炎 (Prostatitis)	1,20,84,428
Term 2	膀胱炎 (Cystitis)	36,16,898	肾结石 (Nephrolithiasis)	36,13,030	前列腺癌 (Prostate cancer)	93,56,312
Term 3	前列腺癌 (Prostate cancer)	27,19,162	肾癌 (Kidney cancer)	32,98,004	前列腺增生 (Benign prostate hyperplasia)	22,11,278
Health care services and commodities						
Term 1	山西医科大学 (Shanxi Medical School)	4,62,288	北京大学 (Paikin University)	4,80,608	华中科技大学 (Huazhong University of Science and Technology)	7,48,616

Category and terms	Bladder cancer	Baidu Search Index	Renal cancer	Baidu Search Index	Prostate cancer	Baidu Search Index
Term 2	哈尔滨医科大学 (Harbin medical university)	4,17,328	吉林大学 (Jilin university)	2,53,870	男科医院 (Andrology hospital)	2,96,596
Term 3	问医生 (Ask Doctor.com)	1,86,720	复旦大学 (Fudan University)	2,07,778	百度健康 (Baidu Health)	2,61,804
Diagnosis confirmation						
Term 1	甲流感症状有哪些 (What are the symptoms of influenza A)	1,94,65,776	肾衰竭的早期症状表现 (Early symptoms of renal failure)	1,29,43,182	前列腺癌症状有哪些 (What are the symptoms of prostate cancer)	70,24,984
Term 2	膀胱炎是什么症状表现 (What are the symptoms of cystitis)	1,27,22,312	肾炎的症状是什么 (What are the symptoms of nephritis)	1,02,09,324	胰腺癌的早期症状 (What are the symptoms of pancreas cancer)	21,36,332
Term 3	膀胱癌早期是什么症状 (What are the early symptoms of bladder cancer)	86,79,854	尿毒症的早期症状 (What are the early symptoms of uremia)	95,76,026	如何判断自己前列腺炎 (How to determine prostatitis by my self)	21,03,682
Test and examination						
Term 1	血氧饱和度 (Blood oxygen saturation)	3,72,030	肌酐高是什么问题 (What are the problems causing high creatine level)	6,23,962	PSA (“PSA”)	7,98,870
Term 2	膀胱镜 (Cystoscope)	3,29,426	肾功能检查哪些项目 (What are the items of renal function test)	3,92,388	穿刺检查是什么意思 (What does “puncture & biopsy” means)	6,59,994
Term 3	血糖正常值 (Normal level of plasma glycose level)	2,99,716	血糖正常值 (Normal level of plasma glycose level)	2,92,004	前列腺炎一杯水自测 (Confirming prostatitis with a cup of water)	5,80,210
Prognosis						
Term 1	前列腺炎有什么症状和危害性 (Symptoms and hazard of prostatitis)	97,45,516	肺癌晚期能活多久 (How long one can live with late phase lung cancer)	8,41,568	前列腺炎有什么症状和危害性 (Symptoms and hazard of prostatitis)	5,15,38,700
Term 2	前列腺癌能活多久 (How long one can live with prostate cancer)	8,44,452	肺癌晚期最怕三个征兆 (The three poorest indications in late phase lung cancer)	3,65,816	前列腺癌能活多久 (How long one can live with prostate cancer)	42,52,848
Term 3	肺癌晚期能活多久 (How long one can live with late phase lung cancer)	7,57,840	不化疗和化疗哪个寿命长 (Chemo, or non-Chemo, which to choose for longer life)	2,77,808	白肺是可以治愈的吗 (Is “white lung” curable ())	10,53,438
TCM[a] >diagnosis						
Term 1	舌苔发黄厚腻是什么原因怎么调理 (What causes thick and yellow tongue coating, how to moderate)	60,686	舌苔厚白是什么原因引起的怎么解决 (What causes thick and white tongue coating, how to moderate)	1,67,564	风热感冒和风寒感冒的症状区别 (Difference between “Feng heat” and “Feng cold” fever)	3,24,112
Term 2	飞机打多了属于阴虚还是阳虚 (Is too much masturbation causing essence deficiency in “Yin” or “Yang”)	33,180	肾阴虚 (Deficiency in “Shen” essence of “Yin”)	68,140	湿热疹 (Rash of humid heat)	2,01,220
Term 3	脾虚 (Deficiency in spleen essence)	24,362	肾阳虚 (Deficiency in “Shen” essence of “Yang”)	60,174	肾阴虚 (Deficiency in Shen” essence of “yin”)	33,262

Category and terms	Bladder cancer	Baidu Search Index	Renal cancer	Baidu Search Index	Prostate cancer	Baidu Search Index
TCM diagnosis confirmation						
Term 1	肾虚的表现症状有哪些 (What are the symptoms of deficiency in “Shen” essence)	6,83,020	肾虚的表现症状有哪些 (What are the symptoms of deficiency in “Shen” essence)	32,80,446	肾虚的表现症状有哪些 (What are the symptoms of deficiency in “Shen” essence)	1,66,796
Term 2	脾虚的表现和症状 (What are the symptoms of deficiency in “Pi” essence)	1,36,110	脾虚的表现和症状 (What are the symptoms of deficiency in “Pi” essence)	1,29,000	湿气重怎么排湿最有效 (How to expel humid “Qi” effectively when bearing too much humid “Qi”)	1,12,650
Term 3	肠胃不好怎么调理最有效 (How to moderate gastrointestinal function)	48,930	飞机打多了该怎么补回来 (How to compensate by supplement after loads of masturbation)	1,16,264	脾胃虚弱怎么调理最快 (What is the fastest way to moderate feeble “Pi and Wei”)	64,500
TCM regimen						
Term 1	金银花的功效与作用 (The efficacy of Honeysuckle)	5,70,726	山药的功效与作用 (The efficacy of Chinese yam)	2,01,052	茯苓的功效与作用 (The efficacy of poria)	3,72,518
Term 2	黑枸杞的作用与功效 (The efficacy of Lycium ruthenicum Murr)	3,86,932	甘草 (liquorice)	1,56,030	姜的功效与作用 (The efficacy of ginger)	2,59,726
Term 3	五味子 (Schisandra chinensis (Turcz.) Baill.)	2,83,100	蒲公英的功效与作用 (The efficacy of dandelion)	1,36,652	西洋参 (American ginseng)	2,08,448
TCM remedy and materials						
Term 1	右归丸的作用和功效 (The efficacy of “YouGuiWan” pill)	70,252	云南白药气雾剂的作用与功效 (The efficacy of “baiyao”， Yunnan)	41,140	六味地黄丸有什么功效与作用 (The efficacy of “LiuWeiDiHuangWan” pill)	3,43,428
Term 2	桂枝茯苓丸 (“GuizhiFulingWan” Pill)	62,076	四君子汤的功效与作用 (The efficacy of “Sijunzitang” potion)	32,698	金水宝胶囊的功效与主治 (The efficacy of “Jinshuibao” Capsule)	1,50,424
Term 3	龙胆泻肝丸 (“LongDanXieGanWan” Pill)	61,236	百令胶囊功效与作用 (The efficacy of “Bailing” Capsule)	30,136	玉屏风颗粒的功效与作用 (The efficacy of “Yupingfeng” electuary)	1,40,016

[a]TCM: traditional Chinese medicine.

Discussion

Principal Findings

To the best of our knowledge, this study is the first infodemiology research to explore patients’ awareness and demand for primary urologic cancers (bladder, kidney, and prostate) within China’s vast population, particularly from clinical and health care perspectives [12,22]. By analyzing the most widely used local search platform, with billions of daily active queries, we identified consistent shifts in search trends, dominant regions for searches, and key demographic groups. Furthermore, we examined the most sought-after topics, reflecting user-initiated care-seeking behaviors and decision-making patterns [13].

We observed that while the overtime search trends fluctuate, the overall search volume for each cancer type has shown a general increase, as indicated by the overtime AAPC. Across all urologic cancer themes, the 4 main topics—“complaint,” “inquiry,” “treatment,” and “prognosis”—remain consistent, indicating that user queries focus on key aspects of disease-related decision-making. Notably, there is a strong demand for information on diagnostic criteria, etiology, treatment options, and realistic expectations for each cancer. Among tumor-related keywords, symptom-related searches account for 43.95%, 49.97%, and 36.89% of the total search volume for bladder, kidney, and prostate cancers, respectively. Inquiries, comprising 23 search keywords, account for 46.30%, 46.23%, and 42.64% of the total search volume for bladder, kidney, and prostate cancers, respectively. In terms of treatment-related searches, prostate cancer leads with 14.51% of the total search volume, followed by bladder cancer at 7.49% and kidney cancer at 2.41%. The prominence of searches related to complaints and inquiries is unsurprising, as the complaint

category primarily includes diagnostic keywords, reflecting users' initial, exploratory searches when they are uncertain about the specific information they require [22].

The main concerns in user inquiries revolve around early signs, primary indications, etiologies, specific symptoms, and other cancer-related issues. These topics represent a comprehensive collection of the most common challenges patients face during health care–seeking sessions, particularly in cancer-related scenarios.

In bladder cancer, search keywords frequently reflect concerns about symptoms, particularly in the early or late stages, with hematuria being a primary focus [23]. While hematuria does not definitively indicate bladder cancer, prompt attention to this visible symptom facilitates early detection and diagnosis.

For prostate cancer, user queries often center on symptoms, staging, and metastasis. Lower urinary tract symptoms are commonly reported but frequently stem from benign causes, making symptom-based identification challenging. Prostate cancer detection primarily relies on prostate-specific antigen-magnetic resonance imaging-biopsy combinations, though the screening sensitivity (0.93) and specificity (0.20) underscore limitations and adherence challenges [24-27]. Furthermore, concerns about skeletal metastasis and staging highlight the importance of enhanced education and communication strategies.

Kidney cancers, characterized by diverse carcinomas, are often identified through imaging as "space-occupying lesions," resulting in ambiguous search terms like "mass (肿物)." Compared with bladder and prostate cancers, users frequently lack precise diagnostic information [28]. Advances in imaging technology have increased renal cell carcinoma detection rates by 3.1% annually over the past decade [29,30]. However, diagnostic accuracy remains limited by tumor size and the high cost of imaging. Emerging machine learning systems offer promise, with diagnostic precision ranging from 84.18% to 90.83%, supporting cancer identification and surgical decision-making [29,30]. Addressing misdiagnosis and delays in renal carcinoma detection remains a critical priority for health care providers.

Life expectancy is a primary concern in users' inquiries about the 3 cancers. The standard query, "How long can one live with the diagnosis?" highlights that the most significant need for patients with cancer, in terms of treatment or intervention, is to minimize the negative impact of cancer on life expectancy. Currently, the growing population in Mainland China exacerbates the challenges posed by the morbidity and mortality of urologic cancers, affecting both the national health care system and personal lives [2]. As of 2019, the death rates for bladder cancer, kidney cancer, and prostate cancer in China were 50.39/1000, 40.09/1000, and 23.95/1000, respectively, with mortality rates gradually increasing with age [2]. Predictive models suggest that the morbidity and mortality rates for these 3 genitourinary cancers will continue to rise, although health care facilities are becoming more accessible in traditionally less-developed regions [31]. With improvements in perioperative management and increasing proficiency among surgeons in primary health care facilities, internet users have expressed concerns that could be addressed more promptly and validated by local health care professionals [31,32].

The search trend patterns across each cancer theme showed similarities, with an overall upward trend primarily driven by the topics of complaint and inquiry. The demand for cancer treatment and prognosis has generally declined, with the only exception being observed in prostate cancer treatment topics. Specifically, the search trends for bladder and kidney cancers increased until 2019, followed by a decline until a resurgence in 2021. For prostate cancer, turning points occurred in 2015 and 2020. The rising search trends may partly reflect real-world cancer incidence. Data from the GBD 2019 database indicated that the counts for bladder, kidney, and prostate cancers in 2019 were at least four times higher, with age-standardized incidence rates at least double those of 1990 [2]. This period also marked Baidu's "golden age," during which it became the primary source of information for internet users [13]. Although Baidu's marketing strategies for channeling information have been criticized, they do not appear to have affected its dominant role in providing cancer-related information to users [13].

Despite the COVID-19 pandemic's significant impact on regular medical activities and the shift in public attention toward outbreak management, online demand for information on tumor diseases, including genitourinary cancers, persisted and even increased in 2019 [6,33]. This demand was temporarily suppressed during the 2020-2021 pandemic peak but rebounded as government policies and health care responses evolved [34,35]. Search terms highlighted unresolved issues and user concerns about genitourinary cancers, emphasizing Baidu's role as a trusted source of health-related information. These trends underscore the need to address user-identified health care gaps in future health policy optimization.

Geographically, searches for the 3 cancers were led by East China, accounting for approximately 30%, followed by North China at around 18%. Other regions showed similar levels of interest, each at approximately 10%, while the northwest region ranked last with about 8%.

Socioeconomic inequalities significantly influence health-seeking behavior and online health information searches [33,36]. In China, disparities in medical center density, health care costs, social security policies, and household finances shape patients' decisions, with individuals from disadvantaged backgrounds often hesitating to seek treatment due to financial constraints [37]. Similarly, socioeconomic and educational factors affect online search behavior: higher-income individuals tend to exhibit greater digital literacy, use precise medical terminology, and seek reliable sources [38,39]. By contrast, lower-income groups face barriers such as limited internet access and lower trust in online information, often relying on symptom-based searches [40]. These differences underscore the impact of socioeconomic status on health awareness and access to accurate medical information [39,41].

Demographic data revealed similar interest in cancer topics across genders, though females expressed greater concern about kidney cancer prognosis. Bladder and kidney cancers have significantly higher incidence rates in males, at 5-fold and 2.5-fold, respectively. Interest in prostate cancer, despite being

male-specific, indicates engagement extending beyond patients alone [2]. Agewise, searches for bladder and kidney cancers were common among users aged 20-39 years, with kidney cancer treatment queries peaking in the 40-49-year age group. Prostate cancer searches followed a similar pattern, except prognosis-related queries, which were predominantly from users aged 40-49 years. Bladder and kidney cancer incidence rates in individuals aged 20-39 years (25.48/100,000 and 1.51/100,000, respectively) were half of those observed in the 40-59-year age group, while prostate cancer incidence peaked in individuals over 60 years [2].

The disparity between incidence rates and concern levels across gender and age groups highlights the importance of health consciousness and awareness. The high cost, complex nursing requirements, and strict follow-up schedules associated with cancer treatment underscore that the disease burden affects not only individuals but also entire families [42]. Research indicates that women play a prominent leadership role within families, making approximately 80% of health care–related decisions [43-45]. The concerns and demands of women, as key decision makers within families, should not be underestimated. Although older adults are increasingly integrating internet technology into their daily routines, including health care–seeking activities, the persistent digital divide among older users remains a significant challenge [46]. Therefore, joint decision-making involving family members, rather than focusing solely on the patients' needs, should be taken into account.

Our study found that only 15.74% of user search queries were categorized as relevant, reflecting a focus on trending topics, nonhealth issues, or vague terms, such as medication-related words. This low percentage suggests that public attention often shifts toward less scientifically grounded information, driven by societal trends, misinformation, or curiosity about lifestyle, diet, or alternative medicine [47,48]. The findings highlight the need for more effective dissemination of accurate cancer information and targeted educational campaigns to enhance public understanding of critical cancer issues. Furthermore, many users rely on general or indirect language in their searches, which may prevent their queries from being classified as relevant. This underscores the importance of promoting more precise health communication.

Overall, frequency, diagnosis, and symptom confirmation issues ranked among the top categories, accounting for 30.17% and 30.16%, respectively. Regarding popularity, interest in symptom confirmation ranked first, followed by diagnosis and health-related issues, each comprising approximately 15%. Similar patterns were observed in bladder and kidney cancers. However, in prostate cancer, while interest in diagnosis and symptom confirmation was dominant in terms of frequency, health care–related issues in prognosis garnered the most actual popularity.

In the topics of complaints and etiology inquiries, the most common concerns related to bladder and kidney cancers were hematuria and questions about its causes. As a marker of cancer risk, hematuria is often associated with underlying urologic malignancies. The standardized incidence ratio for overall urologic cancer risk peaks within the first 3 months following a hematuria diagnosis, reaching 14.15% [49]. Specifically, the standard incidence ratios for bladder, kidney, and prostate cancers at 3 months are 186.43%, 81.40%, and 14.18%, respectively [49]. Reports indicate that delayed responses to initial hematuria often lead to missed cancer diagnoses. Therefore, heightened awareness and timely investigation of its causes are recommended to prevent delays in cancer diagnosis [50,51].

We observed that several differential diagnoses were listed within the diagnosis topic. For bladder cancer, cystitis and prostate cancer were frequently mentioned. In the context of kidney cancer and prostate cancer, the most commonly referred differential diagnoses included renal cysts, renal kidney disease, prostatitis, and benign prostatic hyperplasia. These conditions must be ruled out before confirming a tumor diagnosis, suggesting the potential for users to engage in self-assisted cyber diagnosis [52].

Accurate evidence is essential for the correct diagnosis of cancer, as it helps identify tumor characteristics. In some cancer cases, even experienced experts may find judgment and decision-making complicated due to varying test results [53]. The risk of internet self-diagnosis in cancer warrants greater attention, as diagnosis delays and misleading treatments can have serious consequences [52]. For users with limited medical knowledge, it becomes even more challenging to identify accurate and useful information while maintaining reasonable and objective expectations [54]. Therefore, online health care information should be more instructive, neutral, and objective, enabling potential patients to better follow guidance from medical professionals [55].

Limitations

Several limitations of this study should be acknowledged. First, although Baidu is the largest search engine in mainland China, its dominance is increasingly challenged by emerging search platforms and social media, which restricts the scope of online search behavior captured. Consequently, the daily BSI may not fully reflect the demands of all internet users. Second, regional differences in user preferences and political regulations result in the exclusion of searches by individuals legally accessing international platforms. Furthermore, as Baidu is a Chinese platform, searches conducted in foreign languages or by ethnic minorities may be systematically omitted, potentially introducing bias. Third, the absence of a real-time public cancer database is a significant limitation, as it prevents the integration of search data with clinical data for more accurate disease prediction and forecasting. These factors underscore biases in internet access and usage across different demographic groups, which should be considered when interpreting the results in the context of broader cancer awareness and digital health information–seeking behavior.

Future Directions

This study is the first to address public concerns regarding the 3 major genitourinary cancers within the Chinese-speaking population. These cancers were selected due to their complexity and subtle onset, which often lead to delayed diagnoses and severe outcomes. By analyzing online search trends, this

research provides valuable insights into patient perceptions and needs, offering a broader understanding of public demand. Future infodemiology studies should incorporate data from multiple search engines, social media platforms, and multilingual or minority groups to achieve a more comprehensive analysis of public health trends. Integrating such a system into a national cancer database could significantly enhance disease tracking, forecasting, and public health decision-making.

Conclusions

Online searches and inquiries related to genitourinary cancers are on the rise. The depth of users' information demands appears to be influenced by regional economic levels. Cancer treatment decision-making may often involve a family-centered approach. Insights from internet search data are potentially beneficial for medical professionals to better understand public interests and concerns, enabling them to provide more targeted and reliable health care services.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Use of Artificial Intelligence

Artificial intelligence or chatbot tools were used exclusively for language editing in this manuscript. No content generation or ideation was performed by the artificial intelligence tools or chatbot.

Authors' Contributions

Conceptualization was carried out by SZW. Formal analysis was performed by XW and XDX. The investigation was conducted by XDX and SQL, while methodology was developed by SZW and XW. Project administration was handled by SZW and XBL, with resources provided by HCC and XDX. The original draft of the manuscript was written by HCC, XDX, and SZW, Administration support was provided by LXB, and all authors contributed to reviewing and editing the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of keywords used in composite search index.
[DOCX File , 18 KB - cancer_v11i1e57414_app1.docx]

References

1. Gammall J, Lai AG. Pan-cancer prognostic genetic mutations and clinicopathological factors associated with survival outcomes: a systematic review. NPJ Precis Oncol 2022 Apr 20;6(1):27 [FREE Full text] [doi: 10.1038/s41698-022-00269-5] [Medline: 35444210]
2. Huang Q, Zi H, Luo L, Li X, Zhu C, Zeng X. Secular trends of morbidity and mortality of prostate, bladder, and kidney cancers in China, 1990 to 2019 and their predictions to 2030. BMC Cancer 2022 Nov 11;22(1):1164 [FREE Full text] [doi: 10.1186/s12885-022-10244-9] [Medline: 36368976]
3. Zi H, He S, Leng X, Xu X, Huang Q, Weng H, et al. Global, regional, and national burden of kidney, bladder, and prostate cancers and their attributable risk factors, 1990-2019. Mil Med Res 2021 Nov 24;8(1):60 [FREE Full text] [doi: 10.1186/s40779-021-00354-z] [Medline: 34819142]
4. Mitchell KA, Williams H. Emerging genomic biomarkers for improving kidney, prostate, and bladder cancer health disparities outcomes. Urol Oncol 2022 Apr;40(4):126-132. [doi: 10.1016/j.urolonc.2019.04.024] [Medline: 31239186]
5. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
6. Jiang L, Ma Q, Wei S, Che G. Online public attention of COVID-19 vaccination in Mainland China. Digit Health 2022;8:20552076211070454 [FREE Full text] [doi: 10.1177/20552076211070454] [Medline: 35096408]
7. Li K, Liu M, Feng Y, Ning C, Ou W, Sun J, et al. Using Baidu search engine to monitor AIDS epidemics inform for targeted intervention of HIV/AIDS in China. Sci Rep 2019 Jan 23;9(1):320 [FREE Full text] [doi: 10.1038/s41598-018-35685-w] [Medline: 30674890]
8. Saegner T, Austys D. Forecasting and surveillance of COVID-19 spread using Google Trends: literature review. Int J Environ Res Public Health 2022 Sep 29;19(19):12394 [FREE Full text] [doi: 10.3390/ijerph191912394] [Medline: 36231693]
9. Gong X, Han Y, Hou M, Guo R. Online public attention during the early days of the COVID-19 pandemic: infoveillance study based on Baidu Index. JMIR Public Health Surveill 2020 Oct 22;6(4):e23098 [FREE Full text] [doi: 10.2196/23098] [Medline: 32960177]

10. Carrera PM, Kantarjian HM, Blinder VS. The financial burden and distress of patients with cancer: understanding and stepping-up action on the financial toxicity of cancer treatment. CA Cancer J Clin 2018 Mar;68(2):153-165 [FREE Full text] [doi: 10.3322/caac.21443] [Medline: 29338071]
11. Luengo-Fernandez R, Leal J, Gray A, Sullivan R. Economic burden of cancer across the European Union: a population-based cost analysis. Lancet Oncol 2013 Nov;14(12):1165-1174. [doi: 10.1016/S1470-2045(13)70442-X] [Medline: 24131614]
12. Wei S, Ma M, Wu C, Yu B, Jiang L, Wen X, et al. Using search trends to analyze web-based interest in lower urinary tract symptoms-related inquiries, diagnoses, and treatments in Mainland China: infodemiology study of Baidu Index data. J Med Internet Res 2021 Jul 06;23(7):e27029 [FREE Full text] [doi: 10.2196/27029] [Medline: 34255683]
13. Wei S, Ma M, Wen X, Wu C, Zhu G, Zhou X. Online public attention toward premature ejaculation in Mainland China: infodemiology study using the Baidu Index. J Med Internet Res 2021 Aug 26;23(8):e30271 [FREE Full text] [doi: 10.2196/30271] [Medline: 34435970]
14. Xu C, Wang Y, Yang H, Hou J, Sun L, Zhang X, et al. Association between cancer incidence and mortality in web-based data in China: infodemiology study. J Med Internet Res 2019 Jan 29;21(1):e10677 [FREE Full text] [doi: 10.2196/10677] [Medline: 30694203]
15. Clement J. Market share of leading desktop search engines worldwide from January 2015 to January 2025. Statista. 2025 Jan 23. URL: https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ [accessed 2025-01-28]
16. Xie T, Tan T, Li J. An extensive search trends-based analysis of public attention on social media in the early outbreak of COVID-19 in China. Risk Manag Healthc Policy 2020;13:1353-1364 [FREE Full text] [doi: 10.2147/RMHP.S257473] [Medline: 32943953]
17. Liu Y, Wang Z, Rader B, Li B, Wu C, Whittington JD, et al. Associations between changes in population mobility in response to the COVID-19 pandemic and socioeconomic factors at the city level in China and country level worldwide: a retrospective, observational study. The Lancet Digital Health 2021 Jun;3(6):e349-e359. [doi: 10.1016/s2589-7500(21)00059-5]
18. Wang Z, He J, Jin B, Zhang L, Han C, Wang M, et al. Using Baidu Index data to improve chickenpox surveillance in Yunnan, China: infodemiology study. J Med Internet Res 2023 May 16;25:e44186 [FREE Full text] [doi: 10.2196/44186] [Medline: 37191983]
19. Baidu Index. Baidu. URL: https://index.baidu.com/v2/main/index.html#/help [accessed 2025-01-28]
20. Fay MP, Tiwari RC, Feuer EJ, Zou Z. Estimating average annual percent change for disease rates without assuming constant change. Biometrics 2006 Sep;62(3):847-854. [doi: 10.1111/j.1541-0420.2006.00528.x] [Medline: 16984328]
21. SEER*Stat Trends Exercise 1axamine Trends in Rates. National Cancer Institute. URL: https://seer.cancer.gov/seerstat/tutorials/trends1a/webprint/ [accessed 2025-06-01]
22. Guérin S, Khene Z, Peyronnet B, Bensalah K, Zimmern PE. Google searches for stress urinary incontinence treatment options: a worldwide trend analysis in the synthetic mesh controversy era. World J Urol 2023 Aug;41(8):2217-2223. [doi: 10.1007/s00345-023-04474-0] [Medline: 37358599]
23. Lenis AT, Lec PM, Chamie K, Mshs MD. Bladder cancer: a review. JAMA 2020 Nov 17;324(19):1980-1991. [doi: 10.1001/jama.2020.17598] [Medline: 33201207]
24. Hamilton W, Sharp D. Symptomatic diagnosis of prostate cancer in primary care: a structured review. Br J Gen Pract 2004 Aug;54(505):617-621 [FREE Full text] [Medline: 15296564]
25. Merriel SWD, Pocock L, Gilbert E, Creavin S, Walter FM, Spencer A, et al. Systematic review and meta-analysis of the diagnostic accuracy of prostate-specific antigen (PSA) for the detection of prostate cancer in symptomatic patients. BMC Med 2022 Feb 07;20(1):54 [FREE Full text] [doi: 10.1186/s12916-021-02230-y] [Medline: 35125113]
26. Kith G, Lisker S, Sarkar U, Barr-Walker J, Breyer BN, Palmer NR. Defining and measuring adherence in observational studies assessing outcomes of real-world active surveillance for prostate cancer: a systematic review. Eur Urol Oncol 2021 Apr;4(2):192-201 [FREE Full text] [doi: 10.1016/j.euo.2019.06.009] [Medline: 31288992]
27. Miocinovic R, Jones JS, Pujara AC, Klein EA, Stephenson AJ. Acceptance and durability of surveillance as a management choice in men with screen-detected, low-risk prostate cancer: improved outcomes with stringent enrollment criteria. Urology 2011 Apr;77(4):980-984. [doi: 10.1016/j.urology.2010.09.063] [Medline: 21256549]
28. Alaghehbandan R, Siadat F, Trpkov K. What's new in the WHO 2022 classification of kidney tumours? Pathologica 2022 Feb;115(1):8-22 [FREE Full text] [doi: 10.32074/1591-951X-818] [Medline: 36645398]
29. Rossi SH, Klatte T, Usher-Smith J, Stewart GD. Epidemiology and screening for renal cancer. World J Urol 2018 Sep;36(9):1341-1353 [FREE Full text] [doi: 10.1007/s00345-018-2286-7] [Medline: 29610964]
30. Mahmud S, Abbas TO, Mushtak A, Prithula J, Chowdhury MEH. Kidney cancer diagnosis and surgery selection by machine learning from CT scans combined with clinical metadata. Cancers (Basel) 2023 Jun 14;15(12):3189 [FREE Full text] [doi: 10.3390/cancers15123189] [Medline: 37370799]
31. Richters A, Aben KKH, Kiemeney LALM. The global burden of urinary bladder cancer: an update. World J Urol 2020 Aug;38(8):1895-1904 [FREE Full text] [doi: 10.1007/s00345-019-02984-4] [Medline: 31676912]
32. Wang Z, Wang L, Wang S, Xie L. Burden of kidney cancer and attributed risk factors in China from 1990 to 2019. Front Public Health 2022;10:1062504 [FREE Full text] [doi: 10.3389/fpubh.2022.1062504] [Medline: 36589951]
33. Zhou W, Zhong L, Tang X, Huang T, Xie Y. Early warning and monitoring of COVID-19 using the Baidu Search Index in China. J Infect 2022 May;84(5):e82-e84 [FREE Full text] [doi: 10.1016/j.jinf.2022.02.002] [Medline: 35139408]

34. Shen C, Jiang L, Ma H, Che G. Perioperative preparation in thoracic day surgery: battle against COVID-19. Thorac Cancer 2020 Aug;11(8):2376-2379 [FREE Full text] [doi: 10.1111/1759-7714.13500] [Medline: 32558379]
35. COVIDSurg Collaborative. Effect of COVID-19 pandemic lockdowns on planned cancer surgery for 15 tumour types in 61 countries: an international, prospective, cohort study. Lancet Oncol 2021 Nov;22(11):1507-1517 [FREE Full text] [doi: 10.1016/S1470-2045(21)00493-9] [Medline: 34624250]
36. Su M, Lao J, Zhang N, Wang J, Anderson RT, Sun X, et al. Financial hardship in Chinese cancer survivors. Cancer 2020 Jul 15;126(14):3312-3321 [FREE Full text] [doi: 10.1002/cncr.32943] [Medline: 32396242]
37. Zhang H, Fu Y, Chen M, Si L. Socioeconomic inequality in health care use among cancer patients in China: Evidence from the China Health and Retirement Longitudinal Study. Front Public Health 2022;10:942911 [FREE Full text] [doi: 10.3389/fpubh.2022.942911] [Medline: 35983352]
38. Khan AS, Rahman AU, Qazi LT. The relationship between internet usage, socioeconomic status, subjective health and social status. BER 2016 Oct 1;8(Special Edition):67-82. [doi: 10.22547/ber/8.se.5]
39. Willems S, De MS, Deveugele M, Derese A, De MJ. Socio-economic status of the patient and doctor-patient communication: does it make a difference? Patient Educ Couns 2005 Feb;56(2):139-146. [doi: 10.1016/j.pec.2004.02.011] [Medline: 15653242]
40. Elenskaia K, Haidvogel K, Heidinger C, Doerfler D, Umek W, Hanzal E. The greatest taboo: urinary incontinence as a source of shame and embarrassment. Wien Klin Wochenschr 2011 Oct;123(19-20):607-610. [doi: 10.1007/s00508-011-0013-0] [Medline: 21935649]
41. Faber JS, Al-Dhahir I, Reijnders T, Chavannes NH, Evers AWM, Kraal JJ, et al. Attitudes toward health, healthcare, and eHealth of people with a low socioeconomic status: a community-based participatory approach. Front Digit Health 2021;3:690182 [FREE Full text] [doi: 10.3389/fdgth.2021.690182] [Medline: 34713165]
42. Su M, Yao N, Liu L, Cheng J, Sun X, Yue H, et al. Older cancer survivors living with financial hardship in China: a qualitative study of family perspectives. Psychooncology 2022 Apr;31(4):661-670. [doi: 10.1002/pon.5854] [Medline: 34784087]
43. Pang H, Lv J, Xu T, Li Z, Gong J, Liu Q, et al. Incidence and risk factors of female urinary incontinence: a 4-year longitudinal study among 24 985 adult women in China. BJOG 2022 Mar;129(4):580-589 [FREE Full text] [doi: 10.1111/1471-0528.16936] [Medline: 34536320]
44. Matoff-Stepp S, Applebaum B, Pooler J, Kavanagh E. Women as health care decision-makers: implications for health care coverage in the United States. J Health Care Poor Underserved 2014 Dec;25(4):1507-1513. [doi: 10.1353/hpu.2014.0154] [Medline: 25418222]
45. Liu L, Zhang A, Su M, Sun X, Shao D, Cheng J, et al. The development and validation of a patient-reported outcome measure to assess financial hardship among older cancer survivors in China: hardship and recovery with distress survey. Front Oncol 2023;13:1151465 [FREE Full text] [doi: 10.3389/fonc.2023.1151465] [Medline: 37152015]
46. Chee SY. Age-related digital disparities, functional limitations, and social isolation: unraveling the grey digital divide between baby boomers and the silent generation in senior living facilities. Aging Ment Health 2023 Jul 10:1-12. [doi: 10.1080/13607863.2023.2233454] [Medline: 37424361]
47. Bertolami CN. Misinformation? Disinformation? J Oral Maxillofac Surg 2022 Sep;80(9):1455-1457. [doi: 10.1016/j.joms.2022.06.009] [Medline: 36058567]
48. Dunn P, Conard S. Improving health literacy in patients with chronic conditions: a call to action. Int J Cardiol 2018 Dec 15;273:249-251. [doi: 10.1016/j.ijcard.2018.08.090] [Medline: 30193793]
49. Nørgaard M, Veres K, Ording AG, Djurhuus JC, Jensen JB, Sørensen HT. Evaluation of hospital-based hematuria diagnosis and subsequent cancer risk among adults in Denmark. JAMA Netw Open 2018 Nov 02;1(7):e184909 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.4909] [Medline: 30646376]
50. Ghandour R, Freifeld Y, Singla N, Lotan Y. Evaluation of hematuria in a large public health care system. Bladder Cancer 2019;5(2):119-129 [FREE Full text] [doi: 10.3233/BLC-190221] [Medline: 31930164]
51. Nieder AM, Lotan Y, Nuss GR, Langston JP, Vyas S, Manoharan M, et al. Are patients with hematuria appropriately referred to Urology? A multi-institutional questionnaire based survey. Urol Oncol 2010;28(5):500-503. [doi: 10.1016/j.urolonc.2008.10.018] [Medline: 19097811]
52. Hu X, Pan H, He W, Hua H, Yan Z. Evaluation of the content quality of websites for recurrent aphthous ulcers and oral lichen planus. BMC Oral Health 2017 Dec 29;17(1):170 [FREE Full text] [doi: 10.1186/s12903-017-0467-1] [Medline: 29284468]
53. Kim M, Park H, Kho B, Park C, Oh I, Kim Y, et al. Artificial intelligence and lung cancer treatment decision: agreement with recommendation of multidisciplinary tumor board. Transl Lung Cancer Res 2020 Jun;9(3):507-514 [FREE Full text] [doi: 10.21037/tlcr.2020.04.11] [Medline: 32676314]
54. Lewis LF. Exploring the experience of self-diagnosis of autism spectrum disorder in adults. Arch Psychiatr Nurs 2016 Oct;30(5):575-580. [doi: 10.1016/j.apnu.2016.03.009] [Medline: 27654240]
55. Min S, Han J. Topic modeling analysis of diabetes-related health information during the coronavirus disease pandemic. Healthcare (Basel) 2023 Jun 27;11(13):1871 [FREE Full text] [doi: 10.3390/healthcare11131871] [Medline: 37444705]

Abbreviations

AAPC: average annual percentage change
APC: annual percentage change
BSI: Baidu Search Index
GBD: global burden of disease
TCM: traditional Chinese medicine

Edited by N Cahill; submitted 16.02.24; peer-reviewed by M Zhong, A Hassan, LY Sun, B Ofori, C Wang, Y Wang; comments to author 17.04.24; revised version received 09.10.24; accepted 30.12.24; published 14.03.25.

Please cite as:
Lin S, Duan L, Xu X, Cao H, Lu X, Wen X, Wei S
Analyzing Online Search Trends for Kidney, Prostate, and Bladder Cancers in China: Infodemiology Study Using Baidu Search Data (2011-2023)
JMIR Cancer 2025;11:e57414
URL: https://cancer.jmir.org/2025/1/e57414
doi:10.2196/57414
PMID:

Original Paper

Identifying Adverse Events in Outpatients With Prostate Cancer Using Pharmaceutical Care Records in Community Pharmacies: Application of Named Entity Recognition

Yuki Yanagisawa[1], PhD; Satoshi Watabe[1], BSc; Sakura Yokoyama[1], BSc; Kyoko Sayama[1], MSc; Hayato Kizaki[1], PhD; Masami Tsuchiya[1], PhD; Shungo Imai[1], PhD; Mitsuhiro Someya[2], BSc; Ryoo Taniguchi[2], PhD; Shuntaro Yada[3,4], PhD; Eiji Aramaki[3], PhD; Satoko Hori[1], PhD

[1]Division of Drug Informatics, Keio University Faculty of Pharmacy, Tokyo, Japan

[2]Nakajima Pharmacy, Hokkaido, Japan

[3]Nara Institute of Science and Technology, Nara, Japan

[4]Faculty of Library, Information and Media Science, University of Tsukuba, Tsukuba, Japan

Corresponding Author:
Satoko Hori, PhD
Division of Drug Informatics
Keio University Faculty of Pharmacy
1-5-30 Shibakoen
Minato-ku
Tokyo, 105-8512
Japan
Phone: 81 3 5400 2650
Email: satokoh@keio.jp

Abstract

Background: Androgen receptor axis-targeting reagents (ARATs) have become key drugs for patients with castration-resistant prostate cancer (CRPC). ARATs are taken long term in outpatient settings, and effective adverse event (AE) monitoring can help prolong treatment duration for patients with CRPC. Despite the importance of monitoring, few studies have identified which AEs can be captured and assessed in community pharmacies, where pharmacists in Japan dispense medications, provide counseling, and monitor potential AEs for outpatients prescribed ARATs. Therefore, we anticipated that a named entity recognition (NER) system might be used to extract AEs recorded in pharmaceutical care records generated by community pharmacists.

Objective: This study aimed to evaluate whether an NER system can effectively and systematically identify AEs in outpatients undergoing ARAT therapy by reviewing pharmaceutical care records generated by community pharmacists, focusing on assessment notes, which often contain detailed records of AEs. Additionally, the study sought to determine whether outpatient pharmacotherapy monitoring can be enhanced by using NER to systematically collect AEs from pharmaceutical care records.

Methods: We used an NER system based on the widely used Japanese medical term extraction system MedNER-CR-JA, which uses Bidirectional Encoder Representations from Transformers (BERT). To evaluate its performance for pharmaceutical care records by community pharmacists, the NER system was first applied to 1008 assessment notes in records related to anticancer drug prescriptions. Three pharmaceutically proficient researchers compared the results with the annotated notes assigned symptom tags according to annotation guidelines and evaluated the performance of the NER system on the assessment notes in the pharmaceutical care records. The system was then applied to 2193 assessment notes for patients prescribed ARATs.

Results: The F_1-score for exact matches of all symptom tags between the NER system and annotators was 0.72, confirming the NER system has sufficient performance for application to pharmaceutical care records. The NER system automatically assigned 1900 symptom tags for the 2193 assessment notes from patients prescribed ARATs; 623 tags (32.8%) were positive symptom tags (symptoms present), while 1067 tags (56.2%) were negative symptom tags (symptoms absent). Positive symptom tags included ARAT-related AEs such as “pain,” “skin disorders,” “fatigue,” and “gastrointestinal symptoms.” Many other symptoms were classified as serious AEs. Furthermore, differences in symptom tag profiles reflecting pharmacists’ AE monitoring were observed between androgen synthesis inhibition and androgen receptor signaling inhibition.

Conclusions: The NER system successfully extracted AEs from pharmaceutical care records of patients prescribed ARATs, demonstrating its potential to systematically track the presence and absence of AEs in outpatients. Based on the analysis of a large volume of pharmaceutical medical records using the NER system, community pharmacists not only detect potential AEs but also actively monitor the absence of severe AEs, offering valuable insights for the continuous improvement of patient safety management.

(JMIR Cancer 2025;11:e69663) doi:10.2196/69663

KEYWORDS

natural language processing; pharmaceutical care records; androgen receptor axis-targeting agents; adverse events; outpatient care

Introduction

According to the International Agency for Research on Cancer, 20 million new cancer cases were reported in 2022 [1]. In particular, an increasing number of patients are receiving chemotherapy at home [2], driven by the increasing availability of oral anticancer drugs over the past 20 years [3]. Outpatient chemotherapy offers advantages such as reduced invasiveness of administration and fewer hospital visits. However, compared with inpatient chemotherapy, the lack of direct and frequent observation by health care professionals can present safety challenges. In practice, ensuring the safety of oral anticancer drugs prescribed to patients is never easy for treating physicians [4,5]. Additionally, for patients, the adverse events (AEs) experienced after the initiation of treatment may be more burdensome than those reported in clinical trials [6]. Therefore, continuous medical support is essential, extending beyond the hospital setting and involving collaboration with community health care services. Community pharmacies, as the most accessible health care providers, play a crucial role in monitoring AEs in outpatients using oral anticancer drugs [7].

Among outpatient chemotherapy options, endocrine therapy is a common treatment for patients with prostate cancer. In particular, for patients with castration-resistant prostate cancer (CRPC), androgen receptor axis-targeting reagents (ARATs)—abiraterone acetate, apalutamide, darolutamide, and enzalutamide—are taken long term in outpatient settings. Therefore, effective AE monitoring can contribute to prolonged treatment duration [8]. Furthermore, ARATs exhibit distinct AE profiles depending on their pharmacological mechanisms [9,10]. Abiraterone acetate, an androgen synthesis inhibitor, has been reported to cause AEs such as hypertension, gastrointestinal symptoms, fatigue, and liver dysfunction [11]. On the other hand, apalutamide, darolutamide, and enzalutamide, which act through androgen receptor (AR) signaling inhibition, have been associated with AEs such as fatigue and dermatologic disorders [12-15].

In Japan, outpatients prescribed ARATs typically have their prescriptions filled at community pharmacies. Pharmacists at these pharmacies dispense medications, provide patient counseling, monitor for potential AEs, and document pharmaceutical care records. Pharmaceutical care records are commonly written in the SOAP (subjective, objective, assessment, and plan) format [16]. In particular, the assessment section contains pharmacists' evaluations related to the patient's pharmacotherapy and thus may be a fruitful source of information about AEs experienced by patients. By systematically collecting and analyzing the AEs and pharmacists' assessments accumulated in pharmaceutical care records, it could be possible to clarify the AEs experienced outside the hospital by outpatients prescribed ARATs.

However, pharmaceutical care records comprise huge amounts of unstructured text data, including medical terms, accumulated over time for individual patients, making them difficult to analyze manually. Therefore, natural language processing technology, particularly named entity recognition (NER), offers a solution by enabling the extraction of patients' illnesses and symptom data from unstructured medical records [17,18]. Although studies using the NER system to analyze medical texts in hospital inpatients have been reported in the past, few studies have focused on outpatient care, especially in community pharmacy settings, where pharmacists contribute to the safe delivery of pharmacotherapy.

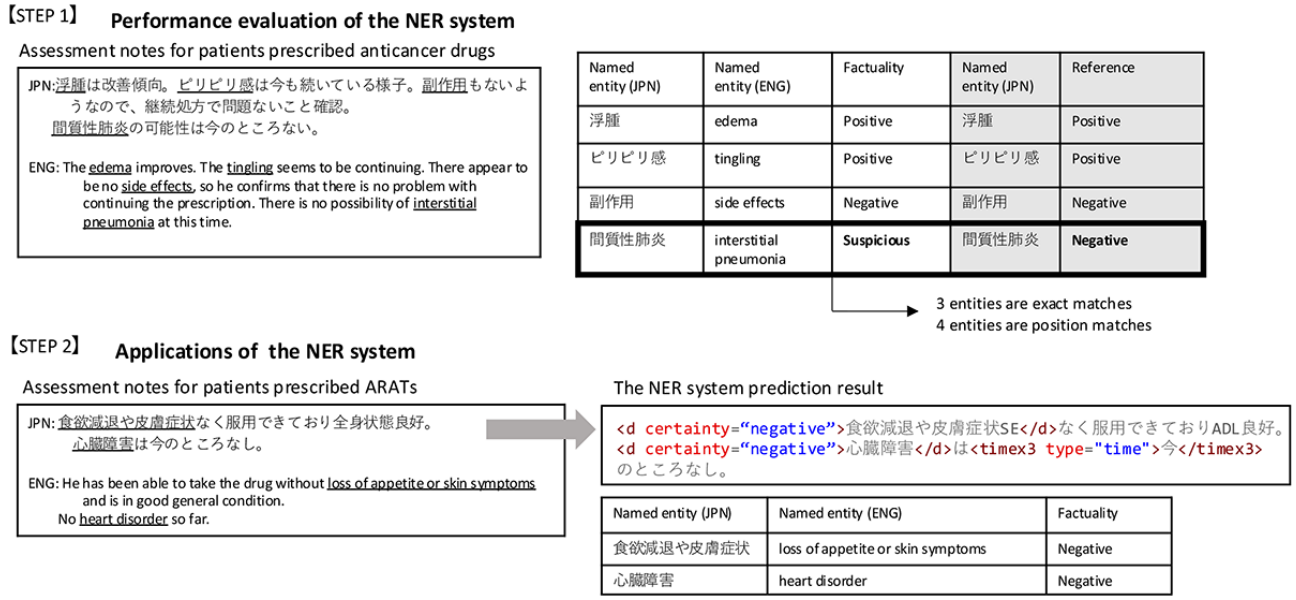
This study aimed to determine whether pharmacotherapy monitoring of outpatients prescribed ARATs can be carried out using NER to systematically collect AEs from pharmaceutical care records, with a particular focus on assessment notes, which may contain detailed records of AEs experienced by patients.

Methods

Outline

An overview of the experimental method is shown in Figure 1. First, we evaluated the performance of the NER system using assessment notes from pharmaceutical care records (STEP 1). The NER system was then used to extract symptoms from the pharmaceutical care records of patients prescribed ARATs (STEP 2).

Figure 1. Outline of the experimental method, including manual annotation on a data set, examination of the model's performance in terms of "position match" and "exact match," and use of the named entity recognition (NER) system to extract symptoms from the pharmaceutical care records of patients prescribed androgen receptor axis-targeting agents (ARATs). ENG: English; JPN: Japanese.



Model Description

System

Our objective was to evaluate the utility of pharmacist-assessed symptoms extracted from pharmaceutical care records kept by community pharmacies. To extract terms, we applied MedNER-CR-JA, an existing Japanese medical term-extraction system based on Bidirectional Encoder Representations from Transformers (BERT) and that was trained using Japanese case reports [19]. Given the input (pharmaceutical care records), the NER system output the symptoms with the factuality (e.g., positive symptom, negative symptom). The factuality was classified into 4 categories as shown in Table 1.

Table 1. Certainty attributes of symptom tags.

Symptom tags	Definition
Positive	The symptom is observed in the patient.
Suspicious	The symptom is suspected to be present in the patient.
Negative	The symptom is not observed in the patient.
General	The symptom is described without reference to the patient's condition.

Materials

This study used data from patients at community pharmacies (n=291,150) belonging to the Nakajima Pharmacy Group in Japan from April 2020 to December 2021. The patients' data consisted of medication orders (structured data) and pharmaceutical care records written in Japanese (unstructured data, n=2,180,902).

First, to evaluate the performance of the NER system (STEP 1), we selected pharmaceutical care records of patients with at least one prescription for anticancer drugs according to the structured data, using the drug codes (YJ codes). YJ codes starting with "42" in Japan indicate anticancer drugs. To evaluate the NER system, we took data recorded during October 2021 through December 2021 at 11 randomly selected pharmacies from the pharmacies with a history of anticancer drug prescriptions. Second, to apply the NER system to pharmaceutical care records (STEP 2), we extracted data for patients prescribed ARATs at least once from April 2020 through December 2021 using individual YJ codes (Multimedia Appendix 1) and used their assessment notes.

In both experiments (STEP 1 and STEP 2), we input the preprocessed text into the NER system. For text preprocessing, we removed line breaks and full-width and half-width spaces from the target text and normalized the text (Unicode). Furthermore, structured sections with template-based input were excluded, and only free-text sections were used.

Performance Evaluations and Metrics for Pharmaceutical Care Records

To evaluate the performance of the NER system on the assessment notes in the pharmaceutical care records, we performed manual annotation on a data set. We verified that the NER system was able to assign symptom tags according to existing annotation guidelines [20] with manual annotation performed by 3 pharmaceutically proficient researchers (SY, YY, KS). These researchers were selected based on their pharmaceutical expertise. Two of them were licensed pharmacists with over 5 years of experience in hospitals or

pharmacies, while the third was a pharmacy student who had completed a clinical internship in both a hospital and a pharmacy. To ensure consistency and reliability in the study, all researchers were expected to be familiar with the existing guidelines regarding the study objectives, task procedures, and evaluation criteria. The existing guidelines were standardized to enable annotation even by nonmedical professionals. To evaluate the consistency and generalizability of the annotation guidelines across annotators, we randomly extracted 100 cases from the target data, and these were annotated. Label agreement was evaluated using the Fleiss kappa (κ) coefficient to assess the concordance rate among the 3 researchers. Let *Po* be the mean of the raters' agreement and *Pe* the degree of concentration for each rating, then the calculation formula is as follows:

For a difficult task like NER, a score of around 0.6 is considered a substantial match, and the obtained evaluations are considered reasonably reliable [21].

After evaluating their agreement, the 3 researchers (SY, YY, KS) annotated the entire data set. The model's performance was evaluated using "position match" and "exact match." Position match is a method that accepts a match with the correct data if the tag's position and name are correct, while exact match requires the tag's position, name, and attributes to match with the correct data. In the position match evaluation, if the position is correct but the predicted tag name is different from the correct tag name, it is considered incorrect. Therefore, in the exact match, we calculated evaluation metrics separately for tags and attributes, while in the position match, we calculated metrics only for the tags. In both methods, the performance was evaluated in terms of precision, recall, and F_1-score:

Ethical Considerations

This study was conducted with anonymized data following approval by the ethics committee of the Keio University Faculty of Pharmacy (approval number: 240618-1) and was conducted in accordance with the relevant guidelines and regulations and the Declaration of Helsinki. Informed consent specific to this study was waived due to the retrospective observational design of the study based on the approval by the ethics committee of the Keio University Faculty of Pharmacy. To respect the will of each stakeholder, however, we provided patients and pharmacists of the pharmacy group with an opportunity to refuse the sharing of their pharmaceutical care records by posting an overview of this study at each pharmacy store or on their websites.

Results

Model Performance and Statistics

The data set used to evaluate the performance of the NER system consisted of 1008 assessment notes. The average word count per text in the target data was 48.3 words, with a median of 42.0 words and minimum and maximum values of 4 and 292 words, respectively. The κ coefficient among the 3 annotators for 100 randomly extracted texts was 0.62.

The F_1-score (NER excluding position matches and attribute classification) was 0.72. This score is the macro-average of the F_1-scores for all attribute classifications (positive, suspicious, negative, and general). For positive and negative symptom tags, the F_1-scores were 0.70 and 0.78, respectively (Table 2). The F_1-scores for all attribute classifications are shown in Multimedia Appendix 2. The position matches showed an excellent F_1-score of 0.86 (Multimedia Appendix 2).

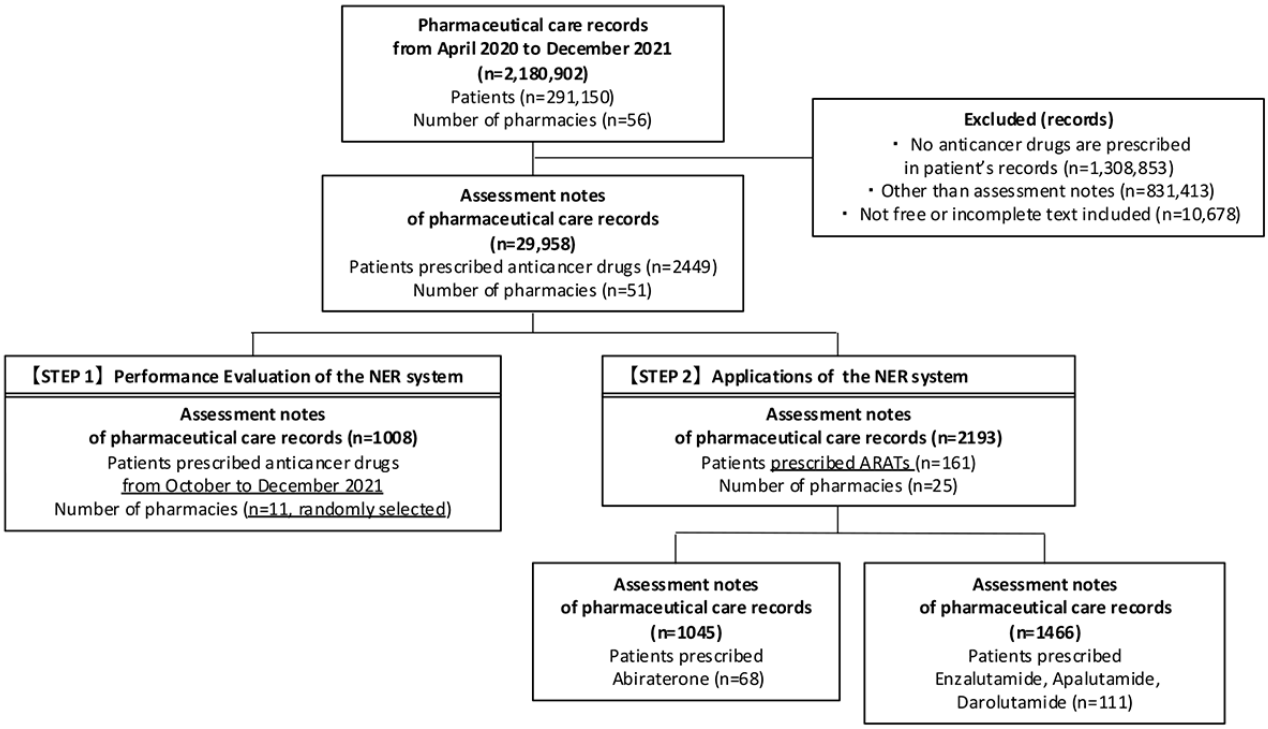
Table 2. Exact matches between the named entity recognition system and annotators.

Tags	Precision	Recall	F_1-score
All symptom tags	0.66	0.78	0.72
Positive symptom tags	0.60	0.85	0.70
Negative symptom tags	0.73	0.83	0.78

Assessment Notes for Patients Prescribed Antiandrogens

Figure 2 shows the flowchart of the procedure for selecting pharmaceutical care records. From April 2020 through December 2021, 161 patients had at least one ARAT prescription and corresponding assessment notes in their pharmaceutical care records. There were 2193 assessment notes recording pharmacists' assessments. Additionally, 68 patients had a history of prescriptions for the androgen synthesis inhibitor abiraterone acetate, with 1045 assessment notes. There were 111 patients with a history of prescription for AR signal transduction inhibitors enzalutamide, apalutamide, and darolutamide, with 1466 assessment notes. On average, each extracted assessment note contained 39.8 words.

Figure 2. Flowchart for selecting pharmaceutical care records to evaluate the performance of the named entity recognition (NER) system (STEP 1) and apply the NER system to pharmaceutical care records (STEP 2). ARAT: androgen receptor axis-targeting agent.



Prediction by the NER System

From 2193 assessment notes of patients prescribed ARATs, the NER system automatically assigned 1900 symptom tags. Among these 1900 symptom tags, 623 (32.8%) were positive symptom tags, predicting the presence of symptoms, while the majority, 1067 (56.2%), were negative symptom tags, predicting the absence of symptoms. Additionally, of the 1900 symptom tags, there were 131 (6.9%) general symptom tags and 79 (4.2%) suspicious symptom tags. Specifically, the positive symptom tags were commonly assigned to notes describing "pain," "skin disorders," "fatigue," and "gastrointestinal symptoms." The negative symptom tags frequently appeared in descriptions indicating general side effects such as "SE" or "side effects," accounting for 13.9% (148/1067) of the total. Furthermore, entities with nonpositive symptom tags were AE expressions related to ARATs, such as gastrointestinal symptoms, neuropsychiatric symptoms, and cardiovascular symptoms (Table 3). Other extracted symptoms included seizures and other neuropsychiatric symptoms, cardiovascular symptoms, and hepatic dysfunction. There were also descriptions of severe AEs, including myelosuppression, interstitial pneumonia, and rhabdomyolysis.

Next, the results were examined according to the pharmacological mechanism of action of the ARATs. For patients prescribed the androgen synthesis inhibitor abiraterone acetate, 876 symptom tags were automatically assigned to 1045 assessment notes. The most common symptom tag was negative, with 488 tags (488/876, 55.7%), while 283 positive symptom tags were assigned (283/876, 32.3%; Multimedia Appendix 3). Of the 876 tags, there were 64 (7.3%) general symptom tags and 41 (4.7%) suspicious symptom tags. Regarding characteristic expressions for patients taking abiraterone acetate, "congestive heart failure" was frequently noted among the positive symptom tags. Entities related to skin disorders, which occur frequently in patients taking AR signal transduction inhibitors, appeared infrequently in the top 20 entities. Additionally, entities such as "congestive heart failure," "drug-induced liver injury," and "hyperglycemia" were extracted among nonpositive symptom tags.

For patients prescribed the AR signal transduction inhibitors enzalutamide, apalutamide, and darolutamide, 1466 symptom tags were automatically assigned to 1274 assessment notes. Negative symptom tags were most common, amounting to 692 (692/1466, 47.2%), while positive symptom tags amounted to 438 (438/1466, 29.9%; Multimedia Appendix 4). Of the 1466 tags, there were 96 (6.5%) general symptom tags and 48 (3.3%) suspicious symptom tags. Among the positive symptom tags, characteristic entities for AR signal transduction inhibitors such as "skin disorders" and "fatigue" were frequently noted. Additionally, entities such as "seizures," "psychiatric symptoms," and "cardiovascular diseases" were extracted among the nonpositive symptom tags.

The differences in symptom entities collected based on the pharmacological mechanism of prescribed ARATs are illustrated in Figure 3.

Table 3. Application of the named entity recognition (NER) system to assessment notes of patients prescribed androgen receptor axis-targeting agents (ARATs), showing the top 20 entities.

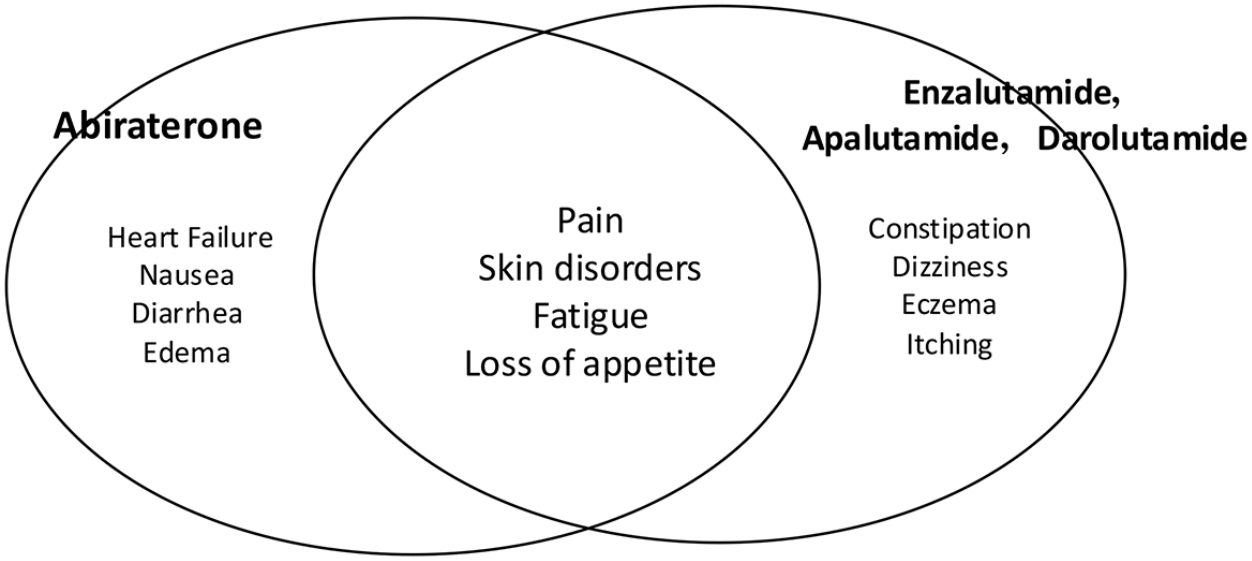
Entity (Japanese)	Results, n (%)
Positive symptom tags (n=623)	
Pain (痛み)	22 (3.5)
Skin disorders (皮膚障害)	22 (3.5)
Fatigue (倦怠感)	16 (2.6)
Loss of appetite (食欲不振)	15 (2.4)
Pain (疼痛)	12 (1.9)
Prostate cancer (前立腺癌)	11 (1.8)
Diarrhea (下痢)	10 (1.6)
Poor compliance (コンプライアンス不良)	10 (1.6)
Gastrointestinal symptoms (消化器症状)	9 (1.4)
Constipation (便秘)	9 (1.4)
Liver function disorders (肝機能障害)	8 (1.3)
Itching (痒み)	8 (1.3)
Congestive heart failure (うっ血性心不全)	8 (1.3)
Hypoglycemia (低血糖)	7 (1.1)
Dizziness (ふらつき)	7 (1.1)
Skin symptoms (皮膚症状)	7 (1.1)
Decreased appetite (食欲低下)	6 (1)
Dry mouth (口渇)	6 (1)
Decreased PSA[a] (PSA低下)	6 (1)
Nausea (嘔気)	6 (1)
Negative symptom tags (n=1067)	
SE (SE)[b]	117 (11)
Gastrointestinal symptoms (消化器症状)	90 (8.4)
Neuropsychiatric symptoms (精神神経症状)	32 (3)
Side effects (副作用)	31 (2.9)
Changes in physical condition (体調変化)	27 (2.5)
Cardiovascular symptoms (循環器症状)	22 (2.1)
Good adherence (アドヒア良好)	21 (2)
Adherence (アドヒア)	17 (1.6)
Unpleasant symptoms (不快な症状)	17 (1.6)
Symptom changes (症状変化)	16 (1.5)
Missed dose (飲み忘れ)	14 (1.3)
Side effect symptoms (副作用症状)	13 (1.2)
Compliance (コンプライアンス)	13 (1.2)
Adverse events (有害事象)	12 (1.1)
Side effect symptoms (SE症状)	12 (1.1)
Progress in physical condition (体調問題)	11 (1)
Bleeding tendency (出血傾向)	11 (1)
Hypoglycemia (低血糖)	11 (1)
Pain (疼痛)	11 (1)

Entity (Japanese)	Results, n (%)
Elevated PSA (PSA上昇)	11 (1)

[a]PSA: prostate-specific antigen.

[b]SE (SE): Side effects.

Figure 3. Main symptom entities extracted as positive for agents with different pharmacological mechanisms of action, including androgen synthesis inhibitors (abiraterone acetate) and androgen receptor signaling inhibitors (enzalutamide, apalutamide, darolutamide), shown in order of the number of adverse events extracted.



Discussion

Overview

We evaluated the performance of the NER system on pharmaceutical care records and successfully extracted symptoms, including AEs, from the free-text assessment notes of community pharmacists. This study is the first to report the use of the NER system to systematically identify AEs from pharmaceutical care records of patients prescribed ARATs.

Predictive Performance of the NER System

The NER system exhibited robust performance on the pharmaceutical care records analyzed. Although the kappa statistics for annotation (0.62) indicate substantial agreement, the following factors were considered to have contributed to the decrease in the kappa. The most common type of disagreement among annotators was related to the selection range of the entity. Specifically, the annotation guidelines [16] we referenced set the standard for selecting the entire range when symptoms were expressed in parallel. However, discrepancies in the interpretation of the range occurred among the annotators. For example, for text such as “gastrointestinal symptoms, hypoglycemia present,” disagreements arose over whether to annotate this as “gastrointestinal symptoms” and “hypoglycemia” or as “gastrointestinal symptoms, hypoglycemia.”

The F_1-score for symptom tagging (NER and attribute classification) was 0.72, and the F_1-scores for positive and negative predictive performance ranged from 0.70 to 0.78. These results indicate that the system is effective for analyzing the pharmaceutical care records used in this study. Notably, the system outperformed prior work by Ohno et al [22], who reported an F_1-score of 0.64 for assessment notes among the SOAP format descriptions in Japanese hospital-based pharmaceutical care records using MedNER-J (NER and positive and negative classification). The NER evaluation method used in this study assumed that entities that completely matched the annotations were correct. In addition, many types of tags were used, and the evaluation was conducted under more stringent conditions.

Several factors may have affected model performance. In particular, there were specific expressions in pharmaceutical care records that we recognized as causing frequent errors in NER predictions during the performance evaluation. First, pharmacist assessment records often include descriptions related to medication adherence. The NER system sometimes incorrectly recognized terms such as “adherence” and “compliance” as symptom entities (for example, the entity “adherence” was predicted as a symptom expression). Second, pharmacist assessment records frequently included descriptions related to AE confirmation, often using nuanced expressions characteristic of Japanese, such as “no suspicion of XX.” As a result, the NER system occasionally predicted “suspicious” for negative symptom expressions (for example, the entity “interstitial pneumonia” was predicted as suspected in the sentence “no suspicion of interstitial pneumonia,” whereas the correct status should be negative). Additionally, the system showed low extraction rates for entities related to hypertension,

which is a side effect specific to enzalutamide acetate, as well as blood glucose–related expressions associated with concomitant corticosteroids and expressions related to hypokalemia, possibly because vital sign–related terms like "blood pressure" and "blood glucose" were treated as test values rather than symptoms. Another challenge was interpreting negation patterns, such as "No symptoms: XX (Symptoms)" and "Symptoms (-)," which may have led to incorrect symptom tagging.

AEs Identified in Pharmaceutical Care Records for Patients Prescribed ARATs

Our analysis of pharmacists' assessment records for outpatients prescribed ARATs revealed that over 90% of the records contained symptom tags. The extracted entities predominantly indicated symptoms related to AEs commonly seen in patients treated with ARATs. Common positive symptom entities included "pain," "skin disorders," "fatigue," "anorexia," and other typical side effects of ARATs. "Pain" was the most frequently identified symptom, reflecting the high prevalence of bone metastases in patients with CRPC requiring ARAT therapy. A study using the Frankfurt Metastatic Cancer Database of the Prostate reported that 78% of patients with CRPC had bone metastases [23].

On the other hand, the fact that the top 20 entities accounted for less than one-third of the total positive symptom tags suggests that many symptoms occur with low frequency, resulting in a long-tail distribution. This indicates that a large portion of symptom tags is dispersed across numerous less-frequent entities. This distribution indicates that the NER system has the sensitivity to capture a wide range of symptoms. However, it also suggests the need for standardization if the aim is to extract specific symptom expressions.

Entities other than positive symptom tags were serious AEs such as "interstitial pneumonia," "myelosuppression," and "rhabdomyolysis." Interestingly, many of the negative symptom tags indicated the absence of severe AEs, suggesting that pharmacists were actively monitoring for signs of AEs and documenting the absence of signs. These negative assessments, particularly for rare but serious AEs, highlight the thoroughness of pharmacists in ensuring the safety of outpatient pharmacotherapy.

Drug-Specific Monitoring by Community Pharmacists

There are two types of ARATs with different pharmacological mechanisms of action: androgen synthesis inhibition and AR signaling inhibition. Each category is associated with distinct AEs, which require tailored care. Despite the common indication for ARATs, this study highlights the differences in pharmacists' monitoring based on drug type.

The most characteristic positive symptom extracted for abiraterone acetate, an androgen synthesis inhibitor, was "heart failure," a critical AE requiring close monitoring. In addition, "cardiovascular symptoms," "congestive heart failure," "rhabdomyolysis," and "drug-induced liver injury" were also identified as nonpositive symptoms.

For patients prescribed AR signaling inhibitors (enzalutamide, apalutamide, darolutamide), the most frequently extracted positive symptoms included "skin problems," "fatigue," and "gastrointestinal symptoms," which are common AEs that are important to monitor. "Skin disorder" was more frequently reported as a symptom for AR signaling inhibitors, whereas it was less commonly reported for abiraterone acetate, an androgen synthesis inhibitor. In addition, "gastrointestinal symptoms" and "neuropsychiatric symptoms" were extracted from the nonpositive symptom tags. "Seizures" were also extracted from the nonpositive symptom tags. Seizures are a serious AE specifically associated with AR signaling inhibitors [15,24,25].

These results indicate that, even though ARATs are used for the same purpose, pharmacists should pay particular attention to certain AEs depending on the pharmacological mechanism of each type of ARAT.

Comparison With Prior Work

The NER system used in this study has previously been reported to be applicable to medical records documented in electronic medical records within hospitals in Japan [22,26,27]. For example, Ohno et al [22] reported an F_1-score of 0.64 for an NER system applied to hospital-based pharmaceutical records. In contrast, this study applied the NER system to pharmaceutical records documented in community pharmacies outside hospitals and achieved an F_1-score of 0.72 under more stringent evaluation conditions.

A key feature of this study is its focus on pharmaceutical care records by community pharmacists, which tend to include more diverse and nuanced expressions of outpatients than electronic medical records. Furthermore, our previous studies have demonstrated that patients with hand-foot syndrome and serious AEs can be identified using NLP applied to pharmaceutical care records by community pharmacies [28]. In this study, the NER system enabled the extraction of a wide range of AEs, not limited to specific side effects. These results highlight advancements in NER methodologies and underscore their applicability in real-world outpatient settings.

Limitations

This study has several limitations. First, the content of pharmaceutical care records may vary depending on the pharmacists' experience, the role of each pharmacy, and the condition of the patients. Second, this study focused on the extraction of AEs and did not include symptom normalization. As a result, synonymous entities such as "pain" were counted separately. Third, the pharmacies included in this study were part of a single pharmacy group, which may limit the generalizability of the findings to other pharmacy settings.

Prospects

The NER system used in this study enabled the automated analysis of a large volume of records accumulated by community pharmacists. This system can extract symptoms from text, determine their factuality, and monitor patients' symptoms over time. Additionally, error analysis of the NER system applied to pharmaceutical medical records by community pharmacists

can contribute to the appropriate and effective use of NER systems in the future.

Our findings support the idea that NER technology offers the potential for real-time and longitudinal monitoring of patients' symptoms through the seamless integration of local health care systems, including community pharmacies. In hospital electronic health records, it has been demonstrated that the accuracy of detecting AEs improves when physicians' notes are combined with records from nurses and pharmacists [23]. Similarly, for outpatients, AEs identified by community pharmacists may play a key role in effective monitoring. This approach would extend beyond the confines of hospital-based electronic medical record systems. By using NER technology, it should be possible to gain detailed insights into drug efficacy and the progression of AEs. Moreover, it should enable earlier interventions and timely adjustments to treatment plans, particularly for outpatients who may otherwise have limited direct interaction with health care providers

Conclusions

We evaluated the performance of the NER system on Japanese medical text, focusing on pharmaceutical care records by community pharmacists. The NER system successfully extracted AEs from pharmaceutical care records. This is the first study to apply NER to the pharmaceutical care records of patients prescribed ARATs in community settings. Our analysis of pharmacists' assessment records for outpatients prescribed ARATs revealed that over 90% of the records contained symptom tags. Notably, ARATs have distinct pharmacological mechanisms and exhibit different AE profiles, and the NER system successfully captured these variations. Furthermore, it highlights the role of community pharmacists in monitoring specific AEs related to ARAT therapy and, importantly, in documenting the absence of severe AEs as well. This study demonstrates that NER can effectively capture symptoms reported by outpatients and documented by community pharmacists, complementing hospital records and contributing to a more comprehensive understanding of AEs in outpatient settings. The NER system is expected to be a valuable tool for enhancing pharmacotherapy monitoring in outpatient settings.

Acknowledgments

This study was supported by Japan Science and Technology Agency CREST (grant number JPMJCR22N1), Japan. GenAI was not used during manuscript writing.

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

YY, SW, and SH designed the study. YY, SY, and KS conducted the annotation. YY, SW, and SY performed the data analysis and conducted all the experiments. MS and RT performed data management and extraction. YS and EA supervised the natural language processing research as specialists. SH supervised the study overall. YY, SW, HK, MT, SI, and SH drafted and finalized the paper. All authors reviewed and approved the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codes for Targeted Pharmaceuticals.
[PDF File (Adobe PDF File), 60 KB - cancer_v11i1e69663_app1.pdf]

Multimedia Appendix 2
Exact and partial matches between the NER system and the annotators.
[PDF File (Adobe PDF File), 77 KB - cancer_v11i1e69663_app2.pdf]

Multimedia Appendix 3
Application of the NER system to assessment notes of patients prescribed abiraterone acetate.
[PDF File (Adobe PDF File), 112 KB - cancer_v11i1e69663_app3.pdf]

Multimedia Appendix 4
Application of the NER system to assessment notes of patients prescribed enzalutamide, apalutamide, darolutamide.
[PDF File (Adobe PDF File), 95 KB - cancer_v11i1e69663_app4.pdf]

References

1. Cancer Today. World Health Organization International Agency for Research on Cancer. URL: https://gco.iarc.who.int/today [accessed 2025-02-26]
2. Evans JM, Qiu M, MacKinnon M, Green E, Peterson K, Kaizer L. A multi-method review of home-based chemotherapy. Eur J Cancer Care (Engl) 2016 Sep;25(5):883-902. [doi: 10.1111/ecc.12408] [Medline: 26545409]
3. Schlichtig K, Dürr P, Dörje F, Fromm MF. New oral anti-cancer drugs and medication safety. Dtsch Arztebl Int 2019 Nov 15;116(46):775-782 [FREE Full text] [doi: 10.3238/arztebl.2019.0775] [Medline: 31920193]
4. van Leeuwen RWF, Brundel DHS, Neef C, van Gelder T, Mathijssen RHJ, Burger DM, et al. Prevalence of potential drug-drug interactions in cancer patients treated with oral anticancer drugs. Br J Cancer 2013 Mar 19;108(5):1071-1078 [FREE Full text] [doi: 10.1038/bjc.2013.48] [Medline: 23412102]
5. Weingart SN, Toro J, Spencer J, Duncombe D, Gross A, Bartel S, et al. Medication errors involving oral chemotherapy. Cancer 2010 May 15;116(10):2455-2464 [FREE Full text] [doi: 10.1002/cncr.25027] [Medline: 20225328]
6. Di Maio M, Basch E, Bryce J, Perrone F. Patient-reported outcomes in the evaluation of toxicity of anticancer treatments. Nat Rev Clin Oncol 2016 May;13(5):319-325 [FREE Full text] [doi: 10.1038/nrclinonc.2015.222] [Medline: 26787278]
7. Egbewande OM, Abdulwasiu MA, Yusuf RO, Durojaye AB, Ashimiyu-Abdulsalam ZI. Roles of community pharmacists in cancer management. Innov Pharm 2022;13(3):1 [FREE Full text] [doi: 10.24926/iip.v13i3.4946] [Medline: 36627904]
8. Hirabatake M, Ikesue H, Iwama Y, Irie K, Yoshino S, Yamasaki T, et al. Pharmacist-urologist collaborative management improves clinical outcomes in patients with castration-resistant prostate cancer receiving enzalutamide. Front Pharmacol 2022;13:901099 [FREE Full text] [doi: 10.3389/fphar.2022.901099] [Medline: 35662711]
9. Mori K, Mostafaei H, Pradere B, Motlagh RS, Quhal F, Laukhtina E, et al. Apalutamide, enzalutamide, and darolutamide for non-metastatic castration-resistant prostate cancer: a systematic review and network meta-analysis. Int J Clin Oncol 2020 Nov;25(11):1892-1900 [FREE Full text] [doi: 10.1007/s10147-020-01777-9] [Medline: 32924096]
10. Pozas J, Álvarez Rodríguez S, Fernández VA, Burgos J, Santoni M, Manneh Kopp R, et al. Androgen receptor signaling inhibition in advanced castration resistance prostate cancer: what is expected for the near future? Cancers (Basel) 2022 Dec 09;14(24):1 [FREE Full text] [doi: 10.3390/cancers14246071] [Medline: 36551557]
11. Fizazi K, Tran N, Fein L, Matsubara N, Rodriguez-Antolin A, Alekseev BY, et al. Abiraterone acetate plus prednisone in patients with newly diagnosed high-risk metastatic castration-sensitive prostate cancer (LATITUDE): final overall survival analysis of a randomised, double-blind, phase 3 trial. Lancet Oncol 2019 May;20(5):686-700. [doi: 10.1016/S1470-2045(19)30082-8] [Medline: 30987939]
12. Chi KN, Chowdhury S, Bjartell A, Chung BH, Pereira de Santana Gomes AJ, Given R, et al. Apalutamide in patients with metastatic castration-sensitive prostate cancer: final survival analysis of the randomized, double-blind, phase III TITAN study. J Clin Oncol 2021 Jul 10;39(20):2294-2303. [doi: 10.1200/JCO.20.03488] [Medline: 33914595]
13. Chi KN, Agarwal N, Bjartell A, Chung BH, Pereira de Santana Gomes AJ, Given R, TITAN Investigators. Apalutamide for metastatic, castration-sensitive prostate cancer. N Engl J Med 2019 Jul 04;381(1):13-24. [doi: 10.1056/NEJMoa1903307] [Medline: 31150574]
14. Smith MR, Hussain M, Saad F, Fizazi K, Sternberg CN, Crawford ED, Montesa-Pino, ARASENS Trial Investigators. Darolutamide and survival in metastatic, hormone-sensitive prostate cancer. N Engl J Med 2022 Mar 24;386(12):1132-1142 [FREE Full text] [doi: 10.1056/NEJMoa2119115] [Medline: 35179323]
15. Hussain M, Fizazi K, Saad F, Rathenborg P, Shore N, Ferreira U, et al. Enzalutamide in men with nonmetastatic, castration-resistant prostate cancer. N Engl J Med 2018 Jun 28;378(26):2465-2474 [FREE Full text] [doi: 10.1056/NEJMoa1800536] [Medline: 29949494]
16. Podder V, Lew V, Ghassemzadeh S. SOAP notes. Treasure Island, FL: StatPearls Publishing; 2024.
17. Aramaki E, Wakamiya S, Yada S, Nakamura Y. Natural language processing: from bedside to everywhere. Yearb Med Inform 2022 Aug 02;31(1):243-253 [FREE Full text] [doi: 10.1055/s-0042-1742510] [Medline: 35654422]
18. Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. Int J Med Inform 2019 May;125:37-46 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.02.008] [Medline: 30914179]
19. Nishiyama T, Nishidani M, Ando A, Yada S, Wakamiya S, Aramaki E. NAISTSOC at the NTCIR-16 Real-MedNLP Task. 2022 Presented at: 16th NTCIR Conference on Evaluation of Information Access Technologies; June 14-17, 2022; Tokyo, Japan URL: https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings16/pdf/ntcir/07-NTCIR16-MEDNLP-NishiyamaT.pdf
20. Yada S, Aramaki E, Tanaka R, Cheng F, Kurohashi S. Clinical text annotation guidelines, version 8. 2011 Jun 29. URL: https://sociocom.naist.jp/real-mednlp/wp-content/uploads/sites/3/2021/07/PRISM_Annotation_Guidelines.pdf [accessed 2025-02-26]
21. Artstein R, Poesio M. Inter-coder agreement for computational linguistics. Computational Linguistics 2008 Dec;34(4):555-596. [doi: 10.1162/coli.07-034-R2]
22. Ohno Y, Kato R, Ishikawa H, Nishiyama T, Isawa M, Mochizuki M, et al. Using the natural language processing system Medical Named Entity Recognition-Japanese to analyze pharmaceutical care records: natural language processing analysis. JMIR Form Res 2024 Jun 04;8:e55798 [FREE Full text] [doi: 10.2196/55798] [Medline: 38833694]

23. Wenzel M, Hoeh B, Humke C, Koll F, Cano Garcia C, Siech C, et al. Influence of metastatic sites and burden on oncological outcomes in patients progressing to metastatic castration resistant prostate cancer. World J Urol 2024 Nov 02;42(1):615. [doi: 10.1007/s00345-024-05341-2] [Medline: 39487858]
24. Fizazi K, Shore N, Tammela TL, Ulys A, Vjaters E, Polyakov S, ARAMIS Investigators. Darolutamide in nonmetastatic, castration-resistant prostate cancer. N Engl J Med 2019 Mar 28;380(13):1235-1246. [doi: 10.1056/NEJMoa1815671] [Medline: 30763142]
25. Smith MR, Saad F, Chowdhury S, Oudard S, Hadaschik BA, Graff JN, SPARTAN Investigators. Apalutamide treatment and metastasis-free survival in prostate cancer. N Engl J Med 2018 Apr 12;378(15):1408-1418 [FREE Full text] [doi: 10.1056/NEJMoa1715546] [Medline: 29420164]
26. Nishiyama T, Yamaguchi A, Han P, Pereira LWK, Otsuki Y, Andrade GHB, et al. Automated system to capture patient symptoms from multitype Japanese clinical texts: retrospective study. JMIR Med Inform 2024 Sep 24;12:e58977 [FREE Full text] [doi: 10.2196/58977] [Medline: 39316418]
27. Mashima Y, Tanigawa M, Yokoi H. Information heterogeneity between progress notes by physicians and nurses for inpatients with digestive system diseases. Sci Rep 2024 Apr 01;14(1):7656 [FREE Full text] [doi: 10.1038/s41598-024-56324-7] [Medline: 38561333]
28. Nishioka S, Watabe S, Yanagisawa Y, Sayama K, Kizaki H, Imai S, et al. Adverse event signal detection using patients' concerns in pharmaceutical care records: evaluation of deep learning models. J Med Internet Res 2024 Apr 16;26:e55794 [FREE Full text] [doi: 10.2196/55794] [Medline: 38625718]

Abbreviations

AE: adverse event
AR: androgen receptor
ARAT: androgen receptor axis-targeting agent
BERT: Bidirectional Encoder Representations from Transformers
CRPC: castration-resistant prostate cancer
NER: named entity recognition
SOAP: subjective, objective, assessment, and plan

Edited by N Cahill; submitted 05.12.24; peer-reviewed by Y Sakamoto, P Wu, S Matsuda; comments to author 07.01.25; revised version received 27.01.25; accepted 20.02.25; published 11.03.25.

Please cite as:
Yanagisawa Y, Watabe S, Yokoyama S, Sayama K, Kizaki H, Tsuchiya M, Imai S, Someya M, Taniguchi R, Yada S, Aramaki E, Hori S
Identifying Adverse Events in Outpatients With Prostate Cancer Using Pharmaceutical Care Records in Community Pharmacies: Application of Named Entity Recognition
JMIR Cancer 2025;11:e69663
URL: https://cancer.jmir.org/2025/1/e69663
doi:10.2196/69663
PMID:

Implementation of a Quality Improvement and Clinical Decision Support Tool for Cancer Diagnosis in Primary Care: Process Evaluation

Sophie Chima[1], MEpi; Barbara Hunter[2], PhD; Javiera Martinez-Gutierrez[1,3], MPH, MD; Natalie Lumsden[2], PhD; Craig Nelson[4], MBBS, PhD; Dougie Boyle[2], PhD; Kaleswari Somasundaram[2], MPhil; Jo-Anne Manski-Nankervis[2,5*], MBBS, PhD; Jon Emery[1*], MBBS, PhD

[1]Department of General Practice and Primary Care, Centre for Cancer Research, University of Melbourne, 305 Grattan St, Melbourne, Australia

[2]Department of General Practice and Primary Care, University of Melbourne, Melbourne, Australia

[3]Department of Family Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

[4]Department of Medicine, Western Health, Melbourne, Australia

[5]Primary Care and Family Medicine, LKC Medicine, Nanyang Technological University, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Sophie Chima, MEpi
Department of General Practice and Primary Care, Centre for Cancer Research, University of Melbourne, 305 Grattan St, Melbourne, Australia

Abstract

Background: For patients with cancer, the pathway to diagnosis will most often begin in general practice. In the absence of strong diagnostic features or in patients with nonspecific symptoms, delays in diagnosis can occur. Initial presentations and routine blood tests are important in determining whether a patient requires further investigation. Quality improvement interventions, including auditing tools and clinical decision support (CDS), have been developed for use in general practice to support this diagnostic process. We conducted a process evaluation of a pragmatic, cluster-randomized trial that evaluated the effectiveness of a new technology, Future Health Today (FHT), implemented in general practice to assist with the appropriate follow-up of patients at risk of undiagnosed cancer.

Objectives: This study aims to understand implementation gaps, explore differences between the general practices involved, provide context to the trial effectiveness outcomes, and understand the mechanisms behind the intervention successes and failures.

Methods: The trial intervention consisted of the FHT tool (with CDS, audit, recall, and quality improvement components), training and educational sessions, benchmarking reports, and ongoing practice support. The 21 general practices in the intervention arm of the trial were included in the process evaluation. Process data were collected using semistructured interviews, usability and educational session surveys, engagement with intervention components, and technical logs. The Medical Research Council's Framework for Developing and Evaluating Complex Interventions was used to analyze and interpret the data.

Results: The uptake of the supporting components of the intervention (training and education sessions, benchmarking reports) was low. Most practices only used the CDS component of the tool, facilitated by active delivery, with general practitioners reporting acceptability and ease of use. Complexity, time, and resources were reported as barriers to the use of the auditing tool. Access to a study coordinator and ongoing practice support facilitated the sustained involvement of practices in the trial, while contextual factors, such as the COVID-19 pandemic and staff turnover, impacted their level of participation. The relevance of the intervention varied between practices, with some practices reporting very low numbers of patients who were flagged for further investigation.

Conclusions: While some components of the intervention, such as the CDS tool, were considered to be acceptable and useful, this process evaluation highlighted barriers such as time and resources, practice differences, and considerations around the optimal amount of support needed when delivering the intervention. Addressing these in future studies may optimize the implementation process. Further work is needed to determine if a scaled-back approach, which meets the time and resource availability of a busy general practice, can effectively facilitate the implementation of CDS tools. Given the variation seen between practices, the use of the FHT cancer module may be better targeted to certain practices based on size, location, and patient demographics.

(JMIR Cancer 2025;11:e65461) doi:10.2196/65461

KEYWORDS

cancer diagnosis; implementation; clinical decision support tool; diagnosis; primary care; process evaluation; quality improvement; intervention; effectiveness; interviews; surveys; cancer care; general practice

Introduction

Diagnosing cancer early can improve patient outcomes and quality of life [1,2]. But, in general practice, the timely detection of cancer can be challenging in the absence of strong diagnostic features, often resulting in prolonged diagnostic intervals [3-5]. In patients presenting to general practice with nonspecific symptoms, the use of routine blood tests can guide decision-making [6]. There is strong evidence that supports the diagnostic utility of abnormal blood tests (eg, iron-deficiency anemia and raised platelets) for multiple cancer types [7-9]. However, suboptimal follow-up and management of abnormal test results have been shown to contribute to delays in diagnosis [10].

Inadequate follow-up of abnormal test results may occur in the case of diagnostic errors, but is also influenced by the general practitioners' (GPs) experience and training; perceptions of cancer care and investigations; patient characteristics; and health system pressures [11,12]. For example, controversy and confusion about prostate-specific antigen (PSA) testing, coupled with changing guidelines and revised thresholds for what is abnormal, contribute to lower follow-up rates in men who have a raised PSA. Surprisingly, there are very few trials that look at modifying the practitioner- and practice-level barriers to following up abnormal results [11].

The general practice electronic medical record (EMR) allows for the integration of novel technologies, where algorithms apply epidemiological data on the underlying risks of undiagnosed cancer based on symptoms and test results to monitor and identify patients who may benefit from further investigation [13]. Clinical decision support (CDS) systems assist in clinical decision-making, where such tools are linked to patient data to produce patient-specific recommendations or prompts for the GP to consider [14,15]. Similarly, auditing tools that use patient information from the EMR enable practice population-level management and review and have the potential to capture patients who are at risk of being lost to follow-up [16,17]. Evidence suggests that tools that highlight patients for review, referral, or further investigation based on evidence-based guidelines can improve patient care, but many of these tools designed to support diagnosis in general practice are met with low uptake and implementation difficulties [18-20].

Complex interventions are used to assess the effectiveness and utility of such tools in general practice. Yet implementing complex interventions can be distinctly difficult, as they involve multiple interrelated components and there are often multiple levels where change is required [21]. Process evaluation can aid in the understanding of the factors that influence how or why a complex intervention succeeds or fails. This study presents the results of a process evaluation of a pragmatic trial, Future Health Today (FHT). This complex intervention consisted of a novel CDS and auditing software, education, quality improvement (QI), and practice support. The pragmatic trial evaluated whether the intervention, which flagged patients with an abnormal blood test that may be indicative of undiagnosed cancer (FHT cancer module), increased the proportion of patients receiving guideline-based care. By gaining process information, we aim to better understand the implementation gaps, explore differences between the general practices involved, understand the interactions between intervention components, and provide context to understand the effectiveness of the intervention.

Methods

Intervention Description and Study Population

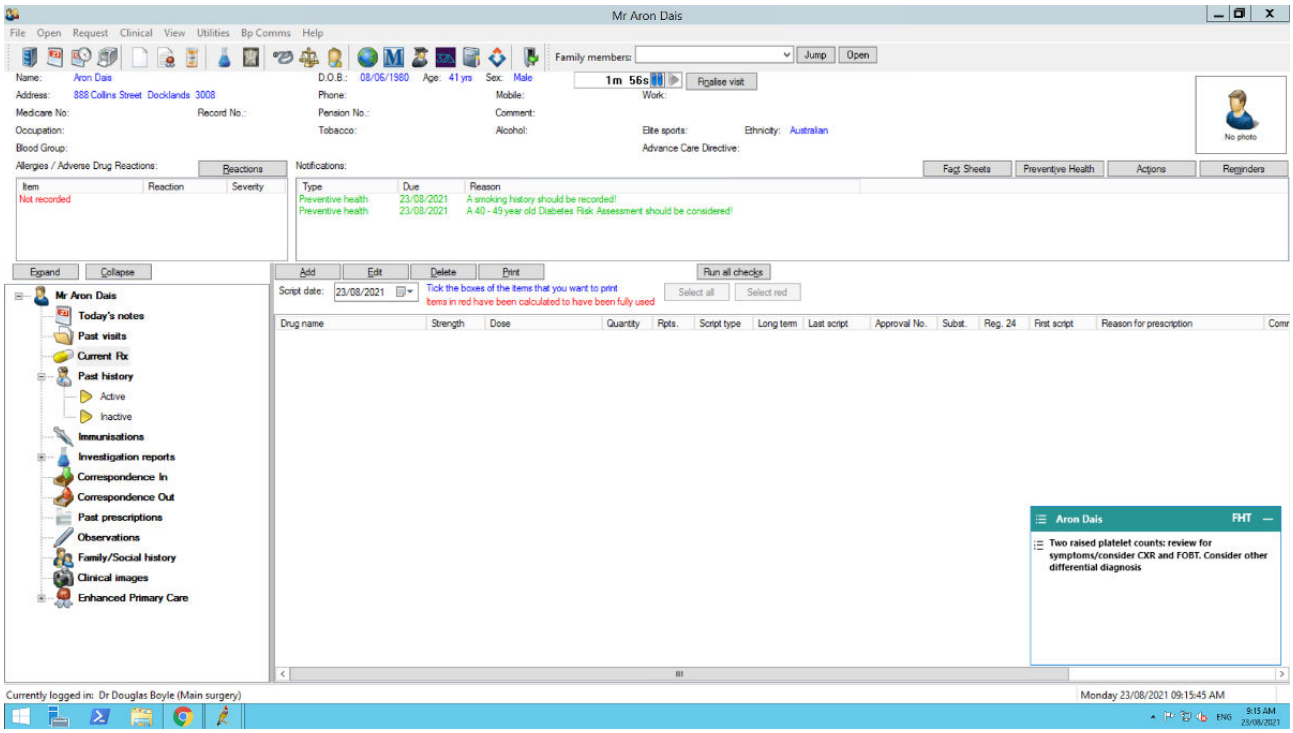
The FHT study was a pragmatic cluster-randomized controlled trial that evaluated the effectiveness of a QI intervention [22]. Pragmatic trials, by definition, are trials that evaluate an intervention in everyday practice, with the aim of measuring the effectiveness of the intervention in routine clinical practice rather than under ideal conditions [23,24]. The implementation of the FHT software and the trial components (including implementation strategies) were applied and adapted to real-world conditions to understand and evaluate how the tool would be used in routine general practice.

The components of the complex intervention included the FHT software, training and educational sessions, benchmarking reports, and practice support. The trial was conducted between October 2021 and September 2022. Practices were randomly allocated to participate in either the intervention (follow-up of patients with abnormal blood test results associated with the risk of undiagnosed cancer) or the active control (which had access to a different FHT module). As the aim of this process evaluation was to explore the factors critical to the implementation of the cancer module, our study population comprises the 21 intervention arm practices only; results for the active control intervention will be reported separately. The study protocol has been published on the Australia and New Zealand Clinical Trial Registry (ACTRN12620000993998) [25].

FHT was integrated within the general practice EMR and consisted of a CDS tool, a web-based audit and feedback tool, and the capacity for general practices to monitor their QI activities [26]. Disease-specific modules were developed for use in FHT. The cancer module used patient information in the EMR (age, sex, previous cancer diagnosis) and results of abnormal tests associated with undiagnosed cancers. The FHT cancer module consisted of 3 central algorithms, designed to assist GPs by flagging patients with abnormal blood test results that are associated with an increased risk of undiagnosed cancer: markers of iron deficiency and anemia, raised PSA, and raised platelet count). The CDS component of the tool activates when the GP or general practice nurse (GPN) opens the patients' medical record, displaying a prompt on screen with guideline-concordant recommendations, such as the review of relevant symptoms or appropriate investigations (Figure 1).

There is also a web-based portal, containing an auditing tool; a QI monitoring tool; and access to resources, guidelines, education, and training, which can be accessed on any computer with FHT installed. Algorithms run each night, extracting data from the practice management software database (eg, Best Practice or Medical Director), processing the data locally by applying FHT algorithms (the data does not leave the practice), and categorizing the results. Examples of a CDS prompt and the audit tool are presented in Multimedia Appendix 1. Further details on the development of the tool and the cancer module explored in this study have been described elsewhere [27-29].

Figure 1. An example of the clinical decision support tool as it appears in the medical record. Simulated patient data are used in this image.



In the pragmatic trial, FHT was installed on general practice computers before study initiation. On the first day of the trial, practices were asked to create 3 cohorts of patients using the FHT auditing tool, one for each abnormal blood test (raised PSA, raised platelets, and markers of anemia). The cohorts included all patients identified by the FHT cancer module who had recommendations for guideline-based follow-up (as part of the trial, practices could then review the patient cohorts and determine if further follow-up was necessary). Cohorts were created again at the 6-month mark, using the audit tool, so that benchmarking information could be determined. After generating the cohorts, practices were invited to use FHT as they chose during the trial.

Implementation of the software was supported by a number of additional intervention components. This multifactorial implementation strategy was informed by the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework, with strategies that were relevant and useful to general practice [30]. These components have previously been shown to increase reach; they are low intensity and high impact, with the purpose of limiting implementation workload while promoting continued engagement with the intervention [31,32]. Training on the use of FHT was offered regularly in the lead up to and in the first month of the trial, and then monthly thereafter. Each practice was assigned a study coordinator, who conducted the Zoom-based training sessions on how to use FHT, assisted with any technological queries, and facilitated requests for support throughout the trial. Practices had access to short training videos on YouTube and a range of short- and long-form written training guides. In addition, 6 Project ECHO (Extension for Community Healthcare Outcomes) [33] educational sessions were run on the topics of cancer diagnosis and QI, each consisting of a 10-minute didactic session, a 10-minute case discussion, followed by an open discussion for approximately 20 - 30 minutes. The ECHO sessions were delivered via a webinar, and general practice staff were invited to attend. Quarterly benchmarking reports were provided to practices to review their progress in the follow-up of patients who had been flagged by the tool, and to compare their progress to other practices in the trial. All practices were required to nominate a practice champion to lead the implementation of FHT in their practice and to be the primary point of contact with the study coordinator during the trial, managing the installation and technical queries, facilitating ongoing use of the tool, identifying staff for process evaluation interviews and to disseminating trial related information to the practice. The goal of the practice champion in this study was to mirror the pragmatic approach of the intervention (eg, they were asked to filter and disseminate information to the practice using an approach that best reflects their individual practice needs and current processes).

Ethical Considerations

The study was approved by the Faculty of Medicine, Dentistry and Health Sciences Human Ethics Sub-Committee at the University of Melbourne (ID:2056564). While practices consented on behalf of all practice staff to participate in the

wider trial, additional written consent was obtained for all interviews. Interview participants were compensated A $100 (US $64.83) for their time. Practice champions also consented separately and were compensated A $200 (US $129.66) for their role as practice champions. All participant data were deidentified and kept anonymous.

Data Collection

Data were collected via qualitative interviews, usability surveys, technical queries, engagement logs, and educational session surveys. For the semistructured interviews, all practice champions were contacted via phone and email to participate in an interview in the first and last months of the trial. The practice champion was most commonly a practice manager (PM) or GPN, but GPs occasionally took on this role during the trial (eg, due to staff changes). The semistructured interviews were conducted over the phone. The interviews were conducted by study researchers (SC, NL, and BH; see the following section on researcher characteristics). The duration of the interviews ranged between 15 and 42 minutes. The interview guides were developed using the Clinical Performance Feedback Intervention Theory framework [34] and were pilot-tested during earlier optimization work on the FHT cancer module [27]. The interviews explored installation, intervention delivery, implementation barriers and facilitators, goals, and usability (see Multimedia Appendix 2 for interview schedule). The interviews explored similar themes at each timepoint, although earlier interviews included questions around goals and intention, and the final interviews explored long-term implementation and sustainability. GPs and GPNs were also recruited for interviews in month 6 of the trial. These interviews have been reported separately [35], as the purpose of the clinical interviews was to explore the acceptability of the clinical recommendations and impact on clinical practice, rather than explore the implementation of the wider intervention.

Usability surveys were sent to practice champions in months 1 and 12 of the trial, with the request to distribute them to the rest of the practice. The survey was delivered via web using REDCap (Research Electronic Data Capture; Vanderbilt University) [36] and included 30 questions (multichoice or free text). This survey was anonymous but captured general demographic information about the user and the general practice in which they work. The survey then explored the use and experience with the intervention (eg, length of time using the tool, what components have been used, and feedback and engagement with the intervention components). The survey also included a System Usability Scale (SUS) a 5-point Likert scale that quantifies the perceived usability of FHT [37]. The usability survey was developed by the study implementation team and is available in full in Multimedia Appendix 3).

Postsession ECHO surveys were sent to all ECHO session participants via REDCap after each educational session and collected both demographic information and feedback on the specific learning outcomes of each webinar. The survey consisted of 23 multiple-choice or free-text questions. An example survey from one of the webinars is included in Multimedia Appendix 4.

Information on the number of installations in each practice, the number of individual users, and recommendation queries (submitted through the technology by the practice) was collected using the FHT technology. Technical reports, including any technical queries by the practice throughout the trial, were recorded by the study coordinator. All engagements between the practice and the study team (study coordinator and technical team), were recorded by the study coordinator and categorized by content (eg, technical queries, training, and administrative items). Implementation diaries were kept by study coordinators to record contextual information (eg, changes in COVID-19 pandemic guidelines, immunization rollout, and general practice initiatives) throughout the trial.

Researcher Characteristics

SC is a PhD candidate at the Department of General Practice and Primary Care, University of Melbourne. BH, a senior qualitative research fellow in the department is the implementation lead for the FHT trial. NL is a postdoctoral research fellow who was the study coordinator for the active control arm of the trial. All are female and experienced in qualitative research and conducting semistructured interviews. Some interview participants were known to the interviewer, given their position in delivering the intervention and supporting the implementation in practices throughout the trial.

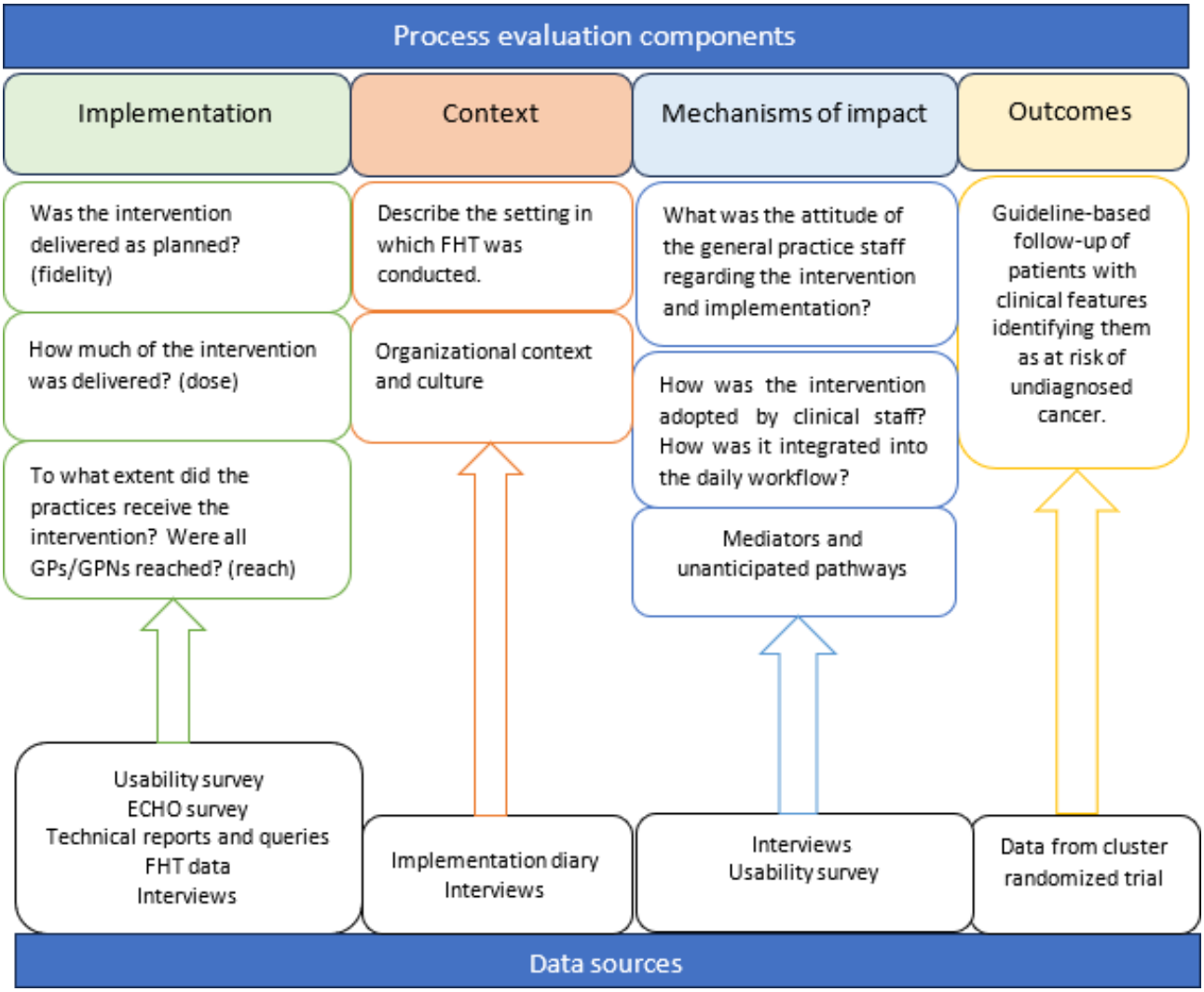
Data Analysis

Recorded interviews were transcribed and imported into NVivo (version 12; Lumivero). Process evaluation data were analyzed independently (SC and BH), prior to trial effectiveness outcomes, so as not to bias the interpretation of the results. Each researcher independently conducted a structured, deductive content analysis of the interview transcripts to extract key themes in the data. The results of the content analysis were collated, and themes were presented to the research team. To promote trustworthiness, analytical codes and emerging concepts and categories were discussed at multiple points in the analysis. Positionality was discussed by the coding team, including how established relationships, biases, and experiences may influence their relationship to the study data, and reflexive notes were kept [38,39]. The interpretation of the key findings and discrepancies in interpretations was discussed with the wider team. The results of the evaluation were then mapped onto the UK Medical Research Council (MRC) framework [40,41].

While several frameworks are available to explore and evaluate the implementation of an intervention, the MRC framework was chosen as it is designed for evaluating complex interventions. It has previously been shown to be useful in evaluating the delivery of new technologies in complex environments and in instances of a multi-faceted implementation approach [42,43]. The framework includes overarching themes of context, implementation, and mechanisms of impact and provides a mechanism for understanding the implementation successes and failures (Figure 2) [40,41]. In the figure, the data sources from the trial are mapped onto the 4 process evaluation components as outlined by the MRC framework (implementation, context, mechanisms of impact, and outcomes). The figure outlines the core components and

questions underpinning each theme, and the process data used to answer these questions.

Figure 2. How the process evaluation data are mapped onto the Medical Research Council framework. ECHO: Extension for Community Healthcare Outcomes; FHT: Future Health Today; GP: general practitioner; GPN: general practice nurse.



Results

Overview

A total of 21 practices participated in the process evaluation. Characteristics of the participating practices are described in Table 1. Characteristics of the interview participants are outlined in Table 2. Participation in other components of the process evaluation (usability survey, ECHO surveys) and additional general practice details are outlined in Multimedia Appendix 5. In summary, 25 interviews were conducted with 19 practice champions in the first and last months of the trial. A total of 12 usability surveys were completed, and 13 post-ECHO session surveys. Usability survey responses included a mix of PMs (n=4), GPNs (n=4), and GPs (n=3), as well as one receptionist (n=1).

Table . General practice characteristics.

Practice characteristics	Practices (n=21), n (%)
State	
Victoria	20 (95)
Tasmania	1 (5)
Relative Socioeconomic Disadvantage Index (Terciles)	
1 (most disadvantaged)	6 (29)
2	6 (29)
3 (least disadvantaged)	9 (42)
Previously participated in QI[a] program	9 (43)
Practice size	
4 or fewer FTE[b] GPs[c]	12 (57)
Greater than 4 FTE GPs	9 (43)
Rurality	
Metro	15 (71)
Rural	6 (29)

[a]QI: quality improvement.
[b]FTE: full-time equivalent.
[c]GP: general practitioner.

Table . Interview participants by timepoint.

	Month 1	Month 12
Role, n (%)		
GP[a]	1 (7)	1 (9)
GPN[b]	2 (14)	4 (36)
PM[c]	11 (79)	5 (46)
Admin	0 (0)	1 (9)
Gender, n (%)		
Women	13 (93)	11 (100)
Men	1 (7)	0 (0)
Rurality, n (%)		
Metro	11 (79)	5 (45)
Rural	3 (21)	6 (55)
Number of interviewees, n	14	11
Number of practices, n	13	9

[a]GP: general practitioner.
[b]GPN: general practice nurse.
[c]PM: practice manager.

Results have been mapped onto the 3 themes of implementation, context, and mechanisms of impact.

Trial Results Summary

The results of the cluster randomized controlled trial did not demonstrate a significant improvement in follow-up in the intervention arm [22]. At 12 months, 76.2% (2820/3709) of patients with abnormal test results in the intervention arm had been followed up compared with 70% in the control arm, with an estimated difference of 2.6% (95% CI −2.8% to 7.9%). No significant differences were identified in the secondary analyses or in the time to follow-up of abnormal tests for patients flagged by the tool. The following results of the process evaluation provide some context for the null outcome of the trial and suggest areas for improvement in the development and

implementation of CDS and audit software for cancer diagnosis in general practice.

Implementation

There were 3 core themes on implementation: intervention delivery, installation, and general practice characteristics, each underpinned by different evaluation data sources. Intervention delivery was supported by data from engagement logs and educational session surveys, installation and general practice characteristics were supported by data from technical reports, and all 3 drew from qualitative interview data.

Intervention Delivery

The intervention consisted of multiple components: the FHT software components (CDS, an auditing tool, and QI monitoring) and the supporting trial components (educational ECHO sessions, zoom-based training sessions, benchmarking reports, and other web-based learning components that practices could opt-in to use). The uptake of the supporting elements of the trial was generally low, except for the initial formal training sessions. GPs, GPNs, and PMs from all intervention practices were invited to the Project ECHO sessions, yet attendance ranged from 2 to 9 people per session, a mix of GPs and GPNs. Three key barriers were assessed as driving the low uptake of these trial components. First, the supporting components of the intervention were promoted via phone calls, newsletters, and regular emails to the practice champion, so it is possible that the knowledge of each session may not have reached the whole practice, dependent on how the practice champion decided to distribute this information to the practice (eg, internal email systems). The second barrier is the time and resource cost associated with each component. For example, attendance at training sessions and ECHO sessions (1 h each), during or after work hours, was not feasible for many clinical staff. The final barrier relates to recognized need and usability, with many practices reporting that they could use the CDS tool and the cancer recommendations adequately, without the need for additional education or training.

> *It's quite straightforward and quite well explained so it didn't need anything extra particularly.* [GP, female, month 1]

Installation

The installation of the software was completed in the month prior to study initiation, with practices having access to a "practice" module on diabetes in the 2 weeks prior to study initiation so any technical issues could be addressed. The installation, which was done remotely and without much interruption to the practice, was reported to be a smooth process for most. For those who required additional assistance, the use of a study coordinator and technical support ensured PMs felt well-supported during this process.

> *I think what really has gone well is how it seamlessly was implemented. There was no - there's no interruption.* [PM, female, month 1]

Due to the pragmatic approach of the trial, practices determined how many workstations in their practice would have FHT installed at the start of the trial. A total of 14 practices had FHT installed on all clinical computers. Five practices had FHT installed on only one computer at trial initiation, and of these, 4 made the decision to add FHT to additional computers later in the trial. Implementation logs and technical reports indicate that 3 practices were offline for a short period of time (range: 2 - 6 wk), although this does not appear to have had a significant impact on the use of the system.

General Practice Characteristics

There was a large variation in the number of patients identified for follow-up across practices. Three inner-city practices, which had a younger and transient patient population, reported that the cancer module may not be useful in their clinic, given the low number of patients flagged by FHT. For example, in one practice, only 14 patients were flagged for follow-up during the entire 12-month trial period. While these practices acknowledged that the FHT cancer module was less useful for them, it did not deter them from continuing to use the tool after the trial, where they would have access to additional FHT modules (see Software Usability section).

> *Actually, it is cancer topic I don't think that it is very suitable for our clinic because our clinic – the majority of our patients are international students, and they are very young.* [PM, female, month 12]

Context

In exploring context, there were 2 prominent themes: the COVID-19 pandemic and staff turnover. Both themes were underpinned by engagement logs, implementation diaries, and qualitative interviews.

COVID-19

The FHT trial was conducted during the COVID-19 pandemic. In Victoria, restrictions were placed on how and when people could leave their homes, with Melbourne experiencing lockdowns for 262 days during the pandemic. There was a major shift in usual care, and many consultations were conducted via telehealth. During 2020, there was an 8% reduction in cancer-related diagnostic tests nationally, with greater reductions seen in Victoria [44]. The trial continued during a nationwide COVID-19 immunization rollout in primary care, and the burden on general practice was high. There were reports throughout the trial that practices could not devote as much time as they would have liked to FHT or to attend the ECHO sessions due to competing webinars related to COVID-19.

> *It's been a time of change, a lot of updates, a lot of new technology with telehealth. Yeah, there's been a lot going on because of COVID.* [PM, male, month 12]

Staff Turnover

Consequently, staff turnover was a common theme throughout the trial, and the resultant loss of information and increased resource pressure featured heavily in the month 12 interviews. A total of 9 practice champions left their practice during the trial, with 2 practices ending the trial with no replacement. Many interviewees talked about the magnitude of staff turnover during the pandemic and how it was a barrier to use and to keep up momentum in the study.

> *We lost two staff, and two doctors at the end of last year. Now we've got two doctors that we're training again. We started off from scratch again.* [GPN, female, month 12]

Mechanisms of Impact

We found 4 mechanisms associated with the delivery of the intervention: adoption and integration, training and support, software usability, and clinical recommendations. The sources of data varied within each theme. Technical reports, usability surveys, and interviews supported adoption and integration. Training and support were underpinned by engagement logs, education session surveys, and interviews. Software usability was supported by the usability survey, interviews, and engagement logs. The final theme of clinical recommendations was elucidated from technical reports (in particular, recommendation queries), which were further explored in the educational sessions and qualitative interviews.

Adoption and Integration

The majority of practices reported that they did not use the QI and audit and recall components of the tool, only the CDS, which was delivered at the point of care. The CDS was considered easy to use and quick to learn and was therefore easily integrated into the clinical workflow by matching the resources available in a busy general practice. However, the audit, recall, and QI components of the tool encountered a number of barriers. First, in comparison to the CDS tool, where recommendations are actively delivered to the GP, the audit and recall tool requires the user to visit a web page and log on to access this part of the tool. Second, there were additional layers of complexity and multiple steps involved in order to identify, review, and recall patients identified in the audit tool.

Training and Support

The level of engagement between the study coordinator and most participating practices was high, and the support provided by the research and technical team facilitated the continued involvement of practices in the study. No practices in the intervention arm withdrew during the study period.

> *The co-operation between the teams and myself was amazing. There were no issues whatsoever and they were always there to help … it was really good.* [PM, female, month 12]

Practice staff who attended training sessions or used the web-based resources found the training adequate enough to use the tool, and practice champions reported that they would be comfortable training other members of the practice who could not attend. However, in most interviews, especially with the GPs who did not attend the training sessions, it became evident that components of training on how to use FHT did not reach the entire practice. For example, many GPs were unaware of the patient deferral button (which allows GPs to pause recommendations for a patient for a specified period of time) or that there are patient resources available. The post-ECHO session surveys highlighted that the education and case discussion components of the ECHO sessions were useful to GPs and GPNs in managing more complex patient scenarios, but did not influence the way in which the tool was used.

Software Usability

Of the 12 usability survey responses, 11 (92%) would recommend FHT to others. As part of the usability survey, respondents filled in a SUS [37]. The results of the survey align with the separate qualitative results from the clinical interviews in that FHT is reported to be easy to use, simple, and intuitive [35]. The average SUS score from the respondents was 74 (out of 100), consistent with an above-average score (average score=68; score >70 is considered good).

Acceptance and perceived usefulness of the FHT software were indicated by the number of practices agreeing to continue using the FHT software posttrial. A total of 18 of the 21 practices opted to continue using the software after the trial ended (practices were offered a 3 mo extension), and 17 practices opted to continue using the tool into 2023‐24.

Clinical Recommendations

The software included a menu option to "report recommendation query" if the GP or GPN thought the recommendation was appearing in error or wanted further information. Five queries about the clinical recommendations in FHT, from 3 practices, were received during the trial. The most frequent recommendation query centered around the clinical recommendations for raised platelets. The risk of undiagnosed cancer increases at a platelet count threshold of 400×10^9/L, but different laboratories report an upper limit of either 400 or 450 $\times 10^9$/L; this caused some confusion among GPs if a patient was flagged with a count in the range of 400‐450x 10^9/L. This issue was addressed in training sessions and regular communications (monthly emails, newsletters), but the perceived error may have impacted some GPs' willingness to use the tool and their trust in the recommendations. Interestingly, there were no queries about the recommendations for raised PSA (the FHT recommendations were based on current Australian guidelines for PSA follow-up with a lower limit of 3ng/mL in men over 50, which contrasts with some laboratories that report a lower limit of the normal range of 4ng/mL). Established referral pathways and familiarity with the abnormal test as a cancer marker (raised platelet is a relatively new marker of cancer) may have been a contributing factor to this difference in response.

Discussion

Overview

In this study, we describe a comprehensive process evaluation exploring the delivery of a complex intervention as part of a pragmatic, randomized trial, where a module to support cancer diagnosis was implemented in general practice. The process evaluation describes implementation gaps and the mechanisms that drive implementation successes and failures in order to provide context to the outcomes from the trial [22].

Principal Findings

The FHT cancer module intervention did not demonstrate a significant improvement in the follow-up of abnormal test results in the patients flagged by the tool. While we hypothesize that the high-performing practices across both arms may have led

to a ceiling effect (ie, there was limited room for improvement given the high rates of follow-up in both arms), an absence of any intervention effect may in part be due to implementation barriers, primarily relating to practice characteristics and contextual factors. There was limited ability for some specific practices to engage with the tool when their patient population was not suited to the FHT module that was implemented. Given this variation in the relevance and usefulness between practices, the use of the FHT cancer module may be better targeted to certain practices based on size, location, and patient demographics.

Comparison to Prior Work

In comparison to interventions with only one component, complex interventions require more time and resources, and are, unsurprisingly, more difficult to implement [31,45]. We found that the uptake of the supportive components of the intervention was low, aside from some initial training on the software. It was also indicated in the interviews that the supporting components were not considered necessary to use the CDS. While the implementation of new software in general practice requires some training and support, the results of this process evaluation indicate that a scaled-back approach to implementation, one which aligns with the time and resources available to general practice, may have been sufficient for the CDS component of the tool [46]. However, given the null outcomes of the trial, the low uptake of the audit tool, and significant contextual factors (COVID-19 pandemic), more work is needed to determine the usefulness of each component, or combination of components, in supporting this type of change in practice.

Implementing new technologies in general practice is a complex and dynamic process, and despite the potential to improve patient outcomes, many tools have low uptake after implementation [47,48]. The trial consisted of a number of implementation strategies that aimed to optimize the uptake of FHT in routine care, and these methods were applied primarily at the professional level (eg, education or training strategies targeting health care professionals and identification of practice champions) [49]. We found that the use of a practice coordinator facilitated the continued involvement and engagement of practices throughout the trial, similar to previously reported successful implementation strategies used in complex evaluations delivered in general practice. One overview of reviews concluded that practice facilitators, who work with practices in areas such as QI, problem-solving, and education, are almost 3 times as likely to adopt evidence-based guidelines, and practice facilitation improved the adoption of guidelines associated with many chronic diseases [32]. But given the large amount of staff turnover, driven by the COVID-19 pandemic, identifying, maintaining, and replacing practice champions was difficult and resulted in a loss of information and a barrier to engagement for some practices.

Strengths and Limitations

This process evaluation was extensive, with a multi-modal approach to collecting process data, including interviews, surveys, technical and software data, engagement logs, and implementation diaries. Interviews and usability surveys were carried out at 2 time points during the trial, to address the dynamic nature of implementation barriers and facilitators and how perceptions of the tool can change over time. This substantive evaluation provides context to a complex intervention and the environment in which it was implemented.

There were, however, some limitations. While all practices were invited to take part or contribute to each component of the process evaluation, there were 3 practices who did not participate in an interview at any timepoint or complete any surveys. The opt-in method for the interviews and surveys meant that we may not have sufficiently captured the views of practices who were less engaged with the intervention. These 3 practices did contribute some data to the process evaluation, through software data, technical information, and engagement logs, which were captured from all practices involved in the trial.

The burden of the COVID-19 pandemic in general practice and the resultant impact on staffing was a core theme throughout the process evaluation and provided context when interpreting the trial results. A second limitation was that the pandemic also likely impacted the time, availability, and resources for general practice staff to participate in the interviews and contributed to the low response rate for the usability survey. To mitigate this, we provided numerous opportunities for users to engage in interviews and respond to surveys throughout the trial and promoted such activities through the continued engagement with each practice champion.

Finally, we had originally planned on including some additional software use statistics to complement the qualitative components of the intervention; however, incomplete data prohibited our ability to do so. Software use data would have allowed us to triangulate users' responses via interviews and surveys with their time using the software, including what parts of the tool they used and when. Future studies would benefit from including software statistics to cross-check the qualitative results.

Implications and Future Research

There are implications for both research and practice. While the FHT cancer module did not increase the proportion of patients followed up according to guidelines, the process evaluation highlighted factors around usability, which facilitated the adoption and integration of the CDS component of the tool. This, coupled with the acceptability findings from separate clinical interviews [35], and the willingness of the majority of practices to continue using the tool after the trial finished, indicates that different modules developed for use in FHT should be explored, as well as CDS tools for cancer diagnosis more broadly. There are also considerations for designing complex interventions that involve the use of a new technology. Given the low uptake of the supporting components of the tool, but indications of use and acceptability of the CDS component of the software, it is unclear whether a multifaceted implementation strategy is useful when implementing new CDS tools, especially if it has been carefully co-designed to meet the needs of users. Future work should be undertaken to determine if a scaled-back approach, which meets the time and resource availability of general practice, could be as effective in supporting the delivery of novel CDS tools.

Conclusions

This process evaluation highlights the implementation and process-related gaps that could be addressed in future studies that aim to implement diagnostic support tools for cancer in general practice. While some of the factors were context-specific (eg, driven by the COVID-19 pandemic), barriers such as time, resources, and practice variations, alongside considerations of design elements, could be built upon to optimize future CDS and QI programs.

Acknowledgments

We would like to thank the general practitioners, health care professionals, and consumers who provided input to the development of Future Health Today (FHT), piloted the tool in their practices, and participated in this trial. We are grateful to the consumer and general practice advisory groups who provided important insights into this trial. We would like to acknowledge the Primary Care Collaborative Cancer Clinical Trials Group (PC4) for their support on this project. Deidentified patient data from the Patron primary care data repository (extracted from consenting general practices), which has been created and is operated by the Department of General Practice and Primary Care, the University of Melbourne, were used to underpin trial feasibility, sample size calculations, and to evaluate the outcomes of the trial. The Paul Ramsay Foundation provided funding for staff costs related to the technical development of FHT and the Implementation and Evaluation team members. They have also provided funding for professional staff support and general practice and consumer advisory groups. The development of the cancer module of FHT was supported by the CanTest Collaborative (funded by Cancer Research UK C8640/A23385) of which JE is an Associate Director. JE is supported by an NHMRC Investigator grant (APP1195320). JMG is supported by a Victorian Cancer Agency Mid-Career Fellowship (MCRF21025).

Data Availability

Deidentified datasets analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

Future Health Today has been developed and managed by the Department of General Practice and Primary Care, University of Melbourne, in collaboration with Western Health.

Multimedia Appendix 1
Examples of the clinical decision support and audit tool.
[PDF File, 398 KB - cancer_v11i1e65461_app1.pdf]

Multimedia Appendix 2
Example interview guide.
[PDF File, 104 KB - cancer_v11i1e65461_app2.pdf]

Multimedia Appendix 3
Usability survey.
[PDF File, 56 KB - cancer_v11i1e65461_app3.pdf]

Multimedia Appendix 4
Educational session evaluation survey.
[PDF File, 46 KB - cancer_v11i1e65461_app4.pdf]

Multimedia Appendix 5
Characteristics of practices, practice staff, and survey respondents.
[PDF File, 78 KB - cancer_v11i1e65461_app5.pdf]

References

1. Pearson C, Fraser J, Peake M, et al. Establishing population-based surveillance of diagnostic timeliness using linked cancer registry and administrative data for patients with colorectal and lung cancer. Cancer Epidemiol 2019 Aug;61:111-118. [doi: 10.1016/j.canep.2019.05.010] [Medline: 31207483]
2. Swann R, McPhail S, Witt J, et al. Diagnosing cancer in primary care: results from the National Cancer Diagnosis Audit. Br J Gen Pract 2018 Jan;68(666):e63-e72. [doi: 10.3399/bjgp17X694169] [Medline: 29255111]
3. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. BMJ Qual Saf 2017 Jun;26(6):484-494. [doi: 10.1136/bmjqs-2016-005401] [Medline: 27530239]

4. Koo MM, Hamilton W, Walter FM, Rubin GP, Lyratzopoulos G. Symptom signatures and diagnostic timeliness in cancer patients: a review of current evidence. Neoplasia 2018 Feb;20(2):165-174. [doi: 10.1016/j.neo.2017.11.005] [Medline: 29253839]
5. Bergin RJ, Emery J, Bollard RC, et al. Rural–urban disparities in time to diagnosis and treatment for colorectal and breast cancer. Cancer Epidemiol Biomarkers Prev 2018 Sep 1;27(9):1036-1046. [doi: 10.1158/1055-9965.EPI-18-0210]
6. Pearson C, Poirier V, Fitzgerald K, Rubin G, Hamilton W. Cross-sectional study using primary care and cancer registration data to investigate patients with cancer presenting with non-specific symptoms. BMJ Open 2020 Jan 10;10(1):e033008. [doi: 10.1136/bmjopen-2019-033008] [Medline: 31924638]
7. Bailey SE, Ukoumunne OC, Shephard EA, Hamilton W. Clinical relevance of thrombocytosis in primary care: a prospective cohort study of cancer incidence using english electronic medical records and cancer registry data. Br J Gen Pract 2017 Jun;67(659):e405-e413. [doi: 10.3399/bjgp17X691109] [Medline: 28533199]
8. Mounce LT, Hamilton W, Bailey SE. Cancer incidence following a high-normal platelet count: cohort study using electronic healthcare records from english primary care. Br J Gen Pract 2020 Sep;70(698):e622-e628. [doi: 10.3399/bjgp20X710957] [Medline: 32719013]
9. Yates JM, Logan ECM, Stewart RM. Iron deficiency anaemia in general practice: clinical outcomes over three years and factors influencing diagnostic investigations. Postgrad Med J 2004 Jul;80(945):405-410. [doi: 10.1136/pgmj.2003.015677] [Medline: 15254305]
10. Singh H, Weingart SN. Diagnostic errors in ambulatory care: dimensions and preventive strategies. Adv Health Sci Educ Theory Pract 2009 Sep;14 Suppl 1(0 1):57-61. [doi: 10.1007/s10459-009-9177-z] [Medline: 19669923]
11. Bastani R, Yabroff KR, Myers RE, Glenn B. Interventions to improve follow-up of abnormal findings in cancer screening. Cancer 2004 Sep 1;101(S5):1188-1200. [doi: 10.1002/cncr.20506]
12. Round T, Steed L, Shankleman J, Bourke L, Risi L. Primary care delays in diagnosing cancer: what is causing them and what can we do about them? J R Soc Med 2013 Nov;106(11):437-440. [doi: 10.1177/0141076813504744] [Medline: 24108536]
13. Murphy DR, Wu L, Thomas EJ, Forjuoh SN, Meyer AND, Singh H. Electronic trigger-based intervention to reduce delays in diagnostic evaluation for cancer: a cluster randomized controlled trial. J Clin Oncol 2015 Nov 1;33(31):3560-3567. [doi: 10.1200/JCO.2015.61.1301] [Medline: 26304875]
14. Moja L, Kwag KH, Lytras T, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. Am J Public Health 2014 Dec;104(12):e12-e22. [doi: 10.2105/AJPH.2014.302164] [Medline: 25322302]
15. Dikomitis L, Green T, Macleod U. Embedding electronic decision-support tools for suspected cancer in primary care: a qualitative study of GPs' experiences. Prim Health Care Res Dev 2015 Nov;16(6):548-555. [doi: 10.1017/S1463423615000109] [Medline: 25731758]
16. Mansell G, Shapley M, Jordan JL, Jordan K. Interventions to reduce primary care delay in cancer referral: a systematic review. Br J Gen Pract 2011 Dec;61(593):e821-e835. [doi: 10.3399/bjgp11X613160] [Medline: 22137419]
17. Nicholson BD, Goyder CR, Bankhead CR, et al. Responsibility for follow-up during the diagnostic process in primary care: a secondary analysis of international cancer benchmarking partnership data. Br J Gen Pract 2018 May;68(670):e323-e332. [doi: 10.3399/bjgp18X695813] [Medline: 29686134]
18. Murphy DR, Laxmisan A, Reis BA, et al. Electronic health record-based triggers to detect potential delays in cancer diagnosis. BMJ Qual Saf 2014 Jan;23(1):8-16. [doi: 10.1136/bmjqs-2013-001874] [Medline: 23873756]
19. Chima S, Reece JC, Milley K, Milton S, McIntosh JG, Emery JD. Decision support tools to improve cancer diagnostic decision making in primary care: a systematic review. Br J Gen Pract 2019 Dec;69(689):e809-e818. [doi: 10.3399/bjgp19X706745] [Medline: 31740460]
20. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3(1):17. [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
21. May C, Finch T, Mair F, et al. Understanding the implementation of complex interventions in health care: the normalization process model. BMC Health Serv Res 2007 Sep 19;7(1):1-7. [doi: 10.1186/1472-6963-7-148] [Medline: 17880693]
22. Chima S, Martinez-Gutierrez J, Hunter B, et al. Future Health Today and patients at risk of undiagnosed cancer: a pragmatic cluster randomised trial of quality- improvement activities in general practice. Br J Gen Pract 2025 Apr 7:BJGP.2024.0491. [doi: 10.3399/BJGP.2024.0491] [Medline: 39567181]
23. Smelt AFH, van der Weele GM, Blom JW, Gussekloo J, Assendelft WJJ. How usual is usual care in pragmatic intervention studies in primary care? an overview of recent trials. Br J Gen Pract 2010 Jul;60(576):e305-e318. [doi: 10.3399/bjgp10X514819] [Medline: 20594432]
24. Roland M, Torgerson DJ. Understanding controlled trials: what are pragmatic trials? BMJ 1998 Jan 24;316(7127):285-285. [doi: 10.1136/bmj.316.7127.285]
25. Manski-Nankervis J. Future health today: a cluster randomised controlled trial of quality improvement activities in general practice. Australian New Zealand Clinical Trials Registry. 2020. URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=380119&isReview=true [accessed 2024-07-10]

26. Hunter B, Biezen R, Alexander K, et al. Future Health Today: codesign of an electronic chronic disease quality improvement tool for use in general practice using a service design approach. BMJ Open 2020 Dec 18;10(12):e040228. [doi: 10.1136/bmjopen-2020-040228] [Medline: 33371024]
27. Chima S, Martinez-Gutierrez J, Hunter B, Manski-Nankervis JA, Emery J. Optimization of a quality improvement tool for cancer diagnosis in primary care: qualitative study. JMIR Form Res 2022 Aug 4;6(8):e39277. [doi: 10.2196/39277] [Medline: 35925656]
28. Hunter B, Davidson S, Lumsden N, et al. Optimising a clinical decision support tool to improve chronic kidney disease management in general practice. BMC Prim Care 2024 Jun 19;25(1):220. [doi: 10.1186/s12875-024-02470-w] [Medline: 38898462]
29. Hunter B, Alexander K, Biezen R, et al. The development of Future Health Today: piloting a new platform for identification and management of chronic disease in general practice. Aust J Prim Health 2022;29(1):8-15. [doi: 10.1071/PY22022]
30. Glasgow RE, Harden SM, Gaglio B, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019;7:64. [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
31. Lau R, Stevenson F, Ong BN, et al. Achieving change in primary care—causes of the evidence to practice gap: systematic reviews of reviews. Implementation Sci 2015 Dec;11(1):1-39. [doi: 10.1186/s13012-016-0396-4]
32. Lau R, Stevenson F, Ong BN, et al. Achieving change in primary care--effectiveness of strategies for improving implementation of complex interventions: systematic review of reviews. BMJ Open 2015 Dec 23;5(12):e009993. [doi: 10.1136/bmjopen-2015-009993] [Medline: 26700290]
33. Arora S, Geppert CMA, Kalishman S, et al. Academic health center management of chronic diseases through knowledge networks: project ECHO. Acad Med 2007 Feb;82(2):154-160. [doi: 10.1097/ACM.0b013e31802d8f68] [Medline: 17264693]
34. Brown B, Gude WT, Blakeman T, et al. Clinical Performance Feedback Intervention Theory (CP-FIT): a new theory for designing, implementing, and evaluating feedback in health care based on a systematic review and meta-synthesis of qualitative research. Implement Sci 2019 Apr 26;14(1):40. [doi: 10.1186/s13012-019-0883-5] [Medline: 31027495]
35. Chima S, Hunter B, Martinez-Gutierrez J, et al. Adoption, acceptance, and use of a decision support tool to promote timely investigations for cancer in primary care. Fam Pract 2024 Dec 2;41(6):1048-1057. [doi: 10.1093/fampra/cmae046] [Medline: 39425610]
36. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
37. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum Comput Interact 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
38. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing among Five Approaches: Sage Publications; 2016.
39. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017;16(1). [doi: 10.1177/16094069177338]
40. Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258. [doi: 10.1136/bmj.h1258] [Medline: 25791983]
41. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]
42. Lakshman R, Griffin S, Hardeman W, Schiff A, Kinmonth AL, Ong KK. Using the Medical Research Council framework for the development and evaluation of complex interventions in a theory-based infant feeding intervention to prevent childhood obesity: the baby milk intervention and trial. J Obes 2014;2014(1):646504. [doi: 10.1155/2014/646504] [Medline: 25143830]
43. Bobrow K, Farmer A, Cishe N, et al. Using the Medical Research Council framework for development and evaluation of complex interventions in a low resource setting to develop a theory-based treatment support intervention delivered via SMS text message to improve blood pressure control. BMC Health Serv Res 2018 Jan 23;18(1):33. [doi: 10.1186/s12913-017-2808-9] [Medline: 29361934]
44. The impact of COVID-19 on cancer-related medical services and procedures in australia in 2020. Cancer Australia. URL: https://www.canceraustralia.gov.au/the-impact-of-COVID-19-on-cancer-related-medical-services-and-procedures-in-Australia-in-2020 [accessed 2024-03-01]
45. Siebenhofer A, Paulitsch MA, Pregartner G, et al. Cluster-randomized controlled trials evaluating complex interventions in general practices are mostly ineffective: a systematic review. J Clin Epidemiol 2018 Feb;94:85-96. [doi: 10.1016/j.jclinepi.2017.10.010] [Medline: 29111470]
46. Squires JE, Sullivan K, Eccles MP, Worswick J, Grimshaw JM. Are multifaceted interventions more effective than single-component interventions in changing health-care professionals' behaviours? An overview of systematic reviews. Implement Sci 2014 Oct 6;9(1):152. [doi: 10.1186/s13012-014-0152-6] [Medline: 25287951]
47. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]

48. Price S, Spencer A, Medina-Lara A, Hamilton W. Availability and use of cancer decision-support tools: a cross-sectional survey of UK primary care. Br J Gen Pract 2019 Jul;69(684):e437-e443. [doi: 10.3399/bjgp19X703745] [Medline: 31064743]
49. Francke AL, Smit MC, de Veer AJE, Mistiaen P. Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review. BMC Med Inform Decis Mak 2008 Sep 12;8(1):1-11. [doi: 10.1186/1472-6947-8-38] [Medline: 18789150]

Abbreviations

CDS: clinical decision support
ECHO: Extension for Community Healthcare Outcomes
EMR: electronic medical record
FHT: Future Health Today
GP: general practitioner
GPN: general practice nurse
MRC: Medical Research Council
PM: practice manager
PSA: prostate-specific antigen
QI: quality improvement
REDCap: Research Electronic Data Capture
SUS: System Usability Scale

Edited by N Cahill; submitted 15.08.24; peer-reviewed by O Jones, R Calitri; revised version received 25.03.25; accepted 27.03.25; published 12.06.25.

Please cite as:
Chima S, Hunter B, Martinez-Gutierrez J, Lumsden N, Nelson C, Boyle D, Somasundaram K, Manski-Nankervis JA, Emery J
Implementation of a Quality Improvement and Clinical Decision Support Tool for Cancer Diagnosis in Primary Care: Process Evaluation
JMIR Cancer 2025;11:e65461
URL: https://cancer.jmir.org/2025/1/e65461
doi:10.2196/65461

The Role of Online Support, Caregiving, and Gender in Preventative Cancer Genetic Testing Participation: Cross-Sectional Study From a National Study

Lavlin Agrawal[1*], PhD; Richelle Oakley DaSouza[1*], PhD; Pavankumar Mulgund[2*], PhD; Pankaj Chaudhary[1*], PhD

[1]Department of Business Information Systems and Analytics, Willie A. Deese College of Business and Economics, North Carolina A&T State University, 1601 E. Market Street, 337 Merrick Hall, Greensboro, NC, United States

[2]Department of Management Information Systems, Fogelman College of Business & Economics, University of Memphis, Memphis, TN, United States

[*]all authors contributed equally

Corresponding Author:
Lavlin Agrawal, PhD
Department of Business Information Systems and Analytics, Willie A. Deese College of Business and Economics, North Carolina A&T State University, 1601 E. Market Street, 337 Merrick Hall, Greensboro, NC, United States

Abstract

Background: Despite its potential to predict and detect early cancer risks, genetic testing remains underused by the public. This study, guided by the health belief model (HBM), examined key factors influencing an individual's willingness to undergo genetic testing for cancer, with a particular focus on gender, caregiver status, and participation in online social support groups.

Objective: This study aimed to explore the factors that can influence the individual's decision to undergo preventative genetic testing for cancer so that more informed action can be taken to encourage the individuals to engage in preventative health behavior.

Methods: This study uses data collected from the 2020 Health Information National Trends Survey (HINTS 5 Cycle 4), which included 2947 respondents representing 199,510,996 US adults aged 18 years and older. Multivariable logistic regression and survey-weighted generalized linear models were applied to examine the relationship between cancer genetic testing and caregiver status, participation in online support groups, gender, and constructs associated with the HBM, while controlling for sociodemographic and health-related characteristics.

Results: Our findings show that women are more likely to undergo cancer genetic testing, with gender moderating the influence of perceived susceptibility (β=2.54, *P*=.03) and severity (β=0.94, *P*<.050) on testing decisions. In line with the HBM, perceived benefits (β=0.19, *P*=.03) and cues to action (β=2.86, *P*<.001) increase the likelihood of testing. Results also show that caregivers of patients with cancer (β=1.25, *P*=.04) and those actively participating in online health support groups (β=0.47, *P*=.04) are also more likely to engage in cancer genetic testing.

Conclusions: Cancer remains a significant health challenge in the United States, with 1.8 million new cases and 606,520 deaths annually. Early detection is vital for treatment success. This study investigates factors influencing the decision to undergo genetic testing for cancer. The examination of caregiver status and online support groups as influencing factors, along with the HBM, provided a significant theoretical contribution to the health care research domain. Results indicated that caregivers and men should be directly targeted with messaging on genetic cancer screening as a proactive health behavior. Additionally, online support groups can promote early detection and encourage participation in genetic testing. Future research should further explore implementing proactive outreach strategies to encourage wider adoption of genetic testing for cancer.

(JMIR Cancer 2025;11:e67650) doi:10.2196/67650

KEYWORDS

genetic testing; cancer; health belief model; caregiver status; online social support; gender differences

Introduction

Overview

Cancer remains one of the most urgent health challenges in the United States, with approximately 2 million new cases and 611,000 deaths projected to occur in the United States in 2024 [1]. These figures highlight the importance of early detection and timely intervention, which are critical for improving treatment outcomes and reducing cancer-related mortality [2]. Early detection, particularly through predictive methods, enables health care providers to identify cancers at earlier stages, where treatment options can significantly improve the chances of successful outcomes. Consequently, the development and widespread adoption of effective screening strategies have become paramount in the fight against cancer.

Among these strategies, genetic testing has emerged as a pivotal tool in cancer prevention and early detection [3]. By analyzing an individual's DNA, genetic testing can identify mutations that may predispose individuals to various forms of cancer [4]. This predictive capability empowers individuals with knowledge that allows them to take preventive actions or pursue early treatments. As medical science and technology continue to advance, the scope of genetic testing has expanded to detect a wide range of inherited disorders, making it an indispensable approach for reducing cancer risk [5,6].

The health belief model (HBM) offers a valuable framework to understand why individuals engage in preventive behaviors like genetic testing. The HBM suggests that individuals' health behaviors are shaped by their perceptions of susceptibility, severity, and the benefits of action [7]. This model has been widely applied to the study of various preventive health actions, including vaccinations for H1N1 [8] and COVID-19 [9], as well as screening for conditions like diabetes [10] and several types of cancer [11-14]. Alongside the HBM, three key factors—caregiver status, participation in online social support groups, and gender—are likely to influence preventive behaviors related to genetic testing.

Caregivers, who are often family members, gain a deeply personal and intimate understanding of cancer's devastating impact. Due to shared genetic ties and family history, these caregivers frequently face a higher risk of developing cancer themselves [15,16]. Their close connection to the disease makes them a critical demographic for studying behaviors related to genetic testing. Given their heightened risk and firsthand exposure to cancer, caregivers may be particularly motivated to pursue genetic testing as a preventive measure.

In addition to caregiver status, online social support groups have been shown to play a crucial role in promoting preventive behaviors. Hwang et al [17] demonstrated that participation in online peer support groups significantly increased motivation for colorectal cancer screening, reinforcing participants' belief in the effectiveness of early detection. Lastly, gender has a profound influence on preventive health actions. Research indicates that women tend to be more proactive in engaging in preventive health behaviors [18-20]. For instance, women with low social support are less likely to participate in breast cancer screening, further illustrating the interplay between social support and gender in health decision-making behaviors [21].

The remainder of this paper is organized as follows. First, we provide an overview of the literature on genetic testing and the HBM. Next, we develop hypotheses and propose our research model. We then outline the research methodology and present the results. Following this, we discuss the findings, exploring both theoretical and practical implications. Finally, we conclude with a summary of the key insights gained from this research and address its limitations.

Background

Advancements in identifying gene mutations have made genetic testing a crucial tool in reducing illness and death by enabling early detection and preventive measures [5]. Predictive genetic testing, which assesses an individual's risk of developing diseases like cancer, focuses on identifying genetic mutations linked to disease susceptibility [22]. However, challenges persist in diagnosing hereditary breast and ovarian cancers, particularly in terms of genetic literacy and interpreting test results [23]. Nelson et al [24] emphasize the critical role of genetic counseling in BRCA-related cancer testing, highlighting the importance of providing informed decision-making support to patients.

Prior research also suggests that people with a family history of cancer often overestimate their risk. Their decisions about genetic testing are often influenced more by their subjective perceptions of vulnerability than by objective data [25]. This cognitive bias can lead to increased anxiety. It may cause individuals to rush into testing without enough information or avoid testing out of fear of the results. Despite these misperceptions, genetic testing consistently offers significant benefits, regardless of the outcome. Whether the results confirm a genetic predisposition or provide reassurance, testing allows individuals to make informed health decisions, take preventive measures, and access appropriate counseling and support services [26]. This highlights the importance of education and counseling in helping individuals to more accurately interpret genetic risks and use the information to make effective health care choices.

Caregivers of patients with cancer often play a pivotal role in health care decision-making, including decisions about genetic testing [27]. As primary support figures, caregivers are frequently involved in gathering health-related information, navigating complex medical choices, and encouraging preventative behaviors. Social support groups, both in-person and online, have been shown to provide critical informational and emotional support during these processes [28]. With the rise of online communities, individuals considering genetic testing now have greater access to peer support and shared experiences. Silence [29] demonstrated that online cancer communities facilitate advice-seeking and information-sharing, creating a valuable space for individuals to navigate genetic testing options. Similarly, Ruco et al [30] found that online social interactions significantly increase participation in cancer screening, highlighting the potential of these platforms to influence health-related behaviors. Given their influence in health decisions and the growing prevalence of online support networks, understanding the combined impact of caregiver status and participation in online social support groups is essential for accurately capturing the social and behavioral drivers behind genetic testing decisions.

The HBM is frequently used to examine health behaviors by considering factors such as perceived susceptibility, severity, cues to action, benefits, barriers, and demographic variables [7]. Bunn et al [31] and Hartman [32] successfully applied the HBM to predict decisions about preventive screenings, like colon cancer screening and mammography. The HBM provides a robust framework for understanding and influencing health-related decision-making, particularly in the context of preventive interventions such as genetic testing.

Theoretical Model

Health Belief Model—Perceived Susceptibility

Perceived susceptibility is an individual's belief about their likelihood of experiencing a health condition and is a key driver of health-related behavior according to the HBM [33,34]. Research consistently demonstrates that higher perceived susceptibility significantly influences medical decisions. For instance, Champion and Skinner [35] found that women with greater perceived susceptibility to breast cancer were more likely to undergo mammography. Similarly, Irigoyen-Camacho et al [36] observed that older adults with heightened perceived susceptibility to COVID-19 engaged more in preventive behaviors. Meta-analyses further confirm the strong correlation between perceived susceptibility and health behavior change [37,38]. Thus, we hypothesize the following:

Hypothesis H1a: Perceived susceptibility positively influences an individual's decision to undergo genetic testing for cancer.

Further, the impact of perceived susceptibility varies by gender. Studies suggest that women generally perceive greater health risks and, therefore, are more likely to adopt health-promoting behaviors, such as regular screenings and healthier lifestyles [39]. For example, Lisha et al [40] found that women with similar levels of physical activity as men had lower alcohol consumption. Based on this, we hypothesize the following:

Hypothesis H1b: Gender moderates the relationship between perceived susceptibility and an individual's decision to undergo genetic testing for cancer.

Health Belief Model—Perceived Severity

Perceived severity refers to an individual's belief about the seriousness of contracting an illness [34]. According to the HBM, higher perceived severity increases motivation to engage in health-promoting behaviors [35]. Studies by Witte and Allen [41] and Brewer et al [37] show that individuals who perceive health threats as severe are more likely to adopt preventive measures, such as vaccination. Similarly, Irigoyen-Camacho et al [36] found that the perceived severity of COVID-19 significantly influenced compliance with stay-at-home guidelines. He et al [42] also reported that individuals with high perceived severity of colorectal cancer were more likely to undergo colonoscopy. These findings highlight the pivotal role of perceived severity in driving medical action. Thus, we hypothesize the following:

Hypothesis H2a: Perceived severity positively influences an individual's decision to undergo genetic testing for cancer.

Further, extant research found that women perceive risks higher and engage in more preventive health behaviors [43]. Women generally report higher levels of perceived severity regarding health issues compared to men, and this heightened perception is linked to greater engagement in preventive health behaviors [39]. Sattler et al [44] also demonstrated that women were more likely than men to perceive the severity of health threats such as COVID-19, which translated into a higher likelihood of adopting recommended health behaviors. Thus, we extend this notion and hypothesize the following:

Hypothesis H2b: Gender moderates the relationship between perceived severity and an individual's decision to undergo genetic testing for cancer.

Health Belief Model—Perceived Benefit

Perceived benefits refer to an individual's belief in the positive outcomes of a health action, and they play a crucial role in motivating behavior change [35]. For instance, individuals are more likely to engage in physical activity if they believe it will lead to significant health improvements [45]. Similarly, Chen et al [46] found that perceived benefits significantly impact decisions to get vaccinated. Chen et al [47] further confirmed the association between perceived benefits and preventive behaviors. Bosompra et al [48] found that perceived benefits significantly impacted decisions to undergo genetic testing for cancer susceptibility. Based on this, we hypothesize the following:

Hypothesis H3: Perceived benefits positively influence an individual's decision to undergo genetic testing for cancer.

Health Belief Model—Perceived Barrier

Perceived barriers refer to an individual's assessment of obstacles that hinder the adoption of health-related behaviors [34]. Research consistently shows that these barriers negatively impact medical decision-making. According to the HBM, barriers may include factors such as cost, time, inconvenience, and fear of adverse outcomes [35]. A systematic review by Al-Noumani et al [49] identifies perceived barriers as a key predictor of poor adherence to health behavior changes in chronic conditions. Similarly, studies demonstrate that perceived barriers influence compliance with COVID-19 preventive measures over time [50,51]. Building upon these findings, we hypothesize the following:

Hypothesis H4: Perceived barriers negatively influence an individual's decision to undergo genetic testing for cancer.

Health Belief Model—Cues to Action

Cues to action are stimuli that prompt individuals to engage in health-promoting behaviors and are crucial for motivating medical action [35]. In the HBM, cues can be internal, such as experiencing symptoms, or external, like advice from others or health campaigns. Carpenter [52] found that both internal cues, like symptoms, and external cues, such as media messages, significantly increase the likelihood of seeking medical care. Similarly, Glanz et al [53] showed that health communication campaigns effectively acted as external cues, leading to higher vaccination and screening rates. Based on this, we hypothesize the following:

Hypothesis H5: Cues to action positively influence an individual's decision to undergo genetic testing for cancer.

Caregiving

A caregiver provides care and support to someone with health-related needs due to chronic illness, disability, or aging [54]. When a family member is diagnosed with cancer, caregivers are deeply involved in diagnosis, treatment, and survivorship care [55]. During this time, they interact closely with health care providers, gather medical information, and

witness their loved one's experiences. The Center for Disease Control encourages caregivers to practice self-care and engage in preventive health care [56]. Research shows that spousal caregivers are more likely to undergo cancer screenings [57,58], with evidence of increased screening for stomach, breast, and cervical cancer among caregivers [59]. Additionally, caregivers are more likely to adopt health-promoting behaviors [60]. Extending this idea to genetic testing, we hypothesize the following:

Hypothesis H6: Caregiving positively influences an individual's decision to undergo genetic testing for cancer.

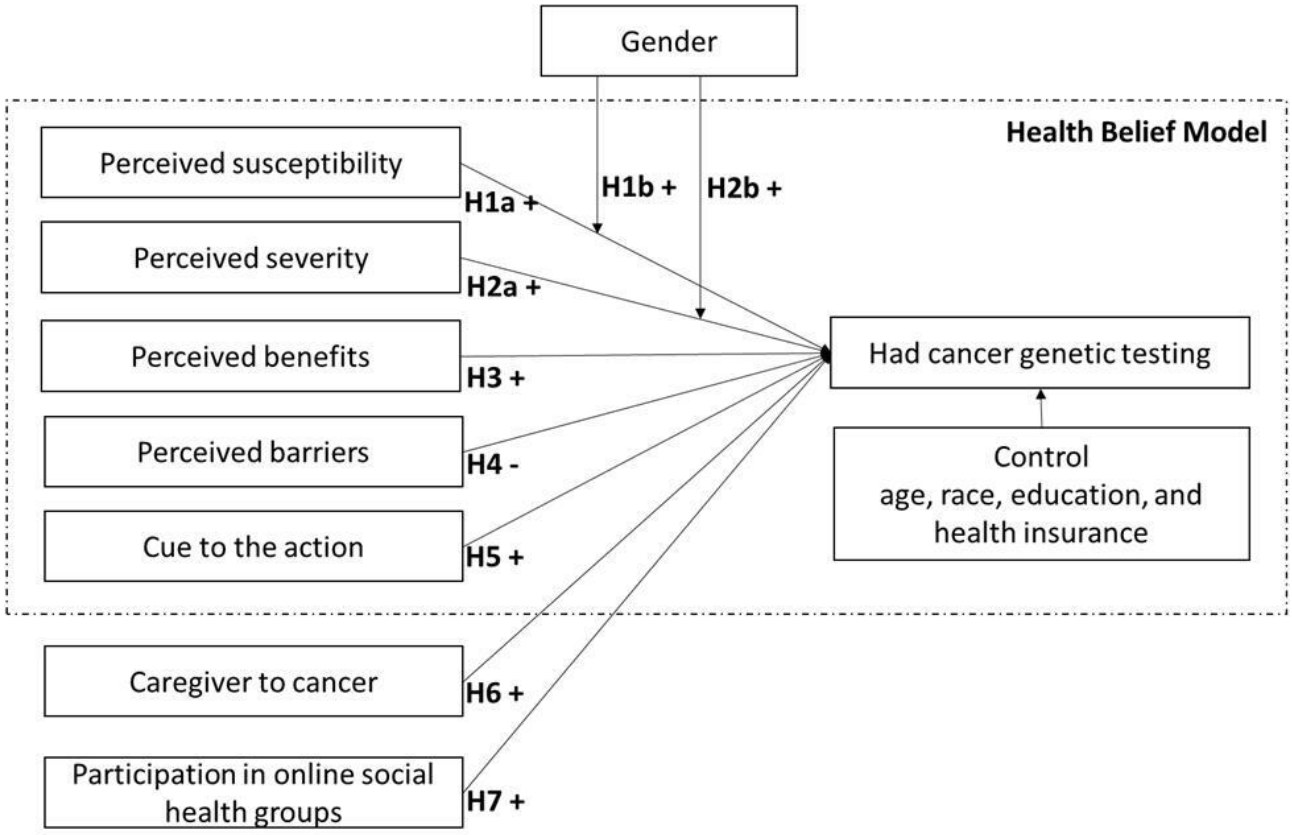
Online Social Support

Online social support groups provide a platform for individuals to share experiences and receive encouragement from others facing similar health challenges. This support fosters motivation and adherence to health goals. Participants often feel more committed to their health plans when they receive positive feedback and peer encouragement [61]. These groups also offer emotional support, reducing stress and anxiety, leading to better adherence to health-promoting behaviors and overall well-being [62]. Additionally, personalized advice from peers who have faced similar issues can be more practical and relevant than generic information, resulting in more effective health behavior changes [63]. A systematic review further supports the effectiveness of social media tools in delivering interventions for cancer prevention and management [64,65]. Based on these insights, we hypothesize the following:

Hypothesis H7: Participation in online health communities positively influences an individual's decision to undergo genetic testing for cancer.

Based on these hypotheses, our proposed research model is presented in Figure 1.

Figure 1. Research model.



Methods

Data

We used data from the 2020 National Cancer Institute Health Information National Trends Survey (HINTS 5 Cycle 4) survey, which is publicly available on the HINTS website [66]. Table 1 presents the correlations between the principal variables with survey weights. High correlations can indicate multicollinearity, but all correlations fall below the threshold of 0.5 [67], so multicollinearity was not an issue. The table also includes the means and SDs for the key variables.

Table . Correlation matrix.

	Mean	SD	PSUS[a]	PSEV[b]	PBEN[c]	CUES[d]	CC[e]	OSS[f]
PSUS	0.732	0.665	1.000					
PSEV	2.509	0.951	0.010	1.000				
PBEN	9.873	1.585	0.060	0.010	1.000			
CUES	0.438	0.704	0.160	−0.120	0.100	1.000		
CC	0.019	0.369	0.030	0.030	−0.090	0.020	1.000	
OSS	0.266	0.746	−0.020	0.040	0.060	0.040	0.010	1.000

[a]PSUS: Perceived susceptibility.

[b]PSEV: Perceived severity.

[c]PBEN: Perceived benefit.

[d]CUES: Cues to action.

[e]CC: Caregiving cancer.

[f]OSS: Online social support.

Measurements

The complete set of questionnaires for the dependent, independent, and control variables can be found in Multimedia Appendix 1. Although many variables are measured using multi-item scales, some are captured through single-item measures. The use of single-item measures is deemed appropriate when the question is straightforward and unambiguous, minimizing the risk of varied interpretation [68]. Moreover, single-item measures are widely recognized and commonly used within the health care domain [60,69,70], further supporting their validity in this study.

Outcome Variable: Cancer Genetic Testing

The main outcome variable "Had cancer genetic test" is a dichotomous variable and coded "1" for an individual who had a genetic test for cancer, otherwise "0." In our data, 142 respondents said they had cancer genetic testing; this represents 8,368,022 in the population when using survey weights. In total, 2805 respondents said they never had cancer genetic testing, which represents 191,142,974 in the population when using survey weights.

We effectively mitigated the potential concern of skewed outcome variables through the application of survey weights. Survey weights are designed to adjust for the complex survey design, including oversampling and nonresponse, ensuring that the results are representative of the target population. Studies have shown that when survey weights are appropriately applied, they can correct for biases introduced by skewed distributions in outcome variables, leading to more accurate and generalizable results [71].

Primary Independent Variables: Cancer Caregiver and Online Social Support

Caregiver cancer is coded as '1' if the individual is caring for or making medical decisions for someone with cancer, and '0' otherwise. Online social support is coded as '2' for individuals who both share health information on social networking sites and participate in an online forum or support group related to similar health or medical issues. If they engage in only one of these activities, it is coded as '1,' and if neither activity is present, it is coded as '0.'

Support Independent Variables: Health Belief Model

Perceived susceptibility is coded as "1" if first- or second-degree biological relatives had cancer, otherwise "0." Perceived severity ranges from "1" to "5" for respondents' general health, varying from "Excellent" to "Poor." Cues to action are coded as "1" if respondents have heard of cancer genetic testing, otherwise "0."

The perceived benefit was obtained by summing respondents' answers to 3 survey questions: How important is knowing a person's genetic information for preventing cancer? How important is knowing a person's genetic information for detecting cancer? and How important is knowing a person's genetic information for treating cancer. These questions have 4 options from "Not at all" to "Very." Several papers have pointed to income as a potential barrier to preventative health care [72,73]. Hence, we consider income as a proxy for perceived barrier, and it has 5 categories "1" to "5." Less than US $20,000 is coded as "1" while US $75,000 or more is coded as "5."

Moderating Variable: Gender

Females were coded as "1" and male as "0."

Control Variables

Race, education, marital status, insurance, and age were used as controls in the model. Whites were coded as "1" and non-Whites were coded as "0." Education of more than high school was coded "1" for individuals, otherwise "0." For married individuals, married was coded as "1", or else "0." Insured individuals were coded as "1" or else "0." We used the natural log of "Age" in our model as "Age" varies from 18 to 102 years.

Statistical Analysis

This study uses survey weights to report national estimates. We used survey-weighted generalized linear models to test the hypotheses in R using the "survey" package. We used HINTS-supplied survey weights using jackknife variance estimation techniques to account for the complex HINTS sampling design and to calculate nationally representative

estimates [74]. Since we used publicly available deidentified data, the Institutional Review Board review was exempted.

Common Method Variance

Data collected with a self-reported single survey may suffer from common method variance (CMV), which hampers the relationship between the variables [75]. To check if our data are suffering from CMV, we used the marker variable technique [76]. A marker variable is a variable that is theoretically unrelated to one or more of the principal variables measured in the study and typically should have a low correlation with the variables of interest.

The theoretically unrelated construct "UnderstandOnlineMedRec" (UOM) was used as a marker variable. The correlation between the marker variable UOM and other principal variables is very low, indicating that CMV is not a problem. The UOM for perceived susceptiblity was 0.03, that for perceived severity was 0.18, that for perceived benefit was −0.03, that for cues to action was −0.21, that for caregiving cancer was 0.02, and that for online social support was 0.02.

Ethical Considerations

The HINTS 5 survey, conducted with the general population, underwent expedited review and received approval from the Westat Institutional Review Board on March 28, 2016 (project no. 6048.14). In addition, on April 25, 2016, the National Institutes of Health Office of Human Subjects Research determined that the survey did not involve human subjects research, providing an exemption (exempt no. 13204) [77]. This analysis used deidentified, publicly available data from the HINTS, which did not constitute human subjects research as defined by 45 CFR 46.102 and, therefore, did not require IRB review. The original consent and IRB approval cover secondary analysis without the need for additional consent. No compensation was provided for participation.

Results

Study Population Characteristics

The data include 3865 civilian, noninstitutionalized US adults aged 18 or older. After filtering for valid responses to the question about genetic testing for high-risk cancer, our final sample consisted of 2947 respondents, representing 199,510,996 US adults. Descriptive statistics of the survey respondents are shown in Table 2 with survey weights applied.

Table . Descriptive statistics.

Demographic characteristics	Sample size, n (%)	With survey weights
	2947 (100)	199,510,996
Gender		
Male	1126 (38.21)	91,565,888
Female	1623 (55.07)	98,228,842
Race		
White	2192 (74.38)	151,626,358
Non-White	594 (20.16)	36,607,304
Education		
Up to high school	597 (20.26)	50,563,767
More than high school	2278 (77.3)	145,310,461
Insured		
Yes	2775 (97.35)	180,753,226
No	172 (2.65)	18,757,770
Married		
Yes	1434 (48.66)	94,007,748
No	1443 (48.96)	101,703,046
Income		
Less than US $20,000	395 (13.4)	23,749,725
US $20,000 to < US $35,000	312 (10.59)	17,936,144
US $35,000 to <US $50,000	344 (11.67)	21,823,964
US $50,000 to < US $75,000	488 (16.56)	35,145,330
US $75,000 or more	1153 (39.12)	87,555,797
Age, years		
Range	18-102	18-102
Mean (SD)	55.34 (16.65)	47.17 (17.26)

Multivariable Logistic Regression Analysis

In Table 3, we present the results from the multivariate survey-weighted generalized linear model that provide significant insights into the factors influencing individuals' decisions to undergo genetic testing for cancer.

Table . Regression results.

Hypothesis and variables	Estimate	SE	*t* value	Pr(>|t|)	Significant
H1a					
Perceived susceptibility	−0.70805	0.76458	−0.926	0.3616	No
H1b					
Perceived susceptibility: female	2.53956	1.15116	2.206	0.0349	Yes
H2a					
Perceived severity	−0.58773	0.38515	−1.526	0.1372	No
H2b					
Perceived severity: female	0.94366	0.43795	2.155	0.0391	Yes
H3					
Perceived benefit	0.18903	0.08685	2.176	0.0373	Yes
H4					
Income 2	−1.11782	0.68477	−1.632	0.1127	No
Income 3	−0.17123	0.64841	−0.264	0.7935	No
Income 4	0.44008	0.59109	0.745	0.4622	No
Income 5	0.05962	0.55642	0.107	0.9154	No
H5					
Cues to action	2.85722	0.55672	5.132	1.47E-05	Yes
H6					
Caregiving cancer	1.25257	0.60046	2.086	0.0453	Yes
H7					
Online social support	0.47306	0.23093	2.049	0.0491	Yes
Female	−3.48317	1.46416	−2.379	0.0237	Yes
Married	−0.1569	0.30504	−0.514	0.6106	No
LogAge	1.17877	0.43007	2.741	0.0101	Yes
White	−0.07895	0.43618	−0.181	0.8575	No
HighSchoolMore	−0.76716	0.35812	−2.142	0.0401	Yes
Insurance	2.18795	1.21807	1.796	0.0822	No

The analysis did not identify a significant direct relationship between perceived susceptibility and the decision to undergo genetic testing (β=−0.70805, *P*=.36), leading to the rejection of hypothesis H1a. However, gender was found to be a significant moderator in this relationship. Specifically, women who perceive themselves as susceptible to cancer are significantly more likely to pursue genetic testing (β=2.53956, *P*=.03), providing support for hypothesis H1b.

Similarly, no significant direct association was observed between perceived severity and the decision to undergo genetic testing (β=−.58773, *P*=.13), leading to the rejection of hypothesis H2a. However, gender again played a moderating role in this relationship. Women were more likely than men to opt for genetic testing when they perceived cancer as a severe threat (β=.94366, *P*=.03), supporting hypothesis H2b.

Beyond gender effects, the findings indicate that individuals who perceive greater benefits from genetic testing are more inclined to undergo testing (β=.18903, *P*=.03), confirming hypothesis H3. The expected relationship between perceived barriers—represented by income in this study—and genetic testing decisions (hypothesis H4) was not supported (β=−1.11782,−0.17123, 0.44008, 0.05962, *P*>.05).

Cues to action emerged as a significant predictor of genetic testing decisions, with greater exposure to such cues being strongly associated with an increased likelihood of pursuing genetic testing (β=2.85722, *P*<.001), supporting hypothesis H5. Additionally, individuals who serve as caregivers for patients

with cancer were found to be significantly more likely to engage in genetic testing themselves (β=1.25257, *P*=.04), confirming hypothesis H6. Lastly, participation in online social health groups was positively associated with the likelihood of undergoing genetic testing (β=.47306, *P*=.04), supporting hypothesis H7.

Discussion

Summary of Findings

In line with previous research [78,79], our findings indicate that perceived susceptibility alone may not be enough to motivate individuals to undergo genetic testing. This suggests that additional cues to action or contextual influences may play a crucial role in shaping decision-making.

Moreover, the results show that perceived susceptibility has a stronger effect on genetic testing behavior among women compared to men. One possible explanation is that women may be more sensitive to health risks, particularly hereditary cancers such as breast and ovarian cancer, making them more likely to act on their perceived vulnerability.

This finding is consistent with existing literature [80,81], which suggests that while perceived severity is an important factor, it may not be a sufficient motivator for health-related behaviors without the presence of additional reinforcing elements.

The results also emphasize the importance of accounting for gender differences in health risk perceptions and decision-making. Women's stronger response to perceived severity may be attributed to their heightened awareness of specific cancer risks and a greater tendency to engage in proactive health behaviors.

Additionally, our findings reinforce a core principle of the HBM, which asserts that perceived benefits are a key driver of health-related actions.

Results further suggest that traditional barriers, such as cost, are being alleviated through evolving health care policies and financial assistance programs [82,83]. Measures such as subsidized testing, reduced out-of-pocket expenses, and insurance coverage for preventive screenings have helped minimize these obstacles [84].

Cues to action emerged as a significant predictor of genetic testing decisions. One explanation for this finding is the changing health care landscape, particularly in preventative care and personalized medicine. Many providers now recognize the value of genetic testing for early risk identification, prompting efforts to reduce financial and logistical barriers [85]. Major insurers increasingly cover preventative genetic testing, reducing costs for individuals [86]. Additionally, online support groups play a role by informing individuals about available financial assistance and providing emotional support, reducing psychological barriers like fear or anxiety about test results [87]. These developments suggest that the traditional "perceived barriers" construct in the HBM may not be relevant in the context of genetic cancer testing.

Another important result from the study is the role of caregiving in influencing decisions to undergo genetic testing. This finding underscores the role of online communities in promoting health-related behaviors by providing access to information, emotional support, and shared experiences. These groups act as cues to action by reducing uncertainty and raising awareness about the benefits of genetic testing, empowering individuals to make more informed health care decisions, especially in preventative care contexts.

Theoretical Contributions

Our study extends the prior body of research on HBM by examining factors that influence an individual's decision to undergo cancer genetic testing and incorporating variables that have not been widely explored in prior research. One of the key theoretical advancements of this study is the evidence supporting the moderating role of gender in the relationship between perceived susceptibility, perceived severity, and health behavior change. Our results demonstrate that women are significantly more likely to undergo genetic testing when they perceive higher susceptibility or severity of cancer, thus supporting the hypotheses related to gender moderation. Gender disparity in health decision-making behavior is rooted in cultural, psychological, or social factors that make women more responsive to health risks, particularly those related to cancer, such as breast and ovarian cancer. These findings align with prior research indicating that women are generally more responsive to health risks, particularly in the context of cancer (eg, breast and ovarian cancers), where early detection and preventive measures are critical [35,88]. This gender-specific behavior underscores the need for health interventions that are tailored to reflect differences in health perceptions and behaviors between men and women.

This study extends the HBM by incorporating caregiving as a variable that influences health behaviors. Our results indicate that individuals who are caregivers to patients with cancer are significantly more likely to engage in cancer genetic testing. Caregivers, who are often emotionally and practically involved in managing the health of others, are more attuned to genetic risks and motivated to take preventive action for their own health [89]. This finding suggests that future interventions could target caregivers specifically, providing them with information about the benefits of genetic testing and encouraging preventive health behaviors.

Furthermore, the study highlights the growing importance of digital communities, such as online social health groups, in shaping health behaviors. Our analysis shows that participation in these groups is positively associated with the likelihood of undergoing genetic testing. This finding is consistent with existing literature that points to the influence of social networks on health behavior change [90,91] and the growing importance of digital communities in shaping health behaviors. Online platforms can serve as a source of information, support, and motivation for individuals contemplating health-related decisions. In online platforms, individuals can share experiences, advice, and support, building virtual environments that encourage proactive health actions like genetic testing. As such, health practitioners and policy makers could leverage the power

of these digital communities to enhance awareness and promote the benefits of genetic testing and other preventive health measures.

Practical Implications

This study's findings have several practical implications for health care providers and public health campaigns. Our results support the importance of perceived benefits in the decision to get genetic testing. Efforts to promote cancer genetic testing should focus on clearly communicating the benefits of early detection, such as the ability to develop personalized cancer treatment plans using information from genetic testing. Early diagnosis significantly improves cancer outcomes, particularly when care is provided at the earliest possible stage [92]. Especially with cancer genetic tests, there is immense potential to facilitate early detection and personalized treatment strategies [93]. Our results also support the importance of cues to action in the decision to get genetic testing. Timely and effective communication can motivate individuals to take proactive steps in their health care journey. However, the success of these interventions depends on individuals' willingness to engage in preventive health behaviors. Public health officials can pursue changes to health care policy that will incentivize individuals to be proactive about their health and get genetic cancer screenings. Financial incentives have proven to be effective in encouraging individuals to get screenings [94]. Employers already reduce individual monthly insurance premiums by requiring employees to get regular health screenings [95]. Insurance providers can simply add free or reduced-cost genetic cancer screening in the included health care screenings, thus implicitly encouraging and supporting individuals to get screened.

The moderating role of gender highlights the need for improved public health campaigns directed towards men. The Center for Disease Control highlights that "men have higher rates of getting and dying from cancer than women" [96]. There is a lack of online social discourse on genetic testing, which may be the reason for lower male engagement with genetic testing [97]. To address this concern, there has been a notable increase in the number of public health campaigns encouraging men to get screening tests. Recent research has examined how to leverage social media to bring public awareness to the value of genetic testing for prostate cancer [98]. Further efforts in this area could focus on developing messaging on the benefits of early detection using gender-coded language to specifically target men. To improve online social discourse on cancer genetic testing [97], public health organizations can post informative messages on X (formerly Twitter) to encourage discussion of genetic testing, create and participate in regional Facebook groups and pages to develop a community of men interested in genetic testing, and share Facebook and YouTube videos to dispel concerns and address any misconceptions regarding genetic testing. Additionally, our research found the role of caregivers is particularly noteworthy. Our results indicate that caregivers are more likely to seek genetic testing, potentially due to their heightened awareness of cancer risk factors through their caregiving experience [98]. Health care providers should recognize caregivers as a key demographic for targeted interventions. By educating caregivers on the benefits of genetic testing, both for themselves and their families, health care professionals can increase the uptake of this preventive measure. Caregivers, often deeply involved in health care decisions, could also serve as advocates for genetic testing within their broader social networks, further amplifying the reach of these interventions.

In addition, the positive association between online social health group participation and genetic testing uptake suggests that digital platforms can be effective tools for health promotion. These platforms, where individuals can share experiences and seek advice, provide a valuable avenue for disseminating information about the benefits of genetic testing. Public health campaigns that leverage social media and online communities could encourage greater awareness and engagement with preventive health behaviors [99]. Engaging individuals in these groups may also help to reduce the stigma or fear surrounding genetic testing, ultimately facilitating behavior change.

Limitations and Future Research

Despite its contributions, this study is not without limitations. First, the research relied on secondary data, which constrained the analysis to available variables. The use of secondary data limits the flexibility to explore unmeasured constructs that may further elucidate the decision-making process surrounding genetic testing. Furthermore, we used the HINTS data that were collected in 2020, and hence, it may not fully capture the most recent trends and developments. Future research could consider incorporating primary data collection methods to include more contemporary variables and insights reflective of the current landscape.

Though our study examines the role of caregiving and participation in online support groups in influencing the genetic testing decision, it is important to acknowledge certain limitations that may reduce the potential positive impact of these factors. For instance, barriers such as limited digital literacy, unequal access to the Internet, and socioeconomic disparities can hinder individuals from fully benefiting from online health resources and support networks. These barriers may be particularly pronounced in underserved or marginalized populations, where individuals may lack the necessary tools or knowledge to engage in digital health activities effectively. Future research could address these limitations by exploring how cultural contexts, including beliefs, norms, and values, shape the genetic testing decision. Examining how different communities perceive genetic testing and their access to digital resources can provide more nuanced insights into reducing disparities and improving health outcomes across diverse populations.

Conclusions

This study explores the factors influencing the decisions to undergo cancer genetic testing in the US population. Our findings emphasize the significant roles of perceived benefits, cues to action, caregiving for patients with cancer, and participation in social health groups in motivating genetic testing. Additionally, gender moderates the relationship between genetic testing and both perceived susceptibility and severity of cancer risk. Given the hereditary nature of cancer, increasing

awareness of genetic testing benefits is essential for promoting preventive health behaviors.

The study contributes to both theory and practice. Theoretically, it extends the HBM by incorporating cancer-specific constructs and highlighting the role of information systems in health decision-making. Practically, it offers actionable insights on how tailored education and social support can foster proactive health behaviors, particularly among caregivers and social health communities. Targeted campaigns, especially within online support groups or aimed at men, can further promote early detection. Future research can explore these strategies to increase the adoption of genetic testing for cancer.

Acknowledgments

The authors thank the reviewers and editors for their constructive feedback.

Generative AI tools like Grammarly/ChatGPT were used for editing and language enhancement in certain sections of the manuscript. The article remains a fully original work produced by the authors, and they take full responsibility for the content of the publication.

Data Availability

The dataset used for this study is publicly available on the National Cancer Institute Health Information National Trend Survey portal [100].

Authors' Contributions

LA contributed to conceptualization, methodology, formal analysis, writing –original draft preparation; ROD contributed to conceptualization, writing – original draft, writing – review and editing; PM contributed to the conceptualization, writing – original draft, and writing – review and editing; PC contributed to conceptualization, methodology, formal analysis, and writing – review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Operationalization of Constructs (source HINTS 5 cycle 4).
[DOCX File, 25 KB - cancer_v11i1e67650_app1.docx]

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA A Cancer J Clinicians 2024 Jan;74(1):12-49. [doi: 10.3322/caac.21820]
2. Crosby D, Bhatia S, Brindle KM, et al. Early detection of cancer. Science 2022 Mar 18;375(6586):eaay9040. [doi: 10.1126/science.aay9040] [Medline: 35298272]
3. Singh DN, Daripelli S, Elamin Bushara MO, Polevoy GG, Prasanna M. Genetic testing for successive cancer treatment. Cureus 2023 Dec;15(12):e49889. [doi: 10.7759/cureus.49889] [Medline: 38179395]
4. Evans JP, Skrzynia C, Burke W. The complexities of predictive genetic testing. BMJ 2001 Apr 28;322(7293):1052-1056. [doi: 10.1136/bmj.322.7293.1052] [Medline: 11325775]
5. Heshka JT, Palleschi C, Howley H, Wilson B, Wells PS. A systematic review of perceived risks, psychological and behavioral impacts of genetic testing. Genet Med 2008 Jan;10(1):19-32. [doi: 10.1097/GIM.0b013e31815f524f] [Medline: 18197053]
6. Lerman C, Croyle RT, Tercyak KP, Hamann H. Genetic testing: psychological aspects and implications. J Consult Clin Psychol 2002;70(3):784-797. [doi: 10.1037/0022-006X.70.3.784]
7. Strecher VJ, Rosenstock IM. The health belief model. In: Cambridge Handbook of Psychology, Health and Medicine 1997. [doi: 10.1017/CBO9780511543579.022]
8. Coe AB, Gatewood SBS, Moczygemba LR, Goode J, Beckner JO. The use of the health belief model to assess predictors of intent to receive the novel (2009) H1N1 influenza vaccine. Innov Pharm 2012;3(2):1-11. [doi: 10.24926/iip.v3i2.257] [Medline: 22844651]
9. Wong LP, Alias H, Wong PF, Lee HY, AbuBakar S. The use of the health belief model to assess predictors of intent to receive the COVID-19 vaccine and willingness to pay. Hum Vaccin Immunother 2020 Sep 1;16(9):2204-2214. [doi: 10.1080/21645515.2020.1790279] [Medline: 32730103]
10. Jiang L, Liu S, Li H, Xie L, Jiang Y. The role of health beliefs in affecting patients' chronic diabetic complication screening: a path analysis based on the health belief model. J Clin Nurs 2021 Oct;30(19-20):2948-2959. [doi: 10.1111/jocn.15802] [Medline: 33951248]

11. Al-Ani A, Hammouri M, Sultan H, Al-Huneidy L, Mansour A, Al-Hussaini M. Factors affecting cervical screening using the health belief model during the last decade: a systematic review and meta-analysis. Psychooncology 2024 Jan;33(1):e6275. [doi: 10.1002/pon.6275] [Medline: 38282232]
12. Codori AM, Petersen GM, Mighoretti DL, Boyd P. Health beliefs and endoscopic screening for colorectal cancer: potential for cancer prevention. Prev Med 2001 Aug;33(2):128-136. [doi: 10.1016/S0091-7435(01)80009-8]
13. Macrae FA, Hill DJ, St. John DJB, Ambikapathy A, Garner JF. Predicting colon cancer screening behavior from health beliefs. Prev Med 1984 Jan;13(1):115-126. [doi: 10.1016/0091-7435(84)90044-6]
14. Sung JJY, Choi SYP, Chan FKL, Ching JYL, Lau JTF, Griffiths S. Obstacles to colorectal cancer screening in Chinese: a study based on the health belief model. Am J Gastroenterology 2008 Apr;103(4):974-981. [doi: 10.1111/j.1572-0241.2007.01649.x]
15. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. Int J Cancer 2002 May 10;99(2):260-266. [doi: 10.1002/ijc.10332] [Medline: 11979442]
16. Ponder BAJ. Setting up and running a familial cancer clinic. Br Med Bull 1994;50(3):732-745. [doi: 10.1093/oxfordjournals.bmb.a072921]
17. Hwang KO, Ottenbacher AJ, Graham AL, Thomas EJ, Street RL Jr, Vernon SW. Online narratives and peer support for colorectal cancer screening: a pilot randomized trial. Am J Prev Med 2013 Jul;45(1):98-107. [doi: 10.1016/j.amepre.2013.02.024] [Medline: 23790994]
18. Deeks A, Lombard C, Michelmore J, Teede H. The effects of gender and age on health related behaviors. BMC Public Health 2009 Jun 30;9:1-8. [doi: 10.1186/1471-2458-9-213] [Medline: 19563685]
19. Langlie JK. Social networks, health beliefs, and preventive health behavior. J Health Soc Behav 1977 Sep;18(3):244-260. [doi: 10.2307/2136352] [Medline: 903596]
20. Wang T, Wang H, Zeng Y, Cai X, Xie L. Health beliefs associated with preventive behaviors against noncommunicable diseases. Patient Educ Couns 2022 Jan;105(1):173-181. [doi: 10.1016/j.pec.2021.05.024] [Medline: 34030926]
21. Jensen LF, Pedersen AF, Andersen B, Vedsted P. Social support and non-participation in breast cancer screening: a Danish cohort study. J Public Health 2016 Jun;38(2):335-342. [doi: 10.1093/pubmed/fdv051]
22. Cancer.gov. The Genetics of Cancer. 2021. URL: http://www.accessexcellence.org/AE/AEPC/NIH/ [accessed 2024-09-02]
23. Eccles DM, Mitchell G, Monteiro ANA, et al. BRCA1 and BRCA2 genetic testing-pitfalls and recommendations for managing variants of uncertain clinical significance. Ann Oncol 2015 Oct;26(10):2057-2065. [doi: 10.1093/annonc/mdv278] [Medline: 26153499]
24. Nelson HD, Pappas M, Zakher B, Mitchell JP, Okinaka-Hu L, Fu R. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: a systematic review to update the U.S. Preventive Services Task Force recommendation. Ann Intern Med 2014 Feb 18;160(4):255-266. [doi: 10.7326/M13-1684] [Medline: 24366442]
25. Croyle RT, Lerman C. Risk communication in genetic testing for cancer susceptibility. JNCI Monographs 1999 Jan 1;1999(25):59-66. [doi: 10.1093/oxfordjournals.jncimonographs.a024210]
26. Cyr A, Dunnagan TA, Haynes G. Efficacy of the health belief model for predicting intention to pursue genetic testing for colorectal cancer. J Genet Couns 2010 Apr;19(2):174-186. [doi: 10.1007/s10897-009-9271-7] [Medline: 19902341]
27. Hayeems RZ, Luca S, Assamad D, Bhatt A, Ungar WJ. Utility of genetic testing from the perspective of parents/caregivers: a scoping review. Children (Basel) 2021 Mar 27;8(4):259. [doi: 10.3390/children8040259] [Medline: 33801725]
28. Pasek M, Goździalska A, Jochymek M, Caruso R. Social support in a cancer patient-informal caregiver dyad: a scoping review. Cancers (Basel) 2023 Mar 14;15(6):1754. [doi: 10.3390/cancers15061754] [Medline: 36980639]
29. Sillence E. Giving and receiving peer advice in an online breast cancer support group. Cyberpsychol Behav Soc Netw 2013 Jun;16(6):480-485. [doi: 10.1089/cyber.2013.1512] [Medline: 23659724]
30. Ruco A, Dossa F, Tinmouth J, Llovet D, Kishibe T, Baxter NN. Social media and mobile health technology for cancer screening: a systematic review and meta-analysis protocol. BMJ Open 2020 Feb 5;10(2):e035411. [doi: 10.1136/bmjopen-2019-035411] [Medline: 32029500]
31. Bunn JY, Bosompra K, Ashikaga T, Flynn BS, Worden JK. Factors influencing intention to obtain a genetic test for colon cancer risk: a population-based study. Prev Med 2002 Jun;34(6):567-577. [doi: 10.1006/pmed.2002.1031] [Medline: 12052015]
32. Hartman KA. Health belief model and its application to mammography screening in a K-12 school district employee wellness program [Dissertation]. : Montana State University-Bozeman, College of Education, Health & Human Development; 2002.
33. Rosenstock IM. Historical origins of the health belief model. Health Educ Monogr 1974 Dec;2(4):328-335. [doi: 10.1177/109019817400200403]
34. Skinner CS, Tiro J, Champion VL. Background on the health belief model. In: Health Behavior: Theory, Research, and Practice, 5th edition:75-94.
35. Champion VL, Skinner CS, Menon U. Health belief model. In: In Health Behavior and Health Education: Theory, Research, and Practice: Jossey-Bass; 2008:45-65.
36. Irigoyen-Camacho ME, Velazquez-Alva MC, Zepeda-Zepeda MA, Cabrer-Rosales MF, Lazarevich I, Castaño-Seiquer A. Effect of income level and perception of susceptibility and severity of COVID-19 on stay-at-home preventive behavior in

a group of older adults in Mexico city. Int J Environ Res Public Health 2020 Oct 12;17(20):7418. [doi: 10.3390/ijerph17207418] [Medline: 33053788]
37. Brewer NT, Chapman GB, Gibbons FX, Gerrard M, McCaul KD, Weinstein ND. Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. Health Psychol 2007 Mar;26(2):136-145. [doi: 10.1037/0278-6133.26.2.136] [Medline: 17385964]
38. Ferrer RA, Klein WM. Risk perceptions and health behavior. Curr Opin Psychol 2015 Oct 1;5:85-89. [doi: 10.1016/j.copsyc.2015.03.012] [Medline: 26258160]
39. Courtenay WH. Constructions of masculinity and their influence on men's well-being: a theory of gender and health. Soc Sci Med 2000 May;50(10):1385-1401. [doi: 10.1016/S0277-9536(99)00390-1]
40. Lisha NE, Martens M, Leventhal AM. Age and gender as moderators of the relationship between physical activity and alcohol use. Addict Behav 2011 Sep;36(9):933-936. [doi: 10.1016/j.addbeh.2011.04.003] [Medline: 21561719]
41. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 2000 Oct;27(5):591-615. [doi: 10.1177/109019810002700506] [Medline: 11009129]
42. He L, Gao S, Tao S, et al. Factors associated with colonoscopy compliance based on health belief model in a community-based colorectal cancer screening program, Shanghai, China. Int Q Community Health Educ 2020 Oct;41(1):25-33. [doi: 10.1177/0272684X19897356] [Medline: 31876256]
43. Rana IA, Bhatti SS, Aslam AB, Jamshed A, Ahmad J, Shah AA. COVID-19 risk perception and coping mechanisms: does gender make a difference? Int J Disaster Risk Reduct 2021 Mar;55:102096. [doi: 10.1016/j.ijdrr.2021.102096] [Medline: 33777688]
44. Sattler S, Taflinger S, Ernst A, Hasselhorn F. A moderated mediation model explaining the relationship between risk-group membership, threat perception, knowledge, and adherence to COVID-19 behavioral measures. Front Public Health 2022;10:842368. [doi: 10.3389/fpubh.2022.842368] [Medline: 35664099]
45. Orji R, Vassileva J, Mandryk RL. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. User Model User-Adap Inter 2014 Dec;24(5):453-498. [doi: 10.1007/s11257-014-9149-8]
46. Chen MF, Wang RH, Schneider JK, et al. Using the health belief model to understand caregiver factors influencing childhood influenza vaccinations. J Community Health Nurs 2011 Jan;28(1):29-40. [doi: 10.1080/07370016.2011.539087] [Medline: 21279888]
47. Chen Y, Zhou R, Chen B, et al. Knowledge, perceived beliefs, and preventive behaviors related to COVID-19 among Chinese older adults: cross-sectional web-based survey. J Med Internet Res 2020 Dec 31;22(12):e23729. [doi: 10.2196/23729] [Medline: 33293262]
48. Bosompra K, Flynn BS, Ashikaga T, Rairikar CJ, Worden JK, Solomon LJ. Likelihood of undergoing genetic testing for cancer risk: a population-based study. Prev Med 2000 Feb;30(2):155-166. [doi: 10.1006/pmed.1999.0610] [Medline: 10656843]
49. Al-Noumani H, Wu JR, Barksdale D, Sherwood G, AlKhasawneh E, Knafl G. Health beliefs and medication adherence in patients with hypertension: a systematic review of quantitative studies. Patient Educ Couns 2019 Jun;102(6):1045-1056. [doi: 10.1016/j.pec.2019.02.022] [Medline: 30846205]
50. Shahnazi H, Ahmadi-Livani M, Pahlavanzadeh B, Rajabi A, Hamrah MS, Charkazi A. Assessing preventive health behaviors from COVID-19 based on the health belief model (HBM) among people in golestan province: a cross-sectional study in northern iran. In Review. Preprint posted online on 2020. [doi: 10.21203/rs.3.rs-24871/v1]
51. Wright L, Fancourt D. Do predictors of adherence to pandemic guidelines change over time? A panel study of 22,000 UK adults during the COVID-19 pandemic. Prev Med 2021 Dec;153:106713. [doi: 10.1016/j.ypmed.2021.106713] [Medline: 34242662]
52. Carpenter CJ. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. Health Commun 2010 Dec;25(8):661-669. [doi: 10.1080/10410236.2010.521906] [Medline: 21153982]
53. Glanz K, Rimer BK, Viswanath K. Health Behavior: Theory, Research, and Practice: Jossey-Bass; 2015.
54. Eden J, Schulz R, editors. Families Caring for an Aging America: National Academies Press; 2016. [doi: 10.17226/23606]
55. Given BA, Given CW, Sherwood PR. Family and caregiver needs over the course of the cancer trajectory. J Support Oncol 2012;10(2):57-64. [doi: 10.1016/j.suponc.2011.10.003] [Medline: 22222251]
56. Cdc.gov. Caregiving for Family and Friends — A Public Health Issue. 2021. URL: https://www.cdc.gov/aging/caregiving/caregiver-brief.html [accessed 2021-02-25]
57. Son KY, Park SM, Lee CH, et al. Behavioral risk factors and use of preventive screening services among spousal caregivers of cancer patients. Support Care Cancer 2011 Jul;19(7):919-927. [doi: 10.1007/s00520-010-0889-7] [Medline: 20445996]
58. Strecher VJ, Rosenstock IM. The health belief model. In: Cambridge Handbook of Psychology, Health and Medicine 1997, Vol. 113:117.
59. Rha SY, Park Y, Song SK, Lee CE, Lee J. Caregiving burden and health-promoting behaviors among the family caregivers of cancer patients. Eur J Oncol Nurs 2015 Apr;19(2):174-181. [doi: 10.1016/j.ejon.2014.09.003] [Medline: 25456972]
60. Agrawal L, Ndabu T, Mulgund P, Sharman R. Factors affecting the extent of patients' electronic medical record use: an empirical study focusing on system and patient characteristics. J Med Internet Res 2021 Oct 28;23(10):e30637. [doi: 10.2196/30637] [Medline: 34709181]

61. Ruland CM, Andersen T, Jeneson A, et al. Effects of an internet support system to assist cancer patients in reducing symptom distress: a randomized controlled trial. Cancer Nurs 2013;36(1):6-17. [doi: 10.1097/NCC.0b013e31824d90d4] [Medline: 22495503]
62. Domínguez M, Sapiña L. "Others Like Me". An approach to the use of the internet and social networks in adolescents and young adults diagnosed with cancer. J Cancer Educ 2017 Dec;32(4):885-891. [doi: 10.1007/s13187-016-1055-9] [Medline: 27251633]
63. Unger JM, Cook E, Tai E, Bleyer A. The role of clinical trial participation in cancer research: barriers, evidence, and strategies. Am Soc Clin Oncol Educ Book 2016;36:185-198. [doi: 10.1200/EDBK_156686] [Medline: 27249699]
64. Han CJ, Lee YJ, Demiris G. Interventions using social media for cancer prevention and management: a systematic review. Cancer Nurs 2018;41(6):E19-E31. [doi: 10.1097/NCC.0000000000000534] [Medline: 28753192]
65. Plackett R, Kaushal A, Kassianos AP, et al. Use of social media to promote cancer screening and early diagnosis: scoping review. J Med Internet Res 2020 Nov 9;22(11):e21582. [doi: 10.2196/21582] [Medline: 33164907]
66. National cancer institute. Health Information National Trends Survey (HINTS). URL: https://hints.cancer.gov/ [accessed 2025-05-30]
67. Cohen J, Cohen P, West SG, Aiken LS. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences: Routledge; 2013. [doi: 10.4324/9780203774441]
68. Wanous JP, Reichers AE, Hudy MJ. Overall job satisfaction: how good are single-item measures? J Appl Psychol 1997 Apr;82(2):247-252. [doi: 10.1037/0021-9010.82.2.247] [Medline: 9109282]
69. Abdelhamid M, Gaia J, Sanders GL. Putting the focus back on the patient: how privacy concerns affect personal health information sharing intentions. J Med Internet Res 2017 Sep 13;19(9):e169. [doi: 10.2196/jmir.6877] [Medline: 28903895]
70. Angst C, Agarwal R. Adoption of electronic health records in the presence of privacy concerns: the elaboration likelihood model and individual persuasion. MIS Q 2009;33(2):339. [doi: 10.2307/20650295]
71. Heeringa SG, West BT, Berglund PA. Applied Survey Data Analysis: Chapman and Hall/CRC; 2017. [doi: 10.1201/9781315153278]
72. Lazar M, Davenport L. Barriers to health care access for low income families: a review of literature. J Community Health Nurs 2018;35(1):28-37. [doi: 10.1080/07370016.2018.1404832] [Medline: 29323941]
73. Loignon C, Hudon C, Goulet É, et al. Perceived barriers to healthcare for persons living in poverty in Quebec, Canada: the EQUIhealThY project. Int J Equity Health 2015 Jan 17;14:4. [doi: 10.1186/s12939-015-0135-5] [Medline: 25596816]
74. Moser RP, Naveed S, Cantor D, et al. Integrative Analytic Methods Using Population-Level Cross-Sectional Data: Hints. cancer; 2013. URL: https://hints.cancer.gov/docs/hints_ida_report.pdf [accessed 2025-06-02]
75. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
76. Lindell MK, Whitney DJ. Accounting for common method variance in cross-sectional research designs. J Appl Psychol 2001 Feb;86(1):114-121. [doi: 10.1037/0021-9010.86.1.114] [Medline: 11302223]
77. HINTS Institutional Review Board (IRB). National Cancer Institute. 2024. URL: https://hints.cancer.gov/about-hints/institutional-review-board.aspx [accessed 2025-02-25]
78. Karl JA, Fischer R, Druică E, Musso F, Stan A. Testing the effectiveness of the health belief model in predicting preventive behavior during the COVID-19 pandemic: the case of Romania and Italy. Front Psychol 2021;12:627575. [doi: 10.3389/fpsyg.2021.627575] [Medline: 35095628]
79. Tomczyk S, Barth S, Schmidt S, Muehlan H. Utilizing health behavior change and technology acceptance models to predict the adoption of COVID-19 contact tracing apps: cross-sectional survey study. J Med Internet Res 2021 May 19;23(5):e25447. [doi: 10.2196/25447] [Medline: 33882016]
80. Subedi S, Leal Filho W, Adedeji A. An assessment of the health belief model (HBM) properties as predictors of COVID-19 preventive behaviour. J Public Health (Berl) 2023:1-11. [doi: 10.1007/s10389-023-02109-7]
81. Zewdie A, Nigusie A, Wolde M. Structural equation modeling analysis of health belief model-based determinants of COVID-19 preventive behavior of academic staff: a cross-sectional study. BMC Infect Dis 2024 Aug 6;24(1):788. [doi: 10.1186/s12879-024-09697-z] [Medline: 39107716]
82. Makhnoon S, Lee M, Prasad T, et al. Awareness, use, motivations and methods of accessing genetic testing in 2022 in the United States. Front Genet 2024;15:1462831. [doi: 10.3389/fgene.2024.1462831] [Medline: 39582788]
83. Xi Q, Jin S, Morris S. Economic evaluations of predictive genetic testing: a scoping review. PLoS ONE 2023;18(8):e0276572. [doi: 10.1371/journal.pone.0276572] [Medline: 37531363]
84. Smith HS, Lakoma M, Hickingbotham MR, et al. Genetic test utilization and cost among families of children evaluated for genetic conditions: an analysis of USA commercial claims data. Appl Health Econ Health Policy 2025 May;23(3):519-530. [doi: 10.1007/s40258-024-00942-9] [Medline: 39777698]
85. Dusic EJ, Theoryn T, Wang C, Swisher EM, Bowen DJ, Team ES. Barriers, interventions, and recommendations: Improving the genetic testing landscape. Front Digit Health 2022;4:961128. [doi: 10.3389/fdgth.2022.961128] [Medline: 36386046]

86. Bedrick BS, Tipping AD, Nickel KB, Riley JK, Jain T, Jungheim ES. State-mandated insurance coverage and preimplantation genetic testing in the United States. Obstet Gynecol 2022 Apr 1;139(4):500-508. [doi: 10.1097/AOG.0000000000004712] [Medline: 35271533]
87. Smith-Uffen M, Bartley N, Davies G, Best M. Motivations and barriers to pursue cancer genomic testing: a systematic review. Patient Educ Couns 2021 Jun;104(6):1325-1334. [doi: 10.1016/j.pec.2020.12.024] [Medline: 33390305]
88. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
89. Cincidda C, Oliveri S, Sanchini V, Pravettoni G. The role of caregivers in the clinical pathway of patients newly diagnosed with breast and prostate cancer: a study protocol. Front Psychol 2022;13:962634. [doi: 10.3389/fpsyg.2022.962634] [Medline: 36405193]
90. Centola D. Social media and the science of health behavior. Circulation 2013 May 28;127(21):2135-2144. [doi: 10.1161/CIRCULATIONAHA.112.101816] [Medline: 23716382]
91. Valente TW. Social Networks and Health: Models, Methods, and Applications: Oxford University Press; 2010. [doi: 10.1093/acprof:oso/9780195301014.001.0001]
92. Smith RA, Cokkinides V, Brawley OW. Cancer screening in the United States, 2009: a review of current American Cancer Society guidelines and issues in cancer screening. CA Cancer J Clin 2009;59(1):27-41. [doi: 10.3322/caac.20008] [Medline: 19147867]
93. Khoury MJ, McBride CM, Schully SD, et al. The scientific foundation for personal genomics: recommendations from a National Institutes of Health-Centers for Disease Control and Prevention multidisciplinary workshop. Genet Med 2009 Aug;11(8):559-567. [doi: 10.1097/GIM.0b013e3181b13a6c] [Medline: 19617843]
94. Mantzari E, Vogt F, Shemilt I, Wei Y, Higgins JPT, Marteau TM. Personal financial incentives for changing habitual health-related behaviors: a systematic review and meta-analysis. Prev Med 2015 Jun;75:75-85. [doi: 10.1016/j.ypmed.2015.03.001] [Medline: 25843244]
95. Vlaev I, King D, Darzi A, Dolan P. Changing health behaviors using financial incentives: a review from behavioral economics. BMC Public Health 2019 Aug 7;19(1):1059. [doi: 10.1186/s12889-019-7407-8] [Medline: 31391010]
96. Centers for disease control and prevention. Cancer and men. URL: https://www.cdc.gov/cancer/features/cancer-and-men.html [accessed 2025-05-30]
97. Loeb S, Massey P, Leader AE, et al. Gaps in public awareness about BRCA and genetic testing in prostate cancer: social media landscape analysis. JMIR Cancer 2021 Sep 20;7(3):e27063. [doi: 10.2196/27063] [Medline: 34542414]
98. Thakker S, Loeb S, Giri VN, Bjurlin MA, Matulewicz RS. Attitudes, perceptions, and use of cancer-based genetic testing among healthy U.S. adults and those with prostate or breast/ovarian cancer. Urol Pract 2023 Jan;10(1):26-32. [doi: 10.1097/UPJ.0000000000000352] [Medline: 37103438]
99. Chou WYS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48. [doi: 10.2196/jmir.1249] [Medline: 19945947]
100. National cancer institute. Health Information National Trends Survey. URL: https://hints.cancer.gov/data/download-data.aspx#H5C4 [accessed 2025-06-04]

Abbreviations

CMV: common method variance
HBM: health belief model
HINTS: Health Information National Trends Survey
UOM: UnderstandOnlineMedRec

Edited by N Cahill; submitted 17.10.24; peer-reviewed by AS Kushwah, L Godino; revised version received 26.02.25; accepted 26.02.25; published 04.06.25.

Please cite as:
Agrawal L, DaSouza RO, Mulgund P, Chaudhary P
The Role of Online Support, Caregiving, and Gender in Preventative Cancer Genetic Testing Participation: Cross-Sectional Study From a National Study
JMIR Cancer 2025;11:e67650
URL: https://cancer.jmir.org/2025/1/e67650
doi:10.2196/67650

Original Paper

User and Developer Views on Using AI Technologies to Facilitate the Early Detection of Skin Cancers in Primary Care Settings: Qualitative Semistructured Interview Study

Owain Tudor Jones[1], BSc, BMBS, MPhil; Natalia Calanzani[2], BSc, MSc, PhD; Suzanne E Scott[3], BSc, MSc, PhD; Rubeta N Matin[4], PhD; Jon Emery[3,5], DPhil; Fiona M Walter[3,5], MA, MD

[1]Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

[2]Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom

[3]Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom

[4]Department of Dermatology, Churchill Hospital, Oxford, United Kingdom

[5]Centre for Cancer Research, and Department of General Practice and Primary Care, University of Melbourne, Melbourne, Australia

Corresponding Author:
Owain Tudor Jones, BSc, BMBS, MPhil
Department of Public Health and Primary Care
University of Cambridge
East Forvie Building
Robinson Way
Cambridge, CB2 0SZ
United Kingdom
Phone: 44 7737204055
Email: otj24@medschl.cam.ac.uk

Abstract

Background: Skin cancers, including melanoma and keratinocyte cancers, are among the most common cancers worldwide, and their incidence is rising in most populations. Earlier detection of skin cancer leads to better outcomes for patients. Artificial intelligence (AI) technologies have been applied to skin cancer diagnosis, but many technologies lack clinical evidence and/or the appropriate regulatory approvals. There are few qualitative studies examining the views of relevant stakeholders or evidence about the implementation and positioning of AI technologies in the skin cancer diagnostic pathway.

Objective: This study aimed to understand the views of several stakeholder groups on the use of AI technologies to facilitate the early diagnosis of skin cancer, including patients, members of the public, general practitioners, primary care nurse practitioners, dermatologists, and AI researchers.

Methods: This was a qualitative, semistructured interview study with 29 stakeholders. Participants were purposively sampled based on age, sex, and geographical location. We conducted the interviews via Zoom between September 2022 and May 2023. Transcribed recordings were analyzed using thematic framework analysis. The framework for the Nonadoption, Abandonment, and Challenges to Scale-Up, Spread, and Sustainability was used to guide the analysis to help understand the complexity of implementing diagnostic technologies in clinical settings.

Results: Major themes were “the position of AI in the skin cancer diagnostic pathway” and “the aim of the AI technology”; cross-cutting themes included trust, usability and acceptability, generalizability, evaluation and regulation, implementation, and long-term use. There was no clear consensus on where AI should be placed along the skin cancer diagnostic pathway, but most participants saw the technology in the hands of either patients or primary care practitioners. Participants were concerned about the quality of the data used to develop and test AI technologies and the impact this could have on their accuracy in clinical use with patients from a range of demographics and the risk of missing skin cancers. Ease of use and not increasing the workload of already strained health care services were important considerations for participants. Health care professionals and AI researchers reported a lack of established methods of evaluating and regulating AI technologies.

Conclusions: This study is one of the first to examine the views of a wide range of stakeholders on the use of AI technologies to facilitate early diagnosis of skin cancer. The optimal approach and position in the diagnostic pathway for these technologies

have not yet been determined. AI technologies need to be developed and implemented carefully and thoughtfully, with attention paid to the quality and representativeness of the data used for development, to achieve their potential.

(JMIR Cancer 2025;11:e60653) doi:10.2196/60653

KEYWORDS

artificial intelligence; AI; machine learning; ML; primary care; skin cancer; melanoma; qualitative research; mobile phone

Introduction

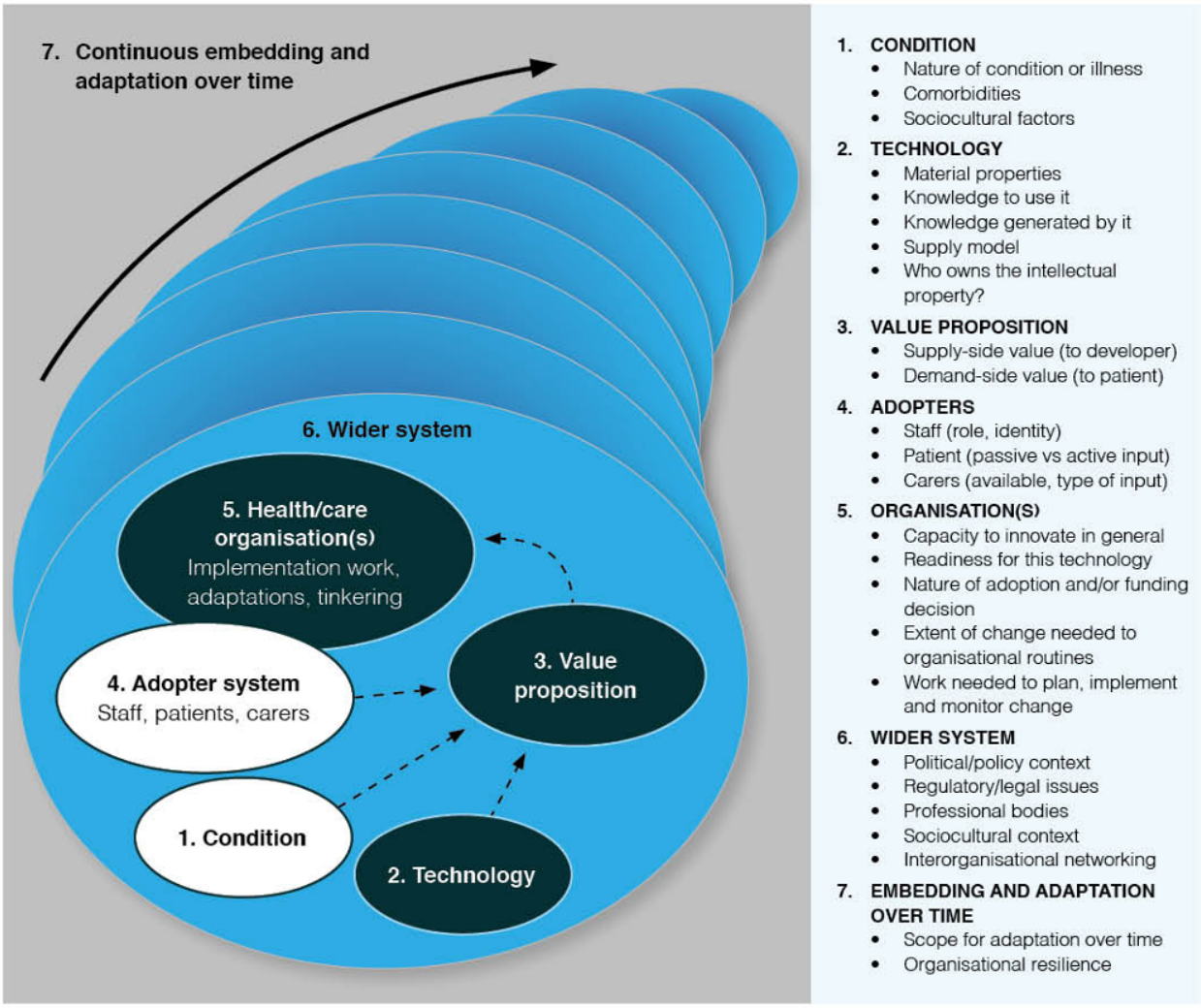
Background

Skin cancers are among the most common cancers worldwide, with increasing incidence in most populations [1,2]. Melanoma is the most lethal skin cancer, but keratinocyte cancers (KCs), which include squamous cell carcinomas and basal cell carcinomas, comprise most skin cancers [1-3]. There were >200,000 skin cancer diagnoses in England in 2021, comprising 193,000 nonmelanoma skin cancers (which include KCs) and nearly 16,000 melanomas [4]. The World Health Organization estimates that 2 to 3 million nonmelanoma skin cancers and 132,000 melanomas occur globally each year [5]. Earlier diagnosis of skin cancers is associated with statistically significantly better outcomes [3]. In the United States, early detection of melanoma is associated with >99% five-year survival but falls to 74% when it has spread to lymph nodes and 35% when spread to distant organs [6].

There has been a substantial interest in applying artificial intelligence (AI) to the diagnosis of skin cancer through visual analysis of skin lesions, either through a smartphone app or uploading images of skin lesions. Many AI technologies have been designed for use by patients and a handful for use by clinicians as clinical decision aids [7,8]. Most of these technologies do not have the appropriate regulatory approvals in place to support their safety and efficacy when used in these settings [9], with limited evidence on their efficacy and accuracy from clinical trials or real-life clinical settings [7,8], although some evidence is emerging [10,11]. There is limited evidence on how users in clinical settings interact with AI technologies and how this might affect patient safety [12], and a lack of qualitative studies reporting public perspectives [13].

Implementation of a diagnostic technology in clinical settings is a complex process and prone to failure—a technology must pass through several stages of development before implementation is likely to be successful [14]. Several frameworks analyze factors around the implementation of new technologies. In this study we chose to use the Nonadoption, Abandonment, and Challenges to Scale-Up, Spread, and Sustainability (NASSS) framework [15] to help understand and interpret the data. This framework incorporates complexity principles and allows researchers to identify and explain the manifestations of complexity in technology-supported change projects (Figure 1) [15]. We believe that these attributes made it best suited to this study compared with other conceptual frameworks for implementation research.

Figure 1. The Nonadoption, Abandonment, and Challenges to Scale-Up, Spread, and Sustainability (NASSS) framework [15].



Objectives

The aim of this study was to consult several stakeholder groups (ie, groups that may have views or concerns about the use of AI to help diagnose skin cancer in primary care) to provide in-depth understanding of barriers and facilitators to the implementation of AI technologies for the early diagnosis of skin cancers in primary care settings. Depending on how AI is implemented in the skin cancer diagnostic pathway, users of the technology could include members of the public, patients, general practitioners (GPs), primary care nurse practitioners (NPs), and dermatologists. Dermatologists also receive most referrals from primary care for suspected skin cancers. AI researchers working in both academic and commercial settings are the primary developers of AI technologies. The views of all these groups were important to understand.

Methods

Design

This was a qualitative study that was performed using semistructured interviews of stakeholders.

Recruitment and Sampling

Four groups were selected for the study: (1) members of the public, (2) patients previously diagnosed with skin cancer, (3) health care professionals (HCPs; including GPs, primary care NPs, and dermatologists), and (4) AI researchers from academic and commercial settings. Members of the public were approached via the Cambridge Biomedical Research Centre Patient and Public Involvement (PPI) group and “snowballing” invitations to their colleagues. Members of the public with a history of skin cancer were included in the patient group, with additional patients approached via Melanoma Focus (a UK melanoma charity that provides information, guidance, and support for patients, carers, and HCPs) [16]. GPs and primary care NPs were identified through the Primary Care Dermatology Society [17], with additional GPs approached via Sermo (medical market-research organization) [18]. Dermatologists were identified from the British Association of Dermatologists (BAD) [19] and through snowballing. AI researchers from academic settings were identified through contacts within academic institutions, and AI researchers from commercial settings were recruited via email from companies identified in 2 reviews [7,8]. In this paper, we use primary care practitioner (PCP) to denote any medical practitioner that works in a primary

care setting and might consult with a patient about a suspicious skin lesion, including GPs, family doctors, NPs, physician assistants, and paramedic practitioners. HCP is used when we refer to the views of wider HCPs, including secondary care HCPs. Staff who work in primary and community care and do not have clinical training and experience in the diagnosis of skin cancer but could potentially use an AI technology with patients with suspicious skin lesions are referred to as allied HCPs; a wide variety of professions could be included in this group, but it certainly includes practice nurses, health care assistants, clinical navigators, pharmacists, and podiatrists.

Participants were sampled to achieve a spread of age, sex, and geographical location within each participant group. Patients and the public were recruited to achieve a spread of ages >60 years and <60 years, reflecting the average age of skin cancer diagnosis. HCPs and AI researchers were recruited to achieve a spread of ages >45 years and <45 years, reflecting the midcareer point of these professions. Patients were sampled to include a range of prior history of skin cancer types. GPs were sampled to include a spread of roles (GP partner, salaried GP, locum GP, GP with extended role in dermatology). Participants were asked how supportive they were of using AI technologies to help diagnose skin cancer in primary care (using a Likert scale of strongly disapprove, disapprove, neutral, approve, and strongly approve). We aimed to recruit at least 1 participant in each group who disapproved and at least 1 who approved of the use of AI in this setting to obtain a range of perspectives.

Ethical Considerations

Ethics approval for this study was granted by the Cambridge Psychology Research Ethics Committee (PRE.2021.098). Participants gave informed written and verbal consent to take part in the interviews. Patient and public participants were invited to bring a friend or family member with them to take part in the study if they wished. All interviews were digitally recorded and transcribed verbatim by a professional transcription company. Transcripts were checked and anonymized before analysis. To facilitate recruitment, a £20 (US $25.24) Apple iTunes or Google Play voucher was offered to all participants.

Data Collection

An interview topic guide was developed with input from our PPI group to explore views on facilitators and barriers to the use of AI technologies to help diagnose skin cancer in primary care. All interviews were conducted by OTJ, who has a clinical background, with guidance from NC, a health services researcher with expertise in qualitative research. Interviews took place on the web using Zoom (Zoom Video Communications) at a time of the participants choosing between September 9, 2022, and May 25, 2023. Interview schedules for patients and members of the public and HCPs and AI researchers are available in Multimedia Appendix 2. Details on the interviews are reported in the COREQ (Consolidated Criteria for Reporting Qualitative Research; Multimedia Appendix 3) checklist. Interviews continued until we had a rich, multifaceted dataset. A reflexivity journal with field notes was kept and was discussed at regular meetings during the study.

Analysis

Interviews were analyzed inductively and deductively using thematic framework analysis [20]. Two researchers (OTJ and NC) repeatedly read the first 5 transcripts to become familiar with the data and generate initial codes. These initial codes were compared with the NASSS framework (Figure 1) [15] to generate a comprehensive list of codes. The remaining transcripts were then read and indexed using NVivo (version 14; Lumivero) [21], with codes modified or created as required where the data did not fit comfortably into the NASSS framework. Coding was completed by OTJ with a sample of transcripts (7/29, 24%) checked by NC. Codes were defined and discussed regularly in team meetings (OTJ, NC, and FMW), and coding files were saved throughout the process to maintain an audit trail of changes to the code tree. Data were charted into Microsoft Excel [22], and the characteristics, similarities, and differences between data were identified. Relationships and connections between categories were then mapped to generate the themes presented in the results. Themes were further refined with guidance from senior team members (FMW and SS [health psychologist]). Together this broad authorship group of clinical academics and behavioral scientists added rigor to data analysis and interpretation.

Results

Participants

A total of 29 interviews were conducted with members of the public (n=6, 21%), patients (n=5, 17%), HCPs (n=13, 45%), and AI researchers (n=5, 17%; Table 1). Participants were recruited via mailing lists and social media; therefore, the denominator is unknown.

Table 1. Participant demographics.

	Overall (n=29), n (%)	Public[a] (n=6), n (%)	Patients[b] (n=5), n (%)	HCPs[c] (n=13), n (%)	AI[d] researchers (n=5), n (%)
Sex					
Male	10 (34)	1 (17)	1 (20)	4 (31)	4 (80)
Female	19 (66)	5 (83)	4 (80)	9 (69)	1 (20)
Age group[e] (y)					
>60	N/A[f]	2 (33)	2 (40)	N/A	N/A
<60	N/A	4 (67)	3 (60)	N/A	N/A
>45	N/A	N/A	N/A	7 (54)	1 (20)
<45	N/A	N/A	N/A	6 (46)	4 (80)
Location					
England	23 (79)	5 (83)	5 (100)	11 (85)	2 (40)
Wales	2 (7)	1 (17)	0	1 (8)	0
Scotland	1 (3)	0	0	1 (8)	0
Outside United Kingdom[g]	3 (10)	0	0	0	3 (60)
Support for using AI to facilitate the early diagnosis of skin cancers in primary care					
Strongly approve	10 (34)	2 (33)	1 (20)	5 (38)	2 (40)
Approve	15 (52)	3 (50)	3 (60)	6 (46)	3 (60)
Neutral	2 (7)	0	1 (20)	1 (8)	0
Disapprove	1 (3)	1 (17)	0	0	0
Missing data	1 (3)	0	0	1 (8)	0
Highest educational qualification[h]					
Foundation or intermediate qualifications	0 (0)	0	0	0	0
Advanced qualifications	1 (3)	0	1 (20)	0	0
Higher qualifications	28 (97)	6 (100)	4 (80)	13 (100)	5 (100)
Ethnicity					
Black African	1 (3)	1 (17)	0	0	0
Middle Eastern	1 (3)	0	0	0	1 (20)
Missing data	3 (10)	0	1 (20)	1 (7)	1 (20)
White British	18 (62)	5 (83)	3 (60)	10 (77)	0
White European	6 (21)	0	1 (20)	2 (15)	3 (60)
History of skin cancer					
Yes	N/A	0	5[i] (100)	N/A	N/A
No	N/A	6 (100)	0	N/A	N/A
Skin cancer in a family member or close friend	N/A	2 (33)	5 (100)	N/A	N/A

[a]Occupations: project manager, lawyer, pharmaceutical industry, music industry, biomedical scientist, or missing data (n=1 for each).

[b]Occupations: fraud services, social researcher, teacher, accountant, and research consultant (n=1 for each).

[c]HCP: Health care professional. HCPs included general practitioners (GPs, n=6), primary care nurse practitioners (n=3), and dermatologists (n=4).

[d]AI: artificial intelligence.

[e]Patients and the public were recruited to achieve a spread of participants aged >60 and <60 years. HCPs and AI researchers were recruited to achieve a spread of ages >45 and <45 years.

[f]N/A: not applicable.

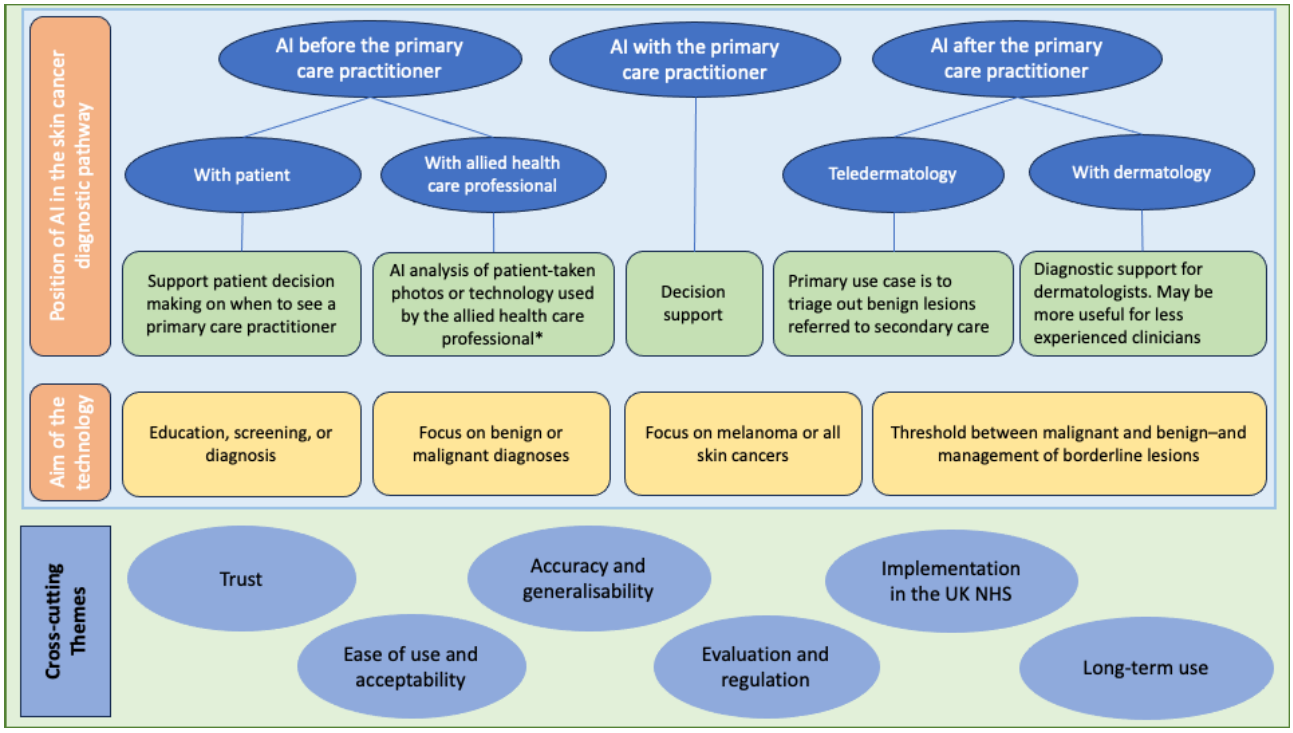
[g]Including the Netherlands and North America.

[h]Educational qualification categories were taken from the UK National Census 2021. Definitions are available on the internet [23].

[i]Basal cell carcinoma (n=3), melanoma (n=1), and Merkel cell carcinoma (n=1).

The *Results* section is structured around the major themes and subthemes generated from the data. Major themes included the "position of AI in the skin cancer diagnostic pathway" and the "aim of the AI technology." There were several cross-cutting themes, including trust, acceptability, generalizability, evaluation, regulation, implementation, and long-term use (Figure 2). Participant quotations are identified by participant group, sex, and age.

Figure 2. Issues around the design and intended positioning of artificial intelligence (AI) technologies for the diagnosis of skin cancer that were identified from the data, and further cross-cutting themes identified. *Allied health care professionals that could use the technology include pharmacists, health care assistants, practice nurses, podiatrists, hairdressers, and potentially many more. UK NHS: UK National Health Service.



Position of AI in the Skin Cancer Diagnostic Pathway

AI Before the PCP

Patients and members of the public felt that giving patients access to AI technology would be more accurate than "random googling" and could help them to decide when to see a physician. However, they expressed concerns about using the AI technology without clinical input. The loss of the human touch in consultations worried them, including how the diagnosis would be communicated and whether patients would be able to get a PCP follow-up appointment to discuss the diagnosis and ask questions. Another concern was whether all patients would be able to effectively use the technology and the potential for it to exclude some groups of patients, including older people. Patients and the public discussed the risk of overuse of the technology by patients; they felt that implementing it in a practice nurse and allied HCP-led clinic would reduce the risk of overuse while also making it easier for patients to access skin lesion assessments.

All HCPs could envision AI technology being used by patients and thought there would be a high demand for this service. Some dermatologists felt this could be a useful approach, particularly for patients at high risk of skin cancer between dermatology clinic appointments; however, they were concerned about the potential for overuse and subsequent workload due to false positives that would inevitably occur. They thought patients might struggle to understand the risks, benefits, and the output of the AI technology and would find it difficult to take high-quality images needed for the AI to analyze. Some HCPs commented on the potential psychological benefits of AI technology, in particular reducing patient anxiety if the technology reduced the number of urgent suspected cancer referrals, although some HCPs felt that it could increase patient anxiety through increasing access to skin lesion checks. Dermatologists and NPs commented on how accurate AI technology could enable a wide variety of allied HCPs to use the technology to triage patients presenting with suspicious skin lesions, which may improve access and enable earlier detection of skin cancer. GPs reported both positive and negative views on the potential of positioning AI technologies in the role of triaging patient-submitted photographs before an appointment with a PCP.

AI researchers from academic backgrounds were broadly skeptical of patients having access to AI technology themselves, highlighting concerns about the diagnostic accuracy of the technology, the risk of false reassurance, and that patients would not necessarily understand the context of when it was safe to use. However, AI researchers from commercial backgrounds felt that, if the AI technology was accurate enough, positioning AI technologies with patients had the greatest potential for

impact. They highlighted various potential positive effects, including reducing the barriers to getting a skin lesion assessed, helping patients make better decisions about when they needed to see an HCP, and educating patients about skin cancer.

AI With the PCP

Patients commented that GP surgeries are accessible for most patients and so are an ideal place to locate the technology. They felt that if they were consulting about a potentially serious skin cancer (ie, melanoma), they would like to be able to speak to a PCP. Combining the clinical judgment of the PCP with the AI technology was generally viewed as being more powerful than the PCP alone and could be more beneficial for PCPs with less experience in skin cancer and reduce the variability in skin cancer triage in primary care. The public discussed whether PCPs would become reliant on the AI technology and lose their clinical judgment and even whether this was an issue if the AI diagnostic accuracy were better than the PCP's.

GPs largely wanted PCPs to have access to the AI technology and were concerned that the implementation of AI technology at other points in the diagnostic pathway could undermine their gatekeeping role. They believed that PCPs are well positioned to triage and monitor skin lesions over time and provide continuity of care and that it would be more efficient for the technology to be used to triage a lesion in a single consultation instead of needing a separate appointment to capture the image. NPs agreed that AI technology would be best placed with a PCP. One GP emphasized that the positioning of the technology in the diagnostic pathway is probably the most important factor in how impactful it will be.

Dermatologists expressed the opinion that something was needed to improve the accuracy of referrals to secondary care but were not sure implementing AI technology in primary care was the best approach. A broad educational program for PCPs was suggested as a better alternative. Whether AI technology is currently accurate enough to be used by PCPs was a concern, as well as the risk of deskilling PCPs in skin lesion recognition. Dermatologists did see benefits in situating AI technology with PCPs to help with rarer skin cancer recognition, to break bad news, and answer patient questions, and to give PCPs more confidence in diagnosing a lesion as benign and not referring on to secondary care.

Most AI researchers expressed strongly that a "human-in-the-loop" approach (ie, where a human clinician is always involved in decision-making alongside the AI) is needed. Potential benefits of this approach included increased safety, allowing for patient interactions, allowing the PCP to focus more on the patient, and increasing knowledge of PCPs. AI researchers also discussed that the decision to biopsy or refer a skin lesion involves many biopsychosocial and clinical factors outside of the appearance of the lesion, which cannot be measured or considered by the AI and thus needs input from a PCP.

AI After the PCP

Use of AI technology to triage referrals from primary care to secondary care in a teledermatology setting was only mentioned by 1 GP and 2 dermatologists. The dermatologists suggested AI technology could help triage obviously benign skin lesions, preventing them from being referred to a dermatology clinic.

Few participants commented on positioning an AI technology with dermatologists. Dermatologists believed that it could be a useful training tool, and in the future, if it were proven to be more accurate than dermatologists, then it could be implemented in dermatology clinics. One AI researcher agreed and discussed how AI technology could help improve dermatologists' consistency in diagnosis (Textbox 1).

Textbox 1. Positioning of artificial intelligence (AI) in the skin cancer diagnostic pathway.

AI before the primary care clinician

- So I think people first self-diagnose via Google a lot anyway so having something that's a little bit more accurate than self-diagnosis is probably helpful. [Patient 4; female, aged <60 years]
- I don't think it would be effective giving it directly to patients, least of all because they won't have the training and the right level of professional knowledge in being able to interact with digital innovations effectively. [Patient 1; male, aged <60 years]
- I would be a little bit concerned about it just being open-ended, because...the potential for over-use by some people and under-use by others would be quite large, I think. So, I think...a nurse-led clinic or something like that might be better. Or give people the option. [Member of the public 3; female, aged >60 years]
- I know from doing online and video consultations during COVID, what was quite evident is that patients have got varying skills in terms of how they use their mobile devices to show you their skin remotely and taking images, and the quality of those images. So I think you'd have to think very carefully about who is taking the image, what they understand about what the clarity of that image needs to be, and what devices they're being used on. Because that's going to affect your interpretation. So I would be hesitant to say as a matter of routine for it to be patient led. [NP1 (nurse practitioner 1); female, aged >45 years]
- And seeing this opportunity, for example, hairdressers, podiatrist, pharmacist, they get asked a lot about these things, about lesions. And from time to time you get referrals because they went to the hairdresser and they spotted a mole... it would be a good thing to implement in these areas, expand beyond the GPs. [Dermatologist 3; female, aged <45 years]
- I would really like to then streamline the process. So we have an ability within our practice that the patient doesn't need an appointment for asking about a mole of concern and they can send a photo. That photo would be reviewed by a clinician. It would be absolutely fantastic if that photo was also able to be reviewed by an AI process and advise whether there were concerning features on it or not. It would be useful if that photo could then be reviewed again six weeks later against a new photo to assess for changes. [GP5 (general practitioner 5); male, aged <45 years]
- Yeah, it (using the app) has especially low barrier to use it so it's way lower than going to a doctor or going to a dermatologist. It is really a first step into being more interested or more concerned about skin cancer. And I think it can help to raise awareness (of skin cancer) and inform people better. [AIR4 (AI researcher 4); male, aged >45 years]

AI with the primary care clinician

- I don't want to think that a doctor would see results on an app and still not think about it at all...But I think with time they might lose their skills. [Member of the public 6; female, aged <60 years]
- I'd love it in my hands. Anything that saves primary care queuing people up in dermatology...I think skin is difficult. And so my feeling is that human error will always exist and I guess if you've got something to support you in making the correct diagnosis in serious skin lesions, then it seems to me like a win-win and a sensible option. [GP2; male, aged >45 years]
- Most patients with a lesion will go straight to their GP. So I think definitely that's where it's going to be best placed and whether that's a specialist nurse within a GP setting or a GP, certainly within that primary care set up I think is most relevant. [NP1 (nurse practitioner 1); female, aged >45 years]
- I think selfishly, I'd be disappointed if I was told you don't need your dermatoscope anymore, because it's an area I'm really interested in and I'm really enjoying...because if it's automated and, therefore, you lose the skill set and you lose motivation around the topic. [NP3; male, aged <45 years]
- I would be worried about de-skilling our GPs by giving them a tool that tells them what to do. [Dermatologist 3; female, aged <45 years]
- The great thing would be if they were able to pick up those skin cancers that they haven't thought about, for example, amelanotic melanomas, or nodular melanomas, that they don't follow the typical a, b, c, d criteria. [Dermatologist 3; female, aged <45 years]
- If AI had a really, really good dataset of benign lesions, that would give a GP confidence to say 'no, that's benign, that doesn't need to be referred in' [Dermatologist 4; female, aged >45 years]
- They (dermatologists) are very experienced and they don't need a tool. But some of those tools might be needed for the people who refer patients, not to make any mistake at that stage for the early detection...So I can think of those two aspects, like speed and a better decision for the practitioner. I see really big opportunities in those kinds of things. [AIR2; male, aged <45 years]

AI after the primary care clinician

- So,...either a patient-generated image, but preferably taken in primary care on a high-quality camera and sent securely to dermatology for triage, then the AI helping with that triage process. [GP3; male, aged <45 years]
- In secondary care it would be really helpful, I think particularly for juniors starting out, to have a list of the differential, including the rarer things that could possibly be consistent with the appearances that the algorithm has identified. But giving you ranking of likelihood. [Dermatologist 2; female, aged >45 years]

Aim of the Technology

Education, Screening, or Diagnosis

Participants mainly discussed AI technology as a diagnostic tool; however, participants from all groups raised alternative uses. Patients discussed skin self-monitoring, including sequential monitoring of skin lesions over time, and patient education about the "red flags" of skin cancer. All groups discussed the potential to raise awareness and educate patients about skin cancer, including skin cancer prevention and how to perform skin self-monitoring.

GPs and NPs discussed the potential for AI technologies to educate PCPs and improve diagnostic skills. They suggested that a potent educational attribute would be if the AI technology could highlight visual features of skin cancer in images of skin lesions; AI researchers commented that these types of features have been developed. Dermatologists discussed the potential use of a patient-facing AI technology as a screening tool for high-risk populations.

Focus on Benign or Malignant Diagnoses

Patients and members of the public stated that the primary aim of the AI technology should be a very high accuracy to avoid missing skin cancers and giving false reassurance; to achieve this, the technology would need to have a high sensitivity. All HCP groups commented that aiming to diagnose benign skin lesions might be a safer approach with less risk of missing skin cancers. They proposed that focusing on diagnosing specific common benign skin lesions, such as seborrheic keratoses and dermatofibromas might reduce unnecessary referrals to secondary care.

Focus on Melanoma or All Skin Cancers

Patients were primarily concerned that any AI technology was accurate for melanoma. Some members of the public had experience of KCs in family or friends and felt AI technologies should also address KCs. GPs were primarily concerned about melanoma because it often affects younger patients and has higher mortality—they felt that AI technologies could be applied to other skin cancers in the longer term. Dermatologists and a NP commented that diagnosing all types of skin lesions, including all skin cancers, was important, especially for technologies that are designed to be used by patients or HCPs with less experience in diagnosing skin cancer. However, some dermatologists thought it could be difficult to train AI to diagnose KCs accurately, because patient history often has greater importance than for melanoma, and AI currently does not always incorporate this.

Threshold Between Malignant and Benign and Management of Borderline Lesions

All groups discussed the difficulty in setting a threshold between benign and malignant lesions for the AI technology, in essence how to translate the continuous risk score generated by the AI into a binary clinical decision of whether to refer a patient or biopsy a skin lesion. A 3-layer management strategy was suggested by several participants with further assessment of lesions that are close to the threshold, either through follow-up assessment in primary or secondary care or through sequential monitoring over time (similar to short-term sequential digital dermoscopy imaging models that already exist [24]). NPs and dermatologists reflected on how this issue demonstrates the complexity of clinical practice and that clear guidelines will be needed about what to do at each risk level (Textbox 2).

Textbox 2. Aim of the artificial intelligence (AI) technology.

Education, screening, or diagnosis

- I guess you always take a picture at a point in time and...say you noticed it change in a couple of months could you go back and say, it's changed from this to this, and almost keeping a progression record...I think that would be a really helpful feature. [Patient 4; female, aged <60 years]
- Give people recommendations of you should not go out unless you're wearing 50 SPF, we think you should not sunbathe between 12 and four, just because it's cloudy does not mean you're not going to burn...put a kid in a sun hat so they don't have sunstroke. [Member of the public 3; female, aged >60 years]
- I say to them (patients) there are mole apps that you can monitor, and that maybe it's not great in terms of getting a diagnosis, but they're more aware and they're keeping attention more to a particular lesion. And definitely it helps to diagnose cancer in the early state than what we saw years ago. [GP1 (general practitioner 1); female, aged <45 years]
- But the way I used AI was to try to train on some features in the image rather than giving me a diagnosis. For instance, to teach a deep learning algorithm to tell me whether that the borders are regular or irregular, 'cause that's something subjective sometimes between readers...So I used AI to reveal the features rather than giving the full diagnosis. [AIR2 (AI researcher 2); male, aged <45 years]
- So if you could develop one that was validated it might be useful for selected patients, maybe not the entire population but particularly high risk patients, maybe patients with lots of moles. [Dermatologist 2; female, aged >45 years]

Focus on benign or malignant diagnoses

- The first aim of the app needs to be to go and get it checked or not, or to go and get it biopsied or not. [Patient 4; female, aged <60 years]
- Just screening out the seborrheic keratosis, the pigmented dermatofibromas, the benign...if it could screen all of those out, which are the vast majority of the two-week-wait referrals that we see, that would be incredibly important for providing a better, more targeted service...plus potentially triaging things that are so unlikely to be a cancer that they don't need to be referred. So it kind of works both ways, early detection and reducing the massive numbers of 2-week-wait referrals that we get. [Dermatologist 2; female, aged >45 years]

Focus on melanoma or all skin cancers

- I think melanoma is the most important because I'd like to think that because your differential features for a squamous cell are quite obvious, that would be referred on anyway. I think it's melanoma and other pigmented lesion differentiation that's really tricky. So my thought would be more melanoma, and certainly that's what I've used it for. [NP1 (nurse practitioner 1); female, aged >45 years]
- People just think it's about diagnosing melanoma, where actually it's not. It's about recognising all skin lesions. You've got to diagnose benign lesions, and you've to diagnose malignant lesions, and it's not just melanomas, you've got BCCs, you've got SCCs, you've got AKs, and then you've got all the benign lesions. [Dermatologist 1; female, aged >45 years]

Threshold between malignant and benign—and management of borderline lesions

- I guess one approach you could take with marginal cases, you could say that we suggest you get re-tested in six months' time or something like that. So, you have a three-layer band. One you definitely need action, one you definitely don't, and then a middle level where you come back for a test after a bit. [Member of the public 3; female, aged >60 years]
- What I suggest would be good...if you're less than 70 per cent sure, it's an arbitrary...whatever number sure, then there is clinician involvement. So take a deeper dive into the history and that person then comes in and is looked at. [NP3; male, aged <45 years]
- So I think it would be really interesting, if when I looked at a lesion I gave the patient a percentage of how right I thought I was or even any diagnosis. I mean, I'd love that, that would be really cool to have that honesty. 'I think you've got a viral chest infection, I'm about 50-50, I'm going to hold off on the antibiotics, but here's some strict safety netting.' Yeah, I don't know whether I'd want to give the patient that information. [NP3 male, aged <45 years]
- So it would come back with a comment like 'this is 80% likely to be a melanoma, this is 20% likely to be a melanoma' and there would have to be some sort of understanding, some sort of cut off, what is the point at which a referral is merited...And I suppose that one of the dangers...where do you cast the net in terms of risk? [Dermatologist 4; female, aged >45 years]

Cross-Cutting Themes

Trust

Patients and members of the public often raised the issue of trust. The newness of the technology, the involvement of private companies and concerns about data privacy, and the diagnostic accuracy being <100% were all felt to be reasons for a lack of trust. However, there were also participants who thought patients would trust a consultation with a PCP more if it involved AI technology, even if it had made no difference to the assessment. GPs worried that patients could demand the AI technology be used in consultations and were concerned about the risk of false reassurance. NPs felt that patients fundamentally trust people more than machines. Dermatologists stated that there is not enough evidence that existing AI technologies are safe and accurate enough to be used in clinical practice and that it is difficult for patients to determine which patient-facing AI technologies they can trust.

AI researchers discussed that the "black box" nature of AI technologies should not be a barrier to trust, as we similarly do not understand the mechanism of action of many medications.

They felt that if clinicians recommended or adopted an AI technology, then patients would be more likely to trust it.

Ease of Use and Acceptability

Ease of use was a major concern for patients and members of the public; participants discussed that a technology that is easy to use enables patients to take high-quality images that are better suited for AI analysis and reduces the risk that patients will be unable to use the technology or use it incorrectly. Several members of the public raised the importance of enabling patients to choose how to consult, rather than mandating a consultation type that is inappropriate for them. GPs and NPs' primary concern was that it should be easy to integrate the AI technology with their current computer and Wi-Fi systems. Dermatologists and patients were concerned about whether the technology would facilitate taking high-quality images for the AI to analyze. AI researchers commented on how practical limitations of a technology can prevent it being used, even if it is potentially very beneficial; therefore, "real-world factors," such as cost, ease of use, and how well it fits into clinical workflows, need to be considered.

Accuracy and Generalizability

Accuracy was raised frequently by all groups and is linked to many of the subthemes. The primary concerns were the false negative rate and the risk of false reassurance and whether the diagnostic accuracy is generalizable to other clinical settings, using different camera technologies, and with other populations and demographics. AI researchers commented that currently AI technologies are often developed on small datasets, which are not representative of the general population and may contain errors and biases and hence, do not generalize well to be accurate in all sections of the population. Members of the public and HCPs were concerned this might mean AI technologies would be less accurate in melanin-rich skin or for rarer skin cancers.

AI researchers discussed how this situation might be improved with close collaboration between clinical and AI researchers and a focus on data quality in a "data-centric" approach. An AI researcher from a commercial background commented on the challenges in collecting a representative dataset for AI development and testing. There are fewer publicly available images of skin lesions in melanin-rich skin, and their app had significantly lower uptake among patients with melanin-rich skin.

Evaluation and Regulation

Patients felt that AI technologies should be evaluated by a mix of professionals before they can be adopted, including independent confirmation of diagnostic accuracy. Patients were concerned that AI technology would be adopted based on novelty and hype or because it is cheaper than clinicians' time when it may not be in the best interests of patients and their clinical care. HCPs felt that significant data from clinical trials would be needed to evaluate AI technologies but recognized that this might take time. AI researchers were concerned that we currently lack good measures to evaluate AI algorithms for use in clinical settings. They stated that the current practice of using simple diagnostic accuracy measures (eg, sensitivity and specificity) is not comprehensive enough to demonstrate the accuracy, benefits, and risks of AI technologies. AI researchers added that, while clinical studies, due diligence, and understanding biases and flaws in an AI system are all important aspects of evaluation, there are other factors, including business sustainability, that need to be considered before a decision on adoption can be made. Many groups commented on the need for health-economic evaluation as part of any evaluation program.

Patients and members of the public often disclosed that they did not understand regulatory processes. Some felt it was important that AI technologies had a Conformité Européene marking and were evaluated by a national body. HCPs were concerned about regulatory processes for AI technologies not being as robust as for medicines and treatments. GPs raised concerns about where the medicolegal responsibility for errors related to the use of AI technologies would lie. Dermatologists wanted the use case for AI technologies to be made clear and for the regulator to assess the technology based on this but acknowledged that regulating AI technologies is challenging. An AI researcher commented that regulatory processes for AI technologies are becoming more complex.

Implementation in the UK National Health Service

Members of the public and GPs discussed variation in National Health Service infrastructure in different regions of the United Kingdom, including wireless connectivity, and how this could make implementation difficult.

Most participant groups commented on the capacity of the health care system to implement a new technology where resources are strained. Current pressures may mean that PCPs have insufficient time to understand and implement a new technology. Some HCP participants highlighted that an AI technology with a low specificity might lead to a significant increase in referrals and worsen workload pressures. Conversely, some NPs and dermatologists commented that use of AI technology could help to reduce the number of referrals to secondary care. AI researchers hoped AI technologies could be used alongside PCPs and dermatologists to ease workload pressures.

All groups commented on the importance of professional bodies in the implementation of new technologies. Patients and HCPs felt that implementation would be greatly helped if professional bodies, such as the National Institute for Health and Care Excellence, National Health Service England, Integrated Care Boards, the BAD, or the Medicines and Health products Regulatory Agency, had recommended or evaluated it. GPs and NPs believed any decision to adopt and fund AI technology would need to come from a higher body rather than individual GP practices. AI researchers commented on the importance of the views of professional bodies for the adoption of AI technologies but also on the associated commercial challenges as these bodies vary internationally.

NPs and dermatologists highlighted the need for adequate training in how to use an AI technology as part of the implementation process and clear guidelines on how the AI technology should be used and interpreted.

Long-Term Use

The unique potential for AI technologies to continue learning after implementation and for diagnostic performance to improve over time was raised by different participant groups. GPs and patients assumed that this feature would be standard practice and that regular updates would sequentially improve the AI technologies performance as it learned from new data. Few participants commented on how this process could be regulated in practice.

A major concern of all groups was how to check that the AI technology was performing accurately. AI researchers were particularly worried about the potential for the performance of the AI technology to deteriorate over time, referred to as "drift in performance." They commented that this could occur because of a change in the way the AI technology is used, a change in the population (for example, using the technology in a population with lower skin cancer prevalence or different demographics compared to the development and testing datasets), a change in the hardware, or a change in the accuracy of the technology over time. All groups suggested "sanity checks" that could be used regularly to detect if an AI technology was not performing accurately. These "sanity checks" included expert systems monitoring image metrics over time, comparing the prediction of the new AI technology to the previous AI gold standard; a clinician reviewing all cases that the AI diagnoses as "likely skin cancer" or where AI has low confidence in the diagnosis; or limiting the use of AI technologies with a "human-in-the-loop" approach. AI researchers added that there needs to be an incentive for the makers of the AI technology to maintain and provide ongoing support long-term (Textbox 3).

Textbox 3. Cross-cutting themes.

Trust

- The fact that it hasn't been the norm within the health care setting, I think will make quite a lot of people feel uncomfortable. So the fact that it's just so up and coming and new will naturally spark a bit of anxiety in terms of people and they think about if it's safe, if it's got the same principles and standards in place in terms of care, safety, confidentiality et cetera. [Patient 1; male, aged <60 years]
- I think there is also a section of the population who are very suspicious of data not being used correctly. So, you would need to have some sort of reassurance about correct data use as well. [Member of the public 4; female, aged >60 years]
- I think people fundamentally put trust in people and we're still wary about putting trust in machines, or computers, because we've been fed sci fi for years that makes us worry, and also I do think that we still think we know best, even though the algorithms that will be in the computer will be absolute gold standard, and they won't be tired, hung over, jaded, they'll be right every time. [NP3 (nurse practitioner 3); male, aged <45 years]
- I think the concern at the moment is that there isn't enough evidence that any of these machines, or machine learning is up to speed to be able to make diagnoses without missing any skin cancer, and that includes rare skin cancers, skin cancers with rare presentations, or in different ethnicities. [Dermatologist 3; female, aged <45 years]
- Yeah. I'm just really unsure in how far, to be honest, machine learning can or will take over a medical setting, or should. The more I see and read, the more I'm getting also suspicious. [AIR3 (AI researcher 3); female, aged <45 years]
- I would question the premise that being a black box is actually a terrible thing or even a novel thing, because clinicians are black boxes. As far as I understand for most medicines we have no idea why they work, and we still use them because statistically it works. And as long as we regularly check that our models work statistically then who cares whether that's a chemical or a computational black box. Obviously, it's nicer if you can also present the clinician with some data that the clinician then can use to make further judgment calls. [AIR1; male, aged <45 years]
- As a field we're only beginning to scratch the surface of what that means, to trust technology. If say my physician tells me, this is a good way to understand more about your skin conditions, there's some element where I trust he or she as a professional and the information relayed is therefore accurate, and therefore that trust extends to the thing they had suggested. [AIR5; male, aged <45 years]

Ease of use and acceptability

- So, being accurate, being honest, easy to use, easy to understand. And easy for the GPs to use as well, the other health care professionals who've got to get the information through that as well, easy for them. Because they're going to have to look at whatever it's come up with and try and make some sense of that before they actually sit down in front of us. [Member of the public 1; female, aged >60 years]
- For me patient choice is so important—so both have a machine in the GP surgery for those who would prefer to do that and have an app for the people who are quite comfortable using those. [Member of the public 1; female, aged >60 years]
- I suppose if it was a lengthy process. So if there was dodgy Wi-Fi or it's hard to get your image or it's taking time to upload...I suppose it's just the practicalities and the ease of the software and hardware that you're using, those will be barriers, if it didn't work. [NP1; female, aged >45 years]
- How easy is it to take the photograph? I think that's really important. Because I get loads of referrals with photographs, but the photographs are completely useless, they're blurred and a complete waste of time. Somebody's ticked the box and said, sent a photograph, but they might as well not be there. [Dermatologist 1; female, aged >45 years]
- If you have a technology that's extremely beneficial for a certain disease but it's really expensive, it takes a really long time, and it's hard to use then no one will ever use it. So there's real world factors there. [AIR5; male, aged <45 years]

Accuracy and generalizability

- I think the key issue is generalisability, because algorithms are developed on a very small set which have very specific properties and there is biases. And then of course, they do not work on a wider range of other images from other scanners from other countries...this is a huge issue, the generalisability. You have all these algorithms that achieve higher numbers in one setting but it doesn't mean that they will work well on new data they have not seen. [AIR3; female, aged <45 years]
- They need to make sure that there is a proper diversity of people in there. Because AI is only ever going to be as good as what you use to train it. So I think they need to be particularly careful about diversity of skin color and capturing the wide variety that's needed. [Member of the public 13; female, aged >60 years]
- For us it's a kind of a chicken and egg problem. So, there's not a lot of data for darker skin available, so training on that...and especially proving the accuracy on dark skin it's almost impossible...We don't have the users, we don't get the data, we don't get the proof of how good we are on darker skin—so that definitely is a loop that we need to break at some point. [AIR4; male, aged >45 years]

Evaluation and regulation

- I have come across many issues because algorithms have been evaluated with measures that are actually not measuring what you would want or need in a clinical setting...So I actually think the most important thing right now would be to set up proper evaluation schemes and to think about how deep learning models should be or can be properly evaluated...The real question is, how do we evaluate them to make sure they will work in a medical clinical setting. [AIR3; female, aged <45 years]
-

So this is definitely a concern, not to send too many users into healthcare. The first concern's accuracy, we don't want to miss too many skin cancers, for sure, but the second concern is definitely also the health economic case. So, we can of course send a huge amount of people into healthcare. We find more skin cancer, we reach our goals of finding more skin cancer, that's fine. But we don't solve a real problem, we make it worse for the healthcare system as well. [AIR4; male, aged >45 years]

- I worry a little bit that if it's something that is a manageable cost it might be overused. It's a way to shift a number of patients who you might otherwise see...They're always on an efficiency drive, they're always under pressure to save money...I don't know whether, if the technology was affordable for GP surgeries, they would just think, oh great...we're short of doctors we can just process people through the hands-off routine...you stop thinking about it in clinical terms and you start thinking about it in financial terms. [Patient 2; female, aged <60 years]
- I think that's really important that it has CE marking; because I see a lot of apps and crap in the digital space that a lot of people are buying and spending money on and they're not evidence-based medicine. [Member of the public 2; female, aged <60 years]
- I know it (regulation) is fairly patchwork. Obviously, when it comes to medication and prescribing, there's fairly robust regulatory systems... When it comes to certain equipment, some patient devices it's a bit piecemeal. But I would be fairly reassured if the MHRA covered this equipment. [GP3; male, aged <45 years]
- I've been indoctrinated by the British Association of Dermatologists. And what they explained. They gave a position statement, I think it was a year or two ago, and we were've told these are medical devices and therefore they need to go through these medical regulatory agencies. I think before Brexit it obviously would have been Europe as well and now I think it's MHRA should be responsible for this. Which makes sense because obviously it's a very important thing for patients and for doctors, medicolegally as well. So it needs to work for its purpose, for what they say it's going to be doing. [Dermatologist 3; female, aged <45 years]
- And also where does it stand legally? If you come to me and you show me a mole and I do my AI thing and I say to you computer says no, and off you go and you keep doing the things that you do, and then two years later you come back to me with a large black blobule, sentinal lymph node biopsy positive...is that my fault? Is that the AI's fault? And if it's the AI's fault, who are you suing? [GP6; female, aged >45 years]

Implementation in the UK National Health Service (NHS)

- I think when you're an external looking in at the NHS it's just a monster beast to try and understand different parts of it...Overall I think integration of digitals is great, but it's not uniform, and also there are massive connectivity issues still in some areas...Some parts of the country are doing amazing things, and other parts are so backward. [Member of the public 2; female, aged <60 years]
- I'm just very much aware that the basics of general practice is under so much pressure that adding a new technology and complication is not everybody's first priority...I appreciate a lot of these new technologies can save time and resources in the long term, but certainly, if I'm thinking now of the start of a difficult winter, it's going to be a tough sell right now. [GP3; male, aged <45 years]
- One of the things I came across is that there was a shortage of dermatologists in general. This means that the practitioner is expected to see a lot of patients in the same day...the practitioner is a human, maybe some stress, or he had a bad day, he might not spot some feature in that image. So those (AI) tools can be something that fixes this gap if in that day he had a bad day and didn't notice some features. [AIR2; male, aged <45 years]
- So it's going to cost something to have this kind of level of equipment, but there's not an infinite amount of money in GP land, if they're spending on that then it's coming out of somewhere else, which might mean we won't have the money or equipment for something that may benefit patients. [NP2; female, aged <45 years]
- It's got to come with the support and the training along with it, it's not just about getting a new piece of kit...I think the important thing is that that cost incorporates training and a clear understanding of the device and what it's used for, and why, and how to use it. It's silly to give a piece of kit and then not have the support to know how to use it best. [NP1; female, aged >45 years]

Long-term use

- In some ways the accuracy would build as the technology was used. So the longer it's been in play the more you can depend on it probably...I assume that you don't just test the heck out of it and say it's fine now and then stop refining it. [Patient 2; female, aged <60 years]
- Let's say the model was trained in a way it works really well when you are taking dermatoscope photos, and it gives you a really good sense of the malignancy potential there and the studies support this. Then a patient sends you a photo...and this photo was taken using a smart phone with different kinds of lighting conditions, in a setting where it was actually never designed for use but that fact is not obvious to anyone who hasn't been thinking about this for a really long time. So with the best of intentions, I think you could still have a situation where the performance starts drifting away from how it was designed for use. [AIR5; male, aged <45 years]
- Maybe scanners will change...It is not clear that if an algorithm works well on the scanner of the last generation, even from the same company, (if) you get a new one that it would work on that as well...They (scanners) will not stay the same, but then we really should think about how do we integrate new modalities into the machine learning models, because it's not sustainable to develop something on a fixed dataset. [AIR3; female, aged <45 years]
- It should be tested before it gets to the GP surgery, and then it should be checked...at regular intervals, to make sure it's still working correctly. Because if you start falsely diagnosing people, that could be a complete waste of time. [Patient 3; female, aged >60 years]
- If we look at a top-level view across all AI it's really hard to summarise in a single sentence whether this will help or hurt. But if developed well, if used well I still do believe there's lots of great potential here. [AIR5; male, aged <45 years]

Discussion

Principal Findings

In this study, many of the discussions centered on 2 simple questions: who is going to use the AI technology, and what is it going to do? Most participants commented on positioning AI technology with patients or allied HCPs (before the PCP) or with the PCP as a decision support tool. Few participants proposed positioning the technology to triage referrals to secondary care or as a decision aid for dermatologists. Participants highlighted several overarching topics as important to them, including trust, acceptability, generalizability, evaluation, regulation, implementation, and long-term use.

The risk of false negatives resulting from an AI technology with poor sensitivity and missing skin cancers was a major concern for all participant groups. Missed skin cancers, especially melanoma, are likely to lead to late diagnosis and worse outcomes for patients [3,6]. One potential benefit of AI technology mentioned by all groups was the reduction in the workload of health care services, primarily through effective triage of patients that need to see PCPs or specialist clinicians. This could be the case if the AI technology has a good specificity with few false positive results. False positives are inevitable, but an AI technology with a low specificity could significantly increase the workload of health care services. This was a concern raised by HCPs, but few other participant groups commented on this risk. It has been suggested that overdiagnosis of melanoma is rising due to increased rates of skin examination, decreased thresholds for biopsying skin lesions, and for labeling morphological changes on histopathological examination as malignant [25]. Implementation of an AI technology with low specificity could increase both rates of skin examination and biopsy rates, both potentially contributing to overdiagnosis.

AI researchers were the most pragmatic, expressing concerns about the generalizability of many AI technologies and the relevance of current testing approaches in preparation for clinical implementation. They were concerned about the robustness of the datasets underlying AI technologies, including the representativeness of skin cancer prevalence and patient demographics in the datasets. They felt that, at the current time, these technologies need to be implemented with a human in the loop. Many HCPs were aware of the lack of evidence and potential risk to patients that has been publicly commented on by the BAD [9]. Patients and the public were aware of potential improvements in patient access, diagnostic accuracy, and reduced workload AI technologies could offer but were concerned about the risk of missing cancers and losing the opportunity for human interaction and to ask questions.

Complexity underpins most of the generated themes. Developing, evaluating, implementing, regulating, and maintaining AI technology in health care settings are all multifaceted, complex processes containing many opportunities for error, as laid out in the CanTest framework [14]. Therefore, it is unsurprising that so many diagnostic technologies fail at the implementation stage [26]. Several studies have discussed the complexity of implementing digital, AI, and machine learning (ML) interventions into health care settings; they recommended a whole-of-system approach with particular focus on how users interact with devices and user training [12,27].

Comparison With Existing Literature

The NASSS framework [15] was chosen to guide this study because it includes a wide range of domains that help capture the complexity of implementing health care interventions [28]. Most of the cross-cutting themes link closely to NASSS domains, specifically “knowledge to use the technology,” “demand-side value (to patient),” “adopters,” and “organizations.” There were several themes raised that did not fit into the NASSS framework (such as the potential for continued learning with AI technologies), as well as themes (such as trust and acceptability) that seemed to lose some of their breadth and nuance by being contained within the NASSS domains. We chose to code both inductively and deductively, which allowed us to better capture participants’ views and build knowledge about applying the NASSS framework to novel AI technologies.

A recent Swedish study used an AI-based melanoma decision support aid for clinicians as an example technology to generate discussions with participants about the implementation of decision aids. In keeping with their findings, many participants in our study discussed the issues of accuracy, safety, data security, liability, ease of use, and integration [29]. Our findings support those of a recent systematic review on the use of ML-based risk prediction models in health care settings [13] in which participants demonstrated largely positive views of AI technologies but identified many barriers. Echoing findings from several recent studies [30-33] we identified concerns about diagnostic accuracy, risk of patient harm, ease and speed of use, HCP overreliance on the technology, legal liability, data protections, data quality, impersonality, and positioning in the diagnostic pathway. In particular, we identified aspects of the consultation participants felt that AI technologies could not replicate, that is, the need for human interaction and clinical experience and judgment. Participants commented on the risk that AI technologies will not be effective in minority populations who are inadequately represented in training and testing datasets and may exclude populations with lower technological literacy, such as older people.

The wide variety in positioning and approach of existing AI technologies [7,8] indicates that the optimal position and approach have yet to be determined. Several participant groups highlighted that they wanted more evidence of the accuracy of technologies in real-life clinical use. This fits with recent reviews of AI technologies aimed at detecting skin cancer [7,8,33], although increasing evidence from clinical trials is emerging [10]. PCPs were keen to have an AI technology to support their diagnostic decision-making, in keeping with findings from a previous study [34]. AI researchers highlighted a growing body of research in AI technology development for health care settings, including the “human-in-the-loop” approach and the “data-centric AI movement” [12,35,36].

Regulation was a topic raised by several participant groups. The fast-moving pace of AI development makes regulation of AI technologies challenging: underregulation risks patient safety while overly zealous regulatory approaches could hinder AI

development pipelines and implementation [37,38]. AI has the unique potential to continue developing over time as it is exposed to more data. Regulatory bodies around the world are attempting to keep up with the rapid developments in AI technologies. In the United States, the Food and Drug Administration has proposed the 510k pathway, which facilitates the approval of software as a medical device if it is substantially based on a previously approved technology [39], and is developing an approach to an AI-ML workflow that would enable continued learning after implementation [40]. There are emerging national and international regulatory policies, including the European AI Act, the United States AI Bill of Rights, and the United Kingdom policy on AI [41-43].

Strengths and Limitations

To the authors' knowledge, this is the first study to report the views of a broad range of stakeholders about the use of AI technologies to facilitate the earlier diagnosis of skin cancer in primary care settings. We had good variation among interviewees in terms of background, age, sex, and geographical location. The study benefited from PPI at every stage, and a strong conceptual framework was used to develop the framing of the interview schedule and data coding.

Aiming to recruit a wide range of stakeholders was a conscious choice, as we felt this was important to achieve a breadth of opinions; the trade-off was that time and resources meant we were only able to include limited numbers in each participant group. The aim was to achieve breadth of opinion without necessarily achieving saturation. AI and clinical implementation are complex subjects, which meant that we were more likely to recruit participants with higher educational attainment and who are engaged with health care research or AI; both these aspects may have affected the balance of views we obtained. We only recruited 1 participant who reported that they disapproved of using AI technologies to help diagnose skin cancer. Skin cancer is more common in melanin-poor skin; however, a key limitation of current AI technologies is their lack of training and testing in populations with melanin-rich skin. We recruited limited numbers of participants with melanin-rich skin in this study, so participant views on this issue may be incomplete. In contrast to other participants, AI researchers were largely based outside the United Kingdom, reflecting the location of the majority of commercial companies developing these technologies. However, it meant their knowledge of United Kingdom clinical practice and diagnostic pathways was sometimes limited.

Implications for Clinical Practice, Research, Adoption, and Policy

Health care services are working under extreme pressures in primary and secondary care [44]. AI technologies aimed at diagnosis or triage of skin lesions could facilitate early diagnosis of skin cancer to improve outcomes for patients and potentially ease some of these pressures. However, before this can happen, research is required to prove their efficacy with real-world clinical populations and to address the questions that remain about the most effective positioning of the technologies in the diagnostic pathway and the optimal approach for their use. Diagnostic technologies that are used in populations that are different from those they were developed and tested in are prone to spectrum bias [45]. Better measures of clinical performance are required to inform these studies, which consider not only diagnostic accuracy but also provide a measure of generalizability and dataset quality [46].

Some of our findings can be used to further develop the NASSS framework, for example, to consider in more depth how users interact with an AI technology and the potential for continued development after implementation. When developing an AI technology aimed at the diagnosis or triage of skin cancer, developers need to consider carefully and be specific about the intended use, including where it will fit into the diagnostic pathway for skin cancer and the approach that it is going to take. Developers must also consider the quality and representativeness of the data they use to develop the AI.

The decision to adopt an AI technology is complex and multifaceted. Clear regulatory processes that consider unique features of AI technologies need to be established, including continued learning, to ensure AI technologies are safe and effective when used in clinical settings. Adopters should also consider what safety nets are in place to identify poor performance and reduce false negative results, such as expert systems and regular "sense checks."

Conclusions

AI technologies are being designed with a wide variety of approaches, and the optimal approach and position in the skin cancer diagnostic pathway for these technologies have not yet been determined. AI technologies have the potential to help detect and diagnose skin cancer, to improve patient experience and outcomes, and to reduce the workload of overstretched health care systems. However, we have identified important concerns surrounding trust, acceptability, usability, generalizability, evaluation, regulation, implementation, and long-term use. These technologies need to be developed carefully and thoughtfully to achieve their potential, guided by evidence-based approaches and appropriate implementation, taking into consideration long-term sustainability and safety.

Acknowledgments

The authors would like to thank our patient and public involvement and engagement group—Margaret Johnson, Jenny Hunt, Sylvia Bailey, and John Lancaster—who contributed to study development, recruitment strategies, study materials, and the interview schedule. The authors would also like to thank Professor Steve Morris, who contributed to the development of the study.

The research was supported by a Cancer Research UK Cambridge Centre Clinical Research Fellowship for OTJ. JE is supported by an NHMRC Investigator Award (grant APP1195302). SES is supported by Barts Charity (grants G-001520 and MRC&U0036). RNM is Chair of the AI Working Party Group at the British Association of Dermatologists. The funding sources had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview schedule for patients and members of the public.
[DOC File , 340 KB - cancer_v11i1e60653_app1.doc]

Multimedia Appendix 2
Interview schedule for health care professionals and artificial intelligence researchers.
[DOC File , 340 KB - cancer_v11i1e60653_app2.doc]

Multimedia Appendix 3
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 94 KB - cancer_v11i1e60653_app3.pdf]

References

1. Garbe C, Keim U, Gandini S, Amaral T, Katalinic A, Hollezcek B, et al. Epidemiology of cutaneous melanoma and keratinocyte cancer in white populations 1943-2036. Eur J Cancer 2021 Jul;152:18-25. [doi: 10.1016/j.ejca.2021.04.029] [Medline: 34062483]
2. Non-melanoma skin cancer statistics. Cancer Research UK. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/non-melanoma-skin-cancer [accessed 2024-12-16]
3. Melanoma skin cancer survival statistics. Cancer Research UK. URL: https://www.cancerresearchuk.org/about-cancer/melanoma/survival [accessed 2024-12-16]
4. National disease registration service. Cancer Registration Statistics, England. 2021. URL: https://digital.nhs.uk/data-and-information/publications/statistical/cancer-registration-statistics/england-2021---full-release [accessed 2024-12-16]
5. Health topics: ultraviolet (UV) radiation and skin cancer. World Health Organisation. URL: https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv)-radiation-and-skin-cancer [accessed 2024-12-16]
6. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024 Jan 17;74(1):12-49 [FREE Full text] [doi: 10.3322/caac.21820] [Medline: 38230766]
7. Wen D, Coombe A, Charalambides M, Jones O, Wietek N, Dinnes J, et al. DS03: a scoping review of commercially available artificial intelligence ‐ based technologies for diagnosis or risk assessment of skin cancer. Br J Dermatol 2022 Jul 05;187(S1):159. [doi: 10.1111/bjd.21446]
8. Jones OT, Matin RN, van der Schaar M, Prathivadi Bhayankaram K, Ranmuthu CK, Islam MS, et al. Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review. Lancet Digit Health 2022 Jun;4(6):e466-e476 [FREE Full text] [doi: 10.1016/S2589-7500(22)00023-1] [Medline: 35623799]
9. BAD position statement v2. British Association of Dermatologists Position Statement on Artificial Intelligence (AI) Interventions. 2022 Feb. URL: https://www.bad.org.uk/clinical-services/artificial-intelligence/ [accessed 2024-12-16]
10. Menzies SW, Sinz C, Menzies M, Lo SN, Yolland W, Lingohr J, et al. Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial. Lancet Digit Health 2023 Oct;5(10):e679-e691 [FREE Full text] [doi: 10.1016/S2589-7500(23)00130-9] [Medline: 37775188]
11. Artificial Intelligence technologies for assessing skin lesions selected for referral on the urgent suspected cancer pathway to detect benign lesions and reduce secondary care specialist appointments: early value assessment. National Institute for Health and Care Excellence (NICE). 2024. URL: https://www.nice.org.uk/guidance/indevelopment/gid-hte10047 [accessed 2024-12-16]
12. Lyell D, Wang Y, Coiera E, Magrabi F. More than algorithms: an analysis of safety events involving ML-enabled medical devices reported to the FDA. J Am Med Inform Assoc 2023 Jun 20;30(7):1227-1236 [FREE Full text] [doi: 10.1093/jamia/ocad065] [Medline: 37071804]
13. Giddings R, Joseph A, Callender T, Janes SM, van der Schaar M, Sheringham J, et al. Factors influencing clinician and patient interaction with machine learning-based risk prediction models: a systematic review. Lancet Digit Health 2024 Feb;6(2):e131-e144 [FREE Full text] [doi: 10.1016/S2589-7500(23)00241-8] [Medline: 38278615]

14. Walter FM, Thompson MJ, Wellwood I, Abel GA, Hamilton W, Johnson M, et al. Evaluating diagnostic strategies for early detection of cancer: the CanTest framework. BMC Cancer 2019 Jun 14;19(1):586 [FREE Full text] [doi: 10.1186/s12885-019-5746-6] [Medline: 31200676]
15. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
16. Home page. Melanoma Focus. URL: https://melanomafocus.org/ [accessed 2024-12-16]
17. Home page. Primary Care Dermatology Society (PCDS). URL: https://www.pcds.org.uk/ [accessed 2024-12-16]
18. Medical market research surveys. Sermo Market Research. URL: https://www.sermo.com/resources/business-market-research/ [accessed 2024-12-16]
19. Home page. British Association of Dermatologists (BAD). URL: https://www.bad.org.uk/ [accessed 2024-12-16]
20. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
21. Home page. Lumivero. URL: https://lumivero.com/ [accessed 2024-12-16]
22. Microsoft Excel. Microsoft Corporation. 2018. URL: https://office.microsoft.com/excel [accessed 2024-12-16]
23. Question development: how we researched and tested the questions for the questionnaire. Office of the National Statistics. URL: https://www.ons.gov.uk/aboutus/whatwedo [accessed 2024-04-29]
24. Menzies SW, Emery J, Staples M, Davies S, McAvoy B, Fletcher J, et al. Impact of dermoscopy and short-term sequential digital dermoscopy imaging for the management of pigmented lesions in primary care: a sequential intervention trial. Br J Dermatol 2009 Dec;161(6):1270-1277 [FREE Full text] [doi: 10.1111/j.1365-2133.2009.09374.x] [Medline: 19747359]
25. Welch HG, Mazer BL, Adamson AS. The rapid rise in cutaneous melanoma diagnoses. N Engl J Med 2021 Jan 07;384(1):72-79. [doi: 10.1056/nejmsb2019760]
26. Verbakel JY, Turner PJ, Thompson MJ, Plüddemann A, Price CP, Shinkins B, et al. Common evidence gaps in point-of-care diagnostic test evaluation: a review of horizon scan reports. BMJ Open 2017 Sep 01;7(9):e015760 [FREE Full text] [doi: 10.1136/bmjopen-2016-015760] [Medline: 28864692]
27. Nadav J, Kaihlanen AM, Kujala S, Laukka E, Hilama P, Koivisto J, et al. How to implement digital services in a way that they integrate into routine work: qualitative interview study among health and social care professionals. J Med Internet Res 2021 Dec 01;23(12):e31668 [FREE Full text] [doi: 10.2196/31668] [Medline: 34855610]
28. Plsek PE, Greenhalgh T. Complexity science: the challenge of complexity in health care. BMJ 2001 Sep 15;323(7313):625-628 [FREE Full text] [doi: 10.1136/bmj.323.7313.625] [Medline: 11557716]
29. Frisinger A, Papachristou P. The voice of healthcare: introducing digital decision support systems into clinical practice - a qualitative study. BMC Prim Care 2023 Mar 13;24(1):67 [FREE Full text] [doi: 10.1186/s12875-023-02024-6] [Medline: 36907875]
30. Buck C, Doctor E, Hennrich J, Jöhnk J, Eymann T. General practitioners' attitudes toward artificial intelligence-enabled systems: interview study. J Med Internet Res 2022 Jan 27;24(1):e28916 [FREE Full text] [doi: 10.2196/28916] [Medline: 35084342]
31. Haggenmüller S, Krieghoff-Henning E, Jutzi T, Trapp N, Kiehl L, Utikal JS, et al. Digital natives' preferences on mobile artificial intelligence apps for skin cancer diagnostics: survey study. JMIR Mhealth Uhealth 2021 Aug 27;9(8):e22909 [FREE Full text] [doi: 10.2196/22909] [Medline: 34448722]
32. Jutzi TB, Krieghoff-Henning EI, Holland-Letz T, Utikal JS, Hauschild A, Schadendorf D, et al. Artificial intelligence in skin cancer diagnostics: the patients' perspective. Front Med (Lausanne) 2020;7:233 [FREE Full text] [doi: 10.3389/fmed.2020.00233] [Medline: 32671078]
33. Venkatesan P. Artificial intelligence and cancer diagnosis: caution needed. Lancet Oncol 2021 Oct;22(10):1364 [FREE Full text] [doi: 10.1016/S1470-2045(21)00533-7] [Medline: 34509184]
34. Samaran R, L'Orphelin JM, Dreno B, Rat C, Dompmartin A. Interest in artificial intelligence for the diagnosis of non-melanoma skin cancer: a survey among French general practitioners. Eur J Dermatol 2021 Aug 01;31(4):457-462. [doi: 10.1684/ejd.2021.4090] [Medline: 34642138]
35. Bakken S. AI in health: keeping the human in the loop. J Am Med Inform Assoc 2023 Jun 20;30(7):1225-1226 [FREE Full text] [doi: 10.1093/jamia/ocad091] [Medline: 37337923]
36. Zha D, Bhat ZP, Lai KH, Yang F, Hu X. Data-centric AI: perspectives and challenges. arXiv. Preprint posted online January 12, [FREE Full text] [doi: 10.1137/1.9781611977653.ch106]
37. Gilbert S, Anderson S, Daumer M, Li P, Melvin T, Williams R. Learning from experience and finding the right balance in the governance of artificial intelligence and digital health technologies. J Med Internet Res 2023 Apr 14;25:e43682 [FREE Full text] [doi: 10.2196/43682] [Medline: 37058329]
38. Guni A, Whiting P, Darzi A, Ashrafian H. The next generation of evidence synthesis for diagnostic accuracy studies in artificial intelligence. Lancet Digit Health 2024 Aug;6(8):e541-e542 [FREE Full text] [doi: 10.1016/S2589-7500(24)00115-8] [Medline: 38926009]

39. Muehlematter UJ, Bluethgen C, Vokinger KN. FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. Lancet Digit Health 2023 Sep;5(9):e618-e626 [FREE Full text] [doi: 10.1016/S2589-7500(23)00126-7] [Medline: 37625896]
40. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD). US Food and Drug Administration. URL: https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf [accessed 2024-12-16]
41. Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. European Commission, Directorate-General for Communications Networks, Content and Technology. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206 [accessed 2024-12-16]
42. Blueprint for an AI bill of rights: making automated systems work for? The American people. White House Office of Science and Technology Policy. 2022. URL: https://www.whitehouse.gov/ostp/ai-bill-of-rights/ [accessed 2024-12-16]
43. AI regulation: a pro-innovation approach. Department for Science, Innovation and Technology and Office for Artificial Intelligence. URL: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach [accessed 2024-12-16]
44. An NHS under pressure. British Medical Association. URL: https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/an-nhs-under-pressure [accessed 2024-12-16]
45. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. BMJ 2016 Jun 22;353:i3139 [FREE Full text] [doi: 10.1136/bmj.i3139] [Medline: 27334281]
46. Recommendations for diversity, inclusivity, and generalisability in artificial intelligence health technologies and health datasets. Standing Together. URL: https://www.datadiversity.org/recommendations [accessed 2024-12-16]

Abbreviations

AI: artificial intelligence
BAD: British Association of Dermatologists
COREQ: Consolidated Criteria for Reporting Qualitative Research
GP: general practitioner
HCP: health care professional
KC: keratinocyte cancer
ML: machine learning
NASSS: Nonadoption, Abandonment, and Challenges to Scale-Up, Spread, and Sustainability
NP: nurse practitioner
PCP: primary care practitioner
PPI: Patient and Public Involvement

Edited by N Cahill; submitted 20.05.24; peer-reviewed by J Makin, Z Su; comments to author 18.06.24; revised version received 24.07.24; accepted 29.10.24; published 28.01.25.

Please cite as:
Jones OT, Calanzani N, Scott SE, Matin RN, Emery J, Walter FM
User and Developer Views on Using AI Technologies to Facilitate the Early Detection of Skin Cancers in Primary Care Settings: Qualitative Semistructured Interview Study
JMIR Cancer 2025;11:e60653
URL: https://cancer.jmir.org/2025/1/e60653
doi:10.2196/60653
PMID:

Original Paper

Social Media as a Platform for Cancer Care Decision-Making Among Women: Internet Survey-Based Study on Trust, Engagement, and Preferences

Anna Rose Johnson[1], MPH, MD; Grace Anne Longfellow[2], BS; Clara N Lee[2], MPP, MD; Benjamin Ormseth[3], MD; Gary B Skolnick[1], MBA; Mary C Politi[4], PhD; Yonaira M Rivera[5], MPH, PhD; Terence Myckatyn[1], MD, FRSC(C)

[1]Division of Plastic and Reconstructive Surgery, Department of Surgery, Washington University School of Medicine, Saint Louis, MO, United States
[2]Department of Surgery, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[3]Department of Plastic Surgery, University of Texas Southwestern Medical Center, Dallas, TX, United States
[4]Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, Saint Louis, MO, United States
[5]Department of Communication, School of Communication & Information, Rutgers University, New Brunswick, NJ, United States

Corresponding Author:
Terence Myckatyn, MD, FRSC(C)
Division of Plastic and Reconstructive Surgery
Department of Surgery
Washington University School of Medicine
1020 N. Mason Rd
Ste 110
Saint Louis, MO, 63141
United States
Phone: 1 3149968800
Fax: 1 314 996 3030
Email: myckatyn@wustl.edu

Abstract

Background: Decision aids improve patient and clinician decision-making but are underused and often restricted to clinical settings.

Objective: Given limited studies analyzing the feasibility of disseminating decision aids through social media, this study aimed to evaluate the acceptability, trust, and engagement of women with social media as a tool to deliver online decision aids for cancer treatment.

Methods: To prepare for potential dissemination of a breast cancer decision aid via social media, a cross-sectional survey in February 2023 was conducted via Prime Panels, an online market research platform, of women aged 35-75 years in the United States. Demographics, health, cancer information-seeking behaviors, social media use, trust in social media for health information, as well as the likelihood of viewing cancer-related health information and clicking on decision aids through social media, were assessed. Statistical analyses included descriptive statistics, correlations, and multivariable ordinal regression.

Results: Of 607 respondents, 397 (65.4%) had searched for cancer information, with 185 (46.6%) using the internet as their primary source. Facebook (Meta) was the most popular platform (511/607, 84.2%). Trust in social media for health information was higher among Black (14/72, 19.4%) and Asian respondents (7/27, 25.9%) than among White respondents (49/480, 10.2%; *P*=.003). Younger respondents aged 35-39 years (17/82, 20.7%) showed higher trust than those aged 70-79 years (12/70, 17.1%; *P*<.001). Trust in social media for health information was linked to a higher likelihood of viewing cancer information and accessing a decision aid online (*P*<.001). Participants who rated social media as “Trustworthy” (n=73) were more likely to view cancer information (61/73, 83.6%) and click on decision aids (61/73, 83.6%) than those who found it “Untrustworthy” (n=277; view: 133/277, 48.0%; click: 125/277, 45.1%). Engagement with social media positively correlated with viewing online cancer information (Spearman ρ=0.20, *P*<.001) and willingness to use decision aids (ρ=0.21, *P*<.*001*). Multivariable ordinal regression analyses confirmed that perception of social media’s trustworthiness is a significant predictor of engagement with decision aids (untrustworthy vs trustworthy β=–1.826, *P*<.001; neutral vs trustworthy β=–0.926, *P*=.007) and of viewing cancer information (untrustworthy vs trustworthy β=–1.680, *P*<.001, neutral vs trustworthy β=–0.581, *P*=.098), while age and employment status were not significant predictors.

Conclusions: This exploratory study suggests that social media platforms may increase access to health information and decision aids. No significant differences were observed between demographic variables and the use or trust in social media for health information. However, trust in social media emerged as a mediating factor between demographics and engagement with cancer information online. Before disseminating decision aids on social media, groups should identify existing trust and engagement patterns with different platforms within their target demographic.

(JMIR Cancer 2025;11:e64724) doi:10.2196/64724

KEYWORDS

shared decision-making; SDM; decision aids; cancer treatment; breast cancer; digital health; social media; health communication; online decision aids; health information-seeking behavior; trust in health information; healthcare accessibility; mhealth

Introduction

An estimated 2,001,140 new cancer cases are expected in 2024 [1]. Shared decision-making (SDM) describes a process between the clinician and patient to facilitate preference-sensitive choices [2]. Decision aids, which can support the SDM process, are evidence-based tools designed to provide patients with information, clarify their preferences, and prepare them to make a choice [3,4]. In this study, we explore the potential of social media as an avenue for engagement with decision-making tools.

SDM has been shown to be important for cancer decision-making, with multiple randomized controlled trials demonstrating that decision aids improve patient knowledge and the quality of decisions [3-11]. Unfortunately, decision aid use has been limited, and their dissemination has been largely confined to clinical settings. A 2010 study revealed that only 24% of clinicians working with patients with cancer used decision aids [12-18]. The focus on clinical settings as the singular forum for decision aid deployment is predicated on clinician buy-in and may restrict the use of decision aids to a select cohort of the population [12-19].

Social media offers a promising means for disseminating decision aids without relying on health care access. It may also provide a more extended and personalized modality for disseminating information [20]. With 81% of Americans using social media, a number that continues to grow, social media platforms present an underused opportunity to disseminate highly accessible decision-making tools [21]. Social media can help to overcome challenges associated with traditional clinical encounters (ie, time, workflow, etc) and can enhance the patient-clinician relationship by promoting empowerment, reducing communication barriers, and increasing knowledge about conditions and treatment options [14,22].

Numerous studies have highlighted the positive impact of web-based decision aids for women, particularly in the context of breast cancer, the leading cause of cancer among females [23]. Despite the potential benefits of social media for decision aid dissemination, it remains uncertain whether females will use cancer-related decision aids available through social media or other online channels. To address this gap, we examined factors influencing engagement with decision aids on social media and explored health information-seeking behaviors across various platforms. This study aimed to evaluate the feasibility, acceptability, trust, and engagement with social media as a tool to deliver online decision aids to women for cancer treatment. By focusing specifically on women, we aimed to address the unique health and decision-making needs of this population and provide insight for future research on breast cancer-related decision aids.

Methods

Survey Design

A cross-sectional survey was designed to assess the use and preferences for social media–advertised decision-making tools for cancer care. The survey involved several key areas of inquiry (Multimedia Appendix 1). Briefly, these areas included “Health-Related Information Behaviors” (5 questions) to assess participants’ behaviors in seeking and using health information; “Sources of Health Information” (3 questions), exploring individuals’ preferences and trust levels in various health information sources; “Social Media Use” (18 questions), examining patterns and motivations behind social media interactions, particularly concerning health information; and “Demographic Data” (10 questions), covering a wide range of personal and socioeconomic factors. Within the “Social Media Use” section, two items related to the main outcomes of the study were embedded, created by the study team, which asked participants to imagine themselves or a loved one deciding about cancer treatment and then assessing their likelihood of viewing cancer treatment information or clicking on a decision aid posted on social media. Other survey questions were adapted from items from the Health Information National Trends Survey [24]. Items assessing reasons to use social media included categories identified in the literature through the uses and gratifications theory and the social media engagement model [25-27].

The response formats varied according to the specific inquiry, including multiple-choice options, checkboxes for applicable answers, and Likert scales for assessing attitudes and opinions. Although all 39 questions could be answered, branching logic was used to tailor the survey based on participants’ responses (eg, only those reporting the use of specific social media platforms were asked follow-up questions about their motivations for use). The survey was designed to be completed within 5 to 10 minutes and included 2 attention-check questions. One single question, with the associated branching logic, was shown on the screen at a time. Participants were unable to skip questions (except for the questions asking the frequency of use and reason for use of social media platforms) and were notified, “Please answer this question” if attempted to skip. Participants were able to go back and change their answers if desired. The order of survey items, and answers, was fixed and not

randomized, as the survey design prioritized logical flow and ease of navigation for participants. This study is reported in a manner that is consistent with the specified Checklist for Reporting Results of Internet E-Surveys (CHERRIES) guidelines (Multimedia Appendix 2) [28].

Population Targeting and Survey Distribution

The survey was designed and hosted on the Qualtrics platform (Qualtrics, Provo) and distributed in February 2023 via Prime Panels, a component of Cloud Research [29]. Prime Panels uses a novel data collection method by aggregation of diverse opt-in market research panels into a comprehensive sampling platform, facilitating the recruitment of participants from existing commercial panel pools. This method supports demographic quotas and specific eligibility criteria, enhancing data representativeness, especially among hard-to-reach populations. Eligible participants were invited to participate through targeted email and dashboard invitations sent by the market research panels within the Prime Panels network, based on the demographic criteria specified for the study [30]. Participants were required to complete the survey in a single sitting, and no reminders were sent to those who did not finish it during that session. Due to the wide distribution on several platforms, response rates and the total number of invitations sent were not calculated by Prime Panels.

We aimed to gather a representative sample of United States females aged 35-75 years for this study, as this age range represents the peak period for breast cancer diagnosis. The study exclusively enrolled female participants to direct focus toward future research efforts related to breast cancer in women. Using 2022 US Census data, the population of females aged 35-75 years were inputted into the Qualtrics sample size calculator with a 99% CI and a 5% margin of error to determine the required sample size [31]. Based on these calculations, the survey targeted approximately 660 female participants, with an additional 15% included to account for potential exclusions due to poor response quality, bringing the total recruitment goal to 750 participants. Study specific eligibility criteria incorporated into the Prime Panels query included female participants aged 35 years or older with US IP addresses. Demographic quotas based on United States Census Bureau parameters were set as: 16% Hispanic, Latino, or Spanish origin; 78% White; 13% Black; 5% Asian; 2% American Indian or Alaskan Native; and 2% other races. Age quotas aimed for an equal distribution between the 35-55 years and 56-75 years age ranges to reflect the demographics of breast cancer survivors. To ensure survey security, Qualtrics options for “Bot Detection,” “RelevantID,” and “Prevent Indexing” were enabled.

Statistical Analyses

Data analysis was performed using Microsoft Excel and SPSS Statistics (version 28.0; IBM Corp). Graphs were constructed via R Statistical Software (version 4.1.2; R Core Team 2021). Descriptive statistics characterized the demographic characteristics, health-seeking behaviors, and social media engagement of the study population. Social media engagement was determined based on respondents’ selections of platforms they actively used, followed by survey questions that assessed the frequency of engagement with each platform. These questions categorized usage frequency into 4 levels: multiple times a day, once a day, at least 3 times a week, or less than 3 times a week. Numerical values ranging from 1 to 4 were assigned to these categories, with 1 indicating the least frequent usage and 4 the most frequent. An “Overall Social Media Engagement Score” was computed by aggregating these values across all platforms used by a respondent. Participants were classified into 4 groups based on their “Overall Social Media Engagement Score” to approximate quartiles for analysis. These groups were defined as follows: “Low Engagement” (scores of >0 and ≤3), “Moderate Engagement” (scores of >3 and ≤4), “Moderate-High Engagement” (scores of >4 and ≤8), and “High Engagement” (scores of >8 and ≤23).

To ensure data integrity, surveys were excluded based on the following criteria: completion times shorter than 3 minutes or longer than 20 minutes, flags from Qualtrics indicating duplicate responses, an Amazon Mechanical Turk fraud score above 50, or incorrect responses to two embedded control multiple-choice questions.

Analysis of Trustworthiness of Social Media

To evaluate associations between demographic factors and the perceived trustworthiness of social media as a reliable source of health information, chi-square, and Fisher’s exact tests were performed. To ensure that there were enough observations in each category for the statistical analysis to be reliable, the response categories “Trustworthy” and “Very Trustworthy” were merged into a single “Trustworthy” category, while “Untrustworthy” and “Very Untrustworthy” were combined into “Untrustworthy.”

Analysis of Social Media Engagement

In terms of the two questions assessing the likelihood of engaging with cancer treatment-related information seen on social media, responses were condensed from a 7-tier scale to 3 categories: “Unlikely” (1-3), “Neutral” (4), and “Likely” (5-7) to ensure a more balanced distribution of responses, as some of the original categories had very few observations. Spearman rank correlation coefficients were calculated to quantify the strength and direction of the association between the “Overall Social Media Engagement Score” and the tiered scores representing the likelihood of interacting with cancer-related information. For visual interpretation, the mean likelihood of respondents interacting with cancer-related information was calculated for each unique “Overall Social Media Engagement Score.” For all analyses, statistical significance was set at a P value of less than .05, using 2-tailed testing.

Analysis of Likelihood to View Cancer-Related Health Information or Click on Decision Aid

Nonparametric tests, specifically the Kruskal-Wallis and Mann-Whitney U tests, were used to evaluate the relationships between demographic characteristics, trust in social media, and the propensity to use decision aids or view cancer-related health information on these platforms. Variables that were found to be significantly related to the use of decision aids or viewing health information at $P \leq .10$ were then checked for multicollinearity via variance inflation factor (VIF) values <5 before inclusion in an ordinal regression model.

Ethical Considerations

This study was reviewed and received approval from the institutional review board at Ohio State University as exempt (protocol 2022E0836). Informed consent was obtained from all participants involved in the study. Participant data were collected anonymously, with no identifying information retained in the dataset. The original informed consent included provisions for the use of deidentified data for research purposes, as reviewed and approved by the institutional review board. Data were stored in a secure, password-protected database accessible only to study investigators. All participants were compensated by Prime Panels in the amount agreed to by the platform through which they entered the survey, which is unknown to study personnel.

Results

A total of 757 responses were initially recorded at the completion of distribution. Of these, 607 met inclusion criteria with a Qualtrics "Response Quality" of 99.0%. Participants completed the survey in a mean SD time of 5.5 (SD 2.6) minutes.

Respondent Demographics

All participants were female, aged 35-75 years (Table 1). Of the 607 respondents, most were non-Hispanic (556/607, 91.6%) and White (480/607, 79.1%). The most common education level was some college or an associate degree (201/607, 33.1%). The most common income range was US $20,000 to US $35,000 (119/607, 18.9%), with over half (327/607, 53.9%) earning less than US $50,000 annually.

Table 1. Demographic characteristics of respondents.

Characteristics	Respondents (N=607), n (%)
Ethnicity	
Hispanic	48 (7.9)
Non-Hispanic	556 (91.6)
Unknown or prefer not to answer	3 (0.5)
Race	
White	480 (79.1)
Black	72 (11.9)
Asian	27 (4.4)
Native American or Alaskan Native	16 (2.6)
Native Hawaiian or Pacific Islander	0 (0.0)
Other or prefer not to answer	11 (1.8)
Age (years)	
35-39	82 (13.5)
40-49	141 (23.2)
50-59	149 (24.5)
60-69	165 (27.2)
70-79	70 (11.5)
Highest level of education achieved	
High school diploma, GED[a], or less	175 (28.8)
Technical training or certificate	46 (7.6)
Some college or associate degree	201 (33.1)
College degree	111 (18.3)
Graduate or professional degree	74 (12.2)
Annual household income (US $)	
<20,000	109 (18.0)
20,000-35,000	115 (18.9)
35,000-50,000	103 (17.0)
50,000-75,000	109 (18.0)
75,000-100,000	73 (12.0)
>100,000	82 (13.5)
Unknown or prefer not to answer	16 (2.6)
Insurance type	
Private	229 (37.7)
Government	301 (49.6)
Uninsured	43 (7.1)
Other	8 (1.3)
More than 1 insurance policy	26 (4.3)
Relationship status	
Single	129 (21.3)
Married	268 (44.2)
Separated or divorced	150 (24.7)
Widowed	57 (9.4)

Characteristics	Respondents (N=607), n (%)
Other or unknown or prefer not to answer	3 (0.5)
Employment status	
Full-time	207 (34.1)
Part-time	71 (11.7)
Not working for pay or unemployed	121 (19.9)
Retired	200 (32.9)
Other or unknown or prefer not to say	8 (1.3)
Country of birth	
United States	555 (91.4)
Outside of the United States	52 (8.6)
Years of US residency	
Less than 15 years	20 (3.3)
More than 15 years	587 (96.7)

[a]GED: graduate educational diploma.

Health-Seeking Behavior

In total, 551 out of 607 participants (90.8%) had sought health or medical information from various sources at some point (Table 2). Out of 551 respondents, the internet was the most common first source of health information (n=441/551, 80.0%), while 75 or 13.6% consulted a doctor or health care provider.

Table 2. Characteristics of health-seeking behaviors.

Question	Respondents (N=607), n (%)
Have you ever looked for information about health or medical topics from any source?	551 (90.8)
The most recent time you looked for information about health or medical topics, where did you go first?	551 (100)
Doctor or health care provider	75 (13.6)
Internet	441 (80)
Brochure or pamphlet, etc.	10 (1.8)
Friend or coworker	3 (0.5)
Family	11 (2)
Cancer organization	2 (0.4)
Newspapers	1 (0.2)
Books	5 (0.9)
Library	2 (0.4)
Telephone information number	1 (0.2)
The most recent time you looked for information about health or medical topics, who was it for?	551 (100)
Self	408 (74)
Someone else	72 (13.1)
Both oneself and someone else	71 (12.9)
Which of the following sources have you used in the last month as a source of news or information about health topics?[a]	592 (100)
Blogs or personal websites	72 (12.2)
Center for disease control and prevention	133 (22.5)
World Health Organization	63 (10.6)
Government	46 (7.8)
Community or faith leaders	19 (3.2)
Online news	256 (43.2)
Email	48 (8.1)
Family and friends	204 (34.5)
Health professionals	282 (47.6)
Radio	22 (3.7)
Podcasts	27 (4.6)
TV	69 (11.7)
Social media	119 (20.1)
Print media	46 (7.8)
Video sharing sites	53 (9)
Have you ever looked for information about cancer from any source?	607 (100)
Yes	397 (65.4)
No	210 (34.6)
In the past 12 months, have you used the internet to look for cancer information for yourself?	397 (100)
Yes	185 (46.6)
No	212 (53.4)
Where do you access your social media accounts?[a]	603 (100)
Computer or laptop	258 (42.8)
iPad or tablet	159 (26.4)

Question	Respondents (N=607), n (%)
Smartphone	497 (82.4)

[a]Participants were able to check all that apply.

Social Media Use and Engagement

In total, 80 out of 607 or 95.6% of respondents used social media. Of these, Facebook was the most popular platform, used by 511 or 84.2%, and was used primarily for social interactions by 338 out of 487 respondents (69.4%). YouTube (Alphabet Inc) and Instagram (Meta) were primarily used for entertainment (189 out of 251 or 75.3% and 110 out of 193 or 57.0%, respectively). 18.5%, or 112 out of 607 respondents, demonstrated a “Low Engagement” pattern regarding social media use.

Trustworthiness of Social Media

The majority of the 607 respondents found social media trustworthy (73/607, 12.0%) or neutral (257/607, 42.3%) for health information. Black or Asian race, younger age, and longer duration of US residency were associated with greater trust in social media. Among Black respondents, 14 out of 72 (19.4%) considered social media trustworthy, compared to 49 out of 480 (10.2%) of White respondents (P=.003). Asian respondents showed even higher trust levels, with 7 out of 27 (25.9%) rating social media as trustworthy. Younger individuals also reported greater trust, with 17 out of 82 (20.7%) of those aged 35-39 years trusting social media compared with 12 out of 70 (17.1%) among those aged 70-79 years (P<.001). In addition, respondents with longer US residency (more than 15 years) showed greater trust, with 272 out of 587 (46.3%) indicating trustworthiness in social media, compared with only 5 out of 20 (25.0%) of those with less than 15 years of residency (P=.004). In total, 277 respondents (45.6%) noted social media to be untrustworthy (Table 3).

Table 3. Factors associated with perceived trustworthiness of social media as a source for health information.

Factors	Trustworthy	Neutral	Untrustworthy	Respondents, n (%)	*P* value
Total	73 (12)	257 (42.3)	277 (45.63)	607 (100)	N/A[a]
Ethnicity					.14
Non-Hispanic	64 (11.5)	233 (41.9)	259 (46.58)	556 (100)	
Hispanic	9 (18.8)	23 (47.9)	16 (33.33)	48 (100)	
Unknown or prefer not to answer	0 (0.08)	1 (33.3)	2 (66.67)	3 (100)	
Race					.003[b]
White	49 (10.2)	194 (40.4)	237 (49.4)	480 (100)	
Black	14 (19.4)	32 (44.4)	26 (36.1)	72 (100)	
Asian	7 (25.9)	15 (55.6)	5 (18.5)	27 (100)	
Native American or Alaskan Native	2 (12.5)	7 (43.8)	7 (43.8)	16 (100)	
Native Hawaiian or other Pacific Islander	0 (0)	0 (0)	0 (0)	0 (100)	
Other or prefer not to answer	1 (8.3)	9 (75)	2 (16.7)	12 (100)	
Age (years)					<.001[b]
35-39	17 (20.7)	33 (40.2)	32 (39)	82 (100)	
40-49	20 (14.2)	68 (48.2)	53 (37.6)	141 (100)	
50-59	13 (8.7)	71 (47.7)	65 (43.6)	149 (100)	
60-69	11 (6.7)	67 (40.6)	87 (52.7)	165 (100)	
70-79	12 (17.1)	18 (25.7)	40 (57.1)	70 (100)	
Highest level of formal education achieved					.40
High school diploma, GED[c], or less	24 (13.7)	81 (46.3)	70 (40)	175 (100)	
Technical training or certificate	5 (10.9)	19 (41.3)	22 (47.8)	46 (100)	
Some years of college or associate degree	23 (11.4)	92 (45.8)	86 (42.8)	201 (100)	
College degree	12 (10.8)	40 (36)	59 (53.2)	111 (100)	
Graduate or professional degree	9 (12.2)	25 (33.7)	40 (54.1)	74 (100)	
Annual household income (US $)					.63
<20,000	11 (10.1)	53 (48.6)	45 (41.3)	109 (100)	
20,000-35,000	17 (14.8)	48 (41.7)	50 (43.5)	115 (100)	
35,000-50,000	15 (14.6)	47 (45.6)	41 (39.8)	103 (100)	
50,000-75,000	12 (11)	43 (39.5)	54 (49.5)	109 (100)	
75,000-100,000	10 (13.7)	24 (32.9)	39 (53.4)	73 (100)	
>100,000	8 (9.8)	35 (42.7)	39 (47.6)	82 (100)	
Unknown or prefer not to answer	0 (0)	7 (43.8)	9 (56.3)	16 (100)	
Insurance type					.16
Private	28 (12.2)	98 (42.8)	103 (45)	229 (100)	
Government	42 (14)	123 (40.9)	136 (45.2)	301 (100)	
Uninsured	1 (2.3)	24 (55.8)	18 (41.9)	43 (100)	
Other insurance or more than 1 policy	2 (5.9)	12 (35.3)	20 (58.8)	34 (100)	
Relationship status					.79
Single	15 (11.6)	58 (45)	56 (43.4)	129 (100)	
Married	33 (12.3)	102 (38.1)	133 (49.6)	268 (100)	
Separated or divorced	18 (12)	70 (46.6)	62 (41.3)	150 (100)	

Factors	Trustworthy	Neutral	Untrustworthy	Respondents, n (%)	*P* value
Widowed	7 (12.3)	25 (43.9)	25 (43.9)	57 (100)	
Other or unknown or prefer not to answer	0 (0)	2 (66.7)	1 (33.3)	3 (100)	
Employment status					.09
Full-time	22 (10.6)	96 (46.4)	89 (43)	207 (100)	
Part-time	10 (14.1)	31 (43.7)	30 (42.3)	71 (100)	
Not working for pay or unemployed	17 (14.1)	58 (47.9)	46 (38)	121 (100)	
Retired	24 (12)	68 (34.0)	108 (54)	200 (100)	
Other or unknown or prefer not to say	0 (0)	1 (50)	1 (50)	2 (100)	
Country of birth					.28
United States	64 (11.5)	233 (41.98)	258 (46.49)	555 (100)	
Outside of the United States	9 (17.3)	24 (46.15)	19 (36.54)	52 (100)	
Years of US residency					.004[b]
Less than 15 years	7 (35)	8 (40)	5 (25)	20 (100)	
More than 15 years	66 (11.2)	249 (42.4)	272 (46.3)	587 (100)	
Platforms used (multiple selections allowed)					N/A
Facebook	65 (12.7)	220 (43.1)	226 (44.2)	511 (100)	
Twitter (rebranded as X)	23 (16.9)	65 (47.8)	48 (35.3)	136 (100)	
Instagram	35 (12.6)	125 (45.1)	117 (42.2)	277 (100)	
YouTube	53 (13)	179 (45.4)	162 (41.1)	394 (100)	
WhatsApp (Meta)	17 (21.3)	34 (42.5)	29 (36.3)	80 (100)	
TikTok (ByteDance)	22 (12.9)	85 (49.7)	64 (37.4)	171 (100)	
Other or unknown	4 (13.3)	10 (33.3)	16 (53.3)	30 (100)	

[a]N/A: not applicable.

[b]$P<.05$.

[c]GED: graduate educational diploma.

Among social media platforms, the highest proportion of trustworthy users was noted among the 80 WhatsApp users (n=17, 21.3%), followed by 23 out of the 136 (16.9%) Twitter users. Amongst the 511 respondents who used Facebook, the most frequently used platform, 65 or 12.7% reported trust in social media for health information.

Use of Cancer Information or Decision Aids Through Social Media

Participants who considered social media "Trustworthy" (n=73) were more likely to view cancer information (n=61, 83.6%) or click on a decision aid through social media (n=61, 83.6%) than the 277 respondents who viewed social media as "Untrustworthy" (view: n=133, 48.0%; click: n=125, 45.1%) (Tables 4 and 5). Younger participants, particularly those aged 35-39 years were more likely to view cancer-related information through social media. Only 10 out of 57 (12.2%) in the 35-39 years age group rated their likelihood as "unlikely," compared with 54 out of 89 (32.7%) aged 60-69 years. Among respondents aged 35-39 years, 54 out of 82 (65.9%) were likely to click on the decision aid, while in the 60-69 years age group, 87 out of 165 (52.7%) indicated they were likely to click.

Table 4. Nonparametric analysis of factors influencing viewing of cancer-related health information on social media: social media trustworthiness and demographic insights.

Variable	Likelihood of viewing cancer-related health information seen on social media			Respondents, n (%)	*P* value
	Unlikely	Neutral	Likely		
Total	141 (23.23)	82 (13.51)	384 (63.26)	607 (100)	N/A[a]
Trustworthiness of social media					<.001[b]
Untrustworthy	99 (35.7)	45 (16.2)	133 (48)	277 (100)	
Neutral	36 (14)	31 (12.1)	190 (73.9)	257 (100)	
Trustworthy	6 (8.2)	6 (8.2)	61 (83.6)	73 (100)	
Ethnicity					.86
Hispanic	130 (23.4)	75 (13.5)	351 (63.1)	556 (100)	
Non-Hispanic	11 (22.9)	6 (12.5)	31 (64.6)	48 (100)	
Unknown or prefer not to answer	0 (0)	1 (33.3)	2 (66.7)	3 (100)	
Race					.17
White	110 (22.9)	68 (14.2)	302 (62.9)	480 (100)	
Black	23 (31.9)	9 (12.5)	40 (55.6)	72 (100)	
Asian	4 (14.8)	2 (7.4)	21 (77.8)	27 (100)	
Native American or Alaskan Native	2 (12.5)	3 (18.8)	11 (68.8)	16 (100)	
Native Hawaiian or Pacific Islander	0 (0)	0 (0)	0 (0)	0 (0)	
Other or prefer not to answer	2 (16.7)	0 (0)	10 (83.3)	12 (100)	
Age (years)					.008[b]
35-39	10 (12.2)	15 (18.3)	57 (69.5)	82 (100)	
40-49	27 (19.1)	20 (14.2)	94 (66.7)	141 (100)	
50-59	31 (20.8)	15 (10.1)	103 (69.1)	149 (100)	
60-69	54 (32.7)	22 (13.3)	89 (53.9)	165 (100)	
70-79	19 (27.1)	10 (14.3)	41 (58.6)	70 (100)	
Highest level of education attained					.65
High school diploma, GED[c], or less	36 (20.6)	25 (14.3)	114 (65.1)	175 (100)	
Technical training or certificate	11 (23.9)	8 (17.4)	27 (58.7)	46 (100)	
Some college or associates degree	43 (21.4)	27 (13.4)	131 (65.2)	201 (100)	
College degree	29 (26.1)	13 (11.7)	69 (62.2)	111 (100.0)	
Graduate or professional degree	22 (29.7)	9 (12.2)	43 (58.1)	74 (100)	
Annual household income (US $)					.83
<20,000	25 (22.9)	20 (18.3)	64 (58.7)	109 (100)	
20,000-35,000	23 (20)	16 (13.9)	76 (66.1)	115 (100)	
35,000-50,000	29 (28.2)	8 (7.8)	66 (64.1)	103 (100.0)	
50,000-75,000	24 (22)	18 (16.5)	67 (61.5)	109 (100)	
75,000-100,000	20 (27.4)	7 (9.6)	46 (63)	73 (100)	
>100,000	16 (19.5)	10 (12.2)	56 (68.3)	82 (100)	
Unknown or prefer not to answer	4 (25)	3 (18.8)	9 (56.2)	16 (100)	
Insurance type					.39
Private	50 (21.8)	28 (12.2)	151 (65.9)	229 (100)	
Government	74 (24.6)	40 (13.3)	187 (62.1)	301 (100)	

Variable	Likelihood of viewing cancer-related health information seen on social media			Respondents, n (%)	*P* value
	Unlikely	Neutral	Likely		
Uninsured	6 (14)	9 (20.9)	28 (65.1)	43 (100)	
Other insurance or more than 1 policy	11 (32.4)	5 (14.7)	18 (52.9)	34 (100)	
Relationship status					.29
Single	34 (26.4)	19 (14.7)	76 (58.9)	129 (100)	
Married	56 (20.9)	32 (11.9)	180 (67.2)	268 (100)	
Separated or divorced	36 (24)	20 (13.3)	94 (62.7)	150 (100)	
Widowed	15 (26.3)	10 (17.5)	32 (56.1)	57 (100)	
Other or unknown or prefer not to answer	0 (0)	1 (33.3)	2 (66.7)	3 (100)	
Employment status					.01[b]
Full-time	37 (17.9)	32 (15.5)	138 (66.7)	207 (100)	
Part-time	19 (26.8)	8 (11.3)	44 (62)	71 (100)	
Not working for pay or unemployed	26 (21.5)	7 (5.8)	88 (72.7)	121 (100)	
Retired	57 (28.5)	34 (17)	109 (54.5)	200 (100)	
Other or unknown or prefer not to say	2 (25)	1 (12.5)	5 (62.5)	8 (100)	
Country of birth					.55
United States	128 (23.1)	79 (14.2)	348 (62.7)	555 (100)	
Outside of the United States	13 (25)	3 (5.8)	36 (69.2)	52 (100)	
Length of US residency					.27
Less than 15 years	3 (15)	2 (10)	15 (75)	20 (100)	
More than 15 years	138 (23.5)	80 (13.6)	369 (62.9)	587 (100)	

[a]N/A: not applicable.

[b]$P<.05$.

[c]GED: graduate educational diploma.

Table 5. Nonparametric analysis of factors influencing clicking on a decision aid seen on social media: social media trustworthiness and demographic insights.

Variable	Likelihood of clicking on a decision aid seen on social media			Respondents, n (%)	*P* value
	Unlikely	Neutral	Likely		
Total	151 (24.88)	101 (16.64)	355 (68.48)	607 (100)	N/A[a]
Trustworthiness of social media					<.001[b]
Untrustworthy	103 (37.2)	49 (17.7)	125 (45.1)	277 (100)	
Neutral	42 (16.3)	46 (17.9)	169 (65.8)	257 (100)	
Trustworthy	6 (8.2)	6 (8.2)	61 (83.6)	73 (100)	
Ethnicity					.72
Hispanic	142 (25.5)	89 (16)	325 (58.5)	556 (100)	
Non-Hispanic	9 (18.8)	11 (22.9)	28 (58.3)	48 (100)	
Unknown or prefer not to answer	0 (0)	1 (33.3)	2 (66.7)	3 (100)	
Race					.19
White	121 (25.2)	79 (16.5)	280 (58.3)	480 (100)	
Black	22 (30.6)	13 (18.1)	37 (51.4)	72 (100)	
Asian	3 (11.1)	4 (14.8)	20 (74.1)	27 (100)	
Native American or Alaskan Native	3 (18.8)	4 (25)	9 (56.2)	16 (100)	
Native Hawaiian or Pacific Islander	0 (0)	0 (0)	0 (0)	0 (0)	
Other or prefer not to answer	2 (18.2)	1 (9.1)	8 (72.7)	12 (100)	
Age (years)					.06
35-39	12 (14.6)	16 (19.5)	54 (65.9)	82 (100)	
40-49	28 (19.9)	25 (17.7)	88 (62.4)	141 (100)	
50-59	35 (23.5)	26 (17.4)	88 (59.1)	149 (100)	
60-69	53 (32.1)	25 (15.2)	87 (52.7)	165 (100)	
70-79	23 (32.9)	9 (12.9)	38 (54.3)	70 (100)	
Highest level of education attained					.48
High school diploma, GED[c], or less	38 (21.7)	33 (18.9)	104 (59.4)	175 (100)	
Technical training or certificate	12 (26.1)	9 (19.6)	25 (54.3)	46 (100)	
some college or associates degree	47 (23.4)	34 (16.9)	120 (59.7)	201 (100)	
College degree	28 (25.2)	15 (13.5)	68 (61.3)	111 (100)	
Graduate or professional degree	26 (35.1)	10 (13.5)	38 (51.4)	74 (100)	
Annual household income (US $)					.43
<20,000	30 (27.5)	22 (20.2)	57 (52.3)	109 (100)	
20,000-35,000	24 (20.9)	20 (17.4)	71 (61.7)	115 (100)	
35,000-50,000	28 (27.2)	14 (13.6)	61 (59.2)	103 (100)	
50,000-75,000	26 (23.9)	24 (22)	59 (54.1)	109 (100)	
75,000-100,000	23 (31.5)	6 (8.2)	44 (60.3)	73 (100)	
>100,000	16 (19.5)	11 (13.4)	55 (67.1)	82 (100)	
Unknown or prefer not to answer	4 (25)	4 (25)	8 (50)	16 (100)	
Insurance type					.54
Private	44 (21.3)	38 (18.4)	125 (60.4)	229 (100)	
Government	13 (18.3)	13 (18.3)	45 (63.4)	301 (100)	

Variable	Likelihood of clicking on a decision aid seen on social media			Respondents, n (%)	*P* value
	Unlikely	Neutral	Likely		
Uninsured	28 (23.1)	15 (12.4)	78 (64.5)	43 (100)	
Other insurance or more than 1 policy	63 (31.5)	34 (17)	103 (51.5)	34 (100)	
Relationship status					.37
Single	39 (30.2)	18 (14)	72 (55.8)	129 (100)	
Married	58 (21.6)	44 (16.4)	166 (61.9)	268 (100)	
Separated or divorced	40 (26.7)	25 (16.7)	85 (56.7)	150 (100)	
Widowed	14 (24.6)	13 (22.8)	30 (52.6)	57 (100)	
Other or unknown or prefer not to answer	0 (0)	1 (33.3)	2 (66.7)	3 (100)	
Employment status					.046[b]
Full-time	44 (21.3)	38 (18.4)	125 (60.4)	207 (100)	
Part-time	13 (18.3)	13 (18.3)	45 (63.4)	71 (100)	
Not working for pay or unemployed	28 (23.1)	15 (12.4)	78 (64.5)	121 (100)	
Retired	63 (31.5)	34 (17)	103 (51.5)	200 (100)	
Other or unknown or prefer not to say	3 (37.5)	1 (12.5)	4 (50)	8 (100)	
Country of birth					>.99
United States	137 (24.7)	94 (16.9)	324 (58.4)	555 (100)	
Outside of the United States	14 (26.9)	7 (13.5)	31 (59.6)	52 (100)	
Length of US residency					.19
Less than 15 years	2 (10)	4 (20)	14 (70)	20 (100)	
More than 15 years	149 (25.4)	97 (16.5)	341 (58.1)	587 (100)	

[a]N/A: not applicable.

[b]*P*<.05.

[c]GED: graduate educational diploma.

The “Overall Social Media Engagement Score” was associated with increased likelihood of viewing cancer treatment-related information (Spearman ρ=0.210, *P*<.001; Figure 1). For instance, among those with an engagement score of 1, only 10 out of 25 (40%) were likely to view cancer information, whereas among those with an engagement score of 8, 28 out of 44 (63.6%) were likely. In addition, the engagement score was also associated with accessing a decision aid on social media (Spearman ρ=0.203, *P*<.001; Figure 2). Among respondents with an engagement score of 1, only 8 out of 25 (32%) were likely to click on the decision aid, whereas for those with a score of 8, 25 out of 44 (56.8%) indicated they were likely.

Figure 1. Mean likelihood of viewing cancer-related health information seen on social media by "Overall Social Media Engagement Score.".

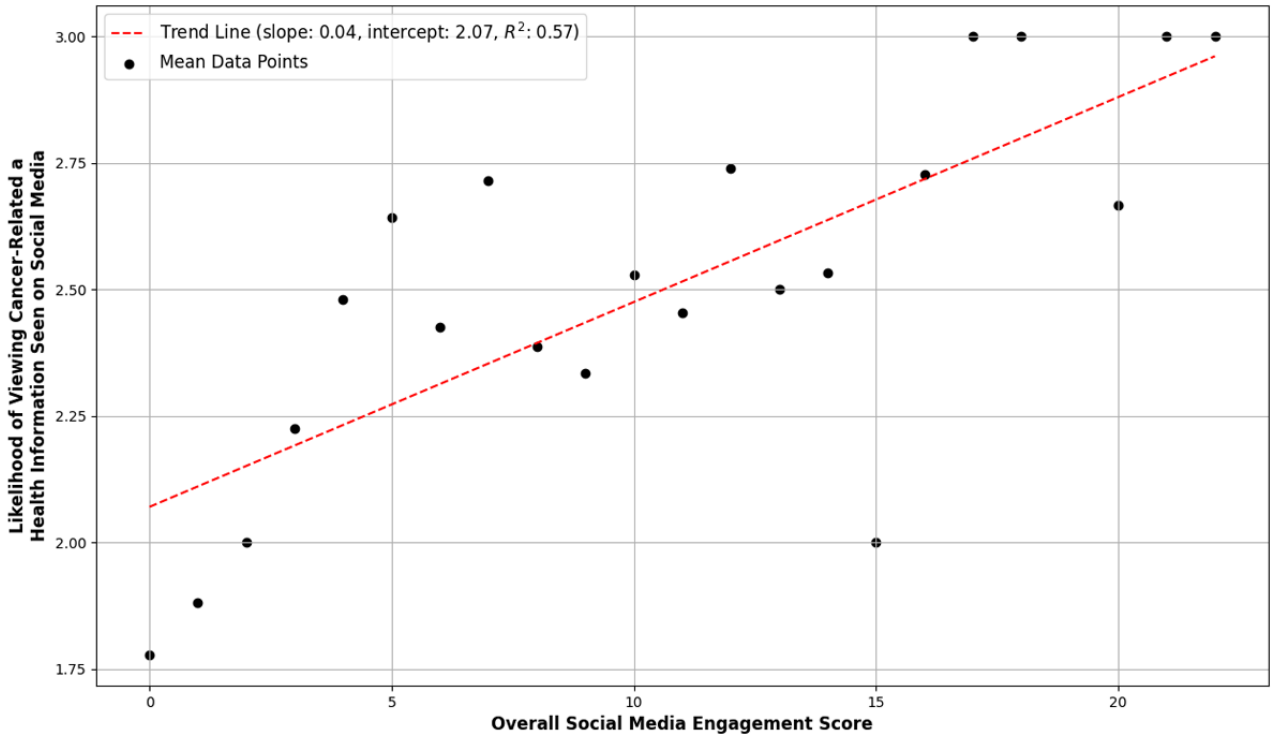
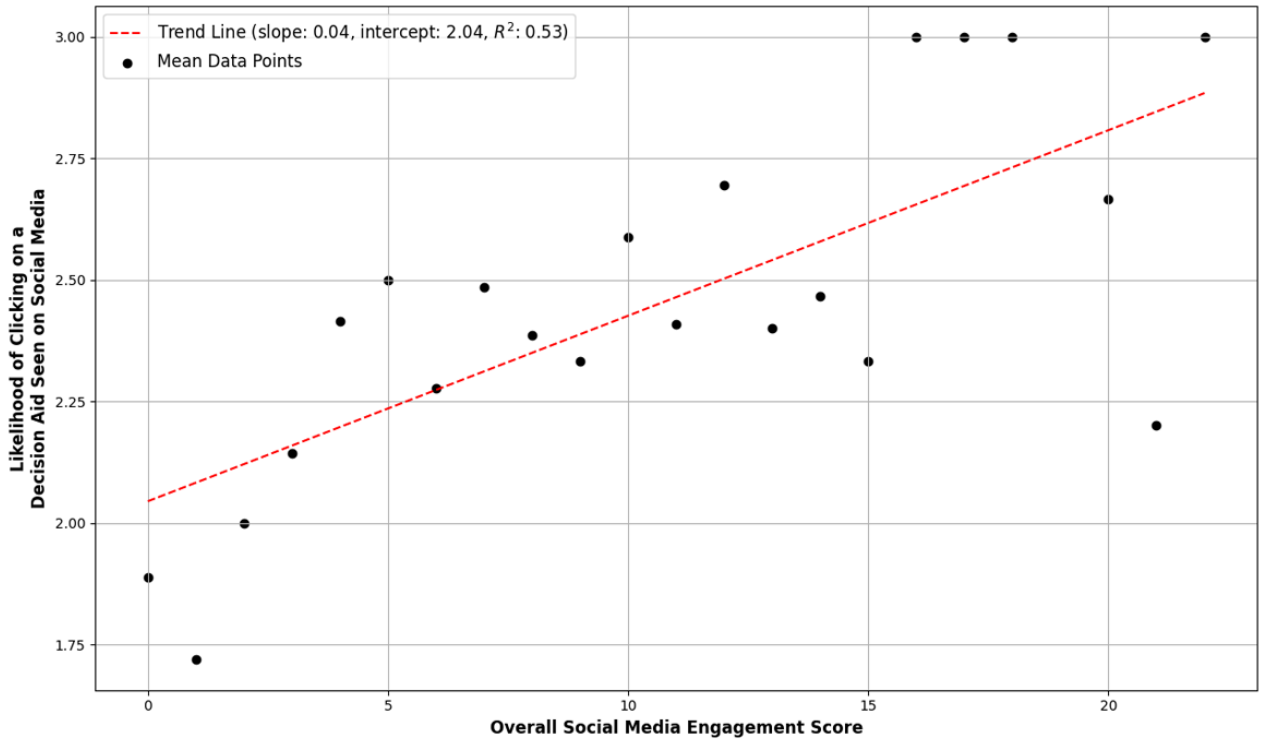


Figure 2. Mean likelihood of clicking on a decision aid seen on social media by "Overall Social Media Engagement Score.".



Age, employment, and perceptions of trustworthiness of social media were included as covariates in the ordinal regression models after significant multicollinearity was ruled out variance inflation factor (VIF≤1.2). The model fit for estimating the likelihood of clicking on decision aids was significant (χ^2_9=60.7, *P*<.001, Nagelkerke R^2=0.113). Respondent perception of the trustworthiness of social media for health information was a significant predictor. Compared with those who found social media trustworthy, respondents who considered social media "untrustworthy" (β=–1.826, Wald χ^2=29.14, *P*<.001) or "neutral" (β=–0.926, Wald χ^2=7.22, *P*<.007) were less likely to click. Age category (*P*=.59) and employment status (*P*=.29) were not significant predictors.

The model for viewing cancer-related health information was also significant (χ^2_9==70.4, P<.001, Nagelkerke R^2=0.133). Relative to those rating social media as trustworthy, those rating it as "untrustworthy" had significantly reduced odds of viewing cancer-related health information (β=–1.680, Wald χ^2=24.31, P<.001) while the reduction in likelihood of those rating it "neutral" did not reach statistical significance (β=–0.581, Wald χ^2=2.74, P=.10). Once again, age (P=.35) and employment status (P=.22) were not significant predictors.

Discussion

Principal Findings

The expanding role of social media in health information dissemination underscores a shift in public health communication. The internet is the most frequently reported source of information for individuals with cancer, especially among women, aligning with our findings [32]. To our knowledge, this is the first study to examine factors influencing engagement with decision aids and cancer-related information amongst women using social media platforms.

These study findings suggest that social media holds potential as a platform for the effective dissemination of cancer decision aids to women. Overall social media usage was high, with almost half of the respondents reporting moderate-high to high engagement. We found that about two-thirds of participants searched for cancer-related information, and nearly half of those used the internet to seek such information for themselves in the past year. Usage patterns varied across platforms: while Facebook emerged as the most used platform, WhatsApp was perceived as the most trustworthy source for health information among our respondents.

Most respondents expressed interest in engaging with cancer treatment information or clicking on a decision aid via social media. Higher frequency of social media use correlated with a higher likelihood of interacting with cancer-related content and decision aids online. In addition, trust in social media appears to be a mediating factor in the relationship between demographics and engagement with cancer information on social media. While younger participants and those who worked full time were more likely to view cancer-related information and click on a decision aid, this effect may be a function of their higher likelihood of trusting social media.

Comparison to Previous Work

Integrating two primary concepts of this study, trust in social media and the likelihood of engaging with health-related content, our findings suggest that individuals who perceive social media as a trustworthy source of health information are more likely to interact with cancer-related treatment information, regardless of their demographic. Numerous consumer studies have highlighted the importance of source credibility in engagement [33-35]. We recognize that trust is a multifaceted construct that is objectively hard to evaluate because it is influenced by factors including demographics, past experiences, and societal and cultural norms [36]. Future studies should focus efforts on better understanding their impact on health information engagement.

Recently, a study by Fridman et al [37] analyzed social media usage and trust in health information among patients with cancer and caregivers, focusing on demographic factors linked to social media use for medical decisions. They also found that a substantial proportion of patients with cancer and caregivers trust social media for health information. Factors associated with higher trust and engagement with social media tools included young age, Black race, and lower education levels. This is consistent with our findings that support trust as a motivating factor for engagement with social media in a medical decision-making context.

Our study was designed to identify trends and general usage patterns across several social media platforms. Facebook (Meta) was the most popular platform among our survey respondents, with 84.2% reporting usage, aligning with 2023 national data indicating that 69% of consumers use Facebook, making it the most used social media site [38]. This is consistent with findings from a recent study which found that Facebook was the most frequently used social media platform for health behavior interventions [39]. However, Facebook is not perceived as being as trustworthy as some of the other platforms. Given the relationship between perceptions of trustworthiness and the likelihood of a respondent using health decision aids on social media, popularity should not be the only factor guiding dissemination. The demographic profile of users also continues to evolve. For example, although less widely used overall, WhatsApp (Meta) is increasingly popular among Latino or Hispanic populations in the United States [40]. These trends highlight the importance of understanding variations in platform use amongst different cohorts when considering platform selection.

In addition, consideration of platforms that can deliver information in diverse formats (eg, text, video, photos, and polls) is important, as each platform's design and interaction style may be better suited for specific sub-audiences. Future research should focus on exploring platform-specific strategies for health information delivery, especially as new platforms emerge (eg, Bluesky [Bluesky PBLLC] and Threads [Meta]) and others become less frequent.

Moreover, the frequency of social media use has a significant impact on the likelihood of engaging with decision aids or accessing cancer-related health information. Frequent social media users may be more likely to perceive others on these platforms as having integrity and competence. They may also report stronger connections with and greater concern for other network users [41]. Frequency of social media use can significantly influence user interactions, such as clicking behavior [42]. However, our Spearman correlation analysis, which focused solely on frequency, accounted for only about 4% of the variance in viewing and clicking behaviors. This suggests that trust, along with other unmeasured factors, likely plays a critical role in these engagement dynamics.

Strengths and Limitations

This exploratory study provides insight into the use of decision aids for health information on social media and highlights the key role of trust. It is an important stepping stone for future research assessing online health behavior among female patients

in cancer. A key strength of our study is the large sample size (N=607) and the inclusion of a cohort that is fairly representative of United States female population demographics based on census data, enhancing the generalizability of our findings. In addition, by recruiting a female-only cohort, our study offers a more nuanced understanding of preferences and engagement patterns among women, which can inform the development of female-specific cancer decision aids tailored for social media.

As a cross-sectional survey-based study, this study was not designed to explore the many complex, nuanced factors associated with online use behavior. First, the survey, although informed by widely used and nationally developed surveys, was not pretested or pilot-tested for face validity. This lack of initial testing may have been associated with increased confusion among participants and could have influenced responses and overall participation in the survey. Relatedly, we were close but unable to meet the overall target population size of over 660 after the application of exclusion criteria. Second, as an internet-based survey, it was subject to selection bias. Participation required English proficiency, internet access, and the ability to navigate an online survey. While incorporating multilanguage options can be considered, facility and comfort with the internet would still be required. Third, we were unable to recruit a racially diverse population that would be matched with corresponding US census data. For instance, the proportion of Hispanic respondents was nearly half of the intended goal (7.9% vs 16%). However, there was more variation in age, education, and income distribution of respondents. Future efforts in larger populations should focus on targeting underrepresented demographics via other survey distribution platforms and recruitment strategies.

Fourth, regarding statistical analysis, although a multivariable regression was performed, our regression models only accounted for roughly 10%-15% of the likelihood of viewing health information or clicking on a decision aid seen on social media. Moreover, potential influences from unmeasured factors, such as medical conditions or personality traits, further complicated attempts to understand these complex dynamics [43,44].

Future Directions

This study offers an initial insight into factors influencing online health information and highlights the role of trust. Future research should explore the potential of social media for the delivery of online decision aids specifically designed for patients seeking cancer information. Our study points to the need for pilot testing health decision tools within the target demographic to help with tool optimization and reliability of findings. Trust is a nuanced concept, and efforts should focus on ways to better quantify and distinguish between trust in both social media platforms and online materials.

Once decision aids have been refined for their target population, continued efforts should consider strategies to promote adoption and optimize engagement. Partnering with reputable health organizations, featuring endorsements from trusted medical professionals, and using verified accounts for content delivery can all be considered. Examining platform-specific formats, such as interactive content on Facebook or visual aids on Instagram, could help increase diffusion. Presenting clear, evidence-based information in user-friendly, visually engaging formats (eg, infographics or explainer videos) may further increase credibility. Incorporating interactive features that allow users to connect with health care providers or support groups on social media could also render greater trust and engagement. Research on these trust-building strategies would offer valuable insight into optimizing social media as a reliable and accessible channel for health information dissemination.

Authors' Contributions

All authors read and approved the final manuscript. ARJ, GAL, GBS, CNL, and TMM conducted formal analysis. ARJ, GAL, and GBS handled visualization. ARJ, GAL, GBS, and MCP, TMM, CNL contributed writing- original draft. ARJ, GAL, GAS, MCP, and YR performed review and editing. YR, MCP, CNL, TMM, and BO handled methodology. MCP, CNL, TMM, and BO conducted to conceptualization. TMM, CNL and BO managed data curation.

Conflicts of Interest

Author MCP was a consultant for Epi-Q, Inc in 2023 and UCB Biopharma in 2024, on topics unrelated to this manuscript. All other authors reported no conflicts of interest.

Multimedia Appendix 1
Survey Used in Study.
[DOCX File , 29 KB - cancer_v11i1e64724_app1.docx]

Multimedia Appendix 2
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 20 KB - cancer_v11i1e64724_app2.docx]

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74(1):12-49 [FREE Full text] [doi: 10.3322/caac.21820] [Medline: 38230766]

2. McGinn TG, McCullagh L, Kannry J, Knaus M, Sofianou A, Wisnivesky JP, et al. Efficacy of an evidence-based clinical decision support in primary care practices: a randomized clinical trial. JAMA Intern Med 2013;173(17):1584-1591. [doi: 10.1001/jamainternmed.2013.8980] [Medline: 23896675]
3. Josfeld L, Keinki C, Pammer C, Zomorodbakhsch B, Hübner J. Cancer patients' perspective on shared decision-making and decision aids in oncology. J Cancer Res Clin Oncol 2021;147(6):1725-1732 [FREE Full text] [doi: 10.1007/s00432-021-03579-6] [Medline: 33682014]
4. Joshi S, Ramarajan L, Ramarajan N, Lee SS, Deshpande O, Fernandes E, et al. Effectiveness of a decision aid plus standard care in surgical management among patients with early breast cancer: a randomized clinical trial. JAMA Netw Open 2023;6(10):e2335941 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.35941] [Medline: 37782500]
5. Shickh S, Leventakos K, Lewis MA, Bombard Y, Montori VM. Shared decision making in the care of patients with cancer. Am Soc Clin Oncol Educ Book 2023;43:e389516 [FREE Full text] [doi: 10.1200/EDBK_389516] [Medline: 37339391]
6. Salwei ME, Ancker JS, Weinger MB. The decision aid is the easy part: workflow challenges of shared decision making in cancer care. J Natl Cancer Inst 2023;115(11):1271-1277. [doi: 10.1093/jnci/djad133] [Medline: 37421403]
7. Stiggelbout AM, Pieterse AH, De Haes JCJM. Shared decision making: concepts, evidence, and practice. Patient Educ Couns 2015;98(10):1172-1179. [doi: 10.1016/j.pec.2015.06.022] [Medline: 26215573]
8. Schubbe D, Yen RW, Leavitt H, Forcino RC, Jacobs C, Friedman EB, et al. Implementing shared decision making for early-stage breast cancer treatment using a coproduction learning collaborative: the SHAIR collaborative protocol. Implement Sci Commun 2023;4(1):79 [FREE Full text] [doi: 10.1186/s43058-023-00453-z] [Medline: 37452387]
9. Waljee JF, Rogers MAM, Alderman AK. Decision aids and breast cancer: do they influence choice for surgery and knowledge of treatment options? J Clin Oncol 2007;25(9):1067-1073. [doi: 10.1200/JCO.2006.08.5472] [Medline: 17369570]
10. Whelan T, Levine M, Willan A, Gafni A, Sanders K, Mirsky D, et al. Effect of a decision aid on knowledge and treatment decision making for breast cancer surgery: a randomized trial. JAMA 2004;292(4):435-441. [doi: 10.1001/jama.292.4.435] [Medline: 15280341]
11. Schapira MM, Hubbard RA, Whittle J, Vachani A, Kaminstein D, Chhatre S, et al. Lung cancer screening decision aid designed for a primary care setting: a randomized clinical trial. JAMA Netw Open 2023;6(8):e2330452 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.30452] [Medline: 37647070]
12. Lin GA, Halley M, Rendle KAS, Tietbohl C, May SG, Trujillo L, et al. An effort to spread decision aids in five california primary care practices yielded low distribution, highlighting hurdles. Health Aff (Millwood) 2013;32(2):311-320. [doi: 10.1377/hlthaff.2012.1070] [Medline: 23381524]
13. Raphael Daniela DB, Russell NS, van Werkhoven E, Immink JM, Westhoff DPG, Stenfert Kroese MC, et al. Implementing a patient decision aid, a process evaluation of a large-scale pre- and post-implementation trial. Breast Cancer Res Treat 2021;185(3):685-695 [FREE Full text] [doi: 10.1007/s10549-020-05975-x] [Medline: 33099691]
14. Tan ASL, Mazor KM, McDonald D, Lee SJ, McNeal D, Matlock DD, et al. Designing shared decision-making interventions for dissemination and sustainment: can implementation science help translate shared decision making into routine practice? MDM Policy Pract 2018;3(2):2381468318808503 [FREE Full text] [doi: 10.1177/2381468318808503] [Medline: 30560204]
15. Padilla Garrido N, Aguado Correa F, Bayo Lozano E, Bayo Calero J, Ortega Moreno M. [Physicians' awareness and assessment of shared decision making in oncology practice.]. Rev Esp Salud Publica 2019;93:e201910066 [FREE Full text] [Medline: 31594916]
16. Brace C, Schmocker S, Huang H, Victor JC, McLeod RS, Kennedy ED. Physicians' awareness and attitudes toward decision aids for patients with cancer. J Clin Oncol 2010;28(13):2286-2292. [doi: 10.1200/JCO.2009.25.2874] [Medline: 20354133]
17. Lane GI, Dupati A, Qi J, Ferrante S, Dunn RL, Paudel R, et al. Factors associated with decision aid use in localized prostate cancer. Urol Pract 2022;9(1):108-115 [FREE Full text] [doi: 10.1097/upj.0000000000000283] [Medline: 35722246]
18. van Tol-Geerdink JJ, van Oort IM, Somford DM, Wijburg CJ, Geboers A, van Uden-Kraan CF, et al. Implementation of a decision aid for localized prostate cancer in routine care: a successful implementation strategy. Health Informatics J 2020;26(2):1194-1207 [FREE Full text] [doi: 10.1177/1460458219873528] [Medline: 31566466]
19. Graham ID, Logan J, Bennett CL, Presseau J, O'Connor AM, Mitchell SL, et al. Physicians' intentions and use of three patient decision aids. BMC Med Inform Decis Mak 2007;7:20 [FREE Full text] [doi: 10.1186/1472-6947-7-20] [Medline: 17617908]
20. Tong G, Geng Q, Wang D, Liu T. Web-based decision aids for cancer clinical decisions: a systematic review and meta-analysis. Support Care Cancer 2021;29(11):6929-6941. [doi: 10.1007/s00520-021-06184-y] [Medline: 33834302]
21. Roundtable on the promotion of health equity and the elimination of health disparities, board on population health and public health practice, health and medicine division, national academies of sciences, engineering, and medicine. In: The Promises and Perils of Digital Strategies in Achieving Health Equity: Workshop Summary [Internet]. Washington (DC): National Academies Press (US); 2016.
22. Stacey D, Légaré F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2014(1):CD001431. [doi: 10.1002/14651858.CD001431.pub4] [Medline: 24470076]

23. Yu L, Li P, Yang S, Guo P, Zhang X, Liu N, et al. Web-based decision aids to support breast cancer screening decisions: systematic review and meta-analysis. J Comp Eff Res 2020;9(14):985-1002 [FREE Full text] [doi: 10.2217/cer-2020-0052] [Medline: 33025800]
24. All HINTS Questions [Internet]. National Cancer Institute: Health Information National Trends Survey. URL: https://hints.cancer.gov/view-questions/all-hints-questions.aspx [accessed 2024-03-12]
25. McCay-Peet L, Quan-Haase A. A model of social media engagement: user profiles, gratifications,experiences. In: O Brien H, Cairns P, editors. Why Engagem Matters Cross-Discip Perspect User Engagem Digit Media Internet. Cham: Springer International Publishing; 2016.
26. McCay-Peet L, Quan-Haase A. SAGE Publications Ltd. The SAGE Handbook of Social Media Research Methods Internet. 2016. URL: https://methods.sagepub.com/book/the-sage-handbook-of-social-media-research-methods [accessed 2018-02-02]
27. Rivera YM, Moran MB, Thrul J, Joshu C, Smith KC. When engagement leads to action: understanding the impact of cancer (Mis)information among latino/a facebook users. Health Commun 2022;37(9):1229-1241 [FREE Full text] [doi: 10.1080/10410236.2021.1950442] [Medline: 34256676]
28. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
29. Chandler J, Rosenzweig C, Moss AJ, Robinson J, Litman L. Online panels in social science research: expanding sampling methods beyond mechanical turk. Behav Res Methods 2019;51(5):2022-2038 [FREE Full text] [doi: 10.3758/s13428-019-01273-7] [Medline: 31512174]
30. Sharma A, Fix B, Hyland A, Quisenberry AJ, Bansal-Travers M, O'Connor RJ. Differences in demographics and behaviors across two web-based survey platforms: observations from a study of risk perceptions of heated tobacco products (HTPs). Prev Med Rep 2023;33:102194 [FREE Full text] [doi: 10.1016/j.pmedr.2023.102194] [Medline: 37223582]
31. United States Population by Age and Sex [Internet]. URL: https://www.census.gov/popclock/data_tables.php?component=pyramid [accessed 2024-11-04]
32. Braun LA, Zomorodbakhsch B, Keinki C, Huebner J. Information needs, communication and usage of social media by cancer patients and their relatives. J Cancer Res Clin Oncol 2019;145(7):1865-1875. [doi: 10.1007/s00432-019-02929-9] [Medline: 31123824]
33. Kareklas I, Muehling DD, Weber TJ. Reexamining health messages in the digital age: a fresh look at source credibility effects. Journal of Advertising 2015;44(2):88-104. [doi: 10.1080/00913367.2015.1018461]
34. Lederman R, Fan H, Smith S, Chang S. Who can you trust? Credibility assessment in online health forums. Health Policy Technol [Internet]. URL: https://www.sciencedirect.com/science/article/pii/S2211883713000853 [accessed 2024-05-09]
35. Jiang T, Wu X, Chen Y, Wang Y. The effects of message framing on online health headline selection of female users: a moderation of approach/avoidance motivation. Int J Med Inform 2021;148:104397. [doi: 10.1016/j.ijmedinf.2021.104397] [Medline: 33503558]
36. Moturu ST, Liu H. Quantifying the trustworthiness of social media content. Distrib Parallel Databases 2011;29(3):239-260. [doi: 10.1007/s10619-010-7077-0]
37. Fridman I, Bylund CL, Elston Lafata J. Trust of social media content and risk of making misinformed decisions: survey of people affected by cancer and their caregivers. PEC Innov 2024;5:100332 [FREE Full text] [doi: 10.1016/j.pecinn.2024.100332] [Medline: 39323933]
38. Jeyaraman M, Ramasubramanian S, Kumar S, Jeyaraman N, Selvaraj P, Nallakumarasamy A, et al. Multifaceted role of social media in healthcare: opportunities, challenges, and the need for quality control. Cureus 2023;15(5):e39111 [FREE Full text] [doi: 10.7759/cureus.39111] [Medline: 37332420]
39. Ghahramani A, de Courten M, Prokofieva M. "The potential of social media in health promotion beyond creating awareness: an integrative review". BMC Public Health 2022;22(1):2402 [FREE Full text] [doi: 10.1186/s12889-022-14885-0] [Medline: 36544121]
40. Gottfried J. Americans? Social Media Use [Internet]. Pew Research Center. URL: https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/ [accessed 2024-06-27]
41. Warner-Søderholm G, Bertsch A, Sawe E, Lee D, Wolfe T, Meyer J, et al. Who trusts social media? Computers in Human Behavior 2018;81:303-315. [doi: 10.1016/j.chb.2017.12.026]
42. Verma N, Fleischmann KR, Koltai KS. Understanding online trustInformation behavior using demographics human values. In: Taylor NG, Christian-Lamb C, Martin MH, Nardi B, editors. Inf Contemp Soc. Cham: Springer International Publishing; 2019.
43. Lee K, Arora R, Freeman G. Examining user?s trust towards healthcare information on Facebook: impacts of information type, disease type, and social consensus. 2018 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); 2018 December 2014; USA URL: https://ieeexplore.ieee.org/document/8947827 [doi: 10.1109/csci46756.2018.00251]
44. Hughes DJ, Rowe M, Batey M, Lee A. A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. Computers in Human Behavior 2012;28(2):561-569. [doi: 10.1016/j.chb.2011.11.001]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
SDM: shared decision-making
VIF: variance inflation factor

Edited by N Cahill; submitted 24.07.24; peer-reviewed by J El Masri, D Spiegel; comments to author 17.10.24; revised version received 03.12.24; accepted 29.01.25; published 05.03.25.

Please cite as:
Johnson AR, Longfellow GA, Lee CN, Ormseth B, Skolnick GB, Politi MC, Rivera YM, Myckatyn T
Social Media as a Platform for Cancer Care Decision-Making Among Women: Internet Survey-Based Study on Trust, Engagement, and Preferences
JMIR Cancer 2025;11:e64724
URL: https://cancer.jmir.org/2025/1/e64724
doi:10.2196/64724
PMID:40053770

Original Paper

Examining How Technology Supports Shared Decision-Making in Oncology Consultations: Qualitative Thematic Analysis

Alan Yung[1*], BSc, MBA, MSc; Tim Shaw[1*], BSc, PhD; Judy Kay[2*], BSc, PhD; Anna Janssen[1*], BA, MA, PhD

[1]Research in Implementation Science and eHealth Group, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[2]Human Centred Technology Research Cluster, School of Computer Science, University of Sydney, Sydney, Australia

[*]all authors contributed equally

Corresponding Author:
Alan Yung, BSc, MBA, MSc
Research in Implementation Science and eHealth Group
Faculty of Medicine and Health
University of Sydney
Charles Perkins Centre
John Hopkins Drive, Camperdown
Sydney, 2006
Australia
Phone: 61 02 86271616
Email: cdss.research@outlook.com

Abstract

Background: Commonly used digital health technologies, such as electronic health record systems and patient portals as well as custom-built digital decision aids, have the potential to enhance person-centered shared decision-making (SDM) in cancer care. SDM is a 2-way exchange of information between at least a clinician and the patient and a shared commitment to make informed decisions. However, there is little evidence in the literature on how technologies are used for SDM or how best they can be designed and integrated into workflows and practice. This may be due to the nature of SDM, which is fundamentally human interactions and conversations that produce desired human outcomes. Therefore, technology must be nonintrusive while supporting the human decision-making process.

Objective: This study examined how digital technologies can help cancer care professionals improve SDM in oncology consultations.

Methods: Health care professionals who treat patients with cancer were invited to participate in online co-design focus group meetings. During these sessions, they shared their experiences using digital technologies for SDM and provided suggestions to improve their use of digital technologies. The session recordings were transcribed and then analyzed using qualitative thematic analysis. The 3-talk SDM model, which consists of 3 steps—team talk, option talk, and decision talk—was used as the guiding framework. This approach was chosen because the 3-talk SDM model has been adopted in Australia. The researchers walked the participants through the SDM model and discussed their routine clinical workflows.

Results: In total, 9 health care professionals with experience treating patients with cancer and using technologies participated in the study. Two focus groups and 2 interviews were conducted in 2024. Three themes and 7 subthemes were generated from the thematic analysis. The findings indicated that various digital technologies, such as electronic health record systems, mobile devices, and patient portals, are used by cancer care professionals to help improve patients' understanding of their disease and available care options. Digital technologies can both improve and undermine SDM. Current systems are generally not designed to support SDM. Key issues such as data integration and interoperability between systems negatively impact the ability of digital technologies to support SDM. Emerging technologies such as generative artificial intelligence were discussed as potential facilitators of SDM by automating information gathering and sharing with patients and between health professionals.

Conclusions: This research indicates that digital technologies have the potential to impact SDM in oncology consultations. However, this potential has not yet been fully realized, and significant modifications are required to optimize their usefulness in person-centered SDM. Although technology can facilitate information sharing and improve the efficiency of consultation workflows, it is only part of a complex human communication process that needs support from multiple sources, including the broader multidisciplinary cancer team.

(JMIR Cancer 2025;11:e70827) doi:10.2196/70827

KEYWORDS

digital health; patient-centered care; person-centered care; shared decision-making; cancer care; oncology; artificial intelligence; AI

Introduction

Background

Shared decision-making (SDM) is defined as a collaborative approach in which patients and health care providers work together to make medical decisions [1]. SDM emphasizes a cooperative relationship between the patient and the physician, characterized by a 2-way exchange of information and a shared commitment to making informed medical decisions [2]. During the SDM conversation, patients and clinicians share information, express preferences, participate in discussions to gain insights, negotiate conflicts, solve problems, and ultimately make decisions [3]. Through this approach, patients can play an active role in their care [4], while physicians gain a better understanding of the unique needs of each patient. Physicians can then make informed and collaborative recommendations that aim to improve patient health outcomes [5]. The use of SDM is particularly crucial in oncology consultations, as the results of treatments are often uncertain. This uncertainty makes treatment decisions complex for patients who often have to choose between aggressive disease management and maintaining their quality of life [6]. Therefore, SDM has been implemented in oncology consultations in several hospitals around the world, and perceptions of its use by cancer care specialists in hospitals have been studied [7-9]. Despite the integration of SDM into health policies and practice standards [10,11], the benefits of SDM are slow to materialize at the operational level [12], and a fragmented health care system can complicate the implementation of SDM.

Efforts have been made to integrate decision aids into electronic health record (EHR) systems used by oncologists [13]. Current EHRs used in oncology practices in hospitals may include functions to facilitate the scheduling of patient consultations and follow-ups, history taking, review of examination results, electronic medication management systems, and care planning [14,15]. However, existing EHRs often do not provide complete details about patients' health values and preferences [16]. This lack of patient details can cause clinicians to misunderstand patient preferences when patients experience cognitive difficulties or when their health conditions worsen too quickly to participate in SDM, which can have significant adverse consequences [16]. The introduction and integration of additional digital tools, such as cancer care dashboards, into EHRs that display patient treatment outcomes and other clinical measurements to monitor patient health status have been developed to increase the ability of both clinicians and patients to visualize results and aid decision-making [17] and to aid the stakeholders during SDM to improve cancer care delivery [18].

Research is ongoing to understand how digital health tools and EHRs can be combined in innovative ways to improve the SDM process [19]. In particular, we need to collect more detailed information to pinpoint where additional digital technology could be developed and used to help the SDM process in the delivery of cancer care. This paper examines how EHRs and other digital tools are used in practice to inform possible future improvements in applied digital technology to facilitate SDM in oncology consultations.

Objectives

Hence, the objective of this study was to explore how health care professionals use digital technology to support SDM in oncology consultations, understand the barriers to technology that support SDM in oncology consultations, and understand the opportunities for future technology to improve SDM in oncology consultations.

Methods

Study Design

This study design was informed by the 3-talk SDM model and the approach of previous studies to develop digital tools to support SDM [20]. The 3-talk model incorporates the principles of team-based collaboration throughout a multistage consultation process and is highly recognized in the health care sector. This model has 3 main components: team talk, option talk, and decision talk [21].

Therefore, to investigate the role of digital technology in SDM in oncology consultations and to achieve the study objectives, we applied the design thinking framework [22]. Design thinking is a creative approach that has been used effectively to address problems in the health care sector [23,24]. It helps to collect user insights to develop efficient products, services, and experiences [23]. Ideas are quickly prototyped and improved through continuous iterations [25]. This study design was chosen because it emphasizes collaboration with end users throughout the problem-solving process. We developed low-fidelity wireframe prototypes of EHRs. This technique was chosen to investigate the potential of EHRs to help oncologists and patients with cancer collaborate on decisions because it has been suggested to be effective in health care management and innovation [26]. Low-fidelity prototypes (Multimedia Appendix 1) were quickly created using affordable graphic software, allowing feedback to be gathered without consuming significant time and resources. We applied co-design and low-fidelity prototyping methods with study participants in focus groups and one-on-one interviews.

Participants and Settings

Health care professionals with opinions on the role of digital technologies in oncology consultations were invited to participate in this study. Specialists in medical and radiation oncology, as well as physicians in training programs, were included. Through existing university connections and local cancer networks, participants were purposefully recruited from 5 cancer care centers in Sydney, Australia. A researcher (TS)

initially contacted key potential participants who collaborated on previous research projects in oncology via email and introduced them to AY. AY then followed up on the communication by providing an information package about the research project and suggesting focus group schedules. The focus groups and interviews were scheduled on Microsoft Teams for remote videoconferencing, and the participants' attendance was recorded.

Data Collection

Guided by the core components of the SDM 3-talk model—team talk, option talk, and decision talk—a focus group and interview topic question guide were developed in advance to shape study inquiries in alignment with the SDM model. The researchers (AY, JK, AJ, and TS) iteratively developed the topic question guide. The topic guide was pretested by running pilot focus group sessions with researchers working on other health care projects within the department. Their feedback helped to refine the topic questions and focus group process. The final version of the topic question guide is shown in Multimedia Appendix 1. The topic questions were used to ask participants about their experience with how technology is used to support SDM within each component of the 3-talk SDM model, particularly if they used the 3 SDM core components in their usual medical practice. The focus groups and interviews were semistructured and guided by the topic questions. The low-fidelity prototypes were presented to participants after discussing the application of technology in their practice, and feedback was sought on the usefulness of the concepts included in the prototype design. The prototypes also served as a trigger for further discussion.

Each focus group and interview concluded by summarizing and reflecting on the discussion and confirming the accuracy of the researcher's understanding of the information provided by the participants while they were still present. This final concluding step was necessary because scheduling busy, working health care professionals providing cancer care to patients for study reviews is difficult.

All interviews and focus groups were recorded in video and audio formats. They took place online between April and May 2024. Author AY led all the focus groups and interviews.

Data Analysis

Three researchers (AY, AJ, and TS) analyzed the qualitative data collected using the reflexive thematic analysis as a framework by Braun and Clarke [27-30]. This method guided the initial coding process applied to the focus group meetings and interview recording transcripts, which were deidentified and anonymized. The researchers first read through the transcripts to fully understand the data. They then proceeded with line-by-line coding, collaboratively compiling and discussing the codes. After completing the coding, the codes were inductively arranged into themes and subthemes. Researcher AY created a codebook, and the researchers engaged in multiple discussions to agree on the identified themes and subthemes. The codebook was tested on 1 transcript. Iterative discussions and consensus resulted in a refinement of the codebook. The final codebook was then used to code the remaining focus groups and interview transcripts. Then, AY used the codebook to code the content of each remaining transcript. Columns in an Excel (Microsoft Corporation) sheet were created to represent different themes and subthemes. AY analyzed the content of each transcript line by line and coded the text. The coded chunks of text were extracted and added to the Excel table according to their alignment with the themes. As new knowledge was found, the codes were refined accordingly. Afterward, AJ reviewed, modified, and confirmed the recategorization of the codes. Eventually, AY finalized the recorded data in the Excel sheet.

Ethical Considerations

The Human Research Ethics Committee of the University of Sydney approved this study (project number: 2023/790). All participants provided written informed consent. Data collected were anonymized and deidentified, and the research data were stored in the university's secure computer systems. All the participants provided their time and information freely without receiving financial compensation.

Positionality of the Research Team

Our research team (TS, JK, and AJ) has extensive experience conducting research on the implementation of digital technologies in health care organizations in Australia from an academic point of view. On the other hand, author AY is a practicing professional with experience in developing and implementing computer software in hospital settings for clinicians. We believe that digital technologies can improve health care. Thus, we are driven to implement the latest innovations in health care.

Results

Participants

The study involved 9 participants who participated in different co-design focus groups and interview sessions. One focus group was attended by 5 (56%) participants; another focus group was attended by 2 (22%) participants. Two interviews were conducted one-on-one. Each session lasted between 30 and 60 minutes. The participant demographics are presented in Table 1.

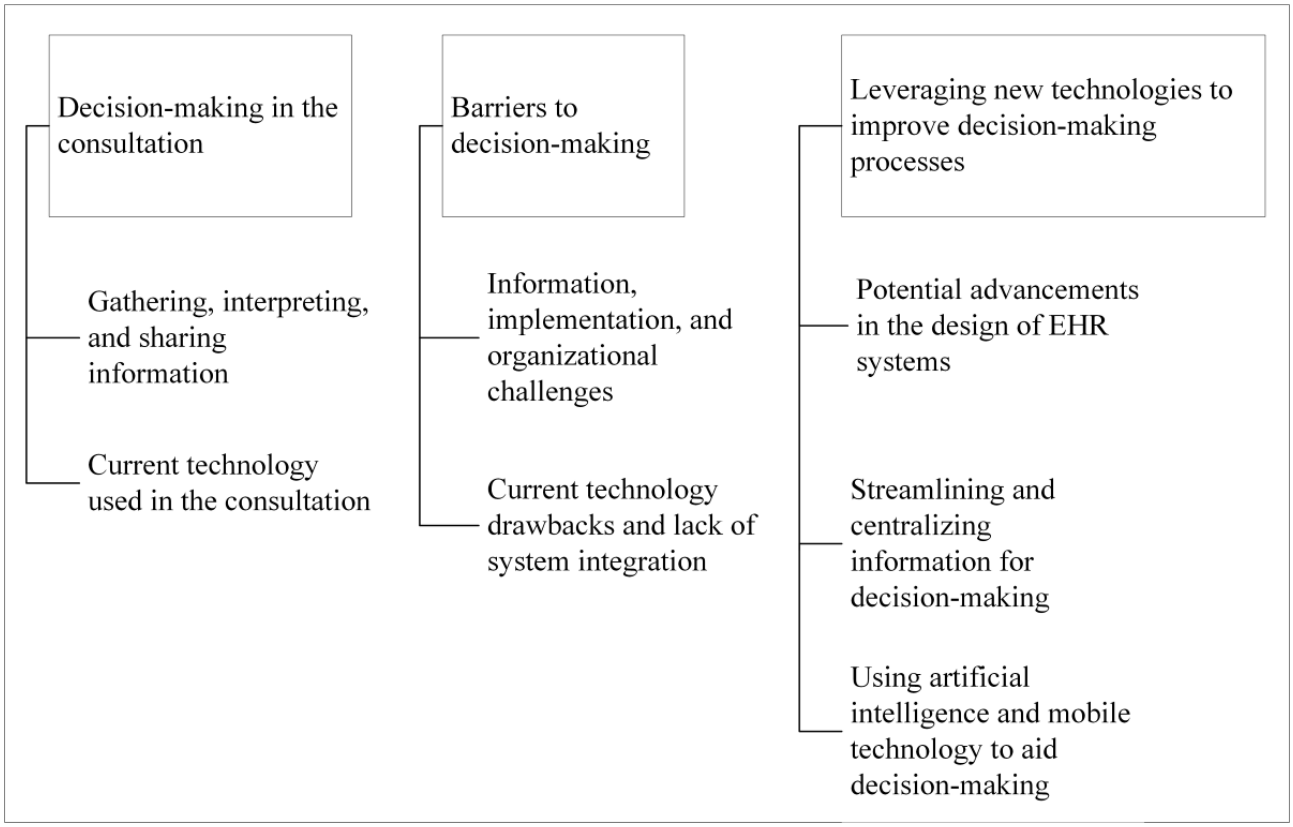
Table 1. Individual participant characteristics.

Participant ID	Session ID	Cancer Care Center ID	Sex	Cancer care stream	Level of experience
P1	A	C1	Female	Breast and lung	Radiation oncologist (consultant)
P2	B	C1	Male	Breast and lung	Radiation oncologist (consultant)
P3	B	C1	Male	Breast and lung	Radiation oncology registrar (in training program)
P4	B	C2	Male	Prostate	Radiation oncology registrar (in training program)
P5	B	C2	Female	Prostate	Radiation oncologist (consultant)
P6	B	C2	Female	Prostate	Radiation oncology registrar (in training program)
P7	C	C3	Male	Lung and head and neck	Medical oncologist (senior consultant and hospital executive)
P8	C	C4	Male	Perioperative	Anesthetic registrar (in training program)
P9	D	C5	Female	Lung	Respiratory specialist (consultant)

Overview of Themes and Subthemes

Three themes and 7 subthemes were generated from the thematic analysis. The three themes are (1) decision-making in the consultation, (2) barriers to decision-making, and (3) leveraging new technologies to improve decision-making processes (Figure 1).

Figure 1. Overview of themes and subthemes. EHR: electronic health record.



Theme 1: Decision-Making in the Consultation

Overview

Participants discussed their decision-making process during consultations and how they felt their clinical workflow aligned with the 3-talk SDM model while being prompted by the wireframe prototypes. The participants appreciated the 3-talk SDM model for breaking the clinical decision-making process into 3 core components:

> *...in general,...this is quite similar to what my approach is in the clinic.* [Participant P4 B C2]
>
> *...it's interesting, and I appreciate this model...breaks it into three pieces.* [Participant P7 C C3]
>
> *...I like team, option, and decision...I hadn't heard of it, but it's exactly how I structure my consultation.* [Participant P9 D C5]

On the basis of their experience, participants highlighted that they did not differentiate between the option talk and the decision talk components:

> *I...in the real world, don't differentiate between option talk and decision talk. So, option talk and decision talk, for me, is the same process. So, as I'm discussing options...I don't sit there and say here are all the options. Now, let's stop and have a discussion about the decision. I blend those two conversations together.* [Participant P7 C C3]

Participants described how they collect information by talking to their patients directly rather than having them fill out responses to a list of questions in advance:

> *I'm actively, kind of, discussing what would be involved in making a decision to go down this pathway. What further information would be needed? So, the implication for digital technology is that it's the same technology I use while discussing options. I don't sort of stop and say now, here's another one I prepared earlier. And let's talk about it this way.* [Participant P7 C C3]

Some participants were concerned about patient privacy:

> *...in the waiting room, I'm not sure...I don't know that I can see an easy way to get personalized digital information in the waiting room, in a safe way. I think that...needs some human and clinical inputs...it could be like a nurse coordinator, someone like that could meet with the patient before going to the consultation. So, there's all those sorts of very personalized differences.* [Participant P9 D C5]

Another important step, in their view, is that the participants noted that they plan activities before patients visit for their consultations:

> *So firstly,...these patients would have been discussed by the multidisciplinary team before they saw me with the surgeons and medical oncologists,...and we would have a plan of action from the MDT.* [Participant P1 A C1]

After the multidisciplinary team planning discussions, the participants described how they would discuss the situation with the patient and involve other professionals in the patient model of care:

> *So, we've looked at performance status, frailty, and pulmonary function. We identified things that are needed. We discussed that at the consultation and asked the care coordinator to link up.* [Participant P1 A C1]

The subthemes of decision-making in consultation include gathering, interpreting, and sharing information through digital and analogue communications.

Subtheme 1.1: Gathering, Interpreting, and Sharing Information

Participants said that decision-making often occurred in multidisciplinary team meetings without input from the patient:

> *...some of our decisions or, you know, consensus, optimal decisions are also influenced by our MDT meetings...in most MDT meetings, the patient is not there....we think we are arriving at a decision that can be communicated to patients, but often, it actually doesn't align with their preferences.* [Participant P5 B C2]

When reflecting on the decision-making component of consultations, participants noted that patient wants and expectations at the point of care they were at were shaped by their previous experiences along the way. Participants remarked that some patients may be ready to decide after meeting with the physician, while others may be hearing about their condition for the first time and feel overwhelmed. The participants know that this approach takes longer. Yet, they prefer having the ability to understand the patient's wishes better:

> *...there's a huge variation in what patients want at this point and what they expect, and it also probably is not independent of what specialty you're in and how they've gotten to you. So, you know, for me, by the time a patient's gotten to me, they may well have been through two or three specialists already. They've got...cancer and, so sometimes, they're already primed. They're ready to make a decision. Other times, it's the exact opposite, and this is the first time they've heard they might have cancer.* [Participant P7 C C3]

> *I use the time when I'm talking to patients, collecting...information to kind of just get to know the person...it takes longer than if they fill in a list of questions in advance and I'm just looking down the list...I'm building a relationship. If I look at a screen...that's not the same as asking those questions and, kind of, building a rapport with a patient.* [Participant P7 C C3]

Different approaches were described for different patient situations and desires for information. Participants said that some patients want to know their treatment plan, while others seek detailed explanations of the decision-making process. The preference for the type of clinical workflow in consultations can also depend on the physician's training, work style, and personality:

> *They've got no idea what's going on....it's the opposite conversation, where they absolutely need to go away and think about it....and I think the things you do to help them in those situations are somewhat different. The first one, those people often already have the information they need. The second one, they absolutely don't.* [Participant P7 C C3]

Several important points related to patient care were covered in the participants' discussions. Participants highlighted the challenges of bringing bad news to patients, the need for better participation of patients in decision-making, and the importance of documentation following decisions:

> *I see them at the start...usually, the person who sees them earlier then has to break the bad news, and then...all the referrals afterward...that's where things*

> *fall apart a little bit...they'll get discussed in MDT...but sometimes that's a little bit delayed. Sometimes, the patient doesn't always get the right information. The right time is the other problem.* [Participant P8 C C4]

The reliability of clinical information sources was raised. It is crucial to always refer to a trusted source of information:

> *...and look, the very important thing in clinical medicine is you go to the primary source for the information; you never make, you should never make a significant clinical decision based on anything but firsthand information.* [Participant P9 D C5]

The value of having care coordinators share the patient care to address patient needs and support them throughout their treatment journey was emphasized:

> *...the need for a care coordinator to triage the patient's care needs, ...it's helpful to have the prostate care nurse who can talk to patients about the radiation therapy and the surgery...they [the patients] get time to make the decision about what they want. We refer them to the men's health physiotherapist as well.* [Participant P6 B C2]

Understanding the specific concerns of the patient is key. It is important to begin the decision-making process with the patient's desired outcome and then work out the appropriate care pathway:

> *...you need to work out the patient's goals first...then work backward from that... "Well, I think it isn't that…" "I definitely don't want radiotherapy" or "I definitely do want radiotherapy because my sister had it. It was good."…you might not get the decision if the patient is still sort of weighing things up...the decision is going to be informed by the goals. It's coming back to the quality of life versus the quantity of life.* [Participant P1 A C1]

The information gathering step is followed by reviewing the patient's results and interpreting the situation before the patient arrives for their visit. One participant described this step as follows:

> *...what I'm talking about is more around interpreting patient results...like the pre-three-talk process...is having the information available...when I prepare for the clinic, I like to have an opportunity to read everything in the pile, and everything is there...that I'm not chasing stuff. So, I'll usually look at my clinic two days ahead of time and make...notes in chronological order to try and figure out firstly whether there is any missing information. Umm…then ensuring that it's adequately documented in a way that is more meaningful to me.* [Participant P2 B C1]

After interpreting the available information step, the participants discussed how they communicated the medical information to patients. They like the way visual aids, as suggested in the wireframe prototypes, help them to clarify and make information more understandable, improving patient understanding and facilitating informed decision-making through effective communication:

> *I find, you know, drawing diagrams and having pictorial, sort of, explanations of things help...I think it breaks through language barriers and understanding of things. Anyway, I'm scrolling through the images and going through the results with the patient, pointing things out, simplifying things, maybe drawing...handwritten...document...to help explain things.* [Participant P4 B C2]

Information and knowledge sharing was discussed in addition to visual aids. Participants mentioned that they often explained results using prognostic calculators that can assess life expectancy, especially in older patients. One participant explained how they discuss different options with patients to help them make treatment decisions:

> *...in some lung cancer patients where there are some poor prognostic factors, and even though they've got technically localized disease that could be curable, you might be a bit worried whether this patient can get through six weeks of umm daily treatment. So, sometimes, we do discuss more palliative options...You give the options; you discuss the harms and benefits of options...but I don't...use the EMR [electronic medical record] apart from the imaging information....I do use...e-prognosis calculators to calculate life expectancy, particularly in older persons.* [Participant P1 A C1]

The participants also discussed the idea of summarizing the consultation decisions:

> *...you know, I appreciate that we don't do it today, but you could imagine a summarized transcript of the consultation generated.* [Participant P7 C C3]

Subtheme 1.2: Current Technology Used in the Consultation

Some participants explained that they do not use digital technology extensively in their consultation workflows. Digital technology is only sometimes used to show patients their medical images or to show images of medications. Videos have been used, but the participants found them too slow. They currently do not have interactive digital tools, but the technology would be useful for discussing treatment options:

> *...in terms of the team talk, how do I use digital technology at this point...mostly show people images, ...I show a lot of scans and X-rays. I usually find the videos are a bit slow for the consultation,...the patients get bored.* [Participant P9 D C5]

Decision aids were discussed. Tools to help predict outcomes of cancer treatments are available on the web for physicians to calculate patients' life expectancy and survival rates. Participants described how they use the decision aids in practice:

> *...I use a predictive tool...I will plug the patient numbers in and print them out for the patient. ...we often use it before we see the patient...in medical oncology, there's one for adjuvant systemic therapy...*

> *" ...without adjuvant chemotherapy, this is your 5 or 10-year survival or recurrence, and with...it's..." they'll show the magnitude of benefit. Then, the patient can decide.* [Participant P1 A C1]

Information sharing was emphasized. Participants described how they provide patients with information about advocacy and treatment protocols and search the internet for basic information, such as images, models, or videos. They share the information they find with patients to educate them. These web-based resources are then used to explain treatment procedures and complex equipment operations, saving them time and effort. The patients are then expected to be able to access and review the same web-based information that they have been introduced to and recommended when at home:

> *...radiation therapy is a technology that most people don't know anything about, ...they get confused....the value of images, models, or video to actually just show what a radiation linear accelerator machine is...you don't have to draw a picture of it. You don't have to waste time taking somebody around to look at the machine...trusted website resources.* [Participant P5 B C2]

> *I found myself doing a lot of...very basic Google images search...the information can be so basic...I think we get lost in explaining things.* [Participant P3 B C1]

Theme 2: Barriers to Decision-Making

Two subthemes were identified under the barriers to decision-making theme. The first subtheme, "information, implementation, organizational challenges," focuses on the participants' perceived challenges regarding access to and the quality of information. The second subtheme, "current technology drawbacks and lack of system integration," deals with the participants' difficulties related to the limitations of the EHRs and the lack of information integration.

Subtheme 2.1: Information, Implementation, and Organizational Challenges

Participants pointed out challenges such as experiencing difficulties when communicating with patients from different cultures and non–English-speaking patients in communities. They also mentioned challenges with patients' lack of health literacy:

> *...meeting patients of non-English backgrounds and cultural and health literacy issues; uh, very significant, and that's very hard together in a very quick clinical environment.* [Participant P2 B C1]

The involvement of the family and interpreters was also raised as a challenging area due to the time needed to understand the needs and priorities of the individuals:

> *...family care as support, and...the interpreter as well, ...can be part of the communication process, which can either assist or umm or slow down dramatically the process...It's hard to think of a solution because it takes time to talk to people and find out what's important to them.* [Participant P5 B C2]

Gathering precise patient information during visits, as patients often forget details, was expressed as a difficulty. Participants noted the need to improve communication methods and understand each patient's needs:

> *There are even times when a patient has had a test done, and it's not until they're, literally, sitting in the clinic room before me, and I go, where did you have this done? Sometimes, I have to ask them three questions to clarify...Umm, it's a common assumption of the patients as well. "Don't you have this information?" And the answer is often no, I don't.* [Participant P6 B C2]

Verifying the accuracy of the information patients provide can be time-consuming, as one participant pointed out the following:

> *...patients come in...and say, oh yes, I had a scan. ...you spent 5 minutes searching all the providers...then you Google where they live and what radiology practice is in their town, and then you find out they did have a scan, but it was an MRI of their ankle. It wasn't actually their chest, but they don't remember.* [Participant P9 D C5]

Subtheme 2.2: Current Technology Drawbacks and Lack of System Integration

Manual processing of information and uploading data into the EHRs is problematic for physicians, especially under time pressure:

> *At the moment, when we upload imaging, it's not the actual images themselves,...to, just, get the image in, I take a screenshot and paste it into a document in the EMR, or I am literally, highlighting and copying the text from the report and pasting it in,...when you are time-pressured, that's just how you get it done.* [Participant P6 B C2]

Obtaining and merging data from various sources presents additional challenges to physicians. Especially the lack of integration among older information systems for data sharing was considered a drawback. This situation caused difficulties in accessing different systems for decision-making tasks:

> *...needing multiple passwords in multiple different information systems or not having access to all the patient results. ...unfortunately, most hospitals, including ours, rely so much on a technology called fax.* [Participant P2 B C1]

The participants said the systems could not provide integrated results even when patients had medical tests conducted in public hospitals:

> *There are already difficulties in accessing scans and results...done even in other public hospitals...patients have blood tests done by multiple providers. Imaging from multiple different providers.* [Participant P6 B C2]

Besides the lack of system integration, 1 participant pointed out that their hospital does not have full access to the facilities of EHRs:

> *I sit in a hospital that does not use an EMR or has a partial EMR. So, the medical notes don't go, for the most part into an EMR, it does in the oncology clinics, but that's not where I work anymore. So, we mostly write on paper in the private clinic. I use my own digital interface and I'm always zooming around to different portals, external radiology, different pathology providers, et cetera.* [Participant P9 D C5]

Poor wireless digital communication network connectivity was also mentioned as another drawback:

> *...it's again getting onto another website, potentially getting password...Terrible Wi-Fi in most cancer centers...I think that is a big barrier.* [Participant P5 B C2]

Theme 3: Leveraging New Technologies to Improve Decision-Making Processes

The theme "leveraging new technologies to improve decision-making processes" encompasses the following subthemes: (1) participants' interest in implementing potential improvements to advance the design of EHR systems; (2) making data more accessible and understandable by streamlining, centralizing, and communicating information for collaborative decision-making; and (3) helping to share evidence data and decisions with patients' care team members outside consultations, as well as analyzing patients' data using artificial intelligence (AI) and mobile technologies.

Subtheme 3.1: Potential Advancements in the Design of EHR Systems

Participants expressed their interest in improving the design of the EHRs. They highlighted the need for improved access to laboratory diagnostic test results and recommended automatically providing reliable medical information from different systems:

> *...if there was some magic like a digital resource that could do all of that detection for me and link me to multiple different providers and go to clinical labs...and pull it all in, I would love it...If it was as good as me, it would be transformative. But you'd have to really be sure and be able to trust it...and then...the reliability of information.* [Participant P9 D C5]

There is interest in decision-making tools to help patients make treatment decisions. Participants said they do not need additional electronic devices to replace what they already have. They want decision-making tools to help patients choose their preferred treatments according to their desires and goals, especially when treatment options are risky:

> *...some, sort of decision tool may help in those situations where radiotherapy is high risk or trying to help people decide about quality versus the longevity of life or some sort of tool where you...answer to some questions..."quality of life is more important to me or length of life is more important" ...it would be good to have a tool where you can...help guide the patients to...their priorities...and...help the decision-making...I don't want any extra devices. I'd do it on the computer and then, maybe, print it out for the patient rather than an iPad type stuff.* [Participant P1 A C1]

Subtheme 3.2: Streamlining and Centralizing Information for Decision-Making

Centralizing and systematically organizing medical information to make it more accessible and easier to interpret is important to some participants. These participants were interested and emphasized that providing the right information to the physician at the point of care would help:

> *...one thing I found very helpful is the centralization of information. ...things like scans, test results from clinics or centers outside of the...health system...something that aggregates that information into...something to sort of centralized or funnel information to us...having patient information presented in a way where...making things more centralized, it would be helpful to us.* [Participant P3 B C1]

Other participants stressed the importance of obtaining comprehensive patient information before the consultation:

> *...I guess what I'm talking about is more around interpreting patient results, which is almost...preempted to the whole three-talk process, really...is having the information available.* [Participant P5 B C2]

The introduction of a patient portal for sharing information with patients is seen as a benefit. This would enhance physicians' ability to maintain communication with patients outside of consultations as they consider treatment options:

> *...if there's a patient portal, they can log in and see things, that could be nice. ...if I could say to them... "I've put all these in...I've put in the options...when you go home, you can log into your patient portal..." I could even imagine they could post some questions.* [Participant P9 D C5]

Subtheme 3.3: Using AI and Mobile Technologies to Aid Decision-Making

The potential use of generative AI was discussed to streamline medical documentation and improve patient care. Participants suggested using basic AI to generate patient reports that can be shared with medical colleagues:

> *...information can be more easily extractable...we use very basic artificial intelligence in our practice where we can generate a patient report, for example, where we pull information from different parts of...and combine it with text that we put in the record and that then goes to the general practitioner. So, I can do a treatment summary on a radiotherapy patient in about a minute, and I only have to type a line or two, and yet, a complex report goes back to the general practitioner, and we do that in medical oncology as well.* [Participant P2 B C1]

However, one other participant disliked the idea of using AI for report writing:

> *Wouldn't use it. I write better than generative AI. I think the kind of language that generative AI produces is boring and opaque, and I'm better than that. So, I wouldn't do it yet.* [Participant P9 D C5]

Discussion

Key Findings

This research examined how health care professionals in Sydney use digital technology to support SDM during oncology consultations. It sought to understand the difficulties they encounter when using technology for SDM and explore potential developments of new technologies that could improve the implementation of SDM in clinical oncology settings. First, the findings of this study emphasize the critical need for oncologists to consolidate health information from patients with cancer to facilitate SDM in oncology consultations. The results also highlight a significant misalignment between the current operations of existing EHRs and the clinical practice workflow in oncology clinics to help clinicians follow the SDM process. Second, the study draws attention to the challenges of access to information due to outdated technologies and communication barriers due to language and the lack of knowledge of the patient about health. Nevertheless, the study participants were interested in developing new technologies that could streamline access to health information and automate administrative processes, thus supporting SDM and ultimately improving the delivery of cancer care.

Current Use of Technology to Support SDM in Oncology Consultations

The study participants stressed the importance of consolidating medical information to improve decision-making in oncology consultations. Studies in similar data-driven cancer care management reinforce these findings of the investigation [31]. Similar to other studies on cancer care, participants in this research study have emphasized the critical role that information and data play in driving SDM processes and improving health service outcomes [18]. As digital technologies transform the health care sector, cancer care is also being transformed [32].

Discussions between health care professionals during the study addressed the 3 key components of the SDM model: team talk, option talk, and decision talk [21]. The prototyped EHRs used to investigate the feasibility of supporting SDM with EHRs demonstrated that some components of the SDM model of care, such as option talk, could be implemented to match established oncology consultation practices and workflows where patients and oncologists usually discuss treatment options. However, the phase sequence of the SDM model did not fit fully into the typical consultation procedures or workflow patterns of the study participants. The health care professionals who participated in this study appreciated the SDM model but stated that, in their routine clinical practice, they frequently combined option and decision discussions. This means that EHRs must be flexible to support cancer care workflows to accommodate the iterative nature of the oncology decision-making process.

Study participants highlighted the importance of direct patient communication to foster relationships and ensure complete information collection before choosing treatments or health care options. Previous research in this area has also emphasized the importance of the relationship and communication between oncologists and patients beyond consultation visits in cancer care management [33]. Several study participants have pointed out that a key to the successful implementation of SDM is the integration of digital systems and EHRs, ensuring accessibility to digital information when needed at the correct point of care for the right patient. However, some participants have also stated that they do not use their digital systems or EHRs extensively to support patient discussions. They may use only part of the system to show diagnostic images to share information with the patient. Other participants use EHRs only to look up patient results or document consultations.

Future Use of Technology to Support SDM in Oncology Consultations

Cancer treatment is based on data, involves multiple disciplines, is a lifelong process, and is increasingly dependent on the smooth digital exchange of clinical information [34]. In this study, the participants identified several key obstacles to SDM in their clinical oncology settings related to access to information, implementation, organization, and limitations of current technology, specifically EHRs. In addition, the participants mentioned communication challenges due to language barriers, emotions, comprehension, low health literacy, participation of patients, difficulties in accessing and integrating patient data, lack of information that often leads to poor data quality and inefficient processes, time pressure, and lack of privacy. Similar barriers have been reported by Steenbergen et al [35]. The participants informed the research about the absence of integrated systems and their continued dependence on outdated technologies in their clinical settings, which hinders information exchange between cancer care facilities. Furthermore, during the investigation, some health care professionals who participated in the study described that their hospitals do not have comprehensive EHRs, leading to a greater dependence on paper records and personal digital interfaces. Researchers in Canadian health systems have also reported on clinician experiences with outdated, ineffective, or inefficient technologies that do not fit their clinical workflows [36]. Therefore, the implementation of better information and communication technologies could eliminate some technological barriers and improve the overall efficiency of cancer care provided by oncologists.

During the study, health care professionals said that they use the information from the EHRs to help in their decision-making process to treat cancer. They focused on integrating digital resources to improve efficiency and support patient care. However, integrating quality health data remains challenging due to the lack of guaranteed interoperability, even between EHRs from the same vendor, as reported in a previous study in the United States [37], although the requirement to improve interoperability among digital health systems was legislated in the United States in 2016 [38], and the Fast Healthcare Interoperability Resources specifications were approved by the Health Level 7 International in August 2019 [39]. In June 2024,

the Canadian government introduced Bill C-72, which requires health IT systems to be interoperable [40]. Therefore, the stated goal of the health care professionals, which is to be able to securely access all the health information of their patients in integrated EHRs, is expected to be achieved in Canada in the future [40]. Therefore, future EHRs in the North American health care systems, designed to make health care information more accessible and transparent to patients and the health care team [41], are expected to be available to provide oncologists with critical cancer care data needed to support the SDM process in oncology consultations.

Furthermore, the study participants were interested in the potential benefits of an integrated web-based portal driven by clinical information designed to simplify access to data from private laboratory tests and automate various clinical documentation processes, such as generating interclinician letters and managing patient diagnostic test results. Petrovskaya et al [42] performed an evaluation of web-based patient portals and emphasized the elements that the study participants seek to help improve patient participation in SDM. The researchers stated that the patient portal is connected to the EHRs of health organizations, providing patients with functionalities such as secure and convenient access to medication lists and the ability to arrange and verify appointment availability and communicate with their health care team securely through SMS text messaging, in addition to access to their laboratory test results [42]. However, in a recent patient portal implementation initiative, Grewal et al [43] found that there are technical challenges in enrolling patients to use the patient portal, but involving nurses in the patient education and enrollment process is a promising approach and reinforces the value of multidisciplinary methods in improving patient care.

During the study, the participants explored the concept of a patient web-based portal that can consolidate health information from multidisciplinary treatment journeys. They emphasized the need for sophistication and proper allocation of resources. The participants envision a web-based portal where patients can access information about care options, ask questions, and review details such as their therapeutic plans and preferences. They believe that this would lead to more streamlined communication, better decision-making, and automation that uses AI capabilities. They perceive that AI innovations could help reduce the double handling of information and miscommunication, as well as prevent patients from falling through the cracks in their care. However, trust in AI systems and the data provided emerged as a significant concern among some participants. In an article on digital transformation in cancer care, Papachristou et al [32] emphasized that ensuring the safety, accuracy, and ethical application of data-driven interventions requires building trust among health care professionals, patients, family members, caregivers, and other stakeholders. Nevertheless, integrating AI into the cancer management workflow has been shown to transform individual treatment planning by accurately predicting responses of patients with cancer to different therapies [44].

Efforts to improve EHRs for better cancer care management are ongoing around the world. Two international workshops focused on technology in cancer care management were held in 2019 and 2020 in Europe [31] and one in 2022 in the United States [38]. These workshops addressed SDM processes, data integration and management, analytics, EHRs, and AI-based clinical decision-making [31,38]. While significant progress has been made in implementing EHRs in public hospitals in Sydney for cancer care [15,45], the full potential of EHRs to consistently improve cancer care quality and patient outcomes has not yet been fully realized [38,45]. Similar to the challenges that the participants of this study encounter with poor EHR usability, lack of fitness with clinical workflows, fragmented data sources, and large amounts of data, researchers from other health care jurisdictions have also described similar experiences [31]. The participants suggested that in addition to using technologies, nurses and other health care professionals could also assist in patient engagement. These additional clinical resources have skills, such as patient education and effective communication, crucial to facilitating patient participation in SDM during clinical oncology consultations and can help improve patient outcomes [46]. The effectiveness of SDM is maximized when health care professionals have experience, strong relationships with patients, and sufficient time for treatment discussions [35]. As reported by Steenbergen et al [35], the exchange of knowledge and the efficient flow of health information between clinicians and patients are essential to facilitate SDM in oncology. Consequently, technological opportunities are tailored to support human interactions [31,38].

Barriers to the effective digitalization of information in oncology have been identified. However, continuous innovations and technological improvements have helped minimize the effects of several major barriers. Technological innovations such as Health Level 7 Fast Healthcare Interoperability Resources [47], the Minimal Common Oncology Data Elements [48], and the Systematized Nomenclature of Medicine–Clinical Terms [49] when combined with legislation, such as the Connected Care for Canadians Act in Canada, make better access to health information possible. Therefore, digital health data in oncology can be shared across health care organizations in a more standardized way that all stakeholders can understand.

Conversely, although AI technologies have been introduced in oncology over numerous decades, a persistent distrust exists toward the suggested technology. The level of trust in AI systems influences the acceptance of these technologies. Therefore, frameworks and guidelines have been suggested to tackle the issues related to the reliability of AI-powered health care systems, such as the FUTURE-AI framework, which defines 6 requirements for trustworthy AI [50]. Accepting AI systems in health care depends on ethical principles, trust dynamics, and rigorous evaluation processes [51].

Tools and protocols are available globally to support SDM in oncology consultations. For example, in the United States, tools include Watson for Oncology [52] and the Adjuvant! Platform [53]. In Australia, EVIQ chemotherapy protocols are available nationally [54,55]. In the United Kingdom, the PREDICT tool aids in breast cancer treatment decisions [56]. In Canada, standards for SDM tools have been developed and are often used as a reference by international researchers [57-59]. Despite multiple trials, the integration of these tools and protocols into practice remains nonroutine, and several programs, such as IBM

Watson for Oncology, have failed to meet expectations [60]. These examples illustrate the ongoing challenges.

In summary, various oncology specialists and health care professionals perceive the usefulness of technology in supporting SDM in oncology consultations differently. A senior medical oncologist preferred face-to-face conversations with patients. In contrast, an anesthetic registrar preferred a high level of computerization and welcomed the possibilities of driving health care delivery with data. Other specialists, especially radiation oncologists, did not see the need to use technology extensively when helping patients make treatment decisions, as their oncology specialization typically involves only one treatment modality. However, they do want technology to accurately and promptly share information provided by other health care professionals. However, young health care professionals are ready to adopt more digitalized medical practices. Most health care professionals recognized the value of technology in supporting access to information for consumers, thereby facilitating informed decision-making.

Limitations and Future Research

The first limitation of the study was that only 9 health care professionals were available to participate in the co-design sessions. The second limitation was that no surgeon was identified to potentially participate in the co-design sessions. It is difficult for practicing physicians to allocate time for research projects and to attend co-design sessions when they are already working overtime and long hours providing patient care. Therefore, physicians who participated in the study may not have fully represented the larger oncology practice community. Only their views and practices on SDM were collected. The third limitation was that oncology consultation involves patients, other oncology specialists, and other health care providers. However, they were not invited to participate in this study due to time constraints. Patients and other health care providers may have provided different perspectives on their experience with SDM and the use of digital technology.

A larger group of oncology specialists, including surgeons, would have represented the larger oncology community and provided more generalized views. Furthermore, patients who have had oncological consultations would have provided their views on decision-making processes, particularly SDM. To mitigate the limitations of this study and obtain more generalizable results, our approach should be replicated in future studies with a larger and more diverse group of cancer health care professionals. This diversity would include many specialty dimensions, including surgeons and other health systems specialists. Furthermore, similar future studies should include patients who have experienced oncology consultations.

Conclusions

The findings of this study indicate that digital health technologies can assist in SDM in oncology consultations. This includes providing concise and consolidated information to support decision-making, tools such as multimedia resources to support patient understanding of cancer and treatments, and patient access to information and data outside of the consultation through tools such as patient portals. Emerging technologies, such as generative AI, may assist SDM by consolidating and personalizing information.

Nevertheless, care needs to be taken to ensure that technology does not erode the development of rapport and trust between a clinician and patient. Although EHRs and other systems are continually improving, there are substantial barriers to realizing the potential of technology to improve SDM, including the lack of data integration between systems and integration of new tools and resources into clinical workflows. However, continuous technological innovations and government efforts through new legislations are eliminating some of the digital system integration and data interoperability difficulties.

In conclusion, the study shows that digital technology can facilitate the exchange of information between independent health care organizations and individual health care providers, thus increasing the efficiency of oncology consultation workflows. However, technology is only part of the support needed for the complex human communication process in oncology. Oncology consultation services need support from a multidisciplinary cancer team, which includes other health care professionals and the patient's family. Health care professionals, such as nurses, must educate and prepare patients for consultations. Allied health professionals are often needed to help with language difficulties. Only through an ecosystem that is fully integrated, interoperable, and seamlessly fits in with the human and social interactions of numerous stakeholders involved in the care of a patient with cancer can the goals of the person-centered model of care be achieved through the implementation of SDM in cancer care.

Acknowledgments

The authors thank the participants who generously shared their experiences to support this research study.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Focus group and interview topic guide questions and low-fidelity prototypes.
[DOCX File , 517 KB - cancer_v11i1e70827_app1.docx]

References

1. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [doi: 10.1016/s0277-9536(96)00221-3] [Medline: 9032835]
2. Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. Soc Sci Med 1999 Sep;49(5):651-661. [doi: 10.1016/s0277-9536(99)00145-8] [Medline: 10452420]
3. Montori VM, Ruissen MM, Branda ME, Hargraves IG, Kunneman M. Problem-based shared decision making: the role of canonical SDM steps. Health Expect 2023 Feb;26(1):282-289 [FREE Full text] [doi: 10.1111/hex.13654] [Medline: 36448245]
4. Resnicow K, Catley D, Goggin K, Hawley S, Williams GC. Shared decision making in health care: theoretical perspectives for why it works and for whom. Med Decis Making 2022 Aug 16;42(6):755-764 [FREE Full text] [doi: 10.1177/0272989X211058068] [Medline: 34784805]
5. Bingaman L. The art of shared decision making. JAAPA 2023 Feb 01;36(2):31-34. [doi: 10.1097/01.JAA.0000902888.46676.73] [Medline: 36701578]
6. Bennett R, DeGuzman PB, LeBaron V, Wilson D, Jones RA. Exploration of shared decision making in oncology within the United States: a scoping review. Support Care Cancer 2022 Dec 31;31(1):94 [FREE Full text] [doi: 10.1007/s00520-022-07556-8] [Medline: 36585510]
7. van Veenendaal H, Voogdt-Pruis H, Ubbink DT, Hilders CG. Effect of a multilevel implementation programme on shared decision-making in breast cancer care. BJS Open 2021 Mar 05;5(2):zraa002 [FREE Full text] [doi: 10.1093/bjsopen/zraa002] [Medline: 33688949]
8. Haltaufderheide J, Wäscher S, Bertlich B, Vollmann J, Reinacher-Schick A, Schildmann J. "I need to know what makes somebody tick …": challenges and strategies of implementing shared decision-making in individualized oncology. Oncologist 2019 Apr;24(4):555-562 [FREE Full text] [doi: 10.1634/theoncologist.2017-0615] [Medline: 30190300]
9. Scholl I, Hahlweg P, Lindig A, Frerichs W, Zill J, Cords H, et al. Evaluation of a program for routine implementation of shared decision-making in cancer care: results of a stepped wedge cluster randomized trial. Implement Sci 2021 Dec 29;16(1):106 [FREE Full text] [doi: 10.1186/s13012-021-01174-4] [Medline: 34965881]
10. Légaré F, Stacey D, Forest PG, Archambault P, Boland L, Coutu MF, et al. Shared decision-making in Canada: update on integration of evidence in health decisions and patient-centred care government mandates. Z Evid Fortbild Qual Gesundhwes 2022 Jun;171:22-29. [doi: 10.1016/j.zefq.2022.04.006] [Medline: 35606312]
11. The NSQHS standards. Australian Commission on Safety and Quality in Health Care. URL: https://www.safetyandquality.gov.au/standards/nsqhs-standards [accessed 2025-05-15]
12. Elwyn G. Shared decision making: what is the work? Patient Educ Couns 2021 Jul;104(7):1591-1595. [doi: 10.1016/j.pec.2020.11.032] [Medline: 33353840]
13. Scalia P, Ahmad F, Schubbe D, Forcino R, Durand MA, Barr PJ, et al. Integrating option grid patient decision aids in the epic electronic health record: case study at 5 health systems. J Med Internet Res 2021 May 03;23(5):e22766 [FREE Full text] [doi: 10.2196/22766] [Medline: 33938806]
14. Yu P, Gandhidasan S, Miller AA. Different usage of the same oncology information system in two hospitals in Sydney--lessons go beyond the initial introduction. Int J Med Inform 2010 Jun;79(6):422-429. [doi: 10.1016/j.ijmedinf.2010.03.003] [Medline: 20371209]
15. Dabliz R, Poon SK, Ritchie A, Burke R, Penm J. Usability evaluation of an integrated electronic medication management system implemented in an oncology setting using the unified theory of the acceptance and use of technology. BMC Med Inform Decis Mak 2021 Jan 06;21(1):4 [FREE Full text] [doi: 10.1186/s12911-020-01348-y] [Medline: 33407411]
16. Lenert L, Dunlea R, Del Fiol G, Hall LK. A model to support shared decision making in electronic health records systems. Med Decis Making 2014 Nov;34(8):987-995 [FREE Full text] [doi: 10.1177/0272989X14550102] [Medline: 25224366]
17. Mohindra NA, Garcia SF, Kircher S, Barnard C, Perry LM, Lyleroehr M, et al. Development of an electronic health record-integrated patient-reported outcome-based shared decision-making dashboard in oncology. JAMIA Open 2024 Oct;7(3):ooae056. [doi: 10.1093/jamiaopen/ooae056] [Medline: 39049991]
18. Rietjens JA, Griffioen I, Sierra-Pérez J, Sroczynski G, Siebert U, Buyx A, et al. Improving shared decision-making about cancer treatment through design-based data-driven decision-support tools and redesigning care paths: an overview of the 4D PICTURE project. Palliat Care Soc Pract 2024 Feb 12;18:26323524231225249 [FREE Full text] [doi: 10.1177/26323524231225249] [Medline: 38352191]
19. Jacobs F, D'Amico S, Zazzetti E, Gaudio M, Benvenuti C, Saltalamacchia G, et al. Digital innovations in breast cancer care: exploring the potential and challenges of digital therapeutics and clinical decision support systems. Digit Health 2024 Nov 03;10:20552076241288821 [FREE Full text] [doi: 10.1177/20552076241288821] [Medline: 39502478]
20. Morken V, Perry LM, Coughlin A, O'Connor M, Chmiel R, Xinos S, et al. Implementation of a co-design strategy to develop a dashboard to support shared decision making in advanced cancer and chronic kidney disease. J Clin Med 2024 Jul 17;13(14):4178 [FREE Full text] [doi: 10.3390/jcm13144178] [Medline: 39064218]

21. Elwyn G, Durand MA, Song J, Aarts J, Barr PJ, Berger Z, et al. A three-talk model for shared decision making: multistage consultation process. BMJ 2017 Nov 06;359:j4891 [FREE Full text] [doi: 10.1136/bmj.j4891] [Medline: 29109079]
22. Kelley D, Brown T. An introduction to design thinking. Institute of Design at Stanford. 2018. URL: https://web.stanford.edu/~mshanks/MichaelShanks/files/509554.pdf [accessed 2025-05-29]
23. Leow MQ, Moosa AS, Salim H, Abdullah A, Lee YK, Ng CJ, et al. Innovation workshop using design thinking framework and involving stakeholders to co-create ideas for management of asthma. NPJ Prim Care Respir Med 2023 Nov 04;33(1):36 [FREE Full text] [doi: 10.1038/s41533-023-00357-4] [Medline: 37925475]
24. Sunjaya AP, Martin A, Jenkins C. A design thinking approach to developing a clinical decision support system for breathlessness in primary care. Stud Health Technol Inform 2022 Jun 06;290:839-843. [doi: 10.3233/SHTI220197] [Medline: 35673136]
25. Ku B, Lupton E. Health Design Thinking: Creating Products and Services for Better Health. Cambridge, MA: MIT Press; 2022.
26. Roberts JP, Fisher TR, Trowbridge MJ, Bent C. A design thinking framework for healthcare management and innovation. Healthc (Amst) 2016 Mar;4(1):11-14. [doi: 10.1016/j.hjdsi.2015.12.002] [Medline: 27001093]
27. Braun V, Clarke V. Conceptual and design thinking for thematic analysis. Qual Psychol 2022 Feb 13;9(1):3-26. [doi: 10.1037/qup0000196]
28. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exercise Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
29. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
30. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. Int J Transgend Health 2023;24(1):1-6 [FREE Full text] [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
31. Kondylakis H, Axenie C, Kiran Bastola D, Katehakis DG, Kouroubali A, Kurz D, et al. Status and recommendations of technological and data-driven innovations in cancer care: focus group study. J Med Internet Res 2020 Dec 15;22(12):e22034 [FREE Full text] [doi: 10.2196/22034] [Medline: 33320099]
32. Papachristou N, Kotronoulas G, Dikaios N, Allison SJ, Eleftherochorinou H, Rai T, et al. Digital transformation of cancer care in the era of big data, artificial intelligence and data-driven interventions: navigating the field. Semin Oncol Nurs 2023 Jun;39(3):151433. [Medline: 37137770]
33. Bomhof-Roordink H, Fischer MJ, van Duijn-Bakker N, Baas-Thijssen MC, van der Weijden T, Stiggelbout AM, et al. Shared decision making in oncology: a model based on patients', health care professionals', and researchers' views. Psychooncology 2019 Jan 08;28(1):139-146. [doi: 10.1002/pon.4923] [Medline: 30346076]
34. Warner JL, Maddux SE, Hughes KS, Krauss JC, Yu PP, Shulman LN, et al. Development, implementation, and initial evaluation of a foundational open interoperability standard for oncology treatment planning and summarization. J Am Med Inform Assoc 2015 May 20;22(3):577-586 [FREE Full text] [doi: 10.1093/jamia/ocu015] [Medline: 25604811]
35. Steenbergen M, de Vries J, Arts R, Beerepoot LV, Traa MJ. Barriers and facilitators for shared decision-making in oncology inpatient practice: an explorative study of the healthcare providers' perspective. Support Care Cancer 2022 May;30(5):3925-3931. [doi: 10.1007/s00520-022-06820-1] [Medline: 35043216]
36. Perez H, Neubauer N, Marshall S, Philip S, Miguel-Cruz A, Liu L. Barriers and benefits of information communication technologies used by health care aides. Appl Clin Inform 2022 Jan 09;13(1):270-286 [FREE Full text] [doi: 10.1055/s-0042-1743238] [Medline: 35263800]
37. Bernstam EV, Warner JL, Krauss JC, Ambinder E, Rubinstein WS, Komatsoulis G, et al. Quantitating and assessing interoperability between electronic health records. J Am Med Inform Assoc 2022 Apr 13;29(5):753-760 [FREE Full text] [doi: 10.1093/jamia/ocab289] [Medline: 35015861]
38. Singh AP, Balogh EP, Carlson RW, Huizinga MM, Malin BA, Melamed A, et al. Re-envisioning electronic health records to optimize patient-centered cancer care, quality, surveillance, and research. JCO Oncol Pract 2025 Feb;21(2):128-135. [doi: 10.1200/op.24.00260]
39. Osterman TJ, Terry M, Miller RS. Improving cancer data interoperability: the promise of the minimal common oncology data elements (mCODE) initiative. JCO Clin Cancer Inform 2020 Nov;4:993-1001. [doi: 10.1200/cci.20.00059]
40. El Sabawy D, Feldman J, Pinto AD. The Connected Care for Canadians Act: an important step toward interoperability of health data. CMAJ 2024 Dec 08;196(42):E1385-E1388 [FREE Full text] [doi: 10.1503/cmaj.241123] [Medline: 39653400]
41. Simmons VM, Boman T, Stewart SA. Impact of the 21st Century Cures Act on patients and the healthcare team. Clin J Oncol Nurs 2024 Jan 22;28(1):21-25. [doi: 10.1188/24.cjon.21-25]
42. Petrovskaya O, Karpman A, Schilling J, Singh S, Wegren L, Caine V, et al. Patient and health care provider perspectives on patient access to test results via web portals: scoping review. J Med Internet Res 2023 Oct 19;25:e43765 [FREE Full text] [doi: 10.2196/43765] [Medline: 37856174]
43. Grewal M, Schlacter J, Martinez MJ, Eaton KP. A multidisciplinary initiative to increase patient portal enrollment. Am J Med Qual 2024;39(6):297-298. [doi: 10.1097/JMQ.0000000000000187] [Medline: 39773727]

44. Qureshi HA, Shah YA, Qureshi SM, Shah SU, Shiwlani A, Ahmad A. The promising role of artificial intelligence in navigating lung cancer prognosis. Int J Multidiscip Res 2024 Aug 27;6(4). [doi: 10.36948/ijfmr.2024.v06i04.26047]
45. Lichtner V, Franklin BD, Dalla-Pozza L, Westbrook JI. Electronic ordering and the management of treatment interdependencies: a qualitative study of paediatric chemotherapy. BMC Med Inform Decis Mak 2020 Aug 14;20(1):193 [FREE Full text] [doi: 10.1186/s12911-020-01212-z] [Medline: 32795356]
46. Acero MX, Minvielle E, Waelli M. Understanding the activity of oncology nurse coordinators: an elaboration of a framework based on an abductive approach. Health Policy 2023 Apr;130:104737 [FREE Full text] [doi: 10.1016/j.healthpol.2023.104737] [Medline: 36791597]
47. Saripalle R, Runyan C, Russell M. Using HL7 FHIR to achieve interoperability in patient health record. J Biomed Inform 2019 Jun;94:103188 [FREE Full text] [doi: 10.1016/j.jbi.2019.103188] [Medline: 31063828]
48. Osterman TJ, Terry M, Miller RS. Improving cancer data interoperability: the promise of the minimal common oncology data elements (mCODE) initiative. JCO Clin Cancer Inform 2020 Oct;4(4):993-1001 [FREE Full text] [doi: 10.1200/CCI.20.00059] [Medline: 33136433]
49. Gallo S. Chapter 18 - ICD-10 and SNOMED CT: the role of healthcare classification systems in digital health. In: Patel D, editor. Digital Health: Telemedicine and Beyond. Cambridge, MA: Academic Press; 2025.
50. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ 2025 Feb 05;388:e081554 [FREE Full text] [doi: 10.1136/bmj-2024-081554] [Medline: 39909534]
51. Herzog C, Blank S, Stahl BC. Towards trustworthy medical AI ecosystems – a proposal for supporting responsible innovation practices in AI-based medical innovation. AI Soc 2024 Oct 16. [doi: 10.1007/s00146-024-02082-z]
52. Jie Z, Zhiying Z, Li L. A meta-analysis of Watson for Oncology in clinical application. Sci Rep 2021 Mar 11;11(1):5792 [FREE Full text] [doi: 10.1038/s41598-021-84973-5] [Medline: 33707577]
53. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J Clin Oncol 2001 Feb 15;19(4):980-991. [doi: 10.1200/JCO.2001.19.4.980] [Medline: 11181660]
54. Langton JM, Pearson SA. eviQ cancer treatments online: how does the web-based protocol system fare in a comprehensive quality assessment? Asia Pac J Clin Oncol 2011 Dec 08;7(4):357-363. [doi: 10.1111/j.1743-7563.2011.01431.x] [Medline: 22151985]
55. Langton JM, Pesa N, Rushton S, Ward RL, Pearson SA. Uptake of a web-based oncology protocol system: how do cancer clinicians use eviQ cancer treatments online? BMC Cancer 2013 Mar 12;13(1):112 [FREE Full text] [doi: 10.1186/1471-2407-13-112] [Medline: 23497080]
56. What is Predict? Predict. URL: https://breast.v3.predict.cam/ [accessed 2025-05-01]
57. Interprofessional Shared Decision Making (IP-SDM) model. The Ottawa Hospital Research Institute. URL: https://decisionaid.ohri.ca/ip-sdm.html [accessed 2024-09-01]
58. Ottawa Decision Support Framework (ODSF). The Ottawa Hospital Research Institute. URL: https://decisionaid.ohri.ca/odsf.html [accessed 2024-09-01]
59. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4(4):CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
60. Yung A, Kay J, Beale P, Gibson KA, Shaw T. Computer-based decision tools for shared therapeutic decision-making in oncology: systematic review. JMIR Cancer 2021 Oct 26;7(4):e31616 [FREE Full text] [doi: 10.2196/31616] [Medline: 34544680]

Abbreviations

AI: artificial intelligence
EHR: electronic health record
SDM: shared decision-making

Edited by N Cahill; submitted 03.01.25; peer-reviewed by DP Touwen, A Jamal, A Okoronkwo; comments to author 25.02.25; revised version received 31.03.25; accepted 28.04.25; published 11.06.25.

Please cite as:
Yung A, Shaw T, Kay J, Janssen A
Examining How Technology Supports Shared Decision-Making in Oncology Consultations: Qualitative Thematic Analysis
JMIR Cancer 2025;11:e70827
URL: https://cancer.jmir.org/2025/1/e70827
doi:10.2196/70827
PMID:

Treatment Outcomes From Erlotinib and Gefitinib in Advanced Epidermal Growth Factor Receptor–Mutated Nonsquamous Non–Small Cell Lung Cancer in Aotearoa New Zealand From 2010 to 2020: Nationwide Whole-of-Patient-Population Retrospective Cohort Study

Phyu Sin Aye[1], PhD; Joanne Barnes[2], PhD; George Laking[3], PhD; Laird Cameron[4], MBChB; Malcolm Anderson[5], MBBS; Brendan Luey[6], PhD; Stephen Delany[7], MBChB; Dean Harris[8], MBChB; Blair McLaren[9], MBChB; Elliott Brenman[10], BM, BS; Jayden Wong[11], MBChB; Ross Lawrenson[12], MBBS; Michael Arendse[13], MMed; Sandar Tin Tin[14], PhD; Mark Elwood[14], DSc; Philip Hope[15]; Mark James McKeage[1], PhD

[1]Department of Pharmacology and Clinical Pharmacology, University of Auckland, Auckland, New Zealand

[10]Cancer and Haematology Services, Te Whatu Ora Health New Zealand Haora a Toi Bay of Plenty, Tauranga Hospital, Tauranga, New Zealand

[11]Cancer Services, Te Whatu Ora Health New Zealand Waikato, Waikato Hospital, Hamilton, New Zealand

[12]Medical Research Centre, University of Waikato, Hamilton, New Zealand

[13]Department of Pathology, Te Whatu Ora Health New Zealand Waikato, Waikato Hospital, Hamilton, New Zealand

[14]Department of Epidemiology and Biostatistics, University of Auckland, Auckland, New Zealand

[15]Lung Foundation New Zealand, Auckland, New Zealand

[2]School of Pharmacy, University of Auckland, Auckland, New Zealand

[3]Te Aka Mātauranga Matepukupuku Centre for Cancer Research, University of Auckland, Auckland, New Zealand

[4]Department of Medical Oncology, Te Pūriri o Te Ora Regional Cancer and Blood Service, Te Whatu Ora Health New Zealand, Auckland City Hospital, Auckland, New Zealand

[5]Department of Medical Oncology, Te Whatu Ora Health New Zealand Te Pae Hauuora o Ruahine o Tararua, Palmerston North Hospital, Palmerston North, New Zealand

[6]Wellington Blood and Cancer Centre, Te Whatu Ora Health New Zealand Capital, Coast and Hutt Valley, Wellington Hospital, Wellington, New Zealand

[7]Department of Oncology, Te Whatu Ora Health New Zealand Nelson Marlborough, Nelson Hospital, Nelson, New Zealand

[8]Oncology Service, Te Whatu Ora—Waitaha Canterbury, Christchurch Hospital, Christchurch, New Zealand

[9]Southern Blood and Cancer Service, Te Whatu Ora Southern, Dunedin Hospital, Dunedin, New Zealand

Corresponding Author:
Phyu Sin Aye, PhD
Department of Pharmacology and Clinical Pharmacology, University of Auckland, Auckland, New Zealand

Abstract

Background: Health care system–wide outcomes from routine treatment with erlotinib and gefitinib are incompletely understood.

Objective: The aim of the study is to describe the effectiveness of erlotinib and gefitinib during the first decade of their routine use for treating advanced epidermal growth factor receptor (*EGFR*) mutation-positive nonsquamous non–small cell lung cancer in the entire cohort of patients treated in Aotearoa New Zealand.

Methods: Patients were identified, and data collated from national pharmaceutical dispensing, cancer registration, and mortality registration electronic databases by deterministic data linkage using National Health Index numbers. Time-to-treatment discontinuation and overall survival were measured from the date of first dispensing of erlotinib or gefitinib and analyzed by Kaplan-Meier curves. Associations of treatment outcomes with baseline factors were evaluated using univariable and multivariable Cox regressions.

Results: Overall, 752 patients were included who started treatment with erlotinib (n=418) or gefitinib (n=334) before October 2020. Median time-to-treatment discontinuation was 11.6 (95% CI 10.8 - 12.4) months, and median overall survival was 20.1 (95% CI 18.1 - 21.6) months. Shorter time-to-treatment discontinuation was independently associated with high socioeconomic deprivation (hazard ratio [HR] 1.3, 95% CI 1.1 - 1.5 compared to the New Zealand Index of Deprivation 1 - 4 group), *EGFR* L858R mutations (HR 1.3, 95% CI 1.1 - 1.6 compared to exon 19 deletion), and distant disease at cancer diagnosis (HR 1.4, 95%

CI 1.2 - 1.7 compared to localized or regional disease). The same factors were independently associated with shorter overall survival. Outcome estimates and predictors remained unchanged in sensitivity analyses.

Conclusions: Outcomes from routine treatment with erlotinib and gefitinib in New Zealand patients with advanced *EGFR*-mutant nonsquamous non–small cell lung cancer are comparable with those reported in randomized trials and other health care system–wide retrospective cohort studies. Socioeconomic status, *EGFR* mutation subtype, and disease extent at cancer diagnosis were independent predictors of treatment outcomes in that setting.

Trial Registration: Australia New Zealand Clinical Trials Registry ACTRN12615000998549; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=368928&isReview=true

International Registered Report Identifier (IRRID): RR2-10.2196/51381

(JMIR Cancer 2025;11:e65118) doi:10.2196/65118

KEYWORDS

non–small cell lung cancer; mutations; epidemiology; target therapy; retrospective cohort study

Introduction

Lung cancer is the most common cause of cancer death in the world today [1]. Most clinical presentations of lung cancer are nonsquamous non–small cell lung cancer (NSCLC) [2]. Epidermal growth factor receptor (*EGFR*)–mutant nonsquamous NSCLC was the first type of lung cancer identified with an oncogenic driver that could be directly targeted by drug treatment [3-5].

The treatment of advanced *EGFR*-mutant nonsquamous NSCLC has evolved rapidly following the results of randomized controlled trials demonstrating improved progression-free survival. Initial randomized controlled trials established the superiority of first-generation *EGFR* kinase inhibitors, erlotinib and gefitinib, over platinum-doublet chemotherapy [6-10]. Subsequent randomized controlled trials established the superiority of second- and third-generation *EGFR* kinase inhibitors, including afatinib, dacomitinib, osimertinib, aumolertinib, and lazertinib, over those first-generation inhibitors [11-15]. Other randomized controlled trials compared erlotinib or gefitinib given alone or in combination with bevacizumab, ramucirumab, or chemotherapy [16-20]. In the 15 aforementioned randomized controlled trials, a total of 2257 patients with advanced *EGFR*-mutant nonsquamous NSCLC were allocated erlotinib or gefitinib monotherapy in control or experimental treatment arms. In those erlotinib or gefitinib monotherapy treatment arms, median progression-free survival ranged from 8.0 to 13.3 months. These clinical trial data provide a point of reference against which real-world studies of outcomes from treatment with erlotinib or gefitinib can be compared.

To fully understand outcomes from treatment with erlotinib and gefitinib in the setting of routine care, large-scale observational studies are required in addition to the extensive data already available from randomized controlled trials. Randomized controlled trials may have overestimated the benefits [21], and underestimated the harms [22], associated with the routine use of erlotinib and gefitinib. Compared to participants in randomized controlled trials, patients presenting for routine treatment with erlotinib or gefitinib are older, are of non-Asian ethnicity, have more comorbidities, have poorer performance status, and more often have brain metastasis. Randomized controlled trials have not evaluated many factors potentially impacting treatment outcomes, such as socioeconomic status. To improve their generalizability and avoid bias, observational studies of real-world outcomes from treatment with erlotinib and gefitinib could include all patients treated within a whole health care system or nation rather than being limited to those from 1 or a few institutions. To aid comparisons to clinical trial data, those observational studies could evaluate progression-free survival or proxies of progression-free survival rather than just overall survival, which is strongly influenced by factors other than treatment with erlotinib or gefitinib.

Only since 2019 have large-scale nationwide or health care system–wide studies reported real-world outcomes from routine treatment with erlotinib and gefitinib in patients with advanced *EGFR*-mutant NSCLC from Canada [23], the United States [24], Taiwan [25,26], Poland [27], Finland [28], and the Netherlands [29,30]. Among those aforementioned studies, 5 studies [24-28] reported progression-free survival or proxies of progression-free survival, such as time-to-treatment failure. In those 5 studies, median progression-free survival or its proxy ranged from 9.7 to 13.1 months. These observational data provide a point of reference against which other real-world studies of outcomes from treatment with erlotinib or gefitinib can be compared.

Starting in 2010, erlotinib and gefitinib were introduced into routine use in Aotearoa New Zealand for treating advanced lung cancer. The overall *EGFR* mutation positivity among patients with nonsquamous NSCLC who were tested was 22.5% in New Zealand [31]. To date, the effectiveness of erlotinib and gefitinib in the general population of New Zealand patients with lung cancer has not been described. With this background, this study aimed to describe the effectiveness of erlotinib and gefitinib during the first decade of their routine use for the treatment of advanced *EGFR*-mutant nonsquamous NSCLC in the entire cohort of patients treated in New Zealand. The study also aimed to evaluate associations between baseline factors and the effectiveness of erlotinib and gefitinib in this real-world setting.

Methods

Study Design and Participants

This was a nationwide, population-based, observational, data-linkage, retrospective cohort study that analyzed routinely collected health and administrative electronic data. The study group was a whole-of-population sample comprising a single group of patients. Patients were eligible for inclusion if they (1) were diagnosed with *EGFR*-mutant lung cancer, (2) dispensed erlotinib or gefitinib first before October 1, 2020, and (3) followed thereafter until death or for at least 1 year. Patients were excluded from the study if they had (1) erlotinib dispensed before January 1, 2014, or gefitinib dispensed before August 1, 2012, when positive *EGFR* mutation test results became mandatory for state-subsidized treatment; (2) no notification of a diagnosis of nonsquamous NSCLC in the New Zealand Cancer Registry; or (3) an unactionable or unknown *EGFR* mutation subtype.

Setting

From 2010 to 2020, New Zealand had a resident population ranging from approximately 4.3 to 5.1 million people, comprising predominately New Zealand European (70%), Māori (17%), Asian (15%), and Pacific people (8%) [32] (the total percent is greater than 100 because some people have more than 1 self-reported ethnicity). New Zealand residents were eligible for state-funded health care, including state-subsidized prescription medicines. Starting in 2010, the *EGFR* kinase inhibitor drugs erlotinib and gefitinib were introduced into routine clinical use in New Zealand for lung cancer treatment [33]. From October 1, 2010, to December 31, 2013, erlotinib was state-funded as a second-line treatment for advanced NSCLC, initially without any requirement for *EGFR* mutation testing. From August 1, 2012, gefitinib was state-funded as a first-line treatment for advanced *EGFR*-mutant NSCLC in New Zealand. On May 1, 2013, the National Health Committee of the New Zealand Ministry of Health issued recommendations for *EGFR* mutation testing in New Zealand, including testing of all patients with nonsquamous NSCLC at diagnosis irrespective of stage as part of standard pathology processes. From January 1, 2014, state funding for erlotinib was restricted to treating advanced *EGFR*-mutant NSCLC in New Zealand. During the first decade of routine use of erlotinib and gefitinib for lung cancer treatment in New Zealand, from 2010 to 2020, no other *EGFR* kinase inhibitor drugs were state-funded for use in New Zealand. During the period of study, treatment with erlotinib or gefitinib was provided by 10 public hospitals, and *EGFR* mutation testing was provided by 3 pathology laboratories in New Zealand.

Ethical Considerations

Ethics approval for this study was obtained from the New Zealand Government Ministry of Health Northern B Health and Disability Ethics Committee (reference 13/NTB/165/AM02). As the research retrospectively analyzed routinely collected data and did not involve direct contact with patients, the participants were not able or required to give informed consent by the ethics committee or governance groups who approved the study. The study used the identifiable data, which were password-protected, stored on the secured University of Auckland managed drive, and only accessible to the research team. The study was registered (ACTRN12615000998549). A study protocol and results of a validation substudy have been published [34].

Data Sources

Patients were identified, and data collated from national electronic pharmaceutical dispensing (Pharmaceutical Information Database [PHARMs]), cancer registration (New Zealand Cancer Registry), and mortality registration (National Mortality Collection) databases. Individual-level data on eligible cohort patients were compiled from these national electronic health databases by deterministic data linkage using each patient's unique National Health Index number. Additional data on eligible cohort patients were sourced from regional laboratory test data repositories, databases, and clinical records to determine the *EGFR* mutation status. A validation substudy demonstrated the feasibility and validity of using these national electronic health databases as the main source of data for this study [34].

Outcomes

The primary effectiveness outcome for this analysis was time-to-treatment discontinuation. Prescribing guidelines [35,36] recommend continuing daily treatment with erlotinib or gefitinib until disease progression, as long as treatment is safe and tolerable. In an analysis of randomized clinical trials submitted to the Food and Drug Administration, time-to-treatment discontinuation correlated well (r=0.91) with progression-free survival for patients with advanced *EGFR*-mutant NSCLC treated with *EGFR* kinase inhibitor drugs [37]. Time-to-treatment discontinuation is also less affected by subsequent cancer treatments and other factors that impact overall survival. Time-to-treatment discontinuation thereby reflects the duration of benefit from treatment with erlotinib or gefitinib. Time-to-treatment discontinuation was defined as the duration between the dates of the first dispensing and the last treatment with erlotinib or gefitinib. The date of last treatment with erlotinib or gefitinib was calculated by adding the number of days erlotinib or gefitinib dispensed for at the last dispensing to the date of the last dispensing, except when death occurred before the calculated date of last treatment, in which case the date of last treatment was the date of death. The secondary effectiveness outcome for this analysis was overall survival, defined as the duration between the date of first dispensing of erlotinib or gefitinib and death from any cause. A validation substudy had demonstrated the feasibility and validity of these methodologies for determining the outcomes of this study [34].

Variables

Baseline variables used for patient characterization included age, sex, ethnicity, geographical region of residence, smoking status, performance status, diagnosis year, NSCLC morphology, basis of NSCLC diagnosis, disease extent at cancer diagnosis, socioeconomic deprivation, rurality, comorbidity, choice of erlotinib or gefitinib for initial treatment, and *EGFR* mutation subtype. Socioeconomic deprivation was determined by mapping domicile codes recorded in the New Zealand Cancer Registry

to the 2006 New Zealand Index of Deprivation and was categorized into deciles with 1 being the least deprived and 10 being the most deprived [38]. Rurality was determined using the same domicile codes applied to Statistics New Zealand's Urban/Rural profile [39]. Comorbidity was assessed using a validated pharmacy-based comorbidity index for patients with cancer [40], modified for this study as previously described [34]. Ethnicity was classified into Asian, Māori, New Zealand European, or Pacific, and prioritized ethnicity was used if a registration listed multiple ethnicities (patients with more than 1 recorded ethnicity were allocated to a single ethnic group in order of priority: Māori, Pacific, Asian, and New Zealand European) [41]. *EGFR* mutation variants were classified according to the system of Koopman et al [42] into the following categories: (1) exon 19 deletion, (2) L858R, (3) uncommon actionable variant, (4) exon 20 insertion, and (5) nonactionable or unknown variant. Since *EGFR* mutation variant categories (4) and (5) were unactionable with erlotinib or gefitinib, patients with those variants were excluded from this study. A validation substudy had demonstrated the feasibility and validity of the methodologies used for determining the variables used for this study [34].

Literature Search

For comparing the results from this study to those from randomized controlled trials and other retrospective observational studies, a literature search was undertaken using a combination of the following MeSH terms: "carcinoma, non-small-cell lung," "ErbB receptors," "erlotinib hydrochloride," "gefitinib," "protein kinase inhibitors," and "mutation." Observational studies were included in this comparison if they were nationwide or health care system–wide studies and reported progression-free survival, or a proxy of progression-free survival, measured from the commencement of treatment with erlotinib or gefitinib [24-28]. Institution-based studies [43] and those not reporting progression-free survival or a proxy of progression-free survival [23,29,30] were excluded from these comparisons.

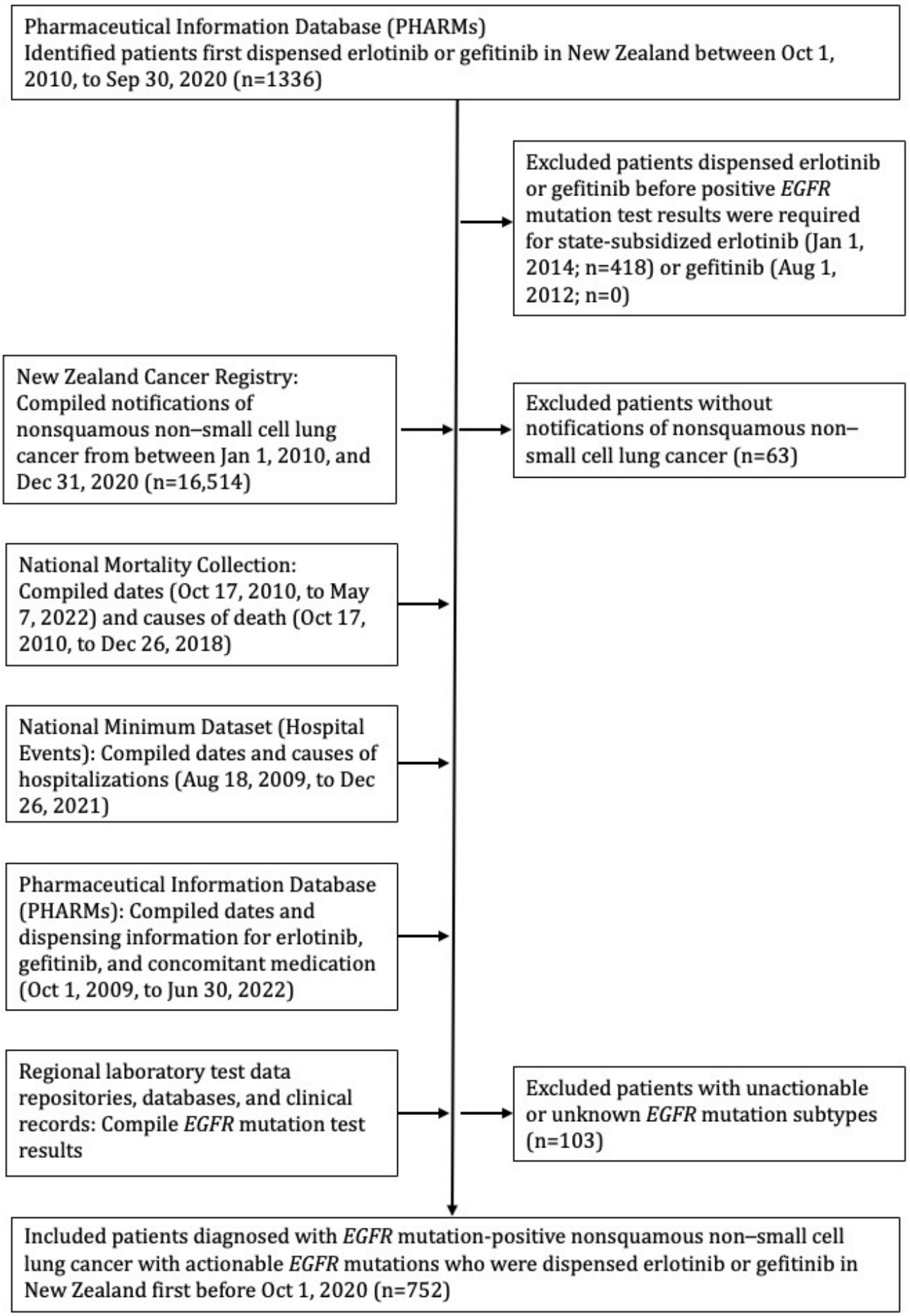
Statistical Analysis

Descriptive statistics were used to analyze the demographic profile and baseline characteristics of the retrospective cohort. Time-to-treatment discontinuation and overall survival were analyzed by Kaplan-Meier curves, and survival differences between subgroups were assessed using log-rank tests. Patients with no known dates of last treatment or death were censored at the date of last follow-up of dispensing (June 30, 2022) or survival (May 7, 2022), respectively. To assess the robustness of estimates of time-to-treatment discontinuation and overall survival, sensitivity analyses were carried out in an expanded study cohort (n=885) that included patients with no registration of nonsquamous NSCLC and those with unknown or nonactionable *EGFR* mutation subtypes, except those with exon 20 insertions. Associations between baseline factors and time-to-treatment discontinuation or overall survival were evaluated by univariable and multivariable Cox regression models to compute hazard ratios and their 95% CIs and *P* values. Baseline factors selected for univariable and multivariable analyses included age, sex, disease morphology, disease extent, *EGFR* mutation subtype, and initial choice of *EGFR* kinase inhibitor drug, which had been identified as independent predictors of outcomes in previous studies [44,45], and ethnicity, comorbidity, socioeconomic deprivation, and residential status (urban vs or rural, and region), which had not been previously evaluated in the New Zealand patient population. There were complete data for all those factors for all 752 cohort patients. Smoking and Eastern Cooperative Oncology Group performance status were excluded from the univariable and multivariable analyses due to high levels of missing data (>50%). Missing extent of disease at cancer diagnosis and ethnicity data were included in univariable and multivariable analyses by adding an unknown category for each of these variables comprising <20% and <1% of patients, respectively. Otherwise, data were complete for all other factors for all 752 patients. Factors were selected for multivariable analyses if they had statistically significant associations with the outcome of interest on univariable analysis. To assess the robustness of the findings of multivariable analyses, sensitivity analyses were carried out using less stringent criteria for factor inclusion. Differences were considered statistically significant when *P* values were less than .05. Data analyses were performed using Stata (version 16; StataCorp LLC).

Results

The assembly of the retrospective cohort and compilation of study data from national electronic health databases was carried out as shown in Figure 1. From the PHARMs, 1336 patients were identified who had been dispensed erlotinib or gefitinib first between October 1, 2010, and September 30, 2020. A total of 418 of those patients were excluded because they were first dispensed erlotinib or gefitinib before positive *EGFR* mutation test results became mandatory for access to state-subsidized erlotinib or gefitinib in New Zealand, leaving 918 potentially eligible patients. From the New Zealand Cancer Registry, 16,516 patients were identified with notifications of nonsquamous NSCLC diagnoses made between January 1, 2010, and December 30, 2020. Of the 918 potentially eligible patients, 63 did not have notifications of diagnoses of nonsquamous NSCLC recorded in the New Zealand Cancer Registry and were excluded, leaving 855 potentially eligible patients. Dates and causes of death and hospitalizations and full dispensing information for erlotinib, gefitinib, and concomitant medications were compiled on those patients from the National Mortality Collection, National Minimum Dataset (Hospital Events), and PHARMs, respectively. *EGFR* mutation test results, smoking status, and Eastern Cooperative Oncology Group performance status were then compiled from regional laboratory test data repositories, databases, and clinical records. Of 855 potentially eligible patients, 103 patients had unactionable or unknown *EGFR* mutation variants, including 33 patients with *EGFR* exon 20 insertions, and were excluded. Finally, 752 patients remained, who had been diagnosed with *EGFR*-mutant nonsquamous NSCLC with actionable *EGFR* mutation variants, and had started treatment with erlotinib or gefitinib prior to October 2020 for inclusion in this study.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the assembly of the cohort and data collation. *EGFR*: epidermal growth factor receptor.



The baseline characteristics of the retrospective 752 patient cohorts are shown in Table 1. Their mean age was 67 (SD 12) years, and 67% (n=504) were female. About one-quarter were Asian, half New Zealand European, one-quarter Pacific or Māori, and 1.1% (n=8) had other or unknown ethnicity. Most had adenocarcinoma and *EGFR* exon 19 deletions or L858R mutations. The extent of disease at cancer diagnosis was available for only 604 (80.3%) patients, most of whom had distant disease at cancer diagnosis. Smoking status was available for only 301 (40.1%) patients, most of whom were nonsmokers. Performance status was available for only 273 (36.6%) patients, most of whom had limited performance status.

Table . Patient characteristics (N=752).

		Values, n (%)
Age (years)[a]		
	<65	318 (42.3)
	65+	434 (57.7)
Sex		
	Male	248 (33)
	Female	504 (67)
Ethnicity		
	Asian	190 (25.3)
	Māori	73 (9.7)
	New Zealand European	392 (52.1)
	Pacific	89 (11.8)
	Other and unknown	8 (1.1)
Region		
	Northern	355 (47.2)
	Midland	103 (13.7)
	Others	294 (39.1)
Smoking		
	Ex-smoker	108 (14.4)
	Nonsmoker	160 (21.3)
	Current smoker	33 (4.4)
	Unknown	451 (59.9)
ECOG[b] **performance status**		
	Fully active (0)	124 (16.5)
	Limited (1-4)	151 (20.1)
	Unknown	477 (63.4)
Diagnosis year		
	2010 - 2013	120 (16)
	2014 - 2016	285 (37.9)
	2017 - 2020	347 (46.1)
Morphology		
	Adenocarcinoma	662 (88)
	Unspecified and other	90 (12)
Basis of diagnosis		
	Histology	459 (61)
	Cytology	275 (36.6)
	Other	18 (2.4)
Extent		
	Localized or regional	136 (18.1)
	Distant	468 (62.2)
	Unknown	148 (19.7)
Deprivation		

		Values, n (%)
	NZDep[c] 1 - 4	305 (40.6)
	NZDep 5 - 7	218 (29)
	NZDep 8 - 10	229 (30.5)
Rurality		
	Urban	652 (86.7)
	Rural	100 (13.3)
Comorbidity		
	No	190 (25.3)
	Yes	562 (74.7)
***EGFR*[d] type**		
	Exon 19 del	424 (56.4)
	Exon 21 L858R	256 (34)
	Uncommon or actionable	72 (9.6)
***EGFR*-TKI[e]**		
	Gefitinib	334 (44.4)
	Erlotinib	418 (55.6)

[a]Mean age 67 (SD 12, range 24 - 92) years.

[b]ECOG: Eastern Cooperative Oncology Group.

[c]NZDep: New Zealand Index of Deprivation.

[d]*EGFR*: epidermal growth factor receptor.

[e]TKI: tyrosine kinase inhibitor.

At the date of the last dispensing follow-up (June 30, 2022), treatment with erlotinib and gefitinib had been discontinued in 724 (96.3%) patients and was continuing in 28 (3.7%) patients. Treatment was discontinued prior to death in 618 patients and at the time of death in 103 patients. Median time-to-treatment discontinuation was 11.6 (95% CI 10.8-12.4) months. The 1-, 2-, and 5-year rates of treatment continuation were 47.3% (95% CI 34.7%-50.9%), 17.4% (95% CI 14%-20.2%), and 3.4% (95% CI 2.2%-5.1%), respectively. Sensitivity analysis in an expanded study cohort (n=885) gave similar results for median time-to-treatment discontinuation (11.1, 95% CI 10.1-11.8 months). Univariable analysis (Table 2) showed that shorter time-to-treatment discontinuation was associated with socioeconomic deprivation, *EGFR* L858R mutations, distant disease at cancer diagnosis, and adenocarcinoma morphology. The initial choice of *EGFR* kinase inhibitor (erlotinib or gefitinib), age, sex, ethnicity, geographical region, year of diagnosis, basis of diagnosis, rurality, and comorbidity were not associated with time-to-treatment discontinuation on univariable analysis. Multivariable analysis showed that shorter time-to-treatment discontinuation was independently associated with socioeconomic deprivation, *EGFR* L858R mutations, and distant disease at cancer diagnosis (Table 2). Sensitivity analyses using less stringent criteria for factor inclusion identified the same independent predictors of time-to-treatment discontinuation (Table S3 in Multimedia Appendix 1).

Table . Univariable and multivariable analysis of time-to-treatment discontinuation.

		Univariable analysis		Multivariable analysis	
		HR[a] (95% CI)	*P* value	HR (95% CI)	*P* value
EGFR[b]**-TKI**[c]					
	Gefitinib	1.1 (1.0 - 1.3)	.07	—[d]	—
	Erlotinib	1.0 (—)	—	—	—
Age (years)					
	<65	1.0 (0.9 - 1.2)	.63	—	—
	65+	1.0 (—)	—	—	—
Sex					
	Male	1.0 (0.8 - 1.1)	.74	—	—
	Female	1.0 (—)	—	—	—
Ethnicity					
	Asian	0.9 (0.7 - 1.0)	.14	—	—
	Māori	1.1 (0.9 - 1.4)	.37	—	—
	New Zealand European	1.0 (—)	—	—	—
	Pacific	1.0 (0.8 - 1.2)	.76	—	—
	Other and unknown	0.8 (0.4 - 1.7)	.58	—	—
Region					
	Northern	1.0 (—)	—	—	—
	Midland	1.1 (0.9 - 1.4)	.45	—	—
	Others	1.0 (0.9 - 1.2)	.95	—	—
Diagnosis year					
	2010 - 2013	1.0 (0.8 - 1.2)	.92	—	—
	2014 - 2016	0.9 (0.8 - 1.1)	.17	—	—
	2017 - 2020	1.0 (—)	—	—	—
Morphology					
	Adenocarcinoma	1.0 (—)	—	1.0 (—)	—
	Unspecified and other	0.8 (0.6 - 1.0)	.04	0.9 (0.7 - 1.1)	.25
Basis of diagnosis					
	Histology	1.0 (—)	—	—	—
	Cytology	1.1 (0.9 - 1.3)	.29	—	—
	Other	0.9 (0.5 - 1.4)	.58	—	—
Extent					
	Localized or regional	1.0 (—)	—	1.0 (—)	—
	Distant	1.5 (1.2 - 1.8)	<.001	1.4 (1.2 - 1.7)	.001
	Unknown	0.9 (0.7 - 1.1)	.38	0.9 (0.7 - 1.1)	.33
Deprivation					
	NZDep[e] 1 - 4	1.0 (—)	—	1.0 (—)	—
	NZDep 5 - 7	1.2 (1.0 - 1.5)	.02	1.2 (1.0 - 1.4)	.06
	NZDep 8 - 10	1.3 (1.1 - 1.6)	.004	1.3 (1.1 - 1.5)	.005
Rurality					
	Urban	1.0 (—)	—	—	—

		Univariable analysis		Multivariable analysis	
		HR[a] (95% CI)	*P* value	HR (95% CI)	*P* value
	Rural	1.1 (0.9 - 1.4)	.39	—	—
Comorbidity					
	No	1.2 (1.0 - 1.4)	.11	—	—
	Yes	1.0 (—)	—	—	—
***EGFR* type**					
	Exon 19 deletion	1.0 (—)	—	1.0 (—)	—
	Exon 21 L858R	1.3 (1.1 - 1.5)	.001	1.3 (1.1 - 1.6)	<.001
	Uncommon or actionable	1.2 (0.9 - 1.6)	.17	1.3 (1.0 - 1.6)	.07

[a]HR: hazard ratio.

[b]*EGFR*: epidermal growth factor receptor.

[c]TKI: tyrosine kinase inhibitor.

[d]Not applicable.

[e]NZDep: New Zealand Index of Deprivation.

At the date of last survival follow-up (May 7, 2022), 614 (81.6%) patients had died, and 138 (18.4%) patients were alive. Median overall survival was 20.1 (95% CI 18.1-21.6) months. The 1-, 2-, and 5-year overall survival rates were 69.2% (95% CI 65.8%-72.4%), 43% (95% CI 37.4%-43.5%), and 13.9% (95% CI 11.2%-17%), respectively. Sensitivity analysis in an expanded study cohort (n=885) gave similar results for median overall survival (19.4, 95% CI 17.8-21.2 months). Univariable analysis (Table 3) showed shorter overall survival in association with socioeconomic deprivation, *EGFR* L858R mutations, distant disease at cancer diagnosis, initial choice of *EGFR* kinase inhibitor of gefitinib (vsversus erlotinib), age >65 years, non-Asian ethnicity, residence outside the Northern or Midlands regions, and adenocarcinoma morphology. Sex, diagnosis year, basis of diagnosis, rurality, and comorbidity were not associated with overall survival on univariable analysis. Multivariable analysis showed that shorter overall survival was independently associated with socioeconomic deprivation, *EGFR* L858R mutations, distant disease at cancer diagnosis, and non-Asian or non-Pacific ethnicities (Table 3). Sensitivity analyses using less stringent criteria for factor inclusion identified the same independent predictors of overall survival (Table S4 in Multimedia Appendix 1).

Table . Univariable and multivariable analysis of overall survival.

		Univariable analysis		Multivariable analysis	
		HR[a] (95% CI)	*P* value	HR (95% CI)	*P* value
EGFR[b]**-TKI**[c]					
	Gefitinib	1.2 (1.0 - 1.4)	.02	1.2 (1.0 - 1.4)	.10
	Erlotinib	1.0 (—[d])	—	1.0 (—)	—
Age (years)					
	<65	0.9 (0.7 - 1.0)	.048	0.9 (0.7 - 1.0)	.14
	65+	1.0 (—)	—	1.0 (—)	—
Sex					
	Male	1.1 (0.9 - 1.3)	.48	—	—
	Female	1.0 (—)	—	—	—
Ethnicity					
	Asian	0.7 (0.5 - 0.8)	<.001	0.7 (0.6 - 0.9)	<.001
	Māori	1.2 (0.9 - 1.5)	.26	1.3 (1.0 - 1.7)	.05
	New Zealand European	1.0 (—)	—	1.0 (—)	—
	Pacific	0.8 (0.6 - 1.0)	.06	0.8 (0.6 - 1.0)	.046
	Other and unknown	0.5 (0.2 - 1.3)	.15	0.5 (0.2 - 1.3)	.16
Region					
	Northern	1.0 (—)	—	1.0 (—)	—
	Midland	1.2 (0.9 - 1.5)	.18	1.1 (0.8 - 1.4)	.57
	Others	1.4 (1.1 - 1.6)	<.001	1.2 (1.0 - 1.5)	.06
Diagnosis year					
	2010 - 2013	1.2 (0.9 - 1.4)	.24	—	—
	2014 - 2016	1.0 (0.8 - 1.2)	.82	—	—
	2017 - 2020	1.0 (—)	—	—	—
Morphology					
	Adenocarcinoma	1.0 (—)	—	1.0 (—)	—
	Unspecified and other	0.7 (0.6 - 1.0)	.02	0.9 (0.7 - 1.2)	.4
Basis of diagnosis					
	Histology	1.0 (—)	—	—	—
	Cytology	1.1 (0.9 - 1.3)	.41	—	—
	Other	1.4 (0.9 - 2.3)	.16	—	—
Extent					
	Localized or regional	1.0 (—)	—	1.0 (—)	—
	Distant	1.7 (1.4 - 2.2)	<.001	1.8 (1.4 - 2.2)	<.001
	Unknown	1.0 (0.8 - 1.3)	.98	1.0 (0.7 - 1.3)	.82
Deprivation					
	NZDep[e] 1 - 4	1.0 (—)	—	1.0 (—)	—
	NZDep 5 - 7	1.4 (1.2 - 1.7)	.001	1.3 (1.1 - 1.6)	.006
	NZDep 8 - 10	1.4 (1.1 - 1.7)	.001	1.4 (1.1 - 1.7)	.004
Rurality					
	Urban	1.0 (—)	—	—	—

		Univariable analysis		Multivariable analysis	
		HR[a] (95% CI)	*P* value	HR (95% CI)	*P* value
	Rural	1.2 (1.0 - 1.5)	.11	—	—
Comorbidity					
	No	1.0 (0.9 - 1.3)	.67	—	—
	Yes	1.0 (—)	—	—	—
***EGFR* type**					
	Exon 19 deletion	1.0 (—)	—	1.0	—
	Exon 21 L858R	1.4 (1.2 - 1.6)	<.001	1.5 (1.2 - 1.7)	<.001
	Uncommon or actionable	1.3 (1.0 - 1.7)	.09	1.2 (0.9 - 1.6)	.18

[a]HR: hazard ratio.

[b]*EGFR*: epidermal growth factor receptor.

[c]TKI: tyrosine kinase inhibitor.

[d]Not applicable.

[e]NZDep: New Zealand Index of Deprivation.

For the purpose of comparison of the results from this study to the existing literature, our literature search identified 15 randomized controlled trials and 5 nationwide or health care system–wide retrospective observational studies. These randomized controlled trials showed that the median progression-free survival for erlotinib and gefitinib monotherapy treatment arms ranged from 8.0 to 13.3 months (Table S1 in Multimedia Appendix 1). The retrospective observational studies showed that the median progression-free survival, or its proxy, ranged from 9.7 to 13.1 months, and the median overall survival ranged from 17.5 to 23.9 months (Table S2 in Multimedia Appendix 1). Our study reports the results following the RECORD (Reporting of Studies Conducted Using Observational Routinely-Collected Health Data) statement checklist (Table S5 in Multimedia Appendix 1).

Discussion

Principal Findings and Comparison to Prior Work

The outcomes from treatment with erlotinib and gefitinib in this study of 752 patients with advanced *EGFR*-mutant nonsquamous NSCLC, treated between 2010 and 2020 in New Zealand, corresponded with those reported in randomized controlled trials and in other large-scale health care system–wide retrospective cohort analyses. The median time-to-treatment discontinuation of 11.6 months found in this study paralleled the median progression-free survival values reported for erlotinib and gefitinib monotherapy treatment arms of 15 randomized controlled trials [6-20], which ranged from 8.0 to 13.3 months (Table S1 in Multimedia Appendix 1). It also paralleled the median progression-free survival values, or its proxy, reported in other nationwide or health care system–wide observational studies of similar patient groups from elsewhere [24-28] (range of median progression-free survival or proxy 9.7 to 13.1 months; Table S2 in Multimedia Appendix 1). The median overall survival of 20.1 months found in this study was also within the range reported in other nationwide or health care system–wide observational studies of similar patient groups [24-28] (range of median overall survival 17.5 to 23.9 months; Table S2 in Multimedia Appendix 1). In this way, this retrospective study has confirmed that the therapeutic benefits expected from erlotinib and gefitinib had been conveyed into the setting of routine care in New Zealand.

EGFR mutation subtype was an independent predictor of outcomes from treatment with erlotinib and gefitinib in this study. Study patients were stratified according to whether their tumors had exon 19 deletions (56%), L858R mutations (34%), or other actionable *EGFR* mutations (10%). Compared to those with exon 19 deletions, study patients with L858R mutations had 30% and 50% increased risks of treatment discontinuation and death, respectively, after commencing treatment with erlotinib or gefitinib. This finding is consistent with those of previous studies exploring outcomes from erlotinib or gefitinib in similar patient groups [45]. *EGFR* mutation subtype may have impacted upon treatment outcomes in this study via the higher pharmacological potency of erlotinib and gefitinib for inhibiting exon 19 deletion *EGFR* oncoproteins compared to those associated with L858R or other *EGFR* mutations [46,47].

Socioeconomic deprivation was an independent predictor of outcomes from treatment with erlotinib and gefitinib in this study. Study patients were stratified into groups with low (41%), intermediate (29%), or high socioeconomic deprivation (30%) based on their residential area. Compared to the study patients from high socioeconomic areas, those from low socioeconomic areas had 30% and 50%, and those from intermediate socioeconomic areas had 20% and 30%, increased risks of treatment discontinuation and death, respectively, after commencing treatment with erlotinib or gefitinib. People from low socioeconomic areas are known to have poorer outcomes from lung cancer due to more limited access to screening and diagnostic services that lead to delayed diagnoses and more advanced disease at presentation [48]. However, few previous studies have evaluated the impacts of socioeconomic deprivation on outcomes from treatment with erlotinib, gefitinib, or other systemic anticancer therapies in patients with *EGFR*-mutant or

other forms of advanced lung cancer. A pooled analysis of SWOG Cancer Research Network clinical trials showed significant associations between socioeconomic deprivation and lower progression-free survival, including in a subgroup of 1307 patients with stage IV NSCLC treated with various platinum-based chemotherapy regimens in randomized clinical trials [49]. Socioeconomic deprivation may have impacted treatment outcomes in this study by limiting access to health care during treatment with erlotinib and gefitinib, directly via other yet to be defined mechanisms or indirectly through correlated predictive factors not accounted for in the multivariable analyses, such as smoking and performance status. Future studies should more closely evaluate the impacts of socioeconomic deprivation on outcomes from the treatment of advanced lung cancer.

Disease extent at cancer diagnosis was an independent predictor of outcomes of treatment with erlotinib and gefitinib in this study. Study patients were stratified according to whether they had localized or regional (18%), distant (62%), or unknown extent of disease (20%) at the time of notification of their diagnosis of nonsquamous NSCLC to the New Zealand Cancer Registry. Compared to those with localized or regional disease extent, study patients with distant disease at diagnosis had 40% and 80% increased risks of treatment discontinuation and death, respectively, after commencing treatment with erlotinib or gefitinib. This finding was consistent with previous studies demonstrating the negative impacts of distant metastasis on outcomes from treatment with erlotinib and gefitinib in similar patient groups [44].

Ethnicity was an independent predictor of overall survival, but not of time-to-treatment discontinuation, in this study. Study patients were categorized as Asian (25%), Māori (10%), New Zealand European (52%), Pacific (12%), or unknown or other ethnicity (1%). Time-to-treatment discontinuation was unchanged among these different ethnic groups when compared to New Zealand European group. However, the risk of death was reduced by 20% and 30%, respectively, in the Pacific and Asian groups but unchanged in the other groups compared to New Zealand European. Overall survival may have been impacted by ethnicity in this study independently of the effectiveness of treatment with erlotinib or gefitinib. Ethnicity may have impacted overall survival indirectly through correlated factors, such as smoking status, that vary between ethnic groups and influence the risk of death [50].

Strengths and Limitations

The strengths of this study include its large population-based sample, internal validity, national generalizability, and unique patient cohort. Only 4 similar analyses [25-28] have included all patients treated in an entire country as far as we are aware. Limitations of the study include those inherent in retrospective study designs or in the use of routinely collected data. The variables available for analysis were limited to those collected routinely during pharmaceutical dispensing and cancer and mortality registration. Some important variables were unavailable or incomplete, such as smoking status, performance status, and clinical stage of disease at the time of commencing treatment with erlotinib and gefitinib, and therefore could not be included in the multivariable analysis. Socioeconomic deprivation was determined by residential area rather than at an individual level, which may have introduced bias. The study did not evaluate the impact of treatments other than erlotinib and gefitinib, which may have influenced overall survival. Safety outcomes were not included in this analysis but will be the subject of subsequent reports.

Conclusions

Outcomes from treatment with erlotinib and gefitinib in this New Zealand cohort of patients with advanced *EGFR*-mutant nonsquamous NSCLC were comparable to those reported in randomized controlled trials and other large-scale health care system–wide retrospective cohort studies. This nationwide study thereby demonstrated that the therapeutic benefits expected from erlotinib and gefitinib had been achieved in the setting of routine care in New Zealand. In that setting, socioeconomic status, *EGFR* mutation subtype, and disease extent at cancer diagnosis were independent predictors of treatment outcomes.

Acknowledgments

The authors thank the patients and their families for contributing to this study. This study has been funded by the Health Research Council of New Zealand project (grant 19-450).

Data Availability

The datasets generated and analyzed during this study are not publicly available due to patient information being identifiable but are available from the corresponding author on reasonable request. For ethics queries, please contact the Health and Disability Ethics Committees at hdecs@health.govt.nz.

Authors' Contributions

PSA, JB, GL, LC, MA, BL, SD, DH, STT, ME, PH, and MJM were involved in conceptualization and funding acquisition. PSA and MJM were involved in validation, writing original draft, and visualization. GL, LC, ME, BL, SD, DH, BM, EB, JW, RL, MA, and MJM were involved in data curation and supervision. JB, GL, and PH were involved in supervision. PSA, STT, and ME were involved in formal analysis and methodology. All authors reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[PDF File, 800 KB - cancer_v11i1e65118_app1.pdf]

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263. [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. Lancet 2021 Aug 7;398(10299):535-554. [doi: 10.1016/S0140-6736(21)00312-3] [Medline: 34273294]
3. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004 May 20;350(21):2129-2139. [doi: 10.1056/NEJMoa040938] [Medline: 15118073]
4. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004 Jun 4;304(5676):1497-1500. [doi: 10.1126/science.1099314] [Medline: 15118125]
5. Pao W, Miller V, Zakowski M, et al. EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci U S A 2004 Sep 7;101(36):13306-13311. [doi: 10.1073/pnas.0405220101] [Medline: 15329413]
6. Maemondo M, Inoue A, Kobayashi K, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. N Engl J Med 2010 Jun 24;362(25):2380-2388. [doi: 10.1056/NEJMoa0909530] [Medline: 20573926]
7. Mitsudomi T, Morita S, Yatabe Y, et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol 2010 Feb;11(2):121-128. [doi: 10.1016/S1470-2045(09)70364-X] [Medline: 20022809]
8. Zhou C, Wu YL, Chen G, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011 Aug;12(8):735-742. [doi: 10.1016/S1470-2045(11)70184-X] [Medline: 21783417]
9. Rosell R, Carcereny E, Gervais R, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012 Mar;13(3):239-246. [doi: 10.1016/S1470-2045(11)70393-X] [Medline: 22285168]
10. Wu YL, Zhou C, Liam CK, et al. First-line erlotinib versus gemcitabine/cisplatin in patients with advanced EGFR mutation-positive non-small-cell lung cancer: analyses from the phase III, randomized, open-label, ENSURE study. Ann Oncol 2015 Sep;26(9):1883-1889. [doi: 10.1093/annonc/mdv270] [Medline: 26105600]
11. Park K, Tan EH, O’Byrne K, et al. Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): a phase 2B, open-label, randomised controlled trial. Lancet Oncol 2016 May;17(5):577-589. [doi: 10.1016/S1470-2045(16)30033-X] [Medline: 27083334]
12. Wu YL, Cheng Y, Zhou X, et al. Dacomitinib versus gefitinib as first-line treatment for patients with EGFR-mutation-positive non-small-cell lung cancer (ARCHER 1050): a randomised, open-label, phase 3 trial. Lancet Oncol 2017 Nov;18(11):1454-1466. [doi: 10.1016/S1470-2045(17)30608-3] [Medline: 28958502]
13. Soria JC, Ohe Y, Vansteenkiste J, et al. Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. N Engl J Med 2018 Jan 11;378(2):113-125. [doi: 10.1056/NEJMoa1713137] [Medline: 29151359]
14. Lu S, Dong X, Jian H, et al. AENEAS: a randomized phase III trial of aumolertinib versus gefitinib as first-line therapy for locally advanced or metastatic non-small-cell lung cancer with EGFR exon 19 deletion or L858R mutations. J Clin Oncol 2022 Sep 20;40(27):3162-3171. [doi: 10.1200/JCO.21.02641] [Medline: 35580297]
15. Cho BC, Ahn MJ, Kang JH, et al. Lazertinib versus gefitinib as first-line treatment in patients with EGFR-mutated advanced non-small-cell lung cancer: results from LASER301. J Clin Oncol 2023 Sep 10;41(26):4208-4217. [doi: 10.1200/JCO.23.00515] [Medline: 37379502]
16. Seto T, Kato T, Nishio M, et al. Erlotinib alone or with bevacizumab as first-line therapy in patients with advanced non-squamous non-small-cell lung cancer harbouring EGFR mutations (JO25567): an open-label, randomised, multicentre, phase 2 study. Lancet Oncol 2014 Oct;15(11):1236-1244. [doi: 10.1016/S1470-2045(14)70381-X] [Medline: 25175099]
17. Nakagawa K, Garon EB, Seto T, et al. Ramucirumab plus erlotinib in patients with untreated, EGFR-mutated, advanced non-small-cell lung cancer (RELAY): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol 2019 Dec;20(12):1655-1669. [doi: 10.1016/S1470-2045(19)30634-5] [Medline: 31591063]
18. Saito H, Fukuhara T, Furuya N, et al. Erlotinib plus bevacizumab versus erlotinib alone in patients with EGFR-positive advanced non-squamous non-small-cell lung cancer (NEJ026): interim analysis of an open-label, randomised, multicentre, phase 3 trial. Lancet Oncol 2019 May;20(5):625-635. [doi: 10.1016/S1470-2045(19)30035-X] [Medline: 30975627]

19. Hosomi Y, Morita S, Sugawara S, et al. Gefitinib alone versus gefitinib plus chemotherapy for non-small-cell lung cancer with mutated epidermal growth factor receptor: NEJ009 study. J Clin Oncol 2020 Jan 10;38(2):115-123. [doi: 10.1200/JCO.19.01488] [Medline: 31682542]
20. Noronha V, Patil VM, Joshi A, et al. Gefitinib versus gefitinib plus pemetrexed and carboplatin chemotherapy in EGFR-mutated lung cancer. J Clin Oncol 2020 Jan 10;38(2):124-136. [doi: 10.1200/JCO.19.01154] [Medline: 31411950]
21. Cramer-van der Welle CM, Peters BJM, Schramel FMNH, et al. Systematic evaluation of the efficacy-effectiveness gap of systemic treatments in metastatic nonsmall cell lung cancer. Eur Respir J 2018 Dec;52(6):1801100. [doi: 10.1183/13993003.01100-2018] [Medline: 30487206]
22. Novello S, Capelletto E, Cortinovis D, et al. Italian multicenter survey to evaluate the opinion of patients and their reference clinicians on the “tolerance” to targeted therapies already available for non-small cell lung cancer treatment in daily clinical practice. Transl Lung Cancer Res 2014 Jun;3(3):173-180. [doi: 10.3978/j.issn.2218-6751.2014.06.10] [Medline: 25806297]
23. Lau SC, Chooback N, Ho C, Melosky B. Outcome differences between first- and second-generation EGFR inhibitors in advanced EGFR mutated NSCLC in a large population-based cohort. Clin Lung Cancer 2019 Sep;20(5):e576-e583. [doi: 10.1016/j.cllc.2019.05.003] [Medline: 31178389]
24. Li YL, Appius A, Pattipaka T, Feyereislova A, Cassidy A, Ganti AK. Real-world management of patients with epidermal growth factor receptor (EGFR) mutation-positive non–small-cell lung cancer in the USA. PLoS ONE 2019;14(1):e0209709. [doi: 10.1371/journal.pone.0209709]
25. Hsieh YY, Fang WT, Lo YW, Chen YH, Chien LN. Comparing the effectiveness of different EGFR-TKIs in patients with EGFR mutant non-small-cell lung cancer: a retrospective cohort study in Taiwan. Int J Cancer 2020 Aug 15;147(4):1107-1116. [doi: 10.1002/ijc.32841] [Medline: 31854456]
26. Chen PY, Wang CC, Hsu CN, Chen CY. Association of EGFR tyrosine kinase inhibitor treatment with progression-free survival among Taiwanese patients with advanced lung adenocarcinoma and EGFR mutation. Front Pharmacol 2021;12:720687. [doi: 10.3389/fphar.2021.720687] [Medline: 34434112]
27. Pluzanski A, Krzakowski M, Kowalski D, Dziadziuszko R. Real-world clinical outcomes of first-generation and second-generation epidermal growth factor receptor tyrosine kinase inhibitors in a large cohort of European non-small-cell lung cancer patients. ESMO Open 2020 Nov;5(6):e001011. [doi: 10.1136/esmoopen-2020-001011] [Medline: 33148621]
28. Manninen O, Puuniemi L, Iivanainen S, Arffman M, Kaarteenaho R, Koivunen JP. Treatment outcomes of non-small cell lung cancers treated with EGFR tyrosine kinase inhibitors: a real-world cohort study. Acta Oncol 2023 Dec 2;62(12):1854-1861. [doi: 10.1080/0284186X.2023.2274481]
29. Gijtenbeek RGP, Damhuis RAM, Groen HJM, van der Wekken AJ, van Geffen WH. Nationwide real-world cohort study of first-line tyrosine kinase inhibitor treatment in epidermal growth factor receptor-mutated non-small-cell lung cancer. Clin Lung Cancer 2020 Nov;21(6):e647-e653. [doi: 10.1016/j.cllc.2020.05.019] [Medline: 32636159]
30. Gijtenbeek RGP, Damhuis RAM, van der Wekken AJ, Hendriks LEL, Groen HJM, van Geffen WH. Overall survival in advanced epidermal growth factor receptor mutated non-small cell lung cancer using different tyrosine kinase inhibitors in the Netherlands: a retrospective, nationwide registry study. Lancet Reg Health Eur 2023 Apr;27:100592. [doi: 10.1016/j.lanepe.2023.100592] [Medline: 36817181]
31. Aye PS, McKeage MJ, Tin Tin S, Khwaounjoo P, Elwood JM. Population-based incidence rates and increased risk of EGFR mutated non-small cell lung cancer in Māori and Pacifica in New Zealand. PLOS ONE 2021;16(5):e0251357. [doi: 10.1371/journal.pone.0251357] [Medline: 33961689]
32. StatsNZ. URL: https://www.stats.govt.nz [accessed 2025-01-07]
33. McKeage M, Elwood M, Tin Tin S, et al. EGFR mutation testing of non-squamous NSCLC: impact and uptake during implementation of testing guidelines in a population-based registry cohort from Northern New Zealand. Target Oncol 2017 Oct;12(5):663-675. [doi: 10.1007/s11523-017-0515-4] [Medline: 28699084]
34. Aye PS, Barnes J, Laking G, et al. Erlotinib or gefitinib for treating advanced epidermal growth factor receptor mutation-positive lung cancer in Aotearoa New Zealand: protocol for a national whole-of-patient-population retrospective cohort study and results of a validation substudy. JMIR Res Protoc 2024 Jul 2;13:e51381. [doi: 10.2196/51381] [Medline: 38954434]
35. Prescribing information Tarceva (erlotinib) tablets. U.S. Food & Drug Administration. 2016. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2016/021743s025lbl.pdf [accessed 2025-01-16]
36. Prescribing information Iressa (gefitinib) tablets. U.S. Food & Drug Administration. 2021. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/206995s004lbl.pdf [accessed 2025-01-16]
37. Blumenthal GM, Gong Y, Kehl K, et al. Analysis of time-to-treatment discontinuation of targeted therapy, immunotherapy, and chemotherapy in clinical trials of patients with non-small-cell lung cancer. Ann Oncol 2019 May 1;30(5):830-838. [doi: 10.1093/annonc/mdz060] [Medline: 30796424]
38. Salmond C, Crampton P, Atkinson J. NZDep2006 Index of Deprivation. : Department of Public Health, University of Otago; 2007 URL: https://www.otago.ac.nz/wellington/otago020348.pdf [accessed 2025-01-11]
39. An urban/rural profile. Statistics New Zealand. URL: https://datafinder.stats.govt.nz/layer/120965-urban-rural-2025/ [accessed 2025-01-11]

40. Sarfati D, Gurney J, Stanley J, Lim BT, McSherry C. Development of a pharmacy-based comorbidity index for patients with cancer. Med Care 2014 Jul;52(7):586-593. [doi: 10.1097/MLR.0000000000000149] [Medline: 24926705]
41. Ethicity data protocols: HISO 10001:2017 version 1.1. Health New Zealand: Te Whatu Ora. 2017. URL: https://www.tewhatuora.govt.nz/assets/Our-health-system/Digital-health/Health-information-standards/HISO-10001-2017-Ethnicity-Data-Protocols.pdf [accessed 2025-01-11]
42. Koopman B, Cajiao Garcia BN, Kuijpers CCHJ, et al. A nationwide study on the impact of routine testing for EGFR mutations in advanced NSCLC reveals distinct survival patterns based on EGFR mutation subclasses. Cancers (Basel) 2021 Jul 20;13(14):3641. [doi: 10.3390/cancers13143641] [Medline: 34298851]
43. Inoue A, Yoshida K, Morita S, et al. Characteristics and overall survival of EGFR mutation-positive non-small cell lung cancer treated with EGFR tyrosine kinase inhibitors: a retrospective analysis for 1660 Japanese patients. Jpn J Clin Oncol 2016 May;46(5):462-467. [doi: 10.1093/jjco/hyw014] [Medline: 26977054]
44. Lin JJ, Cardarella S, Lydon CA, et al. Five-year survival in EGFR-mutant metastatic lung adenocarcinoma treated with EGFR-TKIs. J Thorac Oncol 2016 Apr;11(4):556-565. [doi: 10.1016/j.jtho.2015.12.103] [Medline: 26724471]
45. Lee CK, Wu YL, Ding PN, et al. Impact of specific epidermal growth factor receptor (EGFR) mutations and clinical characteristics on outcomes after treatment with EGFR tyrosine kinase inhibitors versus chemotherapy in EGFR-mutant lung cancer: a meta-analysis. J Clin Oncol 2015 Jun 10;33(17):1958-1965. [doi: 10.1200/JCO.2014.58.1736] [Medline: 25897154]
46. Mulloy R, Ferrand A, Kim Y, et al. Epidermal growth factor receptor mutants from human lung cancers exhibit enhanced catalytic activity and increased sensitivity to gefitinib. Cancer Res 2007 Mar 1;67(5):2325-2330. [doi: 10.1158/0008-5472.CAN-06-4293] [Medline: 17332364]
47. Carey KD, Garton AJ, Romero MS, et al. Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. Cancer Res 2006 Aug 15;66(16):8163-8171. [doi: 10.1158/0008-5472.CAN-06-0453] [Medline: 16912195]
48. Johnson AM, Hines RB, Johnson JA, Bayakly AR. Treatment and survival disparities in lung cancer: the effect of social environment and place of residence. Lung Cancer 2014 Mar;83(3):401-407. [doi: 10.1016/j.lungcan.2014.01.008] [Medline: 24491311]
49. Unger JM, Moseley AB, Cheung CK, et al. Persistent disparity: socioeconomic deprivation and cancer outcomes in patients treated in clinical trials. J Clin Oncol 2021 Apr 20;39(12):1339-1348. [doi: 10.1200/JCO.20.02602] [Medline: 33729825]
50. Blakely T, Fawcett J, Hunt D, Wilson N. What is the contribution of smoking and socioeconomic position to ethnic inequalities in mortality in New Zealand? Lancet 2006 Jul 1;368(9529):44-52. [doi: 10.1016/S0140-6736(06)68813-2] [Medline: 16815379]

Abbreviations

***EGFR*:** epidermal growth factor receptor
HR: hazard ratio
NSCLC: non–small cell lung cancer
PHARMs: Pharmaceutical Information Database
RECORD: Reporting of Studies Conducted Using Observational Routinely-Collected Health Data

Edited by N Cahill; submitted 06.08.24; peer-reviewed by HH Hasbullah, RAM Damhuis; revised version received 24.11.24; accepted 25.11.24; published 03.03.25.

Please cite as:
Aye PS, Barnes J, Laking G, Cameron L, Anderson M, Luey B, Delany S, Harris D, McLaren B, Brenman E, Wong J, Lawrenson R, Arendse M, Tin Tin S, Elwood M, Hope P, McKeage MJ
Treatment Outcomes From Erlotinib and Gefitinib in Advanced Epidermal Growth Factor Receptor–Mutated Nonsquamous Non–Small Cell Lung Cancer in Aotearoa New Zealand From 2010 to 2020: Nationwide Whole-of-Patient-Population Retrospective Cohort Study
JMIR Cancer 2025;11:e65118
URL: https://cancer.jmir.org/2025/1/e65118
doi:10.2196/65118

Predicting Early-Onset Colorectal Cancer in Individuals Below Screening Age Using Machine Learning and Real-World Data: Case Control Study

Chengkun Sun[1], MS; Erin Mobley[2,3], PhD; Michael Quillen[4], MD; Max Parker[4], MD; Meghan Daly[3], MD; Rui Wang[5], PhD; Isabela Visintin[3], MD; Ziad Awad[3], MD; Jennifer Fishe[6], MD; Alexander Parker[2,7], PhD; Thomas George[2,4], MD; Jiang Bian[8], PhD; Jie Xu[1], PhD

[1]Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, 1889 Museum Road, Office 7020, Gainesville, FL, United States

[2]University of Florida Health Cancer Center, University of Florida, Gainesville, FL, United States

[3]Department of Surgery, University of Florida, Jacksonville, FL, United States

[4]Department of Medicine, University of Florida, Gainesville, FL, United States

[5]Center for Data Solutions, University of Florida, Jacksonville, FL, United States

[6]Department of Emergency Medicine, University of Florida, Jacksonville, FL, United States

[7]College of Medicine, University of Florida, Jacksonville, FL, United States

[8]Indiana University Indianapolis, Indianapolis, IN, United States

Corresponding Author:
Jie Xu, PhD
Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, 1889 Museum Road, Office 7020, Gainesville, FL, United States

Abstract

Background: Colorectal cancer is now the leading cause of cancer-related deaths among young Americans. Accurate early prediction and a thorough understanding of the risk factors for early-onset colorectal cancer (EOCRC) are vital for effective prevention and treatment, particularly for patients below the recommended screening age.

Objective: Our study aims to predict EOCRC using machine learning (ML) and structured electronic health record data for individuals under the screening age of 45 years, with the aim of exploring potential risk and protective factors that could support early diagnosis.

Methods: We identified a cohort of patients under the age of 45 years from the OneFlorida+ Clinical Research Consortium. Given the distinct pathology of colon cancer (CC) and rectal cancer (RC), we created separate prediction models for each cancer type with various ML algorithms. We assessed multiple prediction time windows (ie, 0, 1, 3, and 5 y) and ensured robustness through propensity score matching to account for confounding variables including sex, race, ethnicity, and birth year. We conducted a comprehensive performance evaluation using metrics including area under the curve (AUC), sensitivity, specificity, positive predictive value, negative predictive value, and F_1-score. Both linear (ie, logistic regression, support vector machine) and nonlinear (ie, Extreme Gradient Boosting and random forest) models were assessed to enable rigorous comparison across different classification strategies. In addition, we used the Shapley Additive Explanations to interpret the models and identify key risk and protective factors associated with EOCRC.

Results: The final cohort included 1358 CC cases with 6790 matched controls, and 560 RC cases with 2800 matched controls. The RC group had a more balanced sex distribution (2:3 male-to-female) compared to the CC group (2:5 male-to-female), and both groups showed diverse racial and ethnic representation. Our predictive models demonstrated reasonable results, with AUC scores for CC prediction of 0.811, 0.748, 0.689, and 0.686 at 0, 1, 3, and 5 years before diagnosis, respectively. For RC prediction, AUC scores were 0.829, 0.771, 0.727, and 0.721 across the same time windows. Key predictive features across both cancer types included immune and digestive system disorders, secondary malignancies, and underweight status. In addition, blood diseases emerged as prominent indicators specifically for CC.

Conclusions: Our findings demonstrate the potential of ML models leveraging electronic health record data to facilitate the early prediction of EOCRC in individuals under 45 years. By uncovering important risk factors and achieving promising predictive performance, this study provides preliminary insights that could inform future efforts toward earlier detection and prevention in younger populations.

(JMIR Cancer 2025;11:e64506) doi:10.2196/64506

KEYWORDS

prediction; machine learning; ML; rectal cancer; colorectal cancer; CRC; youth; adolescent; middle-aged; United States; Americans; electronic health record; EHR; Shapley Additive Explanations; SHAP; diagnosis; prevention and treatment

Introduction

Colorectal cancer (CRC) is a significant public health challenge, ranking as the third leading cause of cancer-related mortality among both males and females in the United States [1]. It is estimated that in 2023, approximately 153,020 individuals were diagnosed with CRC, and 52,550 succumbed to the disease [1]. While cancer is typically a disease of older age, a concerning trend has emerged—the increasing incidence of early-onset colorectal cancer (EOCRC) in individuals younger than the age of 50 years [1,2]. This increased incidence has led the US Preventive Services Task Force to modify its recommendations, lowering the age to start CRC screening to age 45 [3]. Patients diagnosed with EOCRC tend to present at later stages and face lower disease-specific survival rates, underscoring the need for early detection and treatment initiation [4]. Nevertheless, challenges in addressing EOCRC are compounded by poorly defined risk factors and the role of diagnostic delays. As a result, early prediction and comprehensive understanding of the risk factors of EOCRC are essential for prevention and treatment, particularly for patients who fall below the recommended screening age.

The rapid integration of artificial intelligence and big data analytics has significantly expanded the horizons of medical research and clinical care [5]. Diverse data sources, including imaging and genomic data, have been harnessed for CRC detection through the application of statistical and machine learning (ML) algorithms. Some approaches have included the analysis of tumor DNA and circulating RNA expression profiling data to identify potential pathogenic factors [6,7]. In addition, computer tomography (CT)–based radionics, combined with ML algorithms, have been used to predict the Kirsten rat sarcoma viral oncogene mutation in people with CRC, demonstrating the potential of ML in clinical decision support [8]. Further, a random forest (RF) model trained with standard clinical and pathological prognostic variables, coupled with magnetic resonance imaging (MRI) images, achieved an impressive area under the curve (AUC) score of 0.94 when predicting survival in CRC patients, highlighting the importance of MRI-based texture features in patient survival prediction [9]. However, imaging data produces a small number of unexplainable predictors (around 100), and does not consistently improve diagnostic accuracy and disease prediction, especially when only using imaging data [10]. Furthermore, advanced imaging modalities and genomic data can be costly, with limited accessibility, and lack diversity and representativeness in samples, which could impact timely and accurate diagnosis for all individuals affected by EOCRC or widen already present disparities in patient outcomes.

In contrast to imaging and genomic data, structured data from the electronic health record (EHR) offers a more accessible and cost-effective data source for initial research. Originally designed for administrative and billing purposes, structured EHR data have evolved into valuable tools for health care research, capturing a wealth of patient information, including clinical diagnoses, procedures, medications, and laboratory results, among others [11]. The integration of ML and deep learning with EHR data has demonstrated substantial potential for disease prediction, including Alzheimer disease, gestational diabetes mellitus, and coronary heart disease [12-14]. In the context of CRC, several ML approaches have been used to predict the risk of the disease. For example, Shanbehzadeh et al [15] used structured EHR data and four data mining algorithms to predict CRC risk, identifying critical attributes for the prediction model using the weight statistical χ^2 test. However, the weight statistical χ^2 test assumes independence among variables, which may not hold true in complex datasets where variables are likely correlated. Another study leveraged convolutional neural networks to predict CRC risk based on the structured EHR data from the Taiwan National Health Insurance database [16]. Hussan et al [17] explored multiple ML methods to construct predictive models for CRC among patients aged between 35 and 50 years. However, these studies faced challenges in effectively matching cases and control groups, leading to increased bias and concerns regarding confounding. Furthermore, another limitation across studies is the failure to distinguish between colon cancer (CC) and rectal cancer (RC), despite the differences in clinical presentation, molecular carcinogenesis, pathology, surgical topography and procedures, and multimodal treatment strategies between these 2 cancers [18]. In addition, the lack of model explanations regarding clinical diagnosis of CRC undermined the interpretability and reliability of their strategies. As a result, there is a pressing need for improved methodologies to enhance the reliability and understanding of ML models in EOCRC prediction.

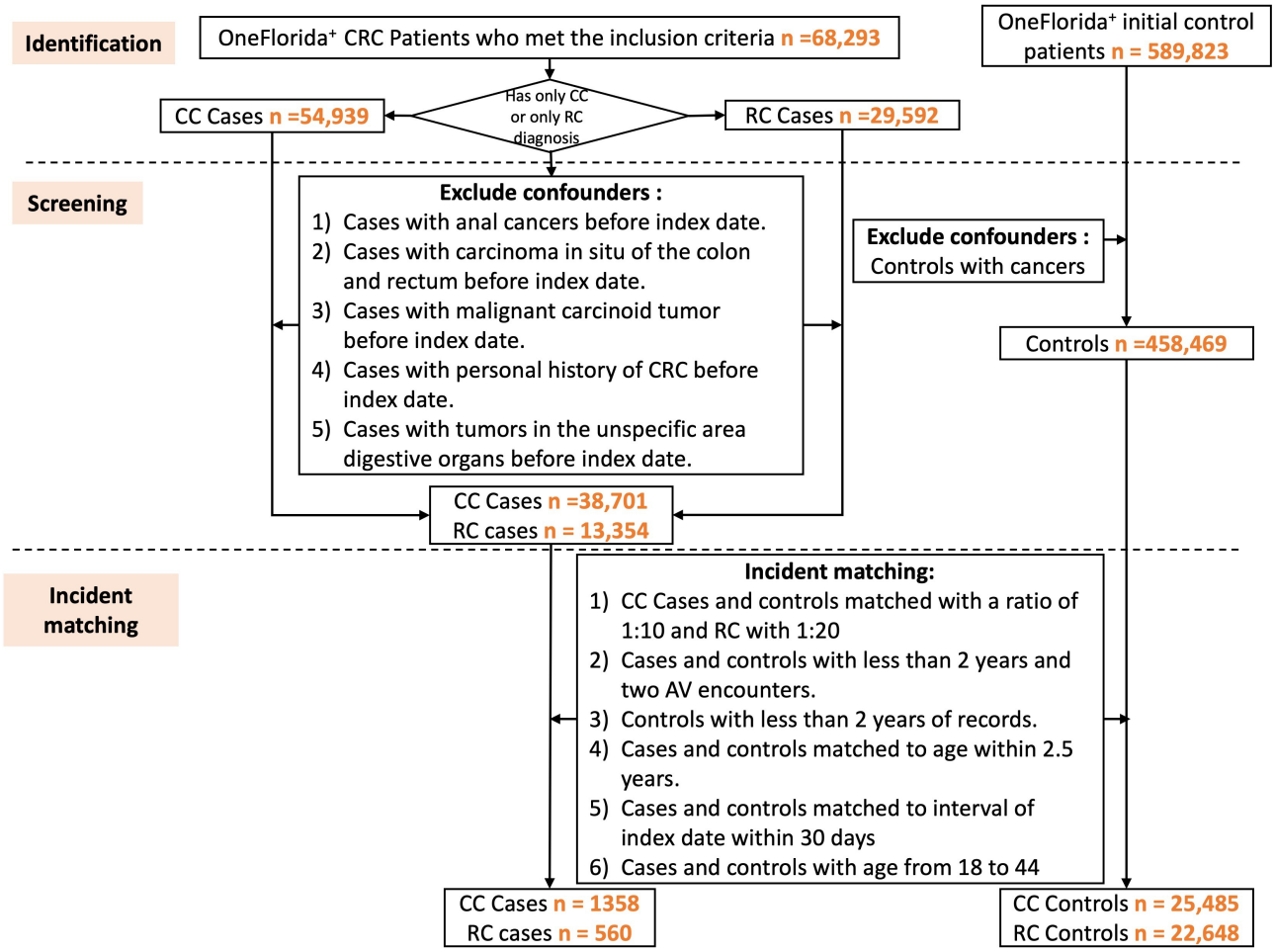
Methods

Data Source and Study Population

This study used deidentified EHR data from the OneFlorida+ Clinical Research Consortium, which operates within the PCORnet Clinical Research Network funded by the Patient-Centered Outcomes Research Institute. PCORnet serves as a national resource dedicated to advancing high-priority health research and improving outcomes through a robust, integrated research infrastructure [19,20]. By combining extensive health data, research expertise, and patient perspectives, it enables network partners to rapidly generate reliable, actionable evidence to support public health and clinical decision-making [19,20]. The OneFlorida+ data encompasses a wide range of patient characteristics from health systems across the southeast, including EHR data collected using the PCORnet Common Data Model [19] regarding demographics, diagnoses, medications, procedures, vital signs, lab tests, and more.

The construction of our study cohort using OneFlorida+ is outlined in Figure 1. OneFlorida+ identified individuals from the OneFlorida+ network, with encounters from January 2012 to January 2023 who met our inclusion criteria as either a case or control. We identified cases of CC using the *International Classification of Diseases, Ninth Revision* (*ICD-9*) code of C18.x or C49A4 or the *International Classification of Diseases, Tenth Revision* (*ICD-10*) code of 153.x, or RC cases with the *ICD-9* code of C19.x, C20.x, C21.0, C21.1, and *ICD-10* code of 154.0 and 154.1. The initial cohort consisted of 68,293 CRC cases (54,939 CC cases and 29,592 RC cases), and 589,823 controls. From those, we excluded patients diagnosed with both CC and RC, other previous cancers, or those who were diagnosed ≥45 years of age. Our final study cohort comprised 1358 CC cases with 25,485 controls and 560 RC cases with 22,648 controls.

Figure 1. Flowchart of patient selection from OneFlorida+. CC: colon cancer; CRC: colorectal cancer; RC: rectal cancer; AV: Ambulatory Visit.

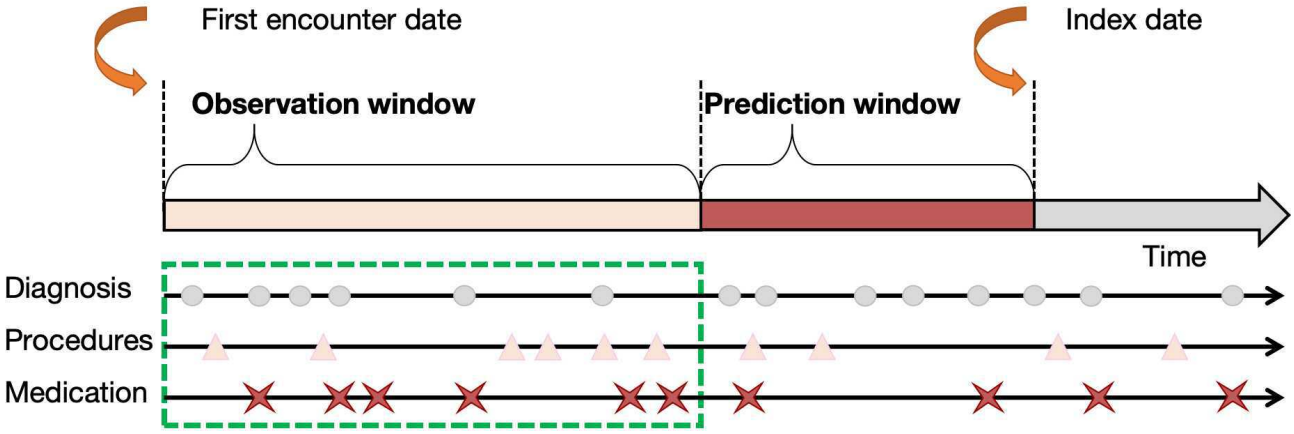


We used an incident matching process to match cases and controls to ensure a fair comparison across these groups. Initially, we retained cases and controls with more than 2 years of records and at least 2 encounters before the first onset date of either CC or RC and ensured that the age gap between matched cases and controls was within 2.5 years. By calculating propensity scores based on race, ethnicity, sex, and birth year (within 2.5 y), we used a narrow caliper of 0.05 with a nearest neighbor approach to achieve a 1:5 case-to-control ratio for each prediction window group [21]. This rigorous methodology ensures a balanced study population for reliable analysis and EOCRC prediction.

Study Setting

We then incorporated a range of different observation periods and prediction windows (Figure 2) to test our prediction algorithms, considering the different use cases. We considered 4 different prediction windows: 0 years, 1 year, 3 years, and 5 years before CRC diagnosis.

Figure 2. Visualization of the observation and prediction windows. For the prediction task. The index date for CRC cases is the date of diagnosis. For the control group, the index date is defined as the closest encounter date to the diagnosis date of the matched case group. The prediction window is the time period before the index date during which CRC cases are predicted. The observation window refers to the specific period during which data is collected or observed for analysis. CRC: colorectal cancer.



Data Preprocessing

The predictors we extracted included data from the demographics, vitals, diagnoses, medications, and procedures tables within the OneFlorida+ Clinical Research Network throughout the observation periods. Age at index date was calculated and categorized into 3 groups (eg, 18 - 29 y, 30 - 39 y, and 40 - 44 y). One-hot encoding [22] was used to represent age groups, race, and sex variables. Statistical analysis shows that the proportion of missing values is approximately 50% (4137/8148 in CC and 1554/3360 in RC). According to [23-25], mean imputation is less sensitive to high proportions of missing data and is more robust compared to other imputation methods, such as median and mode imputations, the indicator method, and regression. Thus, for missing data, we imputed the missing values with the mean of the numerical data derived from the entire sample within each prediction window group. Furthermore, BMI data was categorized into clinically relevant groups, including underweight (≤18.5 y), normal (18.5 - 23 y), overweight (23-30 y), and obese (≥30). Diastolic and systolic measurements were categorized into distinct hypertension stages.

Diagnoses, which were initially represented using *ICD-9* and *ICD-10* codes, were subjected to a data dimensionality reduction process that mapped them into Phecodes [26,27]. Revenue codes and current procedural terminology codes [28] were leveraged to capture billed medical procedures. To integrate these data, we also used the clinical classifications software code [29]. For drug information, National Drug Code [30] and RxNorm codes were used for encoding. National Drug Codes were mapped into RxNorm codes, and further consolidated into anatomical therapeutic chemical classes [31]. To ensure completeness, all features that could not be mapped were retained to prevent any missing information. These steps to transform the data enhanced interpretability and relevance of our predictive models.

Experiments and Validation

We explored several widely used ML models, including linear models such as logistic regression (LR) and the support vector machine (SVM), as well as nonlinear models like XGBoost (Extreme Gradient Boosting) and RF. We adopted two modeling strategies, including (1) prediction without CRC-related features and (2) prediction without cancer-related features, covering the CRC-related features. For the first strategy, features that may be indicative of CRC differential diagnoses (eg, neoplasm of unspecified nature of digestive system) or treatments for CRC (eg, chemotherapy and radiotherapy) were removed from the models and not used as predictors. For the second strategy, we took a more stringent approach by eliminating all diagnoses, drugs, and procedures that could be associated with any cancer from the extracted predictors. This step aimed to identify risk factors while eliminating the influence of other types of cancers, enabling us to focus exclusively on noncancer-related predictors. Regardless of the feature engineering strategy, we maintained a consistent experimental setup. The entire dataset was randomly split into a training dataset and a testing dataset with a ratio of 4:1. Model optimization was conducted on the training set through 5-fold cross-validation, and we fine-tuned hyperparameters using Bayesian optimization. To ensure reproducibility, we fixed the random state seed across all model runs.

To assess the effectiveness of our models comprehensively, we used a battery of evaluation metrics, including AUC, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F_1-score. To mitigate the risk of overfitting and to derive robust CIs, we implemented a bootstrapping strategy. This involved conducting 100 experiments by randomly resampling the training and testing datasets. In addition to traditional performance metrics, we delved into the interpretability of the XGBoost models. Specifically, we computed Shapley Additive Explanations (SHAP) values [32] to gain insights into the inner workings of the ML algorithms and to identify the core contribution predictors. This approach aimed to unveil the high-risk factors associated with EOCRC, shedding light on the most influential features in our prediction model. To further assess generalizability, we performed temporal validation on all CC and RC groups, using data before January 1, 2015, for training and data after for testing. We then trained an XGBoost model to evaluate its performance on the test set.

Ethical Considerations

The study was approved and the requirement to obtain any informed consent was waived by the University of Florida

Institutional Review Board (protocol number IRB202201561). The research does not involve greater than minimal risk for participation. Analyses only involve the secondary analysis of data that are either limited datasets or deidentified. Our research team has no direct contact with human participants. All methods were carried out in accordance with relevant guidelines and regulations.

Results

Table 1 provides an overview of the identified study cohorts after propensity score matching for both CC and RC across various prediction windows. Notably, CC cases outnumber RC cases, with approximately twice as many CC cases. Patients in the RC groups were slightly older compared to those in the CC group. Sex distribution in the RC groups was closer to parity (2:3 male to female) than in the CC group (2:5 male to female). Both RC and CC groups exhibited diverse racial and ethnic representation. In addition, as the prediction window lengthened, the number of cases decreased. Specifically, there were 560, 560, 383, and 225 RC cases, and 1358, 1358, 884, and 532 CC cases in prediction windows for 0 years, 1 year, 3 years, and 5 years, respectively.

Table . Descriptive statistics in case and control groups.

Variables	CC[a] cases (n=1358)	CC controls (n=6790)	RC[b] cases (n=560)	RC controls (n=2800)
Age, mean (SD)	36.54 (5.88)	36.69 (5.73)	37.70 (5.70)	36.80 (5.53)
Sex, n (%)				
Female	938 (69.07)	4461 (65.70)	323 (57.68)	1617 (57.75)
Male	420 (30.93)	2329 (34.30)	237 (42.32)	1183 (42.25)
Race and ethnicity, n (%)				
Hispanic	338 (24.89)	1527 (22.49)	101 (18.04)	514 (18.36)
Non-Hispanic White	554 (40.80)	2893 (42.61)	239(42.68)	1212 (43.29)
Non-Hispanic Black	353 (25.99)	1857 (27.35)	178 (31.79)	887 (31.68)
Other	14 (1.03)	66 (0.97)	4 (0.71)	9 (0.32)
Unknown	99 (7.29)	447 (6.58)	38 (6.79)	178 (6.36)
Vital Signs, missing rate, n (%)				
BMI	475 (34.98)	3393(49.97)	171 (30.54)	1383 (49.39)
Diastolic blood pressure	559 (41.16)	3578 (52.70)	236 (42.14)	1497 (53.46)
Systolic blood pressure	573 (42.19)	3618 (53.28)	240 (42.86)	1517 (54.18)
Top 10 diagnoses, n (%)				
Other tests	1016 (74.82)	5010 (73.78)	45 (8.04%)	83 (2.96)
Abdominal pain	881 (64.87)	3430 (50.52)	133 (23.75)	611 (21.82)
Other symptoms of respiratory system	630 (46.39)	2873 (42.30)	92 (16.43)	189 (6.75)
Overweight, obesity and other hyperalimentation	585 (43.08)	2915 (42.93)	46 (8.21)	187 (6.68)
Nausea and vomiting	581 (42.78)	2095 (30.85)	29 (5.18)	87 (3.11)
Nonspecific chest pain	536 (39.47)	2250 (33.14)	133 (23.75)	347(12.39)
Tobacco use disorder	523 (38.51)	2408 (35.46)	55 (9.82)	121 (4.32)
Acute upper respiratory infections of multiple or unspecified sites	522 (38.44)	2817 (41.49)	1 (0.18%)	4 (0.4)
Other anemias	512 (37.7)	1454 (21.41))	157 (28.04)	77 (2.75)
Hypertension	509 (37.48)	2202 (32.43)	127 (22.68)	557 (19.89)
Top 10 procedures, n (%)				
Other diagnostic procedures	1207 (88.88)	6241 (91.91)	63 (11.25)	100 (3.57)
Dental procedures	1071 (78.87)	5423 (79.87)	351 (62.68)	1909 (68.18)
Microscopic examination (bacterial smear; culture; toxicology)	958 (70.54)	4706 (69.31)	298 (53.21)	1917 (68.46)
Other therapeutic procedures	951 (70.03)	4720 (69.51)	296 (52.86)	1544 (55.14)
General emergency room	845 (62.22)	4097 (60.34)	203 (36.25)	622 (22.21)
Pathology	817 (60.16)	3070 (45.21)	413 (73.75)	2258 (80.64)
Chemistry laboratory-clinical	806 (59.35)	3756 (55.32)	50 (8.93)	266 (9.50)
Hematology laboratory-clinical	800 (58.91)	3740 (55.08)	470 (83.93)	2345 (83.75)
Nonoperative urinary system measurements	792 (58.32)	3784 (55.73)	138 (24.64)	763 (27.25)
General pharmacy	764 (56.26)	3556 (52.37)	47 (8.39)	229 (8.18)

Variables	CC[a] cases (n=1358)	CC controls (n=6790)	RC[b] cases (n=560)	RC controls (n=2800)
Top 10 medications, n (%)				
Other analgesics and antipyretics	769 (56.63)	3301 (48.62)	5 (0.89)	0 (0.00)
Anti-inflammatory and antirheumatic products, nonsteroids	724 (53.31)	3625 (53.39)	65 (11.61)	272 (9.71)
Throat preparations	723 (53.24)	3423 (50.41)	136 (24.29)	725 (25.89)
Anti-infectives	705 (51.91)	3267 (48.11)	51 (9.11)	66 (2.36)
Opioids	687 (50.59)	2779 (40.93)	4 (0.71)	34 (1.21)
Topical products for joint and muscular pain	682 (50.22)	3427 (50.47)	1 (0.18)	1 (0.04)
Stomatological preparations	673 (49.56)	3046 (44.86)	102 (18.21)	377 (13.46)
Other gynecologicals	644 (47.42)	3283 (48.35)	80 (14.29)	361 (12.89)
Other cardiac preparations	577 (42.49)	2949 (43.43)	167 (29.82)	878 (31.36)
Corticosteroids for systemic use, plain	573 (42.19)	2520 (37.11)	121 (21.61)	528 (18.86)

[a]CC: colon cancer.

[b]RC: rectal cancer.

Table 2 presents the results of CC prediction using 2 feature engineering strategies: 1 excluding CRC-related features and the other excluding cancer-related features. Additional evaluation metrics for CC prediction across all settings can be found in Tables S1-S2 in Multimedia Appendix 1. In most cases, tree-based models (XGBoost and RF) outperformed linear models (SVM and LR), yielding higher AUC values. Specifically, after removing CRC-related features, the RF model achieved the highest AUC for the 0-year prediction (0.811, 95% CI 0.808-0.814), while RF performed best for the 1-year (0.748, 95% CI 0.745-0.751), 3-year (0.689, 95% CI 0.684-694), and 5-year (0.686, 95% CI 0.68-0.692) predictions for CC. However, after removing features associated with previous cancers, the model performance decreased: LR achieved AUC values of 0.788 (95% CI 0.786-0.791) for 0-year prediction; RF achieved AUC values of 0.716 (95% CI 0.713-0.719) for 1-year, 0.684 (95% CI 0.679-0.688) for 3-year, and 0.663 (95% CI 0.658-0.668) for 5-year prediction. Performance metrics, including specificity, sensitivity, PPV, NPV, and F_1-score, exhibited similar trends.

Table . AUC[a] comparison for colon cancer prediction using machine learning models across different prediction windows (0, 1, 3, and 5 years).

Feature strategy and model	0-year AUC (95% CI)	1-year AUC (95% CI)	3-year AUC (95% CI)	5-year AUC (95% CI)
Excluding CRC-related[b] features				
LR[c]	0.809 (0.806-0.812)	0.733 (0.73-0.736)	0.683 (0.679-0.688)	0.674 (0.668-0.679)
SVM[d]	0.748 (0.745-0.751)	0.689 (0.685-0.692)	0.614 (0.61-0.618)	0.616 (0.61-0.621)
RF[e]	0.811 (0.808-0.814)	0.748 (0.745-0.751)	0.689 (0.684-0.694)	0.686 (0.68-0.692)
XGBoost[f]	0.802 (0.799-0.806)	0.745 (0.741-0.748)	0.689 (0.684-0.694)	0.657 (0.651-0.663)
Excluding cancer-related features				
LR	0.788 (0.786-0.791)	0.713 (0.71-0.716)	0.669 (0.665-0.674)	0.661 (0.656-0.667)
SVM	0.725 (0.722-0.729)	0.646 (0.643-0.65)	0.604 (0.6-0.608)	0.611 (0.606-0.617)
RF	0.77 (0.767-0.773)	0.716 (0.713-0.719)	0.684 (0.679-0.688)	0.663 (0.658-0.668)
XGBoost	0.76 (0.757-0.764)	0.714 (0.711-0.717)	0.662 (0.657-0.666)	0.643 (0.638-0.648)

[a]AUC: area under the curve.

[b]CRC: colorectal cancer.

[c]LR: logistic regression.

[d]SVM: support vector machine.

[e]RF: random forest.

[f]XGBoost: Extreme Gradient Boosting.

Table 3 provides RC prediction results using the same feature engineering strategies and 4 prediction windows. Additional evaluation metrics for RC prediction across all settings can be found in Tables S3-S4 in Multimedia Appendix 1. Again, after removing CRC-related features, the XGBoost model achieved the highest AUC for the 0-year prediction (0.829, 95% CI 0.825-0.834), while RF performed best for the 1-year (0.771, 95% CI 0.766-0.777), and XGBoost did best for 3-year (0.727, 95% CI 0.721-0.732), and 5-year (0.721, 95% CI 0.713-0.729) predictions for RC. Eliminating cancer-related features resulted in a performance decrease: XGBoost achieved AUC values of 0.811 (95% CI 0.806-0.815) for 0-year prediction; RF achieved AUC values of 0.756 (95% CI 0.751-0.76) for 1-year, 0.724 (95% CI 0.718-0.73) for 3-year, and 0.711 (95% CI 0.704-0.719) for 5-year predictions. Performance metrics exhibited consistent trends. In both the CC and RC prediction tasks, we observed a decline in model performance as the prediction window length increased. Notably, when we removed cancer-related features, the AUC declined. This highlights the pivotal role these features play in enhancing prediction performance.

Table . AUC[a] comparison for rectal cancer prediction using machine learning models across different prediction windows (0, 1, 3, and 5 years).

Feature strategy and model	0-year AUC (95% CI)	1-year AUC (95% CI)	3-year AUC (95% CI)	5-year AUC (95% CI)
Excluding CRC[b]-related features				
LR[c]	0.819 (0.815-0.824)	0.763 (0.758-0.767)	0.722 (0.716-0.728)	0.693 (0.686-0.7)
SVM[d]	0.78 (0.774-0.785)	0.694 (0.689-0.699)	0.656 (0.649-0.662)	0.658 (0.65-0.665)
RF[e]	0.826 (0.822-0.83)	0.771 (0.766-0.777)	0.719 (0.713-0.726)	0.72 (0.712-0.727)
XGBoost[f]	0.829 (0.825-0.834)	0.766 (0.762-0.771)	0.727 (0.721-0.732)	0.721 (0.713-0.729)
Excluding cancer-related features				
LR	0.807 (0.803-0.812)	0.748 (0.743-0.752)	0.709 (0.703-0.715)	0.69 (0.683-0.697)
SVM	0.767 (0.761-0.772)	0.686 (0.68-0.691)	0.653 (0.646-0.659)	0.656 (0.648-0.663)
RF	0.806 (0.802-0.81)	0.756 (0.751-0.76)	0.724 (0.718-0.73)	0.711 (0.704-0.719)
XGBoost	0.811 (0.806-0.815)	0.749 (0.744-0.753)	0.724 (0.718-0.729)	0.679 (0.672-0.687)

[a]AUC: area under the curve.

[b]CRC: colorectal cancer.

[c]LR: logistic regression.

[d]SVM: support vector machine.

[e]RF: random forest.

[f]XGBoost: Extreme Gradient Boosting.

To further evaluate model performance, we integrated XGBoost and RF using soft voting. The AUC fluctuated around 0.01, showing no significant change from the best prevoting performance (Table S9 in Multimedia Appendix 1). For temporal validation, the overall AUC decreased slightly by around 0.02, suggesting potential distribution shifts over time that may have affected generalizability. While the model fit earlier data well, its weaker performance on newer data hints at possible drift (Tables S5-S8 in Multimedia Appendix 1).

To gain deeper insights into the risk factors associated with these findings, we present SHAP summary plots for CC and RC predictions using 2 feature engineering strategies and for 0-year and 3-year prediction windows in Figures 3 and 4. Supplementary SHAP summary plots for all other models can be found in Figures S1-S2 in Multimedia Appendix 1. Within the CC group, several predictors emerged as positively associated with the risk of CC. Notably, several diagnoses involving various tumors, such as suspected cancer, secondary malignant neoplasm, benign neoplasm of uterus, benign neoplasm of skin, neoplasm of uncertain behavior, neoplasm of uncertain behavior of skin, cancer of other female genital organs and myeloproliferative diseases were identified as influential factors. Gastrointestinal symptoms, encompassing conditions like gastrointestinal hemorrhage, other disorders of intestine, other symptoms involving the abdomen and pelvis, noninfectious gastroenteritis, appendiceal conditions, diverticulosis and diverticulitis, intestinal obstruction without hernia, and disorders of the intestine also exhibited a positive association with CC risk. In addition, medical procedures related to gastrointestinal diseases and symptoms, including upper gastrointestinal endoscopy, were significantly associated with the development of CC.

Figure 3. SHAP (Shapley Additive Explanations) summary plot of the top 20 features in CC prediction using best-performing models with 0-year and 3-year prediction windows: (**A**) excluding CRC-related features; (**B**) excluding cancer-related features. The prefix before the "_" in the y-axis labels of plots indicates the source of the corresponding features in the PCORnet data model. Specifically, these sources are: Diagnosis (Diag), Procedure (Proc), Medication (Med), Vital Signs (Vital), and Demographics (Demo). CC: colon cancer; CRC: colorectal cancer; LR: logistic regression; RF: random forest; XGBoost: Extreme Gradient Boosting.

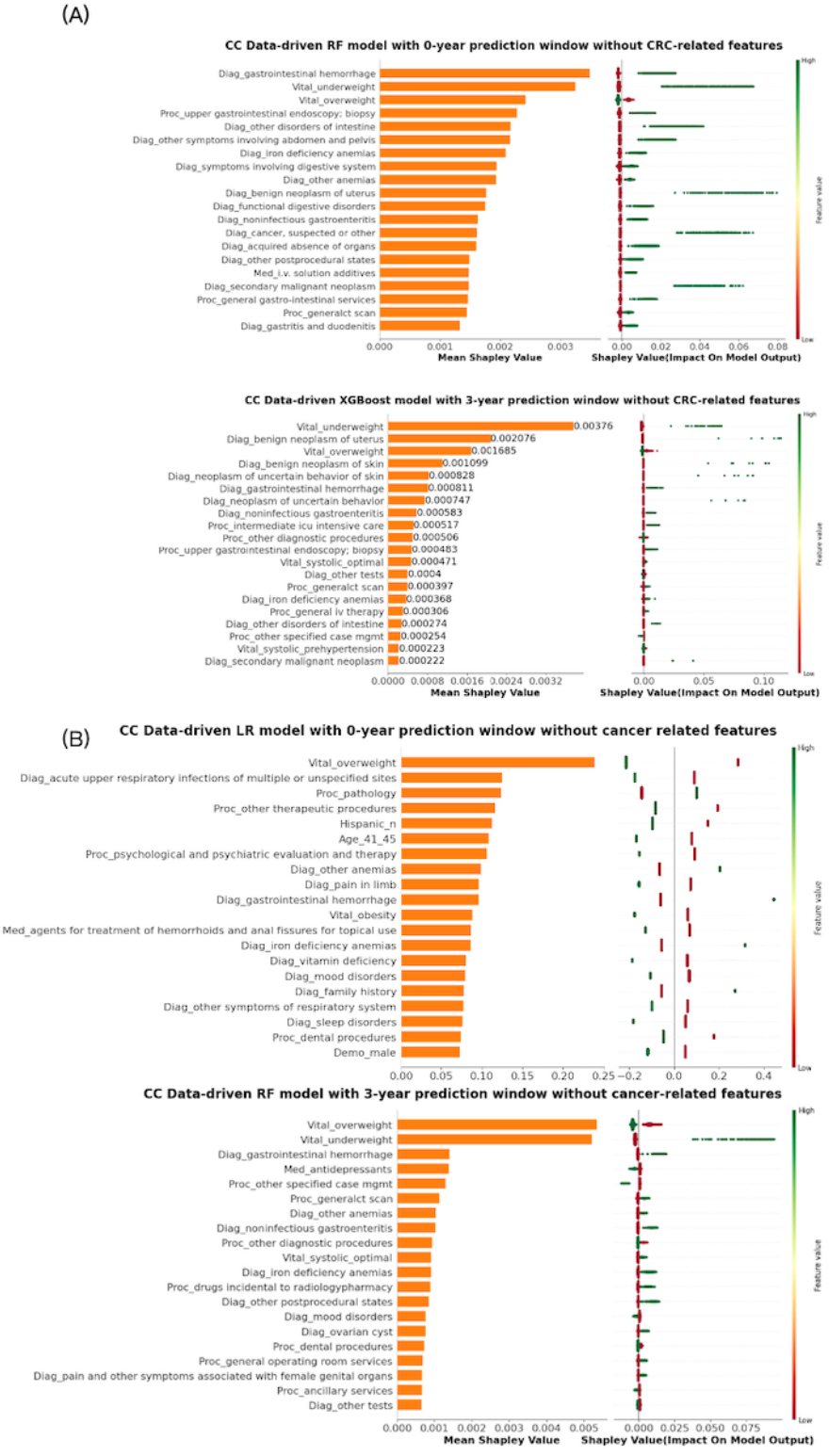
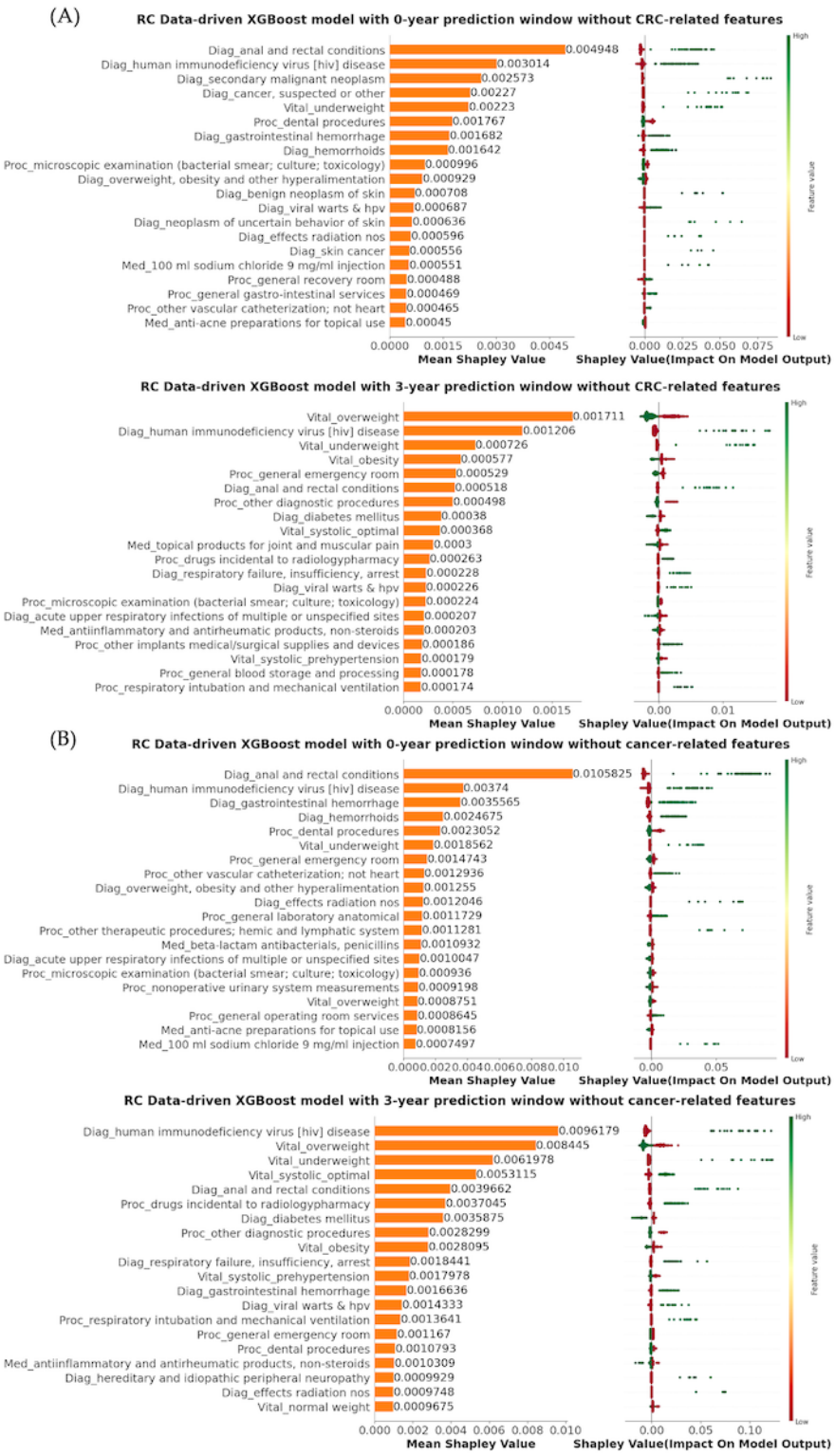


Figure 4. SHAP (Shapley Additive Explanations) summary plot of the top 20 features in RC prediction using best-performing models with 0-year and 3-year prediction windows: (**A**) excluding CRC-related features; (**B**) excluding cancer-related features. CRC: colorectal cancer; RC: rectal cancer; XGBoost: Extreme Gradient Boosting.



In the RC group, similar positive predictors were identified, mirroring the trends observed in the CC group, including gastrointestinal symptoms (eg, gastrointestinal hemorrhage, anal, and rectal conditions) and the presence of other cancers or tumors (eg, secondary malignant neoplasms and benign neoplasms of the uterus or skin). In addition, the presence of autoimmunity, diseases associated with a potentially weakened immune system (eg, HIV, viral warts, and human papillomavirus [HPV]), and conditions like hemorrhoids were linked to a heightened long-term risk of RC. Being underweight was a significant symptom associated with both CC and RC. Conversely, obesity, overweight, and normal weight appeared to be negatively associated with RC development. Importantly, after removing cancer-related features from consideration, the

significance of anemias surged to the forefront in both the CC and RC groups. These included indicators such as iron deficiency anemias and other anemias. Nevertheless, gastrointestinal diseases and immunodeficiency pathological changes remained substantial factors contributing to CC risk, while factors such as HPV and weight retained their significance as primary determinants of RC. The use of anti-inflammatory or antirheumatic medications were associated with decreased risk of RC.

Discussion

Principal Findings

In this study, we used 4 traditional ML algorithms (ie, XGBoost, RF, SVM, and LR) and obtained informative results predicting EOCRC using structured EHR data. In most cases, the tree-based models, (XGBoost and RF) outperformed linear models, achieving the best AUC scores for various prediction windows. In addition, even after excluding cancer diagnosis variables (eg, pancreatic, skin, and thyroid cancer), undergoing cancer-related procedures (eg, liver biopsy and bone marrow biopsy), cancer treatments (eg, cisplatin and doxycycline), our models continued to achieve acceptable AUC scores. Immune and digestive system disorders, blood diseases, and secondary cancers were identified as significant predictors.

Comparison to Previous Work

Most of our experimental findings were consistent with existing published research. Cancer-related diseases and diagnoses emerged as risk factors leading to the diagnosis of EOCRC, both for CC and RC. For example, uterine cancer was identified as a driver of EOCRC, suggesting a potential genetic association between these malignancies in younger patients [33]. Research also demonstrates that the incidence rate of second primary cancers among survivors is significantly higher than cancer in the general population, and survivors experience notable morbidity and mortality from their cancer treatment [34]. In addition, the use of CT scans for other medical reasons could contribute to the incidental identification of EOCRC cases [35]. Notably, we know that some forms of cancer treatment (eg, radiation) predisposes one to an increased risk for secondary malignancies, including EOCRC, particularly in patients surviving childhood cancer [36].

Inflammatory bowel diseases (IBDs) are well established risk factors for CRC, particularly during young adulthood. The chronic inflammation associated with IBD leads to the release of growth cytokines, excess blood flow, and metabolic free radicals, all of which contribute to the heightened risk of developing CRC [37]. Therapies for IBD sometimes involve immune suppression, another known risk factor for cancers. Furthermore, many gastrointestinal diseases can cause malabsorption or malnutrition [38], resulting in patients being underweight which can also contribute to immune dysfunction or suppression [39]. However, overweight patients were at low risk of EOCRC as our analysis demonstrated despite emerging evidence that being overweight may be associated with an increased risk of tumor recurrence and colorectal carcinogenesis [40,41]. The temporal use of antibiotics in relation to subsequent development of EOCRC is an interesting finding as it supports several previously reported roles that the gut microbiome may plan in CRC protection and development [42]. Our analysis highlighted that the diagnosis of iron deficiency anemia predated CC, but had less association with rectal cancers. It is logical, given that CC are situated more proximal in the gastrointestinal tract, causing occult chronic blood loss and subsequent anemia rather than overt gross bleeding as is typically evident from RC.

In addition, our study observed a significantly higher incidence of CRC cases among HIV-infected patients compared to HIV-uninfected individuals [43]. The heightened risk can be attributed to disruptions in immune function caused by immunodeficiency, which exposes individuals to a higher susceptibility against cancer-causing viruses, including HPV, Epstein–Barr virus, Kaposi sarcoma-associated herpesvirus, etc, as evidenced in our analysis [44]. Another notable finding was the association between CC and diseases of myeloproliferative disease. Similar to other cancers, the potential link could be related to genetics, treatments that induce DNA damage that could predispose to EOCRC, and chronic immune dysregulation. Overall, our study sheds light on the complex interplay between IBD, malnutrition, immune function, and specific blood-related diseases in the development of CRC. Understanding these relationships is crucial in advancing our knowledge of EOCRC risk factors and devising targeted interventions for at-risk populations. It can help health care providers identify individuals who may benefit most from screening between the ages of 18 and 44 years. In this case-control study, we identified several factors independently associated with an elevated risk of EOCRC. These findings could inform patient-provider discussions about the need for and approach to CRC screening and support targeted interventions to improve screening uptake among high-risk individuals.

Strengths and Limitations

The strength of our study is to develop an early diagnostic tool that can help identify individuals at higher risk for EOCRC before the onset of clinical symptoms or suspicion. To further clarify, we test the algorithm across different prediction windows—0, 1, 3, and 5 years—meaning we use data from these periods before a patient's first CRC diagnosis to predict whether they will develop CRC in the future. This approach enables us to assess how the algorithm can detect early risk signals well in advance of diagnosis, providing actionable insights for clinicians to consider for individuals who may not yet exhibit symptoms or be under suspicion for CRC. For individuals without clear clinical suspicion (ie, those who are not yet exhibiting symptoms or are below typical screening age), our algorithm could serve as a risk stratification tool. By analyzing real-world data, such as demographic information, medical history, and other relevant factors, the model can help identify patients who may benefit from earlier screening or closer monitoring, even in the absence of overt symptoms. This can be particularly important in populations with no established risk factors for CRC, but who may still be at risk for early-onset cases.

Our study does have several limitations. First, the mechanism through which identified medical factors are associated with

EOCRC is speculative. For example, CT scans contributed significantly to the model's performance, but the specific reasons are unclear. EHRs did not record the reason why patients underwent CT scans. Perhaps some patients obtained CT scans because of symptoms related to undiagnosed CRC while others received CT scans for other reasons with the incidental finding of CRC. It is less likely that CT scans could be associated with causing CRC due to radiation exposure. For that to occur, the cumulative lifetime exposure would need to be very high with exposure over a number of decades for that to occur. Perhaps CT imaging itself is just a surrogate for access to care whereby EOCRC is more likely to be eventually diagnosed as opposed to patients who might expire for other reasons with CRC, but before a diagnosis. Second, the exclusion of confounder samples and features posed difficulties, given the lack of universally accepted standards for phenotype definitions and ambiguous descriptions. These challenges hindered the design of the most optimal experiment [45]. Third, our experiments are carried out based on the EHR data, which inherently contains flaws, including missing values and potential mistakes in records. Efforts were made to fill in missing values, but comprehensive amendments remained challenging. The characteristics of the EHR data, such as temporality, irregularity, sparsity, and data imbalance, can result in abnormal outcomes when applying ML models [46,47]. Fourth, we primarily focused on metrics related to discrimination or classification (eg, AUC), as we believe these provide essential insights into how effectively the model differentiates between cases and noncases. We acknowledge that a more holistic evaluation—including calibration, fairness, stability, and net benefit—would provide a fuller picture of the model's real-world applicability.

Future Directions

Future research should focus on refining the experimental design, exploring alternative feature selection techniques, incorporating large language models based on both ambulatory and inpatient data, and integrating domain knowledge to enhance the performance of the prediction models. Ultimately, these efforts will contribute to early detection and better management of CRC, with the goal of improving patient outcomes. Using techniques like Synthetic Minority Over-sampling Technique or cost-sensitive learning could further improve the model's ability to detect the minority class. These methods were not used in this study but could be considered in future work to explore their potential impact on model performance, especially in terms of improving recall for the minority class.

Conclusion

In conclusion, our study demonstrated the potential of traditional ML algorithms in predicting EOCRC using real-world data for individuals below the screening age guideline. The identification of significant predictors and their consistency with academic research findings provide valuable insights for pursuing additional hypotheses or targeting potential patients at risk for EOCRC. However, addressing the challenges and limitations related to data quality, experimental design, and ML models' development is essential for improving the accuracy and reliability of EOCRC prediction models.

Data Availability

The variables used in this study can be found on GitHub [48].

Authors' Contributions

CS, EM, MQ, MP, RW, and JX were responsible for the overall design, development, and evaluation of this study. CS and JX did the initial drafts and the revisions of the manuscript. CS was responsible for rerunning all experiments and creating Shapley Additive Explanation plots. EM and MD were responsible for providing accurate *International Classification of Diseases* codes to define cases and filter colorectal cancer–related and cancer-related features. All authors reviewed the manuscript critically for scientific content and provided edits, and all authors gave final approval of the manuscript for publication.

Conflicts of Interest

TG serves on the Data and Safety Monitoring Board (DSMB) for Seagen, Nihon Medi-Physics, KAHR Medical, and Arbele and has served as a consultant for Avammune Therapeutics, BillionToOne, Exact Sciences, and Summit Therapeutics. Research support has been provided to his institution by Bristol-Myers Squibb, Merck, AstraZeneca, Lilly, Bayer, Incyte, Ipsen, Genentech, Astellas Pharma, GlaxoSmithKline, Amgen, OncoC4, BillionToOne, Jounce Therapeutics, Elicio Therapeutics, AMAL Therapeutics/Boehringer Ingelheim, Seagen, Regeneron, and Deciphera. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary material.
[DOCX File, 903 KB - cancer_v11i1e64506_app1.docx]

References

1. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. CA Cancer J Clin 2023;73(3):233-254. [doi: 10.3322/caac.21772] [Medline: 36856579]
2. Hofseth LJ, Hebert JR, Chanda A, et al. Early-onset colorectal cancer: initial clues and current views. Nat Rev Gastroenterol Hepatol 2020 Jun 15;17(6):352-364. [doi: 10.1038/s41575-019-0253-4]

3. US Preventive Services Task Force, Davidson KW, Barry MJ, et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. JAMA 2021 May 18;325(19):1965-1977. [doi: 10.1001/jama.2021.6238] [Medline: 34003218]
4. McClelland PT, Liu T, Ozuner G. Early-onset colorectal cancer in patients under 50 years of age: demographics, disease characteristics, and survival. Clin Colorectal Cancer 2022 Jun;21(2):e135-e144. [doi: 10.1016/j.clcc.2021.11.003] [Medline: 34972664]
5. Lv Z, Qiao L. Analysis of healthcare big data. Future Gener Comput Syst 2020 Aug;109:103-110. [doi: 10.1016/j.future.2020.03.039]
6. Luo H, Zhao Q, Wei W, et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. Sci Transl Med 2020 Jan 1;12(524):eaax7533. [doi: 10.1126/scitranslmed.aax7533] [Medline: 31894106]
7. Song W, Fu T. Circular RNA-associated competing endogenous RNA network and prognostic nomogram for patients with colorectal cancer. Front Oncol 2019;9:1181. [doi: 10.3389/fonc.2019.01181] [Medline: 31781492]
8. Porto-Álvarez J, Cernadas E, Aldaz Martínez R, et al. CT-based radiomics to predict KRAS mutation in CRC patients using a machine learning algorithm: a retrospective study. Biomedicines 2023 Jul 29;11(8):2144. [doi: 10.3390/biomedicines11082144] [Medline: 37626641]
9. Daye D, Tabari A, Kim H, et al. Quantitative tumor heterogeneity MRI profiling improves machine learning–based prognostication in patients with metastatic colon cancer. Eur Radiol 2021 Aug;31(8):5759-5767. [doi: 10.1007/s00330-020-07673-0]
10. Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. NPJ Digit Med 2022 Apr 12;5(1):48. [doi: 10.1038/s41746-022-00592-y] [Medline: 35413988]
11. Kim E, Rubinstein SM, Nead KT, Wojcieszynski AP, Gabriel PE, Warner JL. The evolving use of electronic health records (EHR) for research. Semin Radiat Oncol 2019 Oct;29(4):354-361. [doi: 10.1016/j.semradonc.2019.05.010] [Medline: 31472738]
12. Li Q, Yang X, Xu J, et al. Early prediction of Alzheimer's disease and related dementias using real-world electronic health records. Alzheimers Dement 2023 Aug;19(8):3506-3518. [doi: 10.1002/alz.12967] [Medline: 36815661]
13. Artzi NS, Shilo S, Hadar E, et al. Prediction of gestational diabetes based on nationwide electronic health records. Nat Med 2020 Jan;26(1):71-76. [doi: 10.1038/s41591-019-0724-8] [Medline: 31932807]
14. Du Z, Yang Y, Zheng J, et al. Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machine-learning methods: model development and performance evaluation. JMIR Med Inform 2020 Jul 6;8(7):e17257. [doi: 10.2196/17257] [Medline: 32628616]
15. Shanbehzadeh M, Nopour R, Kazemi-Arpanahi H. Comparison of four data mining algorithms for predicting colorectal cancer risk. J Adv Med Biomed Res 2021;29(133):100-108. [doi: 10.30699/jambs.29.133.100]
16. Wang YH, Nguyen PA, Islam MM, Li YC, Yang HC. Development of deep learning algorithm for detection of colorectal cancer in EHR data. Stud Health Technol Inform 2019 Aug 21;264:438-441. [doi: 10.3233/SHTI190259] [Medline: 31437961]
17. Hussan H, Zhao J, Badu-Tawiah AK, et al. Utility of machine learning in developing a predictive model for early-age-onset colorectal neoplasia using electronic health records. PLoS One 2022;17(3):e0265209. [doi: 10.1371/journal.pone.0265209] [Medline: 35271664]
18. Paschke S, Jafarov S, Staib L, et al. Are colon and rectal cancer two different tumor entities? A proposal to abandon the term colorectal cancer. Int J Mol Sci 2018 Aug 30;19(9):2577. [doi: 10.3390/ijms19092577] [Medline: 30200215]
19. Shenkman E, Hurt M, Hogan W, et al. OneFlorida Clinical Research Consortium: linking a clinical and translational science institute with a community-based distributive medical education model. Acad Med 2018 Mar;93(3):451-455. [doi: 10.1097/ACM.0000000000002029] [Medline: 29045273]
20. The National Patient-Centered Clinical Research Network. 2020. URL: https://pcornet.org/ [accessed 2025-04-21]
21. Jacovidis JN, Foelber KJ, Horst SJ. The effect of propensity score matching method on the quantity and quality of matches. J Exp Educ 2017 Oct 2;85(4):535-558. [doi: 10.1080/00220973.2016.1250209]
22. Hinton GE. Learning distributed representations of concepts. University of Toronto Department of Computer Science. 1986. URL: https://www.cs.toronto.edu/~hinton/absps/families.pdf [accessed 2023-08-31]
23. Zhou Y, Shi J, Stein R, et al. Missing data matter: an empirical evaluation of the impacts of missing EHR data in comparative effectiveness research. J Am Med Inform Assoc 2023 Jun 20;30(7):1246-1256. [doi: 10.1093/jamia/ocad066] [Medline: 37337922]
24. Van Roode S, Ruiz-Aguilar JJ, González-Enrique J, Turias IJ. An artificial neural network ensemble approach to generate air pollution maps. Environ Monit Assess 2019 Nov 7;191(12):727. [doi: 10.1007/s10661-019-7901-6] [Medline: 31701254]
25. Zhang Z. Missing data imputation: focusing on single imputation. Ann Transl Med 2016 Jan;4(1):9. [doi: 10.3978/j.issn.2305-5839.2015.12.38] [Medline: 26855945]
26. Bastarache L. Using Phecodes for research with the electronic health record: from PheWAS to PheRS. Annu Rev Biomed Data Sci 2021 Jul 20;4:1-19. [doi: 10.1146/annurev-biodatasci-122320-112352] [Medline: 34465180]
27. PheWAS - Phenome Wide Association Studies. URL: https://phewascatalog.org/ [accessed 2023-10-18]

28. Association AM. Introduction to CPT Coding: Basic Principles to Learning, Understanding, and Applying the CPT Code Set: American Medical Association Press; 2011.
29. Elixhauser A. Clinical classifications for health policy research, version 2: software and user's guide. : Agency for Health Care Policy and Research; 1996.
30. The Food and Drug Administration's National Drug Code Directory: Createspace Independent Publishing Platform; 2018.
31. Anatomical therapeutic chemical (ATC) classification index with defined daily doses (DDDs). WHO Collaborating Centre for Drug Statistics Methodology.: World Health Organization; 2000 URL: https://atcddd.fhi.no/atc/structure_and_principles/
32. Shapley LS. A value for N-person games. : The RAND Corporation; 1952.
33. Denlinger CS, Weinberg DS. Sporadic breast, ovarian, or uterine cancers as risk factors for colorectal cancer. Curr Colorectal Cancer Rep 2009 Jan;5(1):24-29. [doi: 10.1007/s11888-009-0004-y]
34. Clinical practice guidelines in oncology (NCCN guidelines) survivorship. NCCN. URL: https://www.nccn.org/professionals/physician_gls/pdf/survivorship.pdf [accessed 2024-07-02]
35. Brown KGM, Solomon MJ, Mahon K, O'Shannassy S. Management of colorectal cancer. BMJ 2019 Aug 22;366:l4561. [doi: 10.1136/bmj.l4561] [Medline: 31439545]
36. Bhatia S, Sklar C. Second cancers in survivors of childhood cancer. Nat Rev Cancer 2002 Feb;2(2):124-132. [doi: 10.1038/nrc722] [Medline: 12635175]
37. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol 2019;14(2):89-103. [doi: 10.5114/pg.2018.81072] [Medline: 31616522]
38. Beirer A. Malnutrition and cancer, diagnosis and treatment. Memo 2021 Jun;14(2):168-173. [doi: 10.1007/s12254-020-00672-3]
39. Low EE, Demb J, Liu L, et al. Risk factors for early-onset colorectal cancer. Gastroenterology 2020 Aug;159(2):492-501. [doi: 10.1053/j.gastro.2020.01.004] [Medline: 31926997]
40. Ye P, Xi Y, Huang Z, Xu P. Linking obesity with colorectal cancer: epidemiology and mechanistic insights. Cancers (Basel) 2020 May 29;12(6):1408. [doi: 10.3390/cancers12061408] [Medline: 32486076]
41. Hu C, Zhang Q, Jin X, et al. A paradox between preoperative overweight/obesity and change in weight during postoperative chemotherapy and its relationship to survival in stage II and III colorectal cancer patients. Clin Nutr 2021 Apr;40(4):2410-2419. [doi: 10.1016/j.clnu.2020.10.039] [Medline: 33183882]
42. Agus A, Denizot J, Thévenot J, et al. Western diet induces a shift in microbiota composition enhancing susceptibility to Adherent-Invasive E. coli infection and intestinal inflammation. Sci Rep 2016 Jan 8;6:19032. [doi: 10.1038/srep19032] [Medline: 26742586]
43. Coghill AE, Shiels MS, Suneja G, Engels EA. Elevated cancer-specific mortality among HIV-infected patients in the United States. J Clin Oncol 2015 Jul 20;33(21):2376-2383. [doi: 10.1200/JCO.2014.59.5967] [Medline: 26077242]
44. Pai SY, Lurain K, Yarchoan R. How immunodeficiency can lead to malignancy. Immunology 2021 Dec 10;2021(1):287-295. [doi: 10.1182/hematology.2021000261]
45. Brandt PS, Kho A, Luo Y, et al. Characterizing variability of electronic health record-driven phenotype definitions. J Am Med Inform Assoc 2023 Feb 16;30(3):427-437. [doi: 10.1093/jamia/ocac235] [Medline: 36474423]
46. Sarwar T, Seifollahi S, Chan J, et al. The secondary use of electronic health records for data mining: data characteristics and challenges. ACM Comput Surv 2023 Feb 28;55(2):1-40. [doi: 10.1145/3490234]
47. Kohane IS, Aronow BJ, Avillach P, et al. What every reader should know about studies using electronic health record data but may be afraid to ask. J Med Internet Res 2021 Mar 2;23(3):e22219. [doi: 10.2196/22219] [Medline: 33600347]
48. Sunck1/CRC. GitHub. URL: https://github.com/sunck1/CRC [accessed 2025-06-13]

Abbreviations

AUC: area under the curve
CC: colon cancer
CRC: colorectal cancer
CT: computer tomography
EHR: electronic health record
EOCRC: early-onset colorectal cancer
HPV: human papillomavirus
IBD: inflammatory bowel diseases
***ICD-10*:** *International Classification of Diseases, Tenth Revision*
***ICD-9*:** *International Classification of Diseases, Ninth Revision*
LR: logistic regression
ML: machine learning
MRI: magnetic resonance imaging
RC: rectal cancer
RF: random forest

SHAP: Shapley Additive Explanations
SVM: support vector machine
XGBoost: Extreme Gradient Boosting

Edited by N Cahill; submitted 19.07.24; peer-reviewed by J Adeoye, S Zeng; revised version received 29.04.25; accepted 30.04.25; published 19.06.25.

Please cite as:
Sun C, Mobley E, Quillen M, Parker M, Daly M, Wang R, Visintin I, Awad Z, Fishe J, Parker A, George T, Bian J, Xu J
Predicting Early-Onset Colorectal Cancer in Individuals Below Screening Age Using Machine Learning and Real-World Data: Case Control Study
JMIR Cancer 2025;11:e64506
URL: https://cancer.jmir.org/2025/1/e64506
doi:10.2196/64506

Large Language Model Approach for Zero-Shot Information Extraction and Clustering of Japanese Radiology Reports: Algorithm Development and Validation

Yosuke Yamagishi[1], MD, MSc; Yuta Nakamura[2], MD, PhD; Shouhei Hanaoka[1], MD, PhD; Osamu Abe[1], MD, PhD

[1]Division of Radiology and Biomedical Engineering, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[2]Department of Computational Diagnostic Radiology and Preventive Medicine, The University of Tokyo Hospital, Tokyo, Japan

Corresponding Author:
Yosuke Yamagishi, MD, MSc
Division of Radiology and Biomedical Engineering, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Abstract

Background: The application of natural language processing in medicine has increased significantly, including tasks such as information extraction and classification. Natural language processing plays a crucial role in structuring free-form radiology reports, facilitating the interpretation of textual content, and enhancing data utility through clustering techniques. Clustering allows for the identification of similar lesions and disease patterns across a broad dataset, making it useful for aggregating information and discovering new insights in medical imaging. However, most publicly available medical datasets are in English, with limited resources in other languages. This scarcity poses a challenge for development of models geared toward non-English downstream tasks.

Objective: This study aimed to develop and evaluate an algorithm that uses large language models (LLMs) to extract information from Japanese lung cancer radiology reports and perform clustering analysis. The effectiveness of this approach was assessed and compared with previous supervised methods.

Methods: This study employed the MedTxt-RR dataset, comprising 135 Japanese radiology reports from 9 radiologists who interpreted the computed tomography images of 15 lung cancer patients obtained from Radiopaedia. Previously used in the NTCIR-16 (NII Testbeds and Community for Information Access Research) shared task for clustering performance competition, this dataset was ideal for comparing the clustering ability of our algorithm with those of previous methods. The dataset was split into 8 cases for development and 7 for testing, respectively. The study's approach involved using the LLM to extract information pertinent to lung cancer findings and transforming it into numeric features for clustering, using the K-means method. Performance was evaluated using 135 reports for information extraction accuracy and 63 test reports for clustering performance. This study focused on the accuracy of automated systems for extracting tumor size, location, and laterality from clinical reports. The clustering performance was evaluated using normalized mutual information, adjusted mutual information , and the Fowlkes-Mallows index for both the development and test data.

Results: The tumor size was accurately identified in 99 out of 135 reports (73.3%), with errors in 36 reports (26.7%), primarily due to missing or incorrect size information. Tumor location and laterality were identified with greater accuracy in 112 out of 135 reports (83%); however, 23 reports (17%) contained errors mainly due to empty values or incorrect data. Clustering performance of the test data yielded an normalized mutual information of 0.6414, adjusted mutual information of 0.5598, and Fowlkes-Mallows index of 0.5354. The proposed method demonstrated superior performance across all evaluation metrics compared to previous methods.

Conclusions: The unsupervised LLM approach surpassed the existing supervised methods in clustering Japanese radiology reports. These findings suggest that LLMs hold promise for extracting information from radiology reports and integrating it into disease-specific knowledge structures.

(JMIR Cancer 2025;11:e57275) doi:10.2196/57275

KEYWORDS

radiology reports; clustering; large language model; natural language processing; information extraction; lung cancer; machine learning

Introduction

Natural language processing (NLP) is vital in medicine as it allows the interpretation of textual content in medical documents. Radiology reports, written as free text by experienced radiologists, contain detailed information about medical imaging findings. While medical images are valuable, text-based analysis offers unique advantages in terms of computational efficiency and the ability to capture expert interpretations and observations of radiologists that may not be immediately apparent from images. Natural language processing can effectively extract this information, enhance its utilization, and provide new insights into medical imaging.

Advances in radiological NLP applications are driven by the availability of large datasets [1]. For example, the MIMIC Chest X-ray (MIMIC-CXR) includes more than 200,000 images, English-language reports, and structured data [2]. Numerous NLP models have been developed to summarize and extract clinical entities [3,4]. However, the availability of these datasets in languages other than English is limited.

To address this challenge, the NTCIR-16 Real-MedNLP shared task focused on clustering Japanese radiology reports by case basis. It is a set of Japanese radiology reports authored by different radiologists for the same case series of lung cancer, and the task was to cluster reports that describe the same medical case together [5]. This benchmark evaluates the detailed understanding of radiology reports, as NLP systems must extract sufficient information to recognize reports by diagnosing the same image without being affected by different writing styles.

Clustering is a powerful analytical tool in medicine and has been successfully applied in various clinical domains. Studies have demonstrated its effectiveness in clustering patients based on their clinical characteristics to guide medical decisions, ranging from cancer aftercare planning to pulmonary embolism risk assessment [6,7]. Semantic grouping has enabled efficient insight discovery in medical documents [8] and revealed specialty-specific sublanguages in clinical narratives [9]. Radiology reports are particularly suited for such analyses, as they provide high-quality annotated data despite their free-form nature, offering a more tractable alternative to direct image analysis.

While the participants in the NTCIR-16 (NII Testbeds and Community for Information Access Research) shared task used deep-learning models, their clustering performance was constrained by limited training data. Since then, large language models (LLMs) trained on extensive text corpora, such as ChatGPT and LLaMA [10,11], have emerged. These LLMs, which are adaptable to new tasks with minimal instructions or examples, have demonstrated high performance in extracting information from medical documents, even under zero-shot conditions [12].

This study aimed to evaluate the ability of LLM to understand real radiological reports through an information extraction task and apply this information to clustering, which is a clinically meaningful task.

Methods

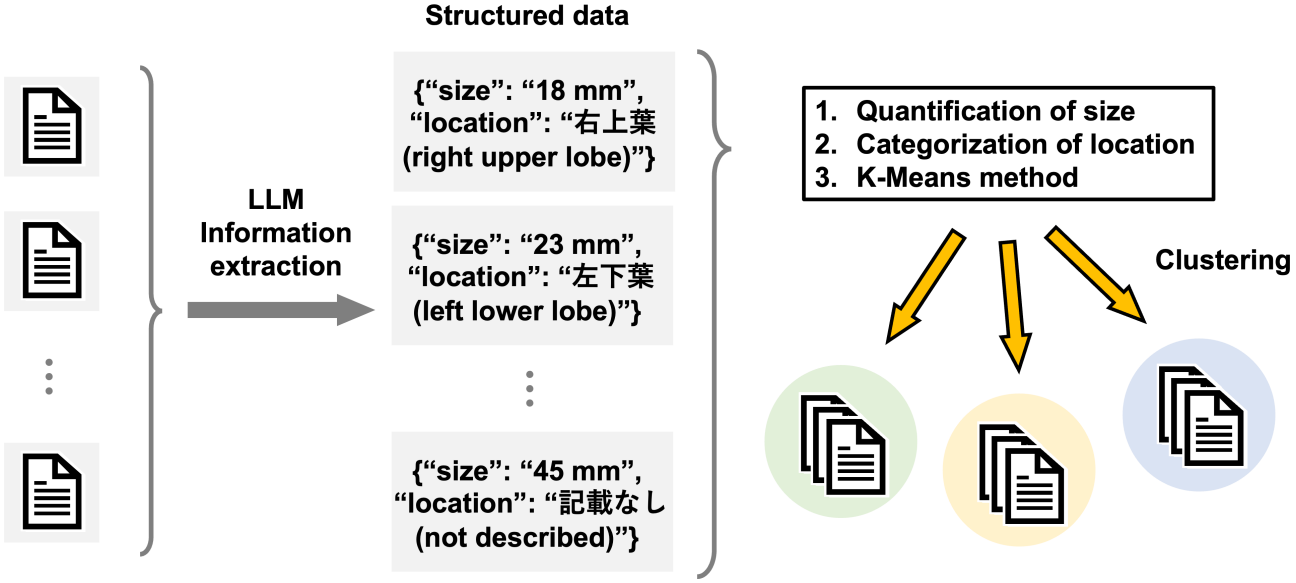
Study Design and Reporting Guidelines

This retrospective observational study followed the relevant items of the checklist for Artificial Intelligence in medical imaging (CLAIM) guidelines for methodology reporting [13,14]. Although this study analyzed text rather than images, CLAIM was followed because it is an established guideline for AI-based research in radiology and is deemed appropriate for NLP [15-17].

Algorithm Overview

The proposed algorithm is illustrated in Figure 1. Using the LLM, key lung cancer findings were extracted from radiology reports and quantified to obtain structured data. The structured data were subsequently used for clustering.

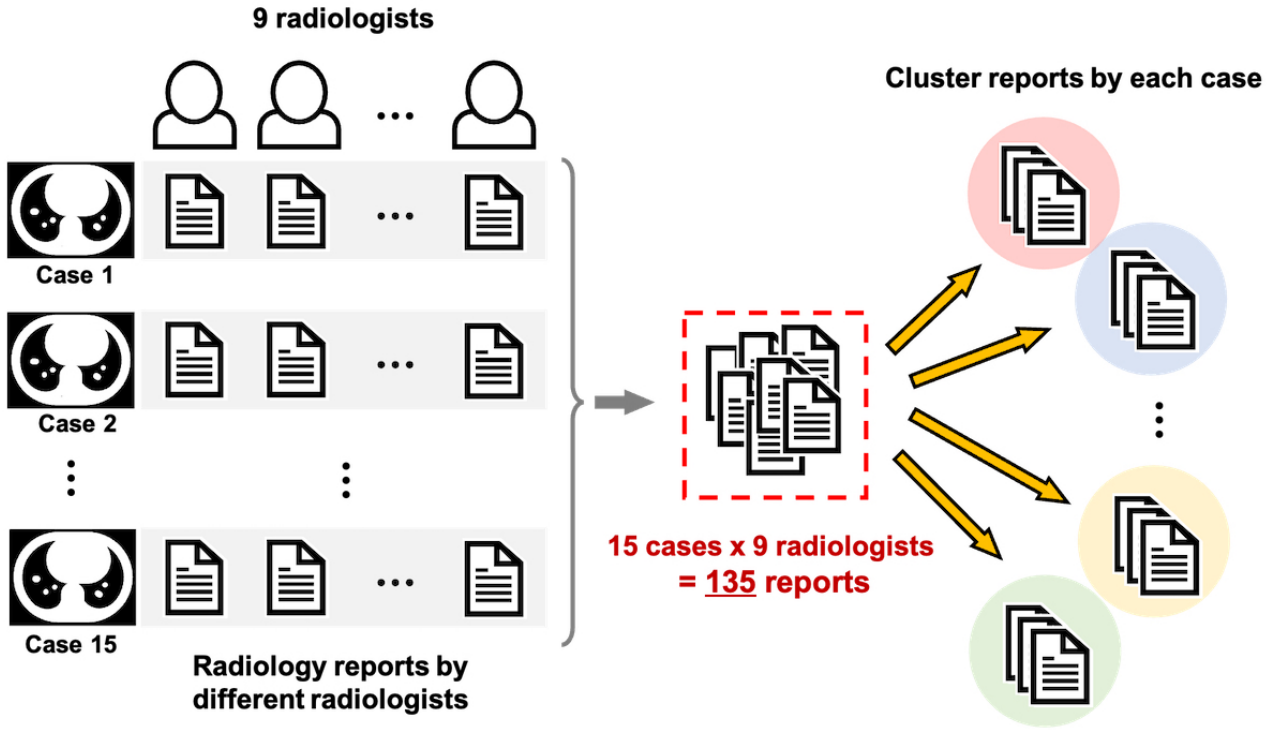
Figure 1. Flowchart of radiology reports clustering using LLM. LLM: large language model.



Dataset

The MedTxt-RR dataset was used in this study [5,18], comprising 135 Japanese radiology reports generated by 9 radiologists who interpreted CT images of 15 lung cancer cases sourced from Radiopaedia [19]. This dataset was used in an NTCIR-16 shared task [5], where participants competed to achieve optimal clustering performance. With each case comprising reports from 9 radiologists, the dataset was suitable for evaluating the clustering performance on a per-case (Figure 2). Eight cases and seven cases were assigned to the development and test sets, respectively. While no model training was conducted using the development set in this study, performance was evaluated on the same data split to facilitate comparison with the shared task results.

Figure 2. Overview diagram of the radiology report clustering task.



LLM Approach

Radiology reports contain confidential patient information; processing them using a cloud-based LLM, such as ChatGPT, could expose sensitive data externally, raising significant medical safety concerns. Therefore, a publicly available offline model was selected as an alternative approach.

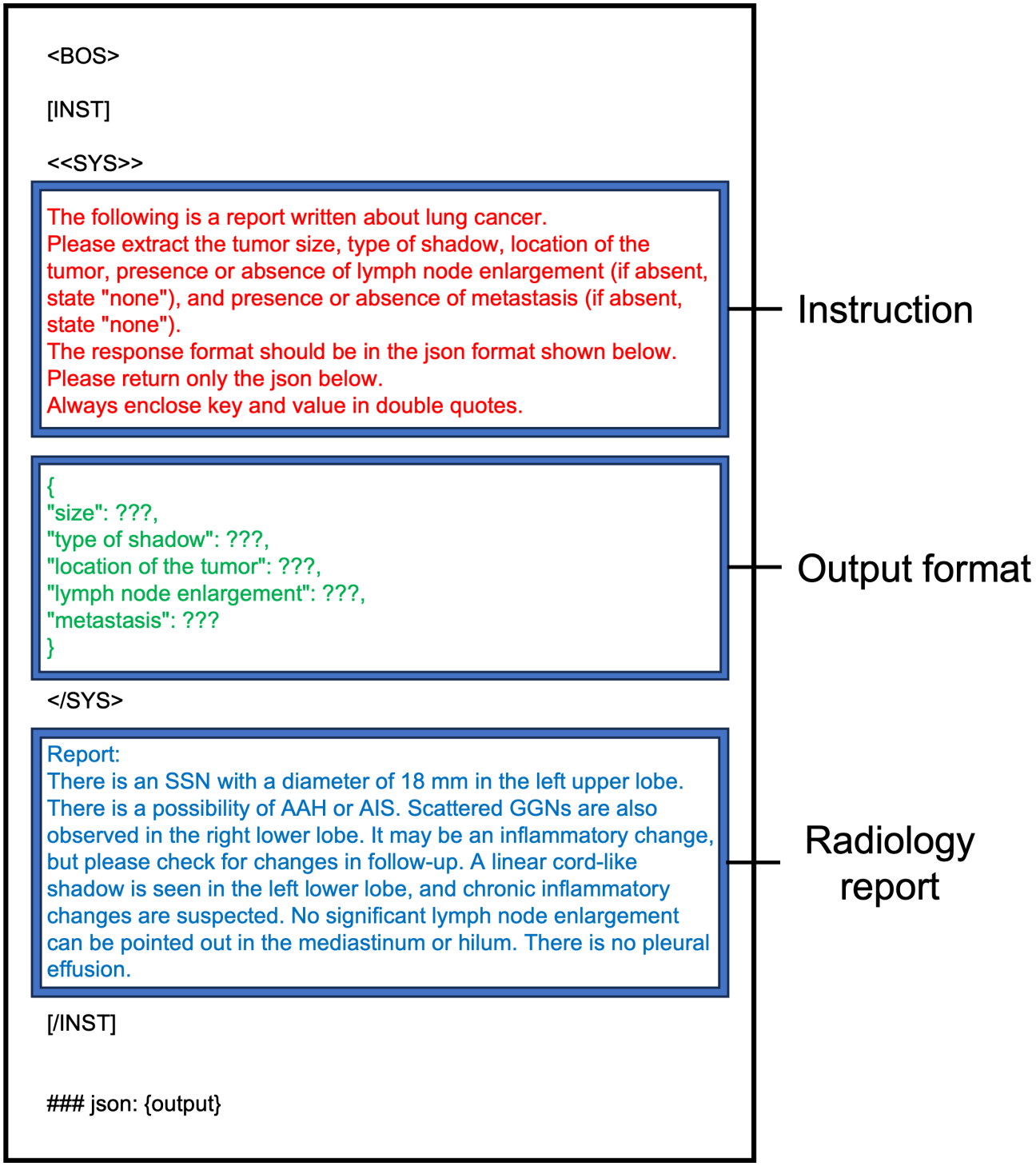
The ELYZA-Japanese-Llama-2-7b-fast-instruct model was employed as the LLM [20]. Adapted from Llama2 and pre-trained using Japanese datasets, this model demonstrated a performance comparable to that of GPT-3.5 on Japanese datasets [21-23].

Information Extraction

The LLM extracted multiple lung cancer staging parameters from radiology reports, including tumor size, tumor location, and the presence or absence of lymph node enlargement, suggesting metastasis and distant metastasis. To determine the optimal combination of features, clustering performance of the development set were repeatedly measured by using certain features. Consequently, sufficient clustering performance was confirmed achievable using only 3 parameters: tumor size, laterality (left or right), and lung location (upper, hilum, or lower region).

The prompt input into the LLM comprises system instructions and output format guidelines using json (JavaScript Object Notation), a standardized text-based format for structured information exchange, where data is organized in key-value pairs, such as {“size”: “45 mm,” “location”: “right upper lobe”}. These system instructions guided the LLM in extracting features from the radiology reports. The details of these prompts are shown in Figure 3 (English version) and Figure S1 in Multimedia Appendix 1 (original Japanese version).

Figure 3. Example of a prompt used as input for the LLM (English translated version). AAH: Atypical Adenomatous Hyperplasia; AIS: Adenocarcinoma in situ; GGN: Ground Glass Nodule; LLM: large language model; SSN: Subsolid Nodule.



The extracted data were converted into integer vectors comprising the tumor size and other categorical values. Unspecified tumor sizes only described as *large* were replaced with 71 mm, corresponding to the highest category in T classification, where T represents tumor categories in cancer staging. The details of this pipeline can be found in the GitHub Repository [24].

Moreover, a rule-based method was employed as the baseline approach and its performance was compared with that of the proposed method. The rule-based method performs context-sensitive word-based information extraction; the detailed algorithm is shown in Figure S2 in Multimedia Appendix 1.

Clustering

The resulting numerical matrices were clustered using the K-means algorithm in the scikit-learn library (version 1.3.1). The number of clusters was set to 8, aligning with expected classifications such as disease type or staging, since it was close to the number of test data cases. Centroid initialization used the k-means++ method, with default values for the centroid seed

and iteration count, because hyperparameter tuning was not conducted in this zero-shot study.

Information Extraction Evaluation

Two independent radiologists, a radiology resident with 1-year experience and a board-certified radiologist with 7 years of experience evaluated the accuracy of extracted information. In cases of discrepancy, the final assessment was determined by consensus. Evaluation focused on three key elements: tumor size, location (upper, hilum, or lower), and laterality (left or right). The performance of the LLM-based approach was compared to that of the rule-based method for information extraction. A detailed error analysis was conducted for cases with errors, categorizing them into missing information, false information generation, and extraction of multiple values.

McNemar's test was performed using Statsmodels (version 0.14.2) [25] to compare performance differences between the LLM-based and rule-based approaches for extracting tumor size and location.

Clustering Performance Evaluation

We assessed clustering performance using three metrics similar to those used in the shared task [5]: (1) normalized mutual information (NMI) that quantifies the mutual dependence between two clusters, normalized to a 0‐1 scale, with 1 indicating perfect clustering; (2) adjusted mutual information (AMI) which is an adjustment that corrects for NMI, accounting for its tendency to increase with the number of clusters; (3) Fowlkes-Mallows index (FM) that measures the similarity between two clusters by calculating the geometric mean of precision and recall, providing a balanced assessment of clustering accuracy.

Ethics Consideration

This study involved analysis of human subject data from publicly available radiology reports. All data were completely de-identified and accessible through MedTxt-RR [26]. In accordance with our institution's policy on research ethics, studies using exclusively de-identified, public datasets are exempt from institutional review board approval [27]. No additional privacy or confidentiality measures were required as the dataset contains no personally identifiable information, with all protected health information having been removed prior to public release.

Results

Information Extraction Performance

The details of the findings targeted at information extraction are summarized in Table 1 . The tumor size was correctly identified in 99 (73.3%) of 135 reports. Among the 36 outputs (26.7%) with errors, 23 (17%) lacked size information in their reports, and 22 (16.3%) contained false size information. The remaining errors were attributed to size inaccuracies or empty values despite size information being mentioned in the reports. Tumor location and laterality were accurately identified in 112 (83%) reports. All 23 (17%) reports with errors contained the necessary information but had empty values for laterality, location, or both, with one output indicating an incorrect location. The detailed error analysis is presented in Table 2.

Table . Summary of lung cancer cases.

Case no.	Side	Lobe	Size (mm)	Lymph node metastasis	Distant metastasis	Data split
Case 1	Left	Upper	18	No	No	Development
Case 2	Right	Lower	12	No	No	Development
Case 3	Left	Upper	28	No	No	Development
Case 4	Left	Upper	40	No	No	Test
Case 5	Left	Upper	48	Yes	No	Test
Case 6	Right	Hilum	Not measurable (due to invasion)	Yes	No	Development
Case 7	Right	Lower	55	Yes	No	Test
Case 8	Left	Upper	Not measurable (due to invasion)	No	No	Test
Case 9	Right	Hilum	43	No	Yes	Development
Case 10	Right	Upper	Not measurable (due to invasion)	No	No	Test
Case 11	Right	Upper	Not measurable (due to invasion)	No	No	Development
Case 12	Right	Lower	Not measurable (due to lung metastasis)	No	Yes	Development
Case 13	Left	Lower	78	Yes	No	Development
Case 14	Left	Upper	85	Yes	No	Test
Case 15	Left	Upper	Not measurable (due to invasion)	Yes	Yes	Test

Table . Detailed error analysis of tumor size, location, and laterality extraction from radiology reports using large language model (LLM) and rule-based methods.

Category	Extraction methods	
	LLM[a], n (%)	Rule-based, n (%)
Tumor size (details)		
Correctly identified	99 (73.3)	93 (68.9)
Errors (total)	36 (26.7)	42 (31.1)
Errors (no size information in reports)	23 (17)	0 (0)
False size information generated	22 (16.3)	0 (0)
T classification extracted instead of size	1 (0.7)	0 (0)
Errors (size mentioned in reports)	13 (9.6)	42 (31.1)
Size inaccuracies	8 (5.9)	3 (2.2)
Empty values	5 (3.7)	39 (28.9)
Tumor location and laterality (details)		
Accurately reported	112 (83)	46 (34.1)
Errors (total)	23 (17)	89 (65.9)
Empty values for laterality	9 (6.7)	0 (0)
Empty values for location	5 (3.7)	0 (0)
Empty values for both	8 (5.9)	80 (59.3)
Incorrect location	1 (0.7)	9 (6.7)

[a]LLM: large language model

The rule-based method correctly identified tumor size in 93 (68.9%) reports, whereas tumor location and laterality were accurately identified in only 46 (34.1%) reports. Among the errors in this method, only 1 case (0.7%) failed to accurately extract size information due to the extraction of multiple sizes. In contrast, for location, the number of errors reached 47 (34.8%) (Figure S3 in Multimedia Appendix 1). Unlike the LLM approach, due to the algorithmic nature of rule-based extraction, there were no cases of false-size information generation. Additionally, as the algorithm extracted laterality and location simultaneously as a single unit, there were no cases where only one of these values was empty; both were either extracted together or left empty.

McNemar's test showed that the LLM approach was significantly superior to the rule-based method in determining location (P<.001) but not size (P=.539).

Clustering Performance

The development data yielded an NMI score of 0.7152, an AMI score of 0.6516, and an FM index of 0.5959, whereas the test data yielded scores of 0.6414 (NMI), 0.5598 (AMI), and 0.5354 (FM).

The proposed method outperformed all previous methods in shared tasks across all evaluation metrics. The detailed results and methods are listed in Table 3. Further details of each method are available in a system paper describing this shared task [28-31].

Table . Clustering scores on the test data.

Method Description (System ID[a])	NMI[b]	AMI[c]	FM[d]	Supervised model	LLM[e]
Developed a matrix from word count in radiology reports and applied user-based collaborative filtering for case similarity and clustering, (D1) [28]	0.3569	0.1988	0.2674	No	No
Used paired radiology reports for BERT[f] input, fine-tuned for same-case identification and clustered based on predictions, (E1) [29]	0.5415	0.1489	0.1814	Yes	No
Generated embeddings from text via multilingual BERT trained on Wikipedia, followed by dimensionality reduction, and K-means clustering, (F1) [30]	0.1744	–0.0117	0.1170	No	No
Labels simplified from the TNM[g] classification of lung cancer were assigned to each document using BERT-based model for training, and in the test data, these predicted labels were used as groups for clustering, (J1) [31]	0.4622	0.3409	0.3622	Yes	No
This study	0.6414	0.5598	0.5354	No	Yes

[a]The System IDs are those used in previously shared tasks with the same dataset [5].

[b]NMI: normalized mutual information

[c]AMI: adjusted mutual information

[d]FM: Fowlkes-Mallows index

[e]LLM: Large language model

[f]BERT: Bidirectional Encoder Representations from Transformers

[g]TNM: Tumor, node, metastasis

Discussion

Principal Findings

The extraction of lung tumor size showed minimal differences compared to the rule-based method, likely because size information is typically accompanied by standardized units (eg, mm or cm). However, the LLM method significantly outperformed the rule-based method in terms of location extraction, achieving over 80% accuracy and reducing the error rate by half. As demonstrated in Figure S3 in Multimedia Appendix 1, the rule-based method frequently generated multiple incorrect location extractions when reports mentioned various anatomical sites, whereas the LLM method successfully identified the correct tumor location. This finding empirically demonstrates the LLM's ability to understand and extract information based on context rather than predefined rules. This capability highlights its value for complex information extraction tasks in medical text analysis, where contextual understanding is crucial.

Comparison to Prior Work

This paper introduces a Japanese LLM algorithm for zero-shot information extraction and clustering that outperforms all previous methods [28-31]. The previous methods (E1, F1, and J1) relied on indirect features extracted by language models, whereas the current approach leverages accurate information extraction through unsupervised learning. The success of this method is particularly notable, given the historically low accuracy of unsupervised methods. By leveraging the LLM's contextual understanding of information extraction, this study demonstrated the potential for effective clustering of medical reports based on various attributes, including disease severity and lesion localization.

Strengths and Limitations

This study has several notable strengths including its methodology and implementation. Accurate information extraction and clustering without supervised learning requirements represent a significant advancement in the field. The flexibility of this method through prompt and algorithmic adjustments suggests broad potential applicability, with potential for further performance improvements through prompt optimization [32]. Furthermore, this method shows particular promise for languages with limited training data compared to English, by converting unstructured reports into language-independent structured data, thereby addressing a crucial gap in current medical text analysis.

However, the limitations must be acknowledged. First, validation was limited to small-scale Japanese datasets. While attempts were made to ensure the representativeness of the dataset by including diverse types of lung cancer cases, this limitation constrained the generalizability of the study findings and should be addressed in future studies through multi-institutional validation. Second, the evaluation focused primarily on clustering tasks; which although is a fundamental task in medical text analysis, its performance in other analytical tasks remains unexplored, suggesting the need for a comprehensive evaluation across various applications. Third, while this method shows promise for languages with limited training data, its generalizability to other languages and medical domains requires further investigation.

Conclusions

The LLM was used to successfully extract important findings from publicly available Japanese radiology reports as highly accurate structured data. By leveraging these structured data, superior results were achieved compared to existing supervised methods for clustering radiology reports. This indicates that employing existing LLMs is effective for solving specific tasks, particularly in languages with a significant shortage of training data compared to English.

Acknowledgments

Prior to submission, the initial draft was edited using ChatGPT by OpenAI for English language improvement, followed by professional editing by Editage. This study did not receive any specific grants from funding agencies in the public, commercial, or non-profit sectors.

Data Availability

The MedTxt-RR training dataset is openly accessible and downloadable via an official website [26]. Access to the test dataset requires approval and can be obtained by directly contacting the data providers.

Authors' Contributions

Research design: YY, YN, SH
Conceptualization: YY, YN, SH
Algorithm development and implementation: YY
Formal analysis: YY, YN
Data curation: YY, YN
Writing – original draft: YY, YN, SH
Writing – review and editing: YN, SH, OA

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of a prompt used as input for the LLM (Japanese original version), pseudo code illustrating the procedure for the rule-based processing, and data representation of extracted information based on the rule-based method.
[DOCX File, 731 KB - cancer_v11i1e57275_app1.docx]

References

1. Irvin J, Rajpurkar P, Ko M, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. AAAI 2019;33(1):590-597. [doi: 10.1609/aaai.v33i01.3301590]
2. Johnson AEW, Pollard TJ, Berkowitz SJ, et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Sci Data 2019 Dec 12;6(1):317. [doi: 10.1038/s41597-019-0322-0] [Medline: 31831740]
3. Delbrouck JB, Varma M, Chambon P, Langlotz C. Overview of the Radsum23 shared task on multi-modal and multi-anatomical radiology report summarization. In: Demner-fushman D, Ananiadou S, Cohen K, editors. The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks: Association for Computational Linguistics; 2023:478-482. [doi: 10.18653/v1/2023.bionlp-1.45]
4. Jain S, Agrawal A, Saporta A, et al. Radgraph: extracting clinical entities and relations from radiology reports. arXiv. Preprint posted online on 2021. [doi: 10.48550/arXiv.2106.14463]
5. Yada S, Nakamura Y, Wakamiya S, Aramaki E. Real-MedNLP: overview of real document-based medical natural language processing task. Presented at: Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies; Jun 14-17, 2022; Tokyo, Japan p. 285-296 URL: https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings16/pdf/ntcir/01-NTCIR16-OV-MEDNLP-YadaS.pdf [accessed 2024-02-10]
6. Beuken MJM, Kanera IM, Ezendam NPM, Braun S, Zoet M. Identification and potential use of clusters of patients with colorectal cancer and patients with prostate cancer in clinical practice: explorative mixed methods study. JMIR Cancer 2022 Dec 27;8(4):e42908. [doi: 10.2196/42908] [Medline: 36574281]
7. Ben Yehuda O, Itelman E, Vaisman A, Segal G, Lerner B. Early detection of pulmonary embolism in a general patient population immediately upon hospital admission using machine learning to identify new, unidentified risk factors: model development study. J Med Internet Res 2024 Jul 30;26:e48595. [doi: 10.2196/48595] [Medline: 39079116]
8. Lin Y, Li W, Chen K, Liu Y. A document clustering and ranking system for exploring MEDLINE citations. J Am Med Inform Assoc 2007;14(5):651-661. [doi: 10.1197/jamia.M2215] [Medline: 17600104]
9. Patterson O, Hurdle JF. Document clustering of clinical narratives: a systematic study of clinical sublanguages. AMIA Annu Symp Proc 2011;2011:1099-1107. [Medline: 22195171]
10. Achiam J, Adler S, Agarwal S, et al. Gpt-4 technical report. arXiv. Preprint posted online on 2023. [doi: 10.48550/arXiv.2303.08774]
11. Touvron H, Lavril T, Izacard G, et al. Llama: open and efficient foundation language models. arXiv. Preprint posted online on 2023. [doi: 10.48550/arXiv.2302.13971]
12. Agrawal M, Hegselmann S, Lang H, Kim Y, Sontag D. Large language models are few-shot clinical information extractors. Presented at: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; Dec 7-11, 2022; Abu Dhabi, United Arab Emirates p. 1998-2022 URL: https://aclanthology.org/2022.emnlp-main [doi: 10.18653/v1/2022.emnlp-main.130]
13. Mongan J, Moy L, Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020 Mar;2(2):e200029. [doi: 10.1148/ryai.2020200029] [Medline: 33937821]
14. Tejani AS, Klontzas ME, Gatti AA, et al. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 update. Radiol Artif Intell 2024 Jul;6(4):e240300. [doi: 10.1148/ryai.240300] [Medline: 38809149]
15. Tejani AS. To BERT or not to BERT: advancing non-invasive prediction of tumor biomarkers using transformer-based natural language processing (NLP). Eur Radiol 2023 Nov;33(11):8014-8016. [doi: 10.1007/s00330-023-10224-y] [Medline: 37740083]
16. Olthof AW, van Ooijen PMA, Cornelissen LJ. Deep learning-based natural language processing in radiology: the impact of report complexity, disease prevalence, dataset size, and algorithm type on model performance. J Med Syst 2021 Sep 4;45(10):91. [doi: 10.1007/s10916-021-01761-4] [Medline: 34480231]
17. Schachner T, Keller R, V Wangenheim F. Artificial Intelligence-Based conversational agents for chronic conditions: systematic literature review. J Med Internet Res 2020 Sep 14;22(9):e20701. [doi: 10.2196/20701] [Medline: 32924957]
18. Nakamura Y, Hanaoka S, Nomura Y, et al. Clinical comparable corpus describing the same subjects with different expressions. Presented at: MEDINFO 2021 p. 253-257. [doi: 10.3233/SHTI220073]
19. Radiopaedia.org, the peer-reviewed collaborative radiology resource. Radiopaedia. URL: https://radiopaedia.org/ [accessed 2024-02-10]

20. Sasaki A, Hirakawa M, Horie S, Nakamura T. ELYZA-japanese-llama-2-7b. Hugging face. URL: https://huggingface.co/elyza/ELYZA-japanese-Llama-2-7b [accessed 2024-02-10]
21. Touvron H, Martin L, Stone K, et al. Llama 2: open foundation and fine-tuned chat models. arXiv. Preprint posted online on 2023. [doi: 10.48550/arXiv.2307.09288]
22. Ortiz Su'arez PJ, Romary L, Sagot B. A monolingual approach to contextualized word embeddings for mid-resource languages. 2020 Presented at: In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics Association for Computational Linguistics p. 1703-1714 URL: https://www.aclweb.org/anthology/2020.acl-main.156 [accessed 2024-02-10]
23. Openai GPT-3 API. OpenAI. URL: https://openai.com/blog/openai-api [accessed 2024-02-10]
24. yamagishi0824/rad-report-clustering GitHub. URL: https://github.com/yamagishi0824/rad-report-clustering [accessed 2024-02-10]
25. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. 2010 Presented at: Python in Science Conference; Austin, Texas. [doi: 10.25080/Majora-92bf1922-011]
26. MedTxt-RR: radiology reports corpus – NAIST social computing lab. Sociocom. URL: https://sociocom.naist.jp/medtxt-en/rr/ [accessed 2024-11-15]
27. Research Ethics Committee of the Faculty of Medicine, The University of Tokyo.: The University of Tokyo, Graduate School of Medicine and Faculty of Medicine URL: https://www.m.u-tokyo.ac.jp/ethics/ethcom/gakugai2/index.html [accessed 2024-12-25]
28. Noguchi R. GunNLP at the NTCIR-16 real-MEDNLP task: collaborative filtering-based similar case identification method via structured data "case matrix". 2022 Presented at: In: Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies p. 349-352 URL: https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings16/pdf/ntcir/11-NTCIR16-MEDNLP-NoguchiR.pdf [accessed 2024-02-10]
29. Nishiyama T, Nishidani M, Ando A, et al. NAISTSOC at the NTCIR-16 real-MedNLP task. 2022 Presented at: In: Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies p. 330-333 URL: https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings16/pdf/ntcir/07-NTCIR16-MEDNLP-NishiyamaT.pdf [accessed 2024-02-10]
30. Ideuchi M, Tsuchiya M, Wang Y, Utiyama M, et al. NICTmed at the NCTIR-16 real-mednlp task. 2022 Presented at: In: Proceedings of the 16th NTCIR conference on evaluation of information access technologies p. 339-344 URL: https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings16/pdf/ntcir/09-NTCIR16-MEDNLP-IdeuchiM.pdf [accessed 2024-02-10]
31. Fujimoto K, Nishio M, Sugiyama O, et al. Approach for named entity recognition and case identification implemented by Zukyo-JA sub-team at the NTCIR-16 real-medNLP task. 2022 Presented at: In: Proceedings of the 16th NTCIR conference on evaluation of information access technologies p. 322-329 URL: https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings16/pdf/ntcir/06-NTCIR16-MEDNLP-FujimotoK.pdf [accessed 2024-02-10]
32. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023 Oct 4;25:e50638. [doi: 10.2196/50638] [Medline: 37792434]

Abbreviations

AMI: adjusted mutual information
CLAIM: Checklist for Artificial Intelligence in Medical Imaging
FM: Fowlkes-Mallows index
JSON: JavaScript Object Notation
LLM: large language model
MIMIC-CXR: MIMIC Chest X-ray
NLP: natural language processing
NMI: normalized mutual information
NTCIR-16: NII Testbeds and Community for Information Access Research

Edited by N Cahill; submitted 11.02.24; peer-reviewed by D Hu, M Jani, R Shraga; revised version received 16.12.24; accepted 18.12.24; published 23.01.25.

Please cite as:
Yamagishi Y, Nakamura Y, Hanaoka S, Abe O
Large Language Model Approach for Zero-Shot Information Extraction and Clustering of Japanese Radiology Reports: Algorithm Development and Validation
JMIR Cancer 2025;11:e57275
URL: https://cancer.jmir.org/2025/1/e57275
doi:10.2196/57275

Leveraging Digital Twins for Stratification of Patients with Breast Cancer and Treatment Optimization in Geriatric Oncology: Multivariate Clustering Analysis

Pierre Heudel[1*], MSc, LLM, MD; Mashal Ahmed[2*], PhD; Felix Renard[2*], PhD; Arnaud Attye[2*], MD, PhD

[1]Department of Medical Oncology, Centre Leon Bérard, 28 rue Laennec, Lyon, France

[2]GEODAISICS, Grenoble, France

[*]all authors contributed equally

Corresponding Author:
Pierre Heudel, MSc, LLM, MD
Department of Medical Oncology, Centre Leon Bérard, 28 rue Laennec, Lyon, France

Abstract

Background: Defining optimal adjuvant therapeutic strategies for older adult patients with breast cancer remains a challenge, given that this population is often overlooked and underserved in clinical research and decision-making tools.

Objectives: This study aimed to develop a prognostic and treatment guidance tool tailored to older adult patients using artificial intelligence (AI) and a combination of clinical and biological features.

Methods: A retrospective analysis was conducted on data from women aged 70+ years with HER2-negative early-stage breast cancer treated at the French Léon Bérard Cancer Center between 1997 and 2016. Manifold learning and machine learning algorithms were applied to uncover complex data relationships and develop predictive models. Predictors included age, BMI, comorbidities, hemoglobin levels, lymphocyte counts, hormone receptor status, Scarff-Bloom-Richardson grade, tumor size, and lymph node involvement. The dimension reduction technique PaCMAP was used to map patient profiles into a 3D space, allowing comparison with similar cases to estimate prognoses and potential treatment benefits.

Results: Out of 1229 initial patients, 793 were included after data refinement. The selected predictors demonstrated high predictive efficacy for 5-year mortality, with mean area under the curve scores of 0.81 for Random Forest Classification and 0.76 for Support Vector Classifier. The tool categorized patients into prognostic clusters and enabled the estimation of treatment outcomes, such as chemotherapy benefits. Unlike traditional models that focus on isolated factors, this AI-based approach integrates multiple clinical and biological features to generate a comprehensive biomedical profile.

Conclusions: This study introduces a novel AI-driven prognostic tool for older adult patients with breast cancer, enhancing treatment guidance by leveraging advanced machine learning techniques. The model provides a more nuanced understanding of disease dynamics and therapeutic strategies, emphasizing the importance of personalized oncology care.

(JMIR Cancer 2025;11:e64000) doi:10.2196/64000

KEYWORDS

digital twins; artificial intelligence; breast cancer; older adult patients with cancer; treatment; geriatric oncology; geriatric; oncology; cancer; clustering analysis; therapeutic; older adult; elder; old; patients with cancer; decision-making tools; decision-making; manifold learning model; chemotherapy; comorbidities; comorbidity; health care

Introduction

Breast cancer is more commonly diagnosed in older populations, particularly among women aged 65 years and older in wealthier countries. In the United States, the average age of breast cancer diagnosis is 62 years, and in 2020, women aged 70 years and older accounted for 30% of all new cases of the disease [1,2]. In the European Union, women older than 65 years made up about 44% of all breast cancer cases [3]. However, treatment approaches for early-stage breast cancer in these older age groups are often inadequate and unclear, largely due to a lack of solid evidence and the unreliability of web-based tools for making decisions about additional therapies, leading to less than ideal treatment outcomes [4,5].

The treatment plan for breast cancer is tailored based on the cancer's characteristics, the patients' overall health status, and their personal preferences. Standard care for early-stage breast cancer usually involves surgery, and may also include radiation, as well as neoadjuvant or adjuvant systemic therapy, used alone or in various combinations. Crafting postsurgical treatment strategies for older patients with breast cancer is complex due to their typically compromised health and the lack of data from

clinical trials, since older adults are seldom participants in such studies and are not well represented in meta-analyses that evaluate the effectiveness of adjuvant chemotherapy in reducing breast cancer mortality and improving survival rates [6,7]. Consequently, artificial intelligence (AI) has been investigated as a potential tool to support decision-making in the context of limited clinical trial evidence.

Early uses of AI in cancer treatment guidance involved knowledge-based systems [8,9]. Recently, a broader spectrum of machine learning methods has been examined to aid both clinicians and patients with breast cancer [10-15]. Nonetheless, most decision support tools are designed for patients aged between 18 and 65 years, reflecting the age group most studied, with limited research focusing on treatment outcomes for older patients with breast cancer [16-19]. The prognostic tool PREDICT [20], although popular, has shown limited effectiveness for older adult patients [21]. Adjutorium [22], which uses extensive datasets from the United Kingdom and the United States, provides more precise prognosis and treatment benefit predictions for breast cancer than PREDICT. Despite this, it primarily includes patients aged between 30 and 65 years, with fewer older patients in its datasets, and omits certain vital tumor information such as progesterone receptor (PR) status [19]. Another established tool, Adjuvant! Online, predicts 10-year overall survival, breast cancer survival, and recurrence rates, commonly used to inform expected outcomes from endocrine therapy and chemotherapy [23]. Its accuracy is questionable for older women with early-stage breast cancer, probably because it was trained on data with a maximum age limit of 69 years [24]. In a review by Engelhardt et al [25], various models could forecast breast cancer outcomes, typically based on genetic risk scores, but only Adjuvant! Online factored in comorbidity status. Yet, none had been thoroughly validated in older adult populations. The more recent PORTRET tool was designed to predict 5-year recurrence, overall mortality, and mortality from other causes in patients older than 65 years with early invasive breast cancer, as well as to estimate the benefits of adjuvant systemic treatment [26]. The tool's authors observed that their treatment effect estimates were based on data from pooled randomized clinical trials, which might not be entirely applicable to older adults due to the typically selective nature of older participants in these trials.

This study aims to develop models that overcome the shortcomings of past research by using cohorts that accurately reflect the demographic of older patients with breast cancer and by leveraging a detailed dataset that includes administrative, biological, treatment, primary tumor, and survival information. Our latest research uses manifold learning, an advanced tool for nonlinear dimensionality reduction that excels in unraveling complex geometric relationships within high-dimensional data, revealing intricate connections between clinical factors.

We introduce a new prognostic and predictive tool tailored for older adult patients with breast cancer, providing postoperative treatment recommendations. This tool is distinctive in its consideration of the interdependencies among variables within a patient population. It acknowledges the relative importance of prognostic factors in a way that many existing models do not. Our findings are set to be extremely beneficial for oncologists when determining suitable adjuvant treatment approaches for older adult patients with breast cancer, taking into account the nuances of both tumor-related and patient-specific characteristics.

Methods

Recruitment

In this retrospective study, we examined pseudonymized data from women aged 70 years and older who received a diagnosis of early-stage breast cancer and underwent surgery with the intent to cure (either lumpectomy or mastectomy, with or without axillary lymph node dissection) at the French Léon Bérard Cancer Center from January 1997 to December 2016. The French Léon Bérard Cancer Center is a 300-bed comprehensive cancer center located in Lyon, France, serving more than 30,000 patients annually, with a multidisciplinary team of 2000 health care professionals and a catchment area covering southeast France.

The inclusion criteria were not limited by the breast cancer's histological or molecular characteristics, the size of the tumor, or the status of the lymph nodes. However, the study did exclude patients who had noninvasive in situ carcinoma without invasive carcinoma, HER2 (human epidermal growth factor receptor 2) positive breast carcinoma, or who presented with distant metastases at the time of surgery. HER2-positive breast cancer cases were excluded because these patients typically receive trastuzumab-based targeted therapies, which dramatically improved their prognosis following its widespread adoption for nonmetastatic breast cancer around 2005. In contrast, chemotherapy protocols for HER2-negative cases remained consistent during the treatment period of the patients included in this study, ensuring uniformity in therapeutic strategies and outcomes across the cohort. The research concentrated on the 5-year survival rates, selecting only those who had at least 5 years of follow-up and whose vital status information was available.

The database was constructed using ConSore, a data-mining application developed by UNICANCER [27]. The ConSore platform extracts data from the electronic health records of the Léon Bérard Cancer Center, integrating patient demographics, clinical variables, and treatment details. To ensure accuracy, each record was also subject to a manual verification process. Data compiled included demographic details and clinical features of patients at diagnosis, alongside comprehensive biological and disease-specific information, and the treatments administered.

We included the following characteristics for patients diagnosed with early-stage breast cancer: age; Eastern Cooperative Oncology Group performance status; BMI; comorbidities such as diabetes, heart failure, coronary artery disease, chronic obstructive pulmonary disease, and cognitive impairments; history of hospitalizations; and polypharmacy. We also gathered biological indicators at the time of diagnosis, which included hemoglobin levels, lymphocyte counts, and creatinine clearance. We extracted data on disease attributes including histological subtype, hormone receptor status, HER2 status,

Scarff-Bloom-Richardson (SBR) grade, tumor count, size of the largest tumor, and the extent of lymph node involvement as per the Tumor," "Nodes," "Metastases (TNM) classification [28]. The statuses of estrogen receptors (ERs), PRs, and HER2 were determined from the histopathological analysis of pretreatment biopsies. Hormone receptor negativity was classified when fewer than 10% of cells were stained for ER and PR. HER2 negativity was assigned when immunohistochemistry staining was below 1+. For tumors scoring 2+, further in situ hybridization tests were conducted to assess HER2 amplification [29]. Treatment data collected encompassed the type of surgery performed, lymph node dissection, and adjuvant treatments including radiotherapy, chemotherapy, and endocrine therapy.

Outcome, Predictors, and Predictive Power

Outcome was overall survival in 5 years. Due to the high percentage of missing values for cause of death, cancer-specific survival was not considered. Nine predictors were selected: age, tumor size (mm), tumor grade (defined as either SBR low: 1 - 2; or high: 3), number of affected ganglions, hormone-receptor status (positive if either estrogen or PRs were immunohistochemically present in ≥10% of tumor cells; otherwise, patients were classified as triple negative), serum hemoglobin (g/dL) and lymphocyte count (G/L), BMI, and the presence of comorbidities.

The initial database, built using ConSore, compiled a range of clinical, biological, and disease-specific data, along with information on administered treatments. We aimed for a predictors representing a mixture of features typically tested before patients undergo treatment plans. Thus, we excluded features regarding treatments as (1) we wanted to gauge prediction accuracies based only on the initial testing of the patient, and (2) the efficacy of treatment strategies was also an outcome of interest in the study. We further excluded features with significant number of missing values so as to limit the loss of usable data. Creatinine was excluded due to its high correlation with patient's age and potential kidney disorders that are not uncommon in the study's demographic. The feature was found to correlate with negative patient outcome, but this was independent of cancer and introduced a bias. Following these steps, 9 predictors were isolated, a list comprising both continuous and categorical variables, as well as an acceptable mixture of relevant biological and clinical features. Random Forest Classification (RFC) and Support Vector Classifier (SVC) were used to evaluate the predictive power of the selected features. We used 5-fold cross-validation to mitigate overfitting and ensure the validity of our results.

Model Development and Validation

Patients in the initial cohort with missing values for any of the 9 predictors were cut from the study. The remaining patients comprised the model development cohort. This was divided into reference and model data.

Reference Data and Digital Twins

The reference data inclusion criteria were positive outcome for survival in 5 years and remission without relapse by the last follow-up. The purpose of this group was to calibrate the our patented algorithm, generating digital twins for future test subjects. Digital twins refer to synthetic patient data derived from the reference group specifically similar in profile to a new test subject. The model uses these synthetic profiles to recognize complex variations within the test profile. Thus, digital twins are generated and used in the model to provide recommendations on a new patient but do not themselves constitute the result that a physician would need to interpret.

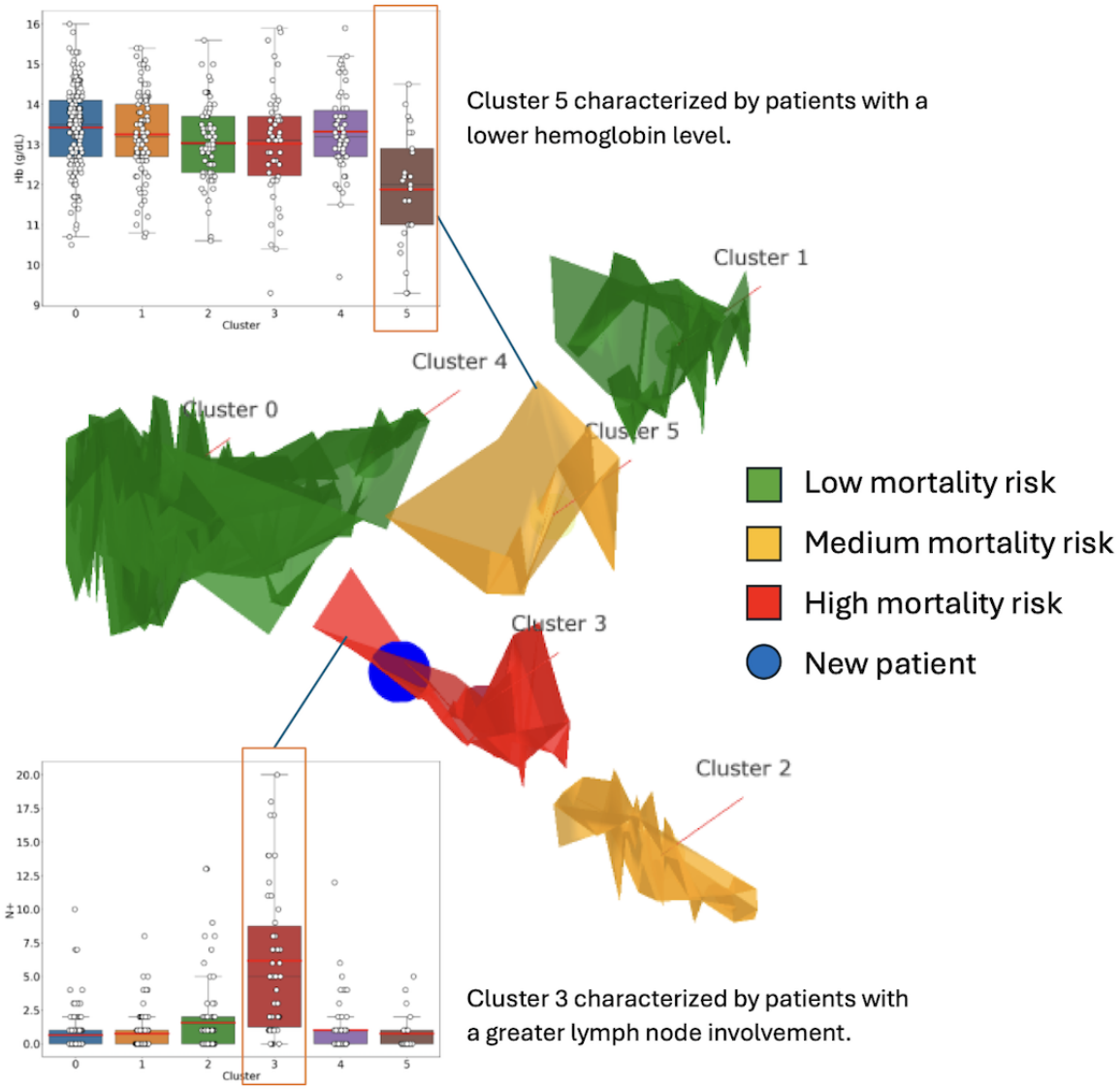
Model Data

The model data, distinct from the reference data to prevent data leaking, are the population that is run through the precalibrated model and scored against the reference group. The data are thus transformed from raw patient data to a numerical and standardized representation of their deviation from the reference group (their digital twins). The purpose of these transformed data is to populate the model with a range of patient profiles that will serve for future prognostic analysis.

PaCMAP, Mean-Shift Clustering, and Manifold Visualization

The transformed model data underwent dimensionality reduction using PaCMAP (Pairwise Controlled Manifold Approximation) [30] to generate 3D data referred to as a manifold, permitting easy visualization. The data were then stratified using mean-shift clustering [31], a nonparametric, density-based clustering algorithm that can be used to identify clusters in a dataset (Figure 1). Each cluster represents a local group of similar patients in the 3D space. Clusters represent typical patient profiles in the overall population. The advantage of clustering is that it captures the variability of subjects of a subgroup for easy analysis. A better understanding of the cluster and its variability allows clinicians to assess whether a new test subject aligns well with the cluster and to identify potential differences. When considering a new patient, estimates of prognosis and expected benefits of adjuvant treatment are ascertained by the examination of cluster-specific treatment outcomes pertinent to the patient's clinical profile.

Figure 1. Graphical representation of the 6 clusters of patients in the 3D manifold space. Patients in the reduced 3D space, or manifold, were grouped into clusters by their spatial distribution and profile similarity. Clusters were then colored based on the overall mortality rate of included patients. A newly tested patient is localized on the manifold and represented by a blue sphere.



Prediction of Chemotherapy Benefit

To estimate the benefit of chemotherapy, the position of a new patient is identified within the 3D manifold. Using the K-nearest neighbors algorithm, the 15 closest chemotherapy-treated patients and the 15 nearest non–chemotherapy-treated patients are pinpointed. Kaplan-Meier (KM) survival curves were plotted for each of these patient groups, providing a visual estimation of chemotherapy benefit for a clinical profile.

Validation of Treatment Benefit Predictions With Kullback-Leibler Divergence

To validate that the distributions of the 2 treatment subgroups are comparable, we used the Probability Density Function, which describes the spread of the data points in the 3D space. To measure the difference between these distributions, we applied the symmetrized Kullback-Leibler (KL) divergence, a statistical method that quantifies how much one distribution differs from another. To assess whether the observed difference was meaningful or just due to random chance, we conducted a permutation analysis. This technique works by randomly shuffling the data multiple times to create many new random comparisons; comparing the real result with the random results allows us to determine whether the observed difference between the distributions was statistically significant. If distributions of 2 different treatment groups were found to be similar, they could be compared to provide a prediction of treatment benefit.

Model Stability Validation

The original model data were split into 2 groups: 70% (327) of every cluster was pooled into the training group, and the remaining 30% (139) was pooled into the test group. A new manifold learning process was applied to the training group, and the test group was then projected onto this newly generated manifold. Patients in the test and model groups from the same cluster of origin were compared to evaluate whether data points would exhibit similar distributions (appear in proximity to each

other) in the new manifold space across 10 different manifold initializations.

Statistical Analysis

Kullback-Leibler Divergence and Permutation Test

The symmetrical KL divergence was used to measure the difference between 2 probability distributions. A permutation test was subsequently conducted to assess the significance of the observed KL divergence. This involved calculating the KL divergence for a large number of permutations of the combined datasets and comparing these values with the original KL divergence. The *P* value is calculated as the proportion of permutations where the KL divergence is as extreme as, or more extreme than, the original KL divergence calculated between the actual groups, thus providing a measure of how likely it was to observe a divergence as extreme as the original, under the null hypothesis of no difference between the distributions. Mathematically, this *P* value is the ratio of the number of permuted KL divergences that are equal to the original KL divergence or greater to the total number of permutations. A low *P*-value suggests that the observed difference in distributions is unlikely to have occurred by chance, thus indicating a significant divergence between the 2 groups.

Survival Analysis using the KM Estimator and Log-Rank Test

The KM estimator was used to generate survival curves for different treatment subgroups. The log-rank test, a nonparametric test, was applied to compare the survival distributions and a *P* value was calculated to determine the statistical significance of the differences observed between the groups. A low *P* value suggests that the observed survival curves are significantly different. The statistical package used for the analysis is Lifelines 0.30.0 (Lifelines Developers) [32].

Ethical Considerations

This retrospective study involving human subjects was reviewed and approved by the French data protection authority, the Commission Nationale de l'Informatique et des Libertés, under authorization number 9191415, dated October 10, 2019. According to institutional and national guidelines, no additional approval from a research ethics board was required, as the data used were previously collected for clinical purposes. No new informed consent was required for this study. The analysis was conducted using data for which participants had provided general consent at the time of data collection. All data were pseudonymized prior to analysis to protect patient confidentiality. No identifiable personal information was retained in the research dataset. No compensation was provided to participants.

Results

Cohort Characteristics

A total of 1229 patients comprised the initial cohort. Of these, 793 (65%) remained after entries with missing values were removed (Figure 2). Eliminating the risk of introducing a bias, the initial cohorts' demographic and clinical characteristics were found to be strictly similar to that of the final cohort and are summarized in Tables 1 and 2.

Figure 2. Flowchart of data construction.

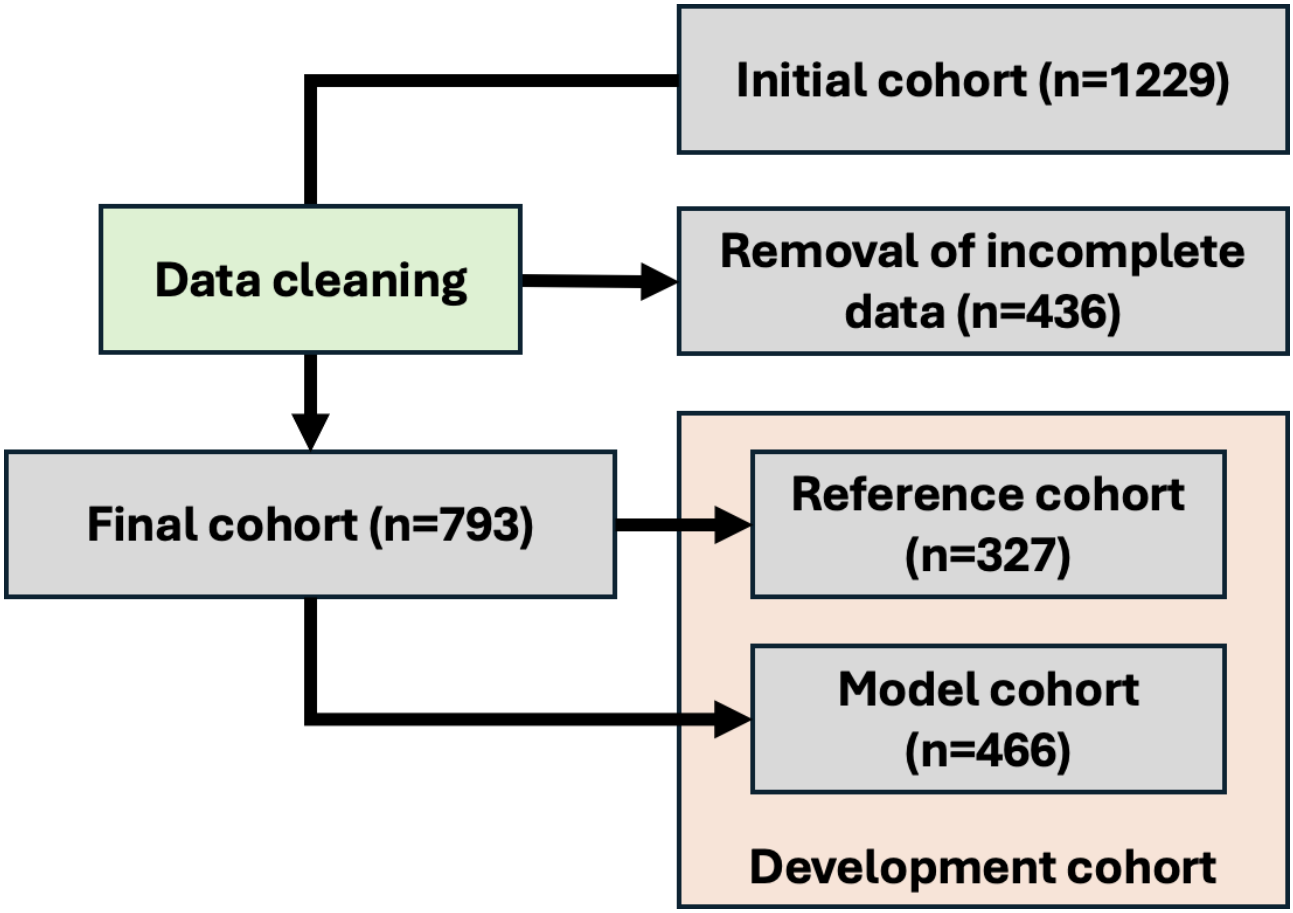


Table . Patient characteristics of the initial cohort (N=1229).

Characteristics	Participants
Age at diagnosis (years), n (%)	
70 - 74	580 (47)
75 - 79	331 (27)
80 - 84	204 (17)
85 - 89	93 (8)
>90	20 (2)
Performance status, n (%)	
0	339 (28)
1	322 (26)
2	48 (4)
3-4	23 (2)
Missing data	497 (40)
BMI, n (%)	
<18.5	32 (3)
18.5 - 25	446 (36)
25 - 30	409 (33)
>30	266 (22)
Missing data	76 (6)
Comorbidities, n (%)	
Creatinine clearance <40 mL/minute	57 (5)
Heart failure	105 (9)
Coronary artery disease	123 (10)
Chronic obstructive pulmonary disease	36 (3)
Diabetes	174 (14)

Table . Cancer characteristics of the initial cohort (N=1229).

Tumor size					
Status	T1	T2	T3	T4	Missing data
Participants, n (%)	567 (46)	286 (23)	36 (3)	250 (20)	90 (7)
Lymph nodes					
Status	N0	N1	N2	N3	Missing data
Participants, n (%)	614 (50)	243 (20)	55 (4)	55 (4)	262 (21)
Grade SBR[a]					
Status	I	II		III	Missing data
Participants, n (%)	188 (15)	648 (53)		281 (23)	112 (9)
Estrogen receptor					
Status	Positive		Negative		Missing data
Participants, n (%)	978 (80)		145 (12)		106 (9)
Progesterone receptor					
Status	Positive		Negative		Missing data
Participants, n (%)	838 (68)		285 (23)		106 (9)

[a]SBR: Scarff-Bloom-Richardson.

Patient demographics and characteristics were evaluated on the date of breast cancer diagnosis (Table 1). Median age was 75 years (range: 70 - 100 years), with 317/1229 (26%) patients aged 80 years or older. Performance status was generally good, as most are categorized as 0 or 1. The main comorbidities were diabetes (174/1229 patients, or 14%), followed by coronary artery disease (123/1229 patents, 10%) and cardiac insufficiency (105/1229 patients, 9%).

The majority presented early-stage tumors (T1 in 567/1229 patients, with a prevalence of 46%), and lymph node involvement was mostly absent (N0 in 614/1229 patients, or 50%). The tumors were typically SBR grade II and 80% (978/1229 patients) were ER-positive. Progesterone receptor positivity was also high at 68% (838/1229 patients). Twelve percent of patients (149/1229) were reported to have received chemotherapy (Table 2).

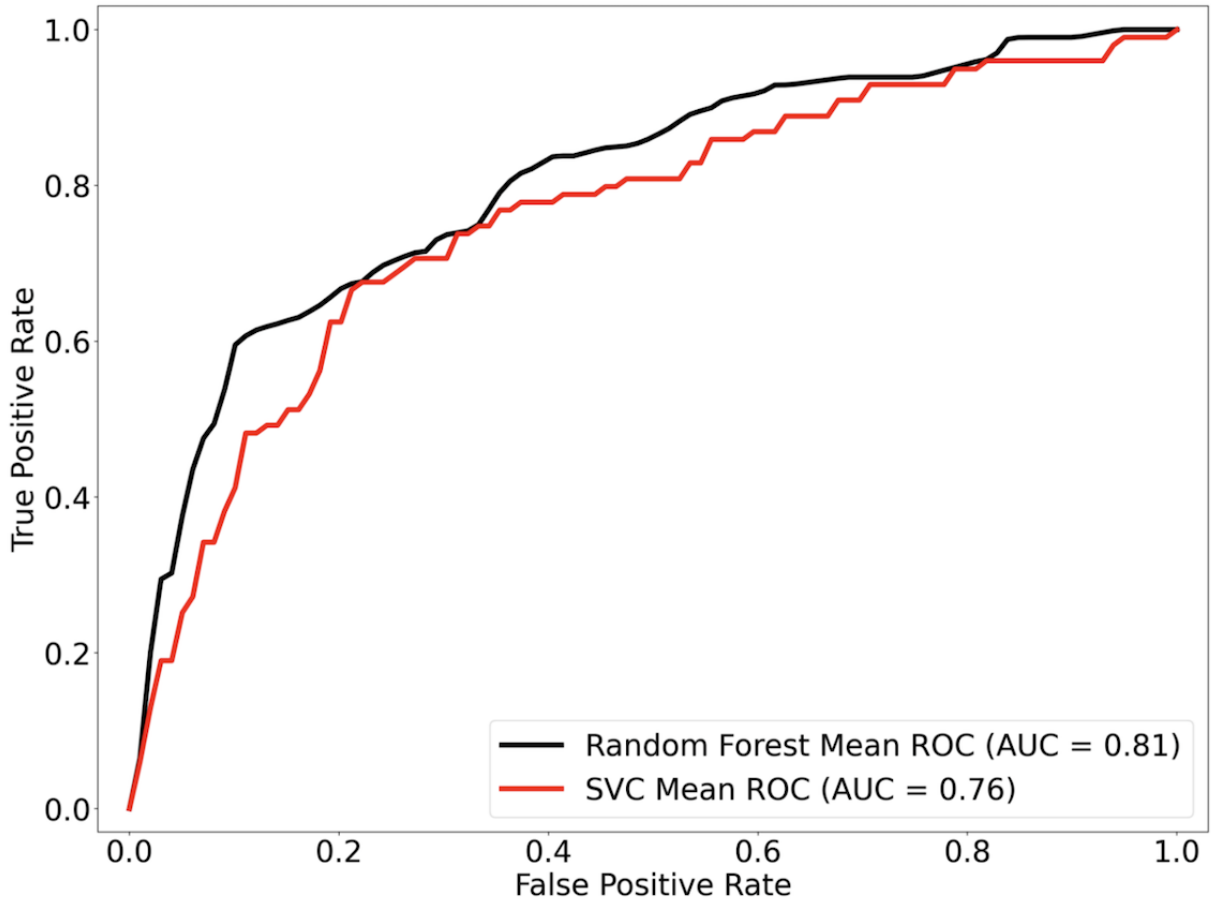
Development Cohort

The final cohort was divided into "reference" and "model" cohorts for model development (Figure 2). A total of 327 patients, that is, 50% of patients meeting the criteria for manifold-estimated derivation training were randomly selected. The purpose of this training group was to calibrate the manifold-estimated derivation–scoring algorithm. The model data comprised all remaining patients (466, 59% of the model development cohort).

Features Performance and Area Under the Curve Scores

In Figure 3, we ascertained the predictive efficacy of the selected variables using RFC and SVC. Analyzing the receiver operating characteristic curves, both models demonstrated commendable predictive capabilities. RFC yielded a mean area under the curve (AUC) of 0.81 (SD 0.06) and a mean accuracy of 0.82 (SD 0.02), while SVC followed closely with a mean AUC of 0.76 (SD 0.05) and a mean accuracy of 0.78 (SD 0.01). The overlapping SDs of these scores suggest that the differences in their performance are not statistically significant.

Figure 3. Receiver operating characteristic curves for 5-year mortality predictive models. The predictive efficacy of the selected features was ascertained using Random Forest Classification and Support Vector Classifier. Results are presented as the mean of ROC and AUC values derived from 5-fold cross-validation. AUC: area under the curve; ROC: receiver operating characteristic; SVC: Support Vector Classifier.



The overall relative importance of variables for the prediction of the 5-year outcome was also determined by RFC (Table 3). Age, tumor size, and hemoglobin were the top predictors, closely followed by lymphocyte count and BMI. Curiously, the cancer grade, axillary lymph nodes involvement, and the presence of comorbidities ranked low in overall importance. This indicates that although typically taken as important factors from a clinical perspective, comorbidities and cancer grade alone are not the best prognostic features in a patient; rather, a patient's overall

biological profile may be more valuable, underscoring the usefulness of manifold learning as a prognostic tool.

Table . Overall importance of predictors according to Random Forest Classification.

Variable	Importance (%)
Age	18.33
Tumor size	17.26
Hemoglobin (g/dL)	16.41
Lymphocytes (g/L)	14.84
BMI	13.06
Lymph nodes involvement	10.39
SBR[a] grade	4.06
ER[b] status	2.88
Comorbidities	2.78

[a]SBR: Scarff-Bloom-Richardson.

[b]ER: estrogen receptor.

Model Stability

Patients in the test and model groups from the same cluster of origin were compared to evaluate whether data points would exhibit similar distributions (appear in proximity to each other) in the new manifold space across 10 different manifold initializations. The distributions of the test group (n=140) consistently matched closely with those of the model group (n=326), with all *P* values being above the threshold of .05 indicating a lack of significant variation between groups (Figure S1 in Multimedia Appendix 1).

Prognostic Ability

The primary objective of our study is to evaluate the prognostic ability of the manifold learning model, as measured by the 5-year survival rate of our population. The 3D clusters in Figure 1 illuminated the landscape of our dataset, representing local groups of patients characterized by distinct clinical and prognostic profiles. Clusters are colored based on the overall mortality rate of included patients: Groups 0, 1, and 4 in green have the best prognosis with a 5-year survival rate of more than 80% while group 3 has the worst prognosis with a 5-year mortality rate of at least 35%.

Table 4 further elucidates the variability in values across the patient clusters, especially in BMI, tumor size (in mm), and median age, underscoring the diversity in our cohort.

Table . Characteristics of the 6 clusters defined by manifold learning.

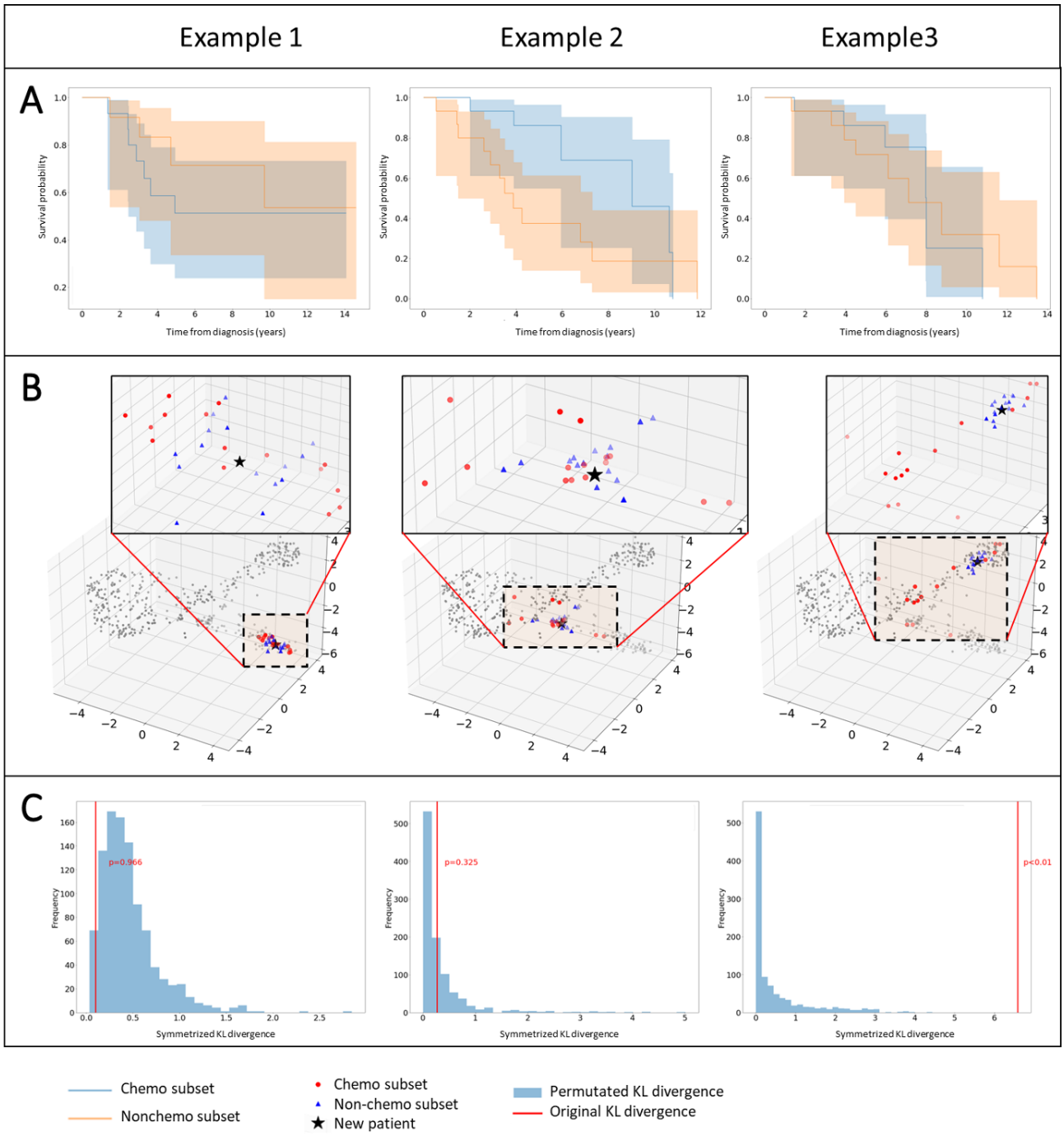
Feature	Cluster					
	0	1	2	3	4	5
Hemoglobin (g/dL)	13.4	13.3	13	13	13.3	11.9
BMI	25	28.4	24.9	28.9	25.6	23
Lymph nodes involved	0.6	0.8	1.6	6.2	1	0.8
Tumor size (mm)	19.6	19.3	26.1	65.8	23.1	30.6
Age (years)	75.8	76.3	77.8	77.4	79.2	80.5
Lymphocytes (g/L)	1.8	2.1	2.1	1.6	3.4	1.6
Comorbidities	0	1	0.4	0.3	0	0.9
Estrogen receptor status	1	1	0	0.8	1	1
SBR[a] (high/low)	0	0.1	0.7	0.5	0.8	0.6

[a]SBR: Scarff-Bloom-Richardson.

Predictive Ability

Next, we attempted to ascertain the individual benefit of performing adjuvant chemotherapy, demonstrated in Figure 4 with 3 examples.

Figure 4. Three case examples assessing the individual benefit of adjuvant chemotherapy. (A) The closest chemotherapy-treated and non–chemotherapy-treated patients to a new patient are identified in the 3D manifold and their survival curves are compared to show the treatment's potential benefit or lack thereof. (B) The new patient's position in the 3D manifold (black star), with the 15 closest patients of each treatment groups are shown, displaying varying distributions of treatment subgroups. (C) To quantify distances between the subgroups, the real calculated KL divergence between the treatment groups' distributions (red line) was compared with that of permutated data (blue histograms) to verify whether observed divergences between treatment subgroups are significant or not. KL: Kullback-Leibler.



When a target patient is localized in the 3D manifold, the closest patient profiles are identified. This is done for 2 treatment groups based on whether the patients received chemotherapy (chemo and nonchemo groups), permitting the visualization of KM survival curves that would show the treatment's potential benefit or lack thereof (Figure 4A).

Figure 4B shows the target patient's position in the 3D manifold (black star), with the 15 closest patients of each treatment groups also marked. In examples 1 and 2, the 2 treatment groups are found to be well "mixed" in the local vicinity of the target, indicating that the target profile is well represented by similar chemotherapy-treated and non–chemotherapy-treated patients. To quantify distances between the subgroups, we used

permutation analysis (Figure 4C). The real calculated KL divergences between the treatment groups' distributions (red line) for examples 1 and 2 fall well within the range of what could be expected by chance (blue histograms) ($P>.1$), indicating that the observed divergences are not significant.

Example 3 showcases a situation where patients from the 2 treatment groups are not well mixed in the local vicinity of the target patient. In this case, the real KL divergence is far right of the histogram ($P<.01$), suggesting a significant difference between the distributions. Thus, the KM survival curves and any conclusion drawn from them must be taken with consideration of the heterogeneity in the profiles of the treatment groups being compared.

Discussion

Principal Findings

From an initial cohort of 1229 patients, we used 793 (65%) to develop a model that clustered patients by their clinical and biological features. These clusters represent a potential prognostic tool for physicians, attributing a risk of mortality in 5 years to patients with consideration to multivariate profiles. The model is further able to indicate the potential benefit or lack thereof of chemotherapy treatment in older adult patients. We found that the predictors used in our model gave a good overall result of 0.81 and 0.76 AUCs with RFC and SVC, respectively.

In summation, our multifaceted approach, blending manifold learning with classical machine learning paradigms and intuitive data visualizations, has unveiled profound insights into the prognosis determinants of early-stage breast cancer in older adults. These revelations bring a more nuanced understanding of the disease and hold promise for tailoring patient-specific therapeutic strategies. Our study's utilization of manifold learning and advanced machine learning algorithms represents a significant contribution to oncology. The accuracy of 81% in differentiating patient subgroups through manifold learning is impressive, showcasing an advancement beyond traditional linear models [33]. This approach is in line with recent trends in personalized medicine [34,35], which discuss the potential of machine learning in cancer prognosis. The high AUC values achieved by RFC and SVC reflect the importance of our combined predictors in medical diagnostics, aligning with the findings of recent studies on the application of machine learning in cancer detection [36,37]. The application of data visualization techniques such as heatmaps and 3D scatterplots in elucidating complex clinical relationships is noteworthy. This approach is supported by advancements in data visualization in medical research, as seen in the study by Borkin et al [38] on how data visualization supports medical decision-making [39,40].

Limitations

The present results should be interpreted in the light of some limitations. First, the monocentric nature of the research may impact the representativeness of the cohort, potentially affecting the generalizability of our findings. Second, the exclusion of specific patient characteristics, such as the ONCODAGE score [41], from our datasets may have limited the comprehensiveness of our prognostic tools. Third, the retrospective design of the study constrains our ability to establish causality between clinical characteristics and patient outcomes. A fourth limitation concerns the fact that patients may present with or have a history of multiple comorbidities. We chose to group together patients with any number of comorbidities for reasons related to (1) the reduction of the sample size for each category of comorbidity, and (2) the potential skewing of patient distribution in the 3D manifold due to multiple related qualitative variables. PaCMAP is susceptible to "overseparate" the population if provided with too many binary features. These reasons in mind, we nonetheless acknowledge that omitting the consideration of multiple comorbidities is a limitation of the study. Other notable limitations include the absence of cancer-specific or treatment-specific survival metrics, a lack of detailed analysis on specific comorbidities, and the need for more data to enhance the less populated clusters. Furthermore, the external validation of our model remains pending, which is crucial for assessing its generalizability.

Future Prospects

Looking forward, the promising application of manifold learning in oncology, as demonstrated in our study, aligns with the burgeoning field of personalized medicine. The integration of machine learning in personalized cancer therapy, as discussed by Danishuddin et al [42], supports the potential of such approaches. The development of advanced AI-driven prognostic tools, particularly for older adult patients who are often underrepresented in clinical trials, could revolutionize treatment guidelines and care approaches. The rapid advancement of machine learning techniques poses a challenge in ensuring the longevity and relevance of models, necessitating continuous updates. This is echoed in the broader context of AI in health care, as discussed in Topol's [43] comprehensive review of AI in medicine. Concerns about the adoption of AI tools due to accuracy, explainability, and ethical considerations are also prevalent, as reflected in the exploration of implementing AI in clinical practice by Char et al [44]. Our findings may open up avenues for the personalized treatment specifically catered to neglected populations in oncology, starting with geriatric patients with breast cancer. We expect our software to provide rapid guidance to physicians in the process of charting treatment plans for their patients, going beyond simple monovariate statistics and instead considering patients' combined clinical and biological profiles.

Conclusions

Our study aimed to further the management of early breast cancer in older adult patients by integrating cutting-edge AI techniques. We proposed a technique that uses patient data to create a visualizable 3D map of pathology profiles that allow rapid prognostic estimations for new patients. These prognostic predictions include the potential benefits of treatment strategies such as chemotherapy, aiding clinical decision-making. It reflects the ongoing evolution in oncology, emphasizing the importance of tailored treatment strategies and highlighting both the potential and the challenges of AI applications in health care. This study also prompts considerations for future research

directions and ethical implications in the rapidly evolving field of AI in medicine.

Data Availability

The data that support the findings of this study are not publicly available due to privacy and confidentiality agreements. However, pseudonymized data may be made available upon reasonable request to the corresponding author and following approval by the French Commission Nationale de l'Informatique et des Libertés under the data-sharing agreements of the Léon Bérard Cancer Center.

Authors' Contributions

FR and PH contributed to conceptualization, methodology, project administration, validation, writing (review and editing). MA contributed to methodology, investigation, data curation, validation, and writing (original draft and editing). AA contributed to conceptualization, supervision, writing (review and editing), and resources.

Conflicts of Interest

The model described in this study was developed by GeodAIsics, who are listed among the authors. PH reports : grants; personal fees; and nonfinancial support from PFIZER, LILLY, DAICHII, and ASTRAZENECA grants, and nonfinancial support from NOVARTIS and ROCHE—personal fees and nonfinancial support from SEAGEN, GILEAD, and MSD—cofounder and chief medical officer of GEODAISICS. AA is the founder and CEO of GeodAIsics. The remaining authors have no conflicts of interest.

Multimedia Appendix 1

Stability analysis of manifold learning applied to clustered data. The original cohort data were divided into 2 groups; 70% of every cluster was pooled into the model group, and the remaining 30% was pooled into the test group. A fresh manifold learning process was applied to the model group, and the test group was then projected onto the newly generated manifold. Patients in the test and model groups from the same cluster of origin were compared to evaluate whether they would exhibit similar distributions (appear in proximity to each other) in the new manifold space. (A) Examples of permutation analysis of clusters 0 and 1. The permutation test determined whether the observed KL (red line) divergence was significantly different from what can be expected from random shuffling of the 2 groups (blue histograms). (B) Table summarizing the median *P* values of the stability tests across 10 different manifold initializations. All *P* values above .05 indicated a lack of significant variation between groups.
[PNG File, 214 KB - cancer_v11i1e64000_app1.png]

References

1. Howlader N, Noone AM, Krapcho M, et al. SEER cancer statistics review, 1975-2017. In: Cronin KA, editor. Based on November 2019 SEER data submission, posted to the SEER web site.: National Cancer Institute URL: https://seer.cancer.gov/archive/csr/1975_2017/ [accessed 2025-05-04]
2. American Cancer Society. Breast cancer facts & figures 2019-2020. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf [accessed 2025-05-04]
3. Ferlay J, Ervik M, Lam F, et al. Global Cancer Observatory: Cancer Today.: Lyon, France: International Agency for Research on Cancer; 2020. URL: https://gco.iarc.fr/today [accessed 2025-05-20]
4. Tesarova P. Specific aspects of breast cancer therapy of elderly women. Biomed Res Int 2016;2016:1381695. [doi: 10.1155/2016/1381695] [Medline: 27807536]
5. Glaser R, Marinopoulos S, Dimitrakakis C. Breast cancer treatment in women over the age of 80: a tailored approach. Maturitas 2018 Apr;110:29-32. [doi: 10.1016/j.maturitas.2018.01.014] [Medline: 29563032]
6. Clarke M, Coates AS, et al. Adjuvant chemotherapy in oestrogen-receptor-poor breast cancer: patient-level meta-analysis of randomised trials. Lancet 2008 Jan 5;371(9606):29-40. [doi: 10.1016/S0140-6736(08)60069-0] [Medline: 18177773]
7. Peto R, Davies C, Early Breast Cancer Trialists' Collaborative Group (EBCTCG), et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Lancet 2012 Feb 4;379(9814):432-444. [doi: 10.1016/S0140-6736(11)61625-5] [Medline: 22152853]
8. Buchanan BG, Shortlife EH, editors. An expert system for oncology protocol management. In: Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project: Reading:Addison-Wesley; 1984:876-881.
9. Lieber J, Bresson B. Case-based reasoning for breast cancer treatment decision helping. In: Blanzieri E, Portinale L, editors. Advances in Case-Based Reasoning EWCBR 2000 Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence): Springer; 2000, Vol. 1898:173-185.
10. Hawley ST, Li Y, An LC, et al. Improving breast cancer surgical treatment decision making: the iCanDecide randomized clinical trial. J Clin Oncol 2018 Mar 1;36(7):659-666. [doi: 10.1200/JCO.2017.74.8442] [Medline: 29364772]

11. Wyatt KD, Jenkins SM, Plevak MF, Venegas Pont MR, Pruthi S. A personalized, web-based breast cancer decision making application: a pre-post survey. BMC Med Inform Decis Mak 2019 Oct 21;19(1):196. [doi: 10.1186/s12911-019-0924-7] [Medline: 31638964]
12. Xu F, Sepúlveda MJ, Jiang Z, et al. Artificial intelligence treatment decision support for complex breast cancer among oncologists with varying expertise. JCO Clin Cancer Inform 2019 Aug;3:1-15. [doi: 10.1200/CCI.18.00159] [Medline: 31419181]
13. Mazo C, Kearns C, Mooney C, Gallagher WM. Clinical decision support systems in breast cancer: a systematic review. Cancers (Basel) 2020 Feb 6;12(2):369. [doi: 10.3390/cancers12020369] [Medline: 32041094]
14. Wishart GC, Azzato EM, Greenberg DC, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res 2010;12(1):R1. [doi: 10.1186/bcr2464] [Medline: 20053270]
15. Jiang X, Wells A, Brufsky A, Neapolitan R. A clinical decision support system learned from data to personalize treatment recommendations towards preventing breast cancer metastasis. PLoS ONE 2019;14(3):e0213292. [doi: 10.1371/journal.pone.0213292] [Medline: 30849111]
16. Hughes KS, Schnaper LA, Berry D, et al. Lumpectomy plus tamoxifen with or without irradiation in women 70 years of age or older with early breast cancer. N Engl J Med 2004 Sep 2;351(10):971-977. [doi: 10.1056/NEJMoa040587] [Medline: 15342805]
17. Clough-Gorr KM, Thwin SS, Stuck AE, Silliman RA. Examining five- and ten-year survival in older women with breast cancer using cancer-specific geriatric assessment. Eur J Cancer 2012 Apr;48(6):805-812. [doi: 10.1016/j.ejca.2011.06.016] [Medline: 21741826]
18. Stotter A, Reed MW, Gray LJ, Moore N, Robinson TG. Comprehensive Geriatric Assessment and predicted 3-year survival in treatment planning for frail patients with early breast cancer. Br J Surg 2015 Apr;102(5):525-533. [doi: 10.1002/bjs.9755] [Medline: 25708660]
19. Alaa AM, Gurdasani D, Harris AL, Rashbass J, van der Schaar M. Machine learning to guide the use of adjuvant therapies for breast cancer. Nat Mach Intell 2012;3(8):716-726. [doi: 10.1038/s42256-021-00353-8]
20. Shachar SS, Muss HB. Internet tools to enhance breast cancer care. NPJ Breast Cancer 2016;2:16011. [doi: 10.1038/npjbcancer.2016.11] [Medline: 28721377]
21. van Maaren MC, van Steenbeek CD, Pharoah PDP, et al. Validation of the online prediction tool PREDICT v. 2.0 in the Dutch breast cancer population. Eur J Cancer 2017 Nov;86:364-372. [doi: 10.1016/j.ejca.2017.09.031] [Medline: 29100191]
22. Van der Schaar Lab. Adjutorium: a machine learning framework for personalized survival prediction. URL: https://vanderschaar-lab.com/adjutorium/ [accessed 2025-05-06]
23. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J Clin Oncol 2001 Feb 15;19(4):980-991. [doi: 10.1200/JCO.2001.19.4.980] [Medline: 11181660]
24. de Glas NA, van de Water W, Engelhardt EG, et al. Validity of Adjuvant! Online program in older patients with breast cancer: a population-based study. Lancet Oncol 2014 Jun;15(7):722-729. [doi: 10.1016/S1470-2045(14)70200-1] [Medline: 24836274]
25. Engelhardt EG, Garvelink MM, de Haes JHCJM, et al. Predicting and communicating the risk of recurrence and death in women with early-stage breast cancer: a systematic review of risk prediction models. J Clin Oncol 2014 Jan 20;32(3):238-250. [doi: 10.1200/JCO.2013.50.3417] [Medline: 24344212]
26. van der Plas-Krijgsman WG, Giardiello D, Putter H, et al. Development and validation of the PORTRET tool to predict recurrence, overall survival, and other-cause mortality in older patients with breast cancer in the Netherlands: a population-based study. Lancet Healthy Longev 2021 Nov;2(11):e704-e711. [doi: 10.1016/S2666-7568(21)00229-4] [Medline: 36098027]
27. Heudel P, Livartowski A, Arveux P, Willm E, Jamain C. The ConSoRe project supports the implementation of big data in oncology. Bull Cancer 2016 Nov;103(11):949-950. [doi: 10.1016/j.bulcan.2016.10.001] [Medline: 27816168]
28. Amin MB, Greene FL, Edge SB, et al. The Eighth Edition AJCC Cancer Staging Manual: continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. CA Cancer J Clin 2017 Mar;67(2):93-99. [doi: 10.3322/caac.21388] [Medline: 28094848]
29. Wolff AC, Hammond MEH, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. J Clin Oncol 2013 Nov 1;31(31):3997-4013. [doi: 10.1200/JCO.2013.50.9984] [Medline: 24101045]
30. Wang Y. GitHub repository. PaCMAP: Pairwise Controlled Manifold Approximation. URL: https://github.com/YingfanWang/PaCMAP [accessed 2025-05-06]
31. Scikit-learn developers. "sklearn.clusterMeanShift," scikit-learn documentation. URL: https://scikit-learn.org/dev/modules/generated/sklearn.cluster.MeanShift.html [accessed 2025-05-06]
32. Davidson-Pilon C. Lifelines: survival analysis in Python," Read the Docs. URL: https://antiphishing.vadesecure.com/v4?[illegible]readthedocs.io%2Fen%2Flatest%2F [accessed 2025-05-06]

33. Bhavnani SK, Zhang W, Visweswaran S, Raji M, Kuo YF. A framework for modeling and interpreting patient subgroups applied to hospital readmission: visual analytical approach. JMIR Med Inform 2022 Dec 7;10(12):e37239. [doi: 10.2196/37239] [Medline: 35537203]
34. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13(13):8-17. [doi: 10.1016/j.csbj.2014.11.005] [Medline: 25750696]
35. Ozer ME, Sarica PO, Arga KY. New machine learning applications to accelerate personalized medicine in breast cancer: rise of the support vector machines. OMICS 2020 May;24(5):241-246. [doi: 10.1089/omi.2020.0001] [Medline: 32228365]
36. Shafiq A, Çolak AB, Sindhu TN, Lone SA, Abushal TA. Modeling and survival exploration of breast carcinoma: A statistical, maximum likelihood estimation, and artificial neural network perspective. Artificial Intelligence in the Life Sciences 2023 Dec;4:100082. [doi: 10.1016/j.ailsci.2023.100082]
37. Wu R, Luo J, Wan H, et al. Evaluation of machine learning algorithms for the prognosis of breast cancer from the Surveillance, Epidemiology, and End Results database. PLoS ONE 2023;18(1):e0280340. [doi: 10.1371/journal.pone.0280340] [Medline: 36701415]
38. Borkin MA, Vo AA, Bylinskii Z, et al. What makes a visualization memorable? IEEE Trans Vis Comput Graph 2013 Dec;19(12):2306-2315. [doi: 10.1109/TVCG.2013.234] [Medline: 24051797]
39. Palmas G, Bachynskyi M, Oulasvirta A, Seidel HP, Weinkauf T. MovExp: a versatile visualization tool for human-computer interaction studies with 3D performance and biomechanical data. IEEE Trans Vis Comput Graph 2014 Dec;20(12):2359-2368. [doi: 10.1109/TVCG.2014.2346311] [Medline: 26356950]
40. Park S, Bekemeier B, Flaxman A, Schultz M. Impact of data visualization on decision-making and its implications for public health practice: a systematic literature review. Inform Health Soc Care 2022 Apr 3;47(2):175-193. [doi: 10.1080/17538157.2021.1982949] [Medline: 34582297]
41. Soubeyran P, Bellera C, Goyard J, et al. Screening for vulnerability in older cancer patients: the ONCODAGE Prospective Multicenter Cohort Study. PLoS One 2014;9(12):e115060. [doi: 10.1371/journal.pone.0115060] [Medline: 25503576]
42. Danishuddin A, Khan S, Kim JJ. From cancer big data to treatment: artificial intelligence in cancer research. J Gene Med 2024 Jan;26(1):e3629. [doi: 10.1002/jgm.3629] [Medline: 37940369]
43. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
44. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. N Engl J Med 2018 Mar 15;378(11):981-983. [doi: 10.1056/NEJMp1714229] [Medline: 29539284]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
COPD: chronic obstructive pulmonary disease
ER: estrogen receptor
HER2: human epidermal growth factor receptor 2
KL: Kullback-Leibler
KM: Kaplan-Meier
PaCMAP: Pairwise Controlled Manifold Approximation
PR: progesterone receptor
RFC: Random Forest Classification
SBR: Scarff-Bloom-Richardson
SVC: Support Vector Classification

Edited by N Cahill; submitted 05.07.24; peer-reviewed by JCL Chow, Z Ping; revised version received 22.01.25; accepted 24.02.25; published 23.05.25.

Please cite as:
Heudel P, Ahmed M, Renard F, Attye A
Leveraging Digital Twins for Stratification of Patients with Breast Cancer and Treatment Optimization in Geriatric Oncology: Multivariate Clustering Analysis
JMIR Cancer 2025;11:e64000
URL: https://cancer.jmir.org/2025/1/e64000
doi:10.2196/64000

A Deep Learning–Enabled Workflow to Estimate Real-World Progression-Free Survival in Patients With Metastatic Breast Cancer: Study Using Deidentified Electronic Health Records

Gowtham Varma[1], MBBS; Rohit Kumar Yenukoti[1], MBBS; Praveen Kumar M[1], MD, DM; Bandlamudi Sai Ashrit[1], MBBS; K Purushotham[1], MBBS; C Subash[1], MBBS; Sunil Kumar Ravi[1], MBBS; Verghese Kurien[1], MBBS; Avinash Aman[2], BTech; Mithun Manoharan[1], MBBS, MMST; Shashank Jaiswal[2], B.Tech; Akash Anand[2], BE; Rakesh Barve[2], BTech, PhD; Viswanathan Thiagarajan[2], B.Tech; Patrick Lenehan[3], MD, PhD; Scott A Soefje[4], BCOP, MBA, PharmD; Venky Soundararajan[3], PhD

[1]Department of Clinical Sciences, Nference, 4th Floor, Indiqube, Golf View Campus Tower-2, 22, 3rd Cross Rd, Murugeshpalya, S R Layout, Bangalore, India

[2]Department of Data Science and Engineering, Nference, Bangalore, India

[3]Department of Clinical Sciences, Nference, Cambridge, MA, United States

[4]Department of Pharmacy, Director of Pharmacy, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Praveen Kumar M, MD, DM
Department of Clinical Sciences, Nference, 4th Floor, Indiqube, Golf View Campus Tower-2, 22, 3rd Cross Rd, Murugeshpalya, S R Layout, Bangalore, India

Abstract

Background: Progression-free survival (PFS) is a crucial endpoint in cancer drug research. Clinician-confirmed cancer progression, namely real-world PFS (rwPFS) in unstructured text (ie, clinical notes), serves as a reasonable surrogate for real-world indicators in ascertaining progression endpoints. Response evaluation criteria in solid tumors (RECIST) is traditionally used in clinical trials using serial imaging evaluations but is impractical when working with real-world data. Manual abstraction of clinical progression from unstructured notes remains the gold standard. However, this process is a resource-intensive, time-consuming process. Natural language processing (NLP), a subdomain of machine learning, has shown promise in accelerating the extraction of tumor progression from real-world data in recent years.

Objectives: We aim to configure a pretrained, general-purpose health care NLP framework to transform free-text clinical notes and radiology reports into structured progression events for studying rwPFS on metastatic breast cancer (mBC) cohorts.

Methods: This study developed and validated a novel semiautomated workflow to estimate rwPFS in patients with mBC using deidentified electronic health record data from the Nference nSights platform. The developed workflow was validated in a cohort of 316 patients with hormone receptor–positive, human epidermal growth factor receptor-2 (HER-2) 2-negative mBC, who were started on palbociclib and letrozole combination therapy between January 2015 and December 2021. Ground-truth datasets were curated to evaluate the workflow's performance at both the sentence and patient levels. NLP-captured progression or a change in therapy line were considered outcome events, while death, loss to follow-up, and end of the study period were considered censoring events for rwPFS computation. Peak reduction and cumulative decline in Patient Health Questionnaire-8 (PHQ-8) scores were analyzed in the progressed and nonprogressed patient subgroups.

Results: The configured clinical NLP engine achieved a sentence-level progression capture accuracy of 98.2%. At the patient level, initial progression was captured within ±30 days with 88% accuracy. The median rwPFS for the study cohort (N=316) was 20 (95% CI 18-25) months. In a validation subset (n=100), rwPFS determined by manual curation was 25 (95% CI 15-35) months, closely aligning with the computational workflow's 22 (95% CI 15-35) months. A subanalysis revealed rwPFS estimates of 30 (95% CI 24-39) months from radiology reports and 23 (95% CI 19-28) months from clinical notes, highlighting the importance of integrating multiple note sources. External validation also demonstrated high accuracy (92.5% sentence level; 90.2% patient level). Sensitivity analysis revealed stable rwPFS estimates across varying levels of missing source data and event definitions. Peak reduction in PHQ-8 scores during the study period highlighted significant associations between patient-reported outcomes and disease progression.

Conclusions: This workflow enables rapid and reliable determination of rwPFS in patients with mBC receiving combination therapy. Further validation across more diverse external datasets and other cancer types is needed to ensure broader applicability and generalizability.

(JMIR Cancer 2025;11:e64697) doi:10.2196/64697

KEYWORDS

real-world evidence; data-driven oncology; real-world progression-free survival; metastatic breast cancer; natural language processing; NLP; survival; cancer; oncology; breast; metastatic; deep learning; machine learning; ML; workflow; report; notes; electronic health record; EHR; documentation

Introduction

Background and Significance

Real-world evidence (RWE) is increasingly accepted to augment traditional clinical trial findings to better understand the effectiveness of oncological interventions. RWE can be leveraged to improve novel therapy development programs and provide better postmarket surveillance of approved therapies [1-3].

Response evaluation criteria in solid tumors (RECIST) is broadly used to ascertain disease progression in clinical trials. However, assessing RECIST in retrospective electronic health record (EHR) data is challenging due to its strict assessment indicators [4]. RECIST considers changes in the size of individual target lesions over time and the presence or absence of new lesions to categorize disease status into complete or partial response, stable disease, or progression [5]. A similar assessment paradigm is adopted in routine clinical practice, where clinicians document the occurrence of progression through serial clinical and radiological examinations. This clinician-confirmed cancer progression in unstructured text (ie, clinical notes) has been shown to serve as a reasonable surrogate for real-world indicators in ascertaining progression endpoints. This is also more practical for real-world studies than purely RECIST-based approaches [6].

In earlier studies across different types of solid tumors, real-world progression (rwP) was captured either through manual abstraction from unstructured data or proxy measures were evaluated based solely on structured drug data [7,8]. Recent studies have also used machine learning models specifically trained to automate the capture and characterization of clinician documentation of tumor response. These specialized models have shown varying accuracies [9,10]. In the past decade, health care natural language processing (NLP) frameworks like Google's Healthcare Natural Language application programming interface (API), Amazon Comprehend Medical, IBM Watson Health, and Microsoft Text Analytics for Health have emerged and shown promise in clinical concept recognition, entity linking, and sentiment analysis. However, these general-purpose NLP frameworks have shown varying degrees of performance on different data sources [11-13]. While large language models (LLMs) are rapidly advancing, they currently have limitations in clinical concept identification and medical relation extraction as structured outputs for direct application. Even specialized clinical LLMs require further fine-tuning for such use cases [14].

We aim to configure a pretrained, general-purpose health care NLP framework to transform free-text clinical notes and radiology reports into structured progression events. By combining these with structured drug records and encounter data, we will compute real-world progression-free survival (rwPFS) for metastatic breast cancer (mBC). This work can also guide other researchers in configuring a general-purpose health care NLP model to capture rwPFS. Developing a standardized and automated path for ascertaining rwP could help scale rwPFS computation across diverse subsets of solid tumors and maintain a better agreement across real-world studies.

Objectives

We aim to (1) develop and validate a novel semiautomated workflow that estimates rwPFS in patients with mBC, (2) explore the essentiality of each model in the general-purpose NLP framework through ablation analysis, and (3) investigate additional factors influencing rwPFS, such as the source of clinician-documented progressions (radiology reports versus routine clinical notes) and the presence of prior or concurrent medications during the observation period.

Method

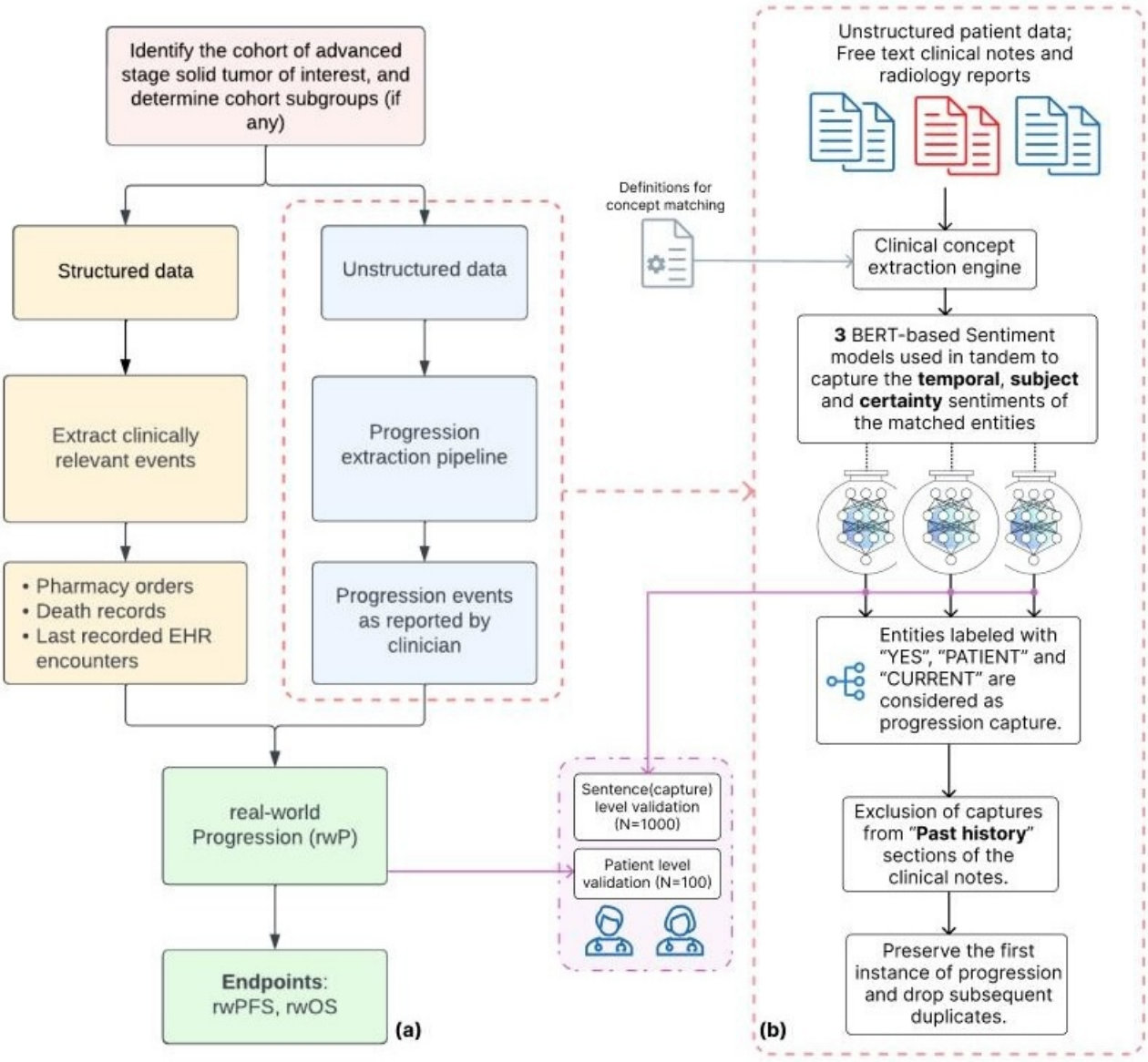
Data Source

This study analyzed deidentified EHR data from a network of tertiary clinical centers tied to an academic medical center (Mayo Clinic) in the United States through the Nference nSights Analytics Platform [15]. In-house tools were used for the deidentification of EHRs. The tool performs with a recall of 0.992 and a precision of 0.979 on the i2b2 2014 dataset at replacing protected health information (PHI) with plausible but fictional surrogates [16]. Overall, the platform hosts data from approximately 7 million patients from across the United States of America, with about 1.8 million patients having a mention of cancer across the structured data. The platform hosts an array of multimodal data, both structured and unstructured, such as clinical notes, radiology reports, Digital Imaging and Communications in Medicine headers, and pathology reports.

Study Design and Definitions

This retrospective observational study demonstrates the estimation of rwPFS with a workflow that integrates clinician-reported progression events from free text (unstructured data) in clinical and radiology documents with structured patient events like drug orders, clinic or hospital encounters, and mortality records. The workflow was developed to leverage the pretrained, general-purpose, deep learning–based health care NLP framework developed at Nference called the clinical NLP engine, which enables clinical concept recognition, sentiment analysis, and linking associated concepts. The models that are a part of the clinical NLP engine were initially trained on human-annotated datasets, and later further augmented by additional ground truth datasets generated by LLM agents from the same parent EHR data source. A high-level overview of the workflow is illustrated in Figure 1.

Figure 1. Methodology flow diagram illustrating the workflow. (**A**) Workflow for real-world progression (rwP) extraction and determining the real-world progression-free survival (rwPFS). (**B**) The methodology for capturing progression from unstructured texts in routine clinical documents and radiology reports using Nference's clinical NLP engine that performs clinical concept recognition, association, and sentiment analysis. BERT: Bidirectional Encoder Representations from Transformers; EHR: electronic health record.



Data Extraction and Augmentation

Breast cancer disease was identified using structured diagnosis codes 174 (*ICD-9* [*International Classification of Diseases, Ninth Revision*]), C50 (*ICD-10* [*International Statistical Classification of Diseases, Tenth Revision*]), and NLP-based positive model confirmations (augmented curation) of the disease-related terms in clinical notes [17]. A similar approach was undertaken for identifying metastatic disease using structured codes 197, 198 (*ICD-9*), C78, and C79 (*ICD-10*). Further cohort attrition for the study population is outlined in Figure 2. Hormone receptor status, human epidermal growth factor receptor-2 (HER-2) status, and Eastern Cooperative Oncology Group scores were captured from clinical notes using the clinical NLP engine. A rule-based approach was used to identify the initiation date of first-line therapy in mBC by analyzing drug orders and administration records. To ensure reliability, only orders appearing for the first time after metastasis diagnosis were included. The same methodology was extended to identify second-line therapies.

Figure 2. Cohort attrition diagram: structured codes 174* (*ICD-9*) and C50* (*ICD-10*) or >4 positive disease sentiments from the augmented curation disease diagnosis model were used for breast cancer. For evidence of metastasis, 197*, 198* (*ICD-9*), C78*, and C79* (*ICD-10*) in conjunction with augmented curation were used; * represents all the children codes within the parent code. ECOG: Eastern Cooperative Oncology Group; EHR: electronic health record; HER-2: human epidermal growth factor receptor-2; HR: hormone receptor; *ICD-9*: *International Classification of Diseases, Ninth Revision*; *ICD-10*: *International Statistical Classification of Diseases, Tenth Revision*; NLP: natural language processing.

Patient Population

The study cohort (N=316) consisted of female patients aged ≥18 years and diagnosed with mBC with hormone receptor–positive and HER-2–negative status with confirmed concurrent exposure to palbociclib and letrozole from January 1, 2015, to December 31, 2021 (study period), and with Eastern Cooperative Oncology Group performance scores of less than 3 around the start of the therapy (±60 d).

Demographics and baseline characterization of the study cohort, such as prior exposure to therapies, disease stage at the start of the study, stage at first cancer diagnosis, and other relevant metrics for the solid tumor of interest, were also documented.

Extracting Progression Events From Unstructured Text

To develop and evaluate our workflow, an initial rule-definition set of 200 cases from the overall mBC cohort (N=10,791) was sampled, and a preliminary manual abstraction was performed. This evaluation aimed to systematically identify a set of rules for configuring the baseline clinical concept extraction model to capture progression. Authors PKM, SKR, VK, and MM used internal clinical document (CD) exploration tools to cluster sentences based on initial pattern matches and iteratively refined these clusters to identify progression-indicative phrases. Independent reviews of oncology and radiology notes were conducted to extract commonly occurring phrases indicative of clinical progression. Authors GV and RHY subsequently collated these identified patterns into a set of regex search patterns. These patterns were tested on clinical notes to ensure they captured the appropriate progression-related contexts while removing duplicate or irrelevant verbatim such as general report headers, unrelated phrases (eg, “CR” for complete response or “PD” for progressive disease without patient-specific context), and noise. Downstream NLP models were applied to validate the extracted patterns by mapping the right set of label combinations that accurately reflected the progression status. This process was repeated iteratively until a consensus was reached among the authors, ensuring a robust set of rules for capturing progression events. The final progression capture configuration is detailed in Table S1 in Multimedia Appendix 1.

The rwP events were captured by configuring the clinical NLP engine. This baseline workflow is an ensemble of deep learning–based multi-BERT (Bidirectional Encoder Representations from Transformers) framework trained on unstructured patient data like CDs and radiology reports to perform named entity recognition and predict the sentiment labels for *subject*, *temporality*, and *certainty* of the captured named entities [18]. The ensemble also infers associations between related entities like disease-severity, drug-disease, and disease-anatomical_structure, among others. These proprietary models are fine-tuned versions of SciBERT-cased [19], a domain-specific transformer model pretrained on scientific text. The base models underwent further supervised fine-tuning for classification tasks on annotated sentences from CD texts of the overall Nference nSights database, but not specifically on the mBC patient note database. The details regarding their architecture, training, and performance of the individual models of the clinical NLP engine are detailed in Note S1 and Table S2 in Multimedia Appendix 1. The clinical NLP engine also uses a section header model to identify the clinical note sections from which the named entities were captured. The rules determined during the initial abstraction were used to capture cancer progression.

rwP Definitions

rwP events were identified by the earliest documentation of disease progression in a clinical or radiology note or by advancement to a new line of therapy. The addition of a new chemotherapy, endocrine therapy, or targeted therapy drug after 60 days of exposure to the initiating therapy of interest is considered line advancement. The following list of drugs were considered potential second-line drug candidates in the study period: tamoxifen, fulvestrant, elacestrant, paclitaxel, carboplatin, abemaciclib, docetaxel, cyclophosphamide, capecitabine, ribociclib, alpelisib, everolimus, doxorubicin, epirubicin, 5-fluorouracil, olaparib, talazoparib, ixabepilone, raloxifene, and toremifene. Censoring events included death, loss of follow-up, and the end of the observation period. Progression events captured within the first 30 days of therapy initiation were excluded as they occurred too early to reflect treatment effectiveness. The rwPFS was also assessed with variations in the origin of unstructured data, comparing radiology reports and CDs as data sources. The key contributing survival variables used for rwPFS were also stratified to understand the source of events.

Progression Capture Validation

Overview

For validation of the clinical progression captures, the manual review and abstraction were performed at 2 levels.

The raw progression captures were evaluated for accuracy independent of their temporality to the observation period. 1000 captures were sampled from the overall pool of progression captures for the sentence-level progression capture analysis.

A stratified sample (mBC validation set; n=100) was selected from the overall study cohort (N=316) to match the progression event occurrence observed in the overall set. This approach ensures that the sample mirrors the broader cohort’s characteristics for a valid comparison in patient-level evaluation for progression capture. These patients were not part of the initial rule-definition set and were evaluated for their first progression events. For the patient-level analysis, the elements of the confusion matrix were defined to account for temporality. We classified the captures into 4 categories: (1) true positives: automated progression captured is within ±30 days of manual capture; (2) false positives: progression was not found through manual capture, but automated progression was captured at any time or automated progression was captured >30 days before manual capture; (3) true negatives: progression was not found through manual review, nor was picked up by automated capture; and (4) false negatives: progression was identified through manual capture, but the automated method has not identified any progression (or) automated method captured progression >30 days after the date captured by manual review.

Ablation Analysis of the Progression Capture Pipeline

To evaluate the contribution of each workflow component to the overall performance, an ablation study was performed at 5 strategic points: (1) temporal model ablation, that is, the removal of the temporality assessment model; (2) subject model ablation, that is, the removal of the subject assessment model; (3) certainty model ablation, that is, the certainty assessment model was removed; (4) all 3 assessment models were removed; and (5) postprocessing ablation where the postprocessing steps, specifically the exclusion of specific note sections and dropping subsequent duplicate mentions of the same note contexts, were removed. Each ablation was analyzed in isolation to quantify its respective contribution to the final output's accuracy, aiding in identifying critical components of the pipeline and potential areas for optimization. This step is further illustrated in Figure S2 in Multimedia Appendix 1.

Validation on the External Dataset

To further assess the generalizability and robustness of the progression capture pipeline, external validation was performed using data from a different partner academic medical center (AMC). In the external dataset, 63 mBC patients on first-line therapy of the metastatic disease with palbociclib with or without aromatase inhibitors were identified for this analysis (see Figure S4 for cohort attrition in Multimedia Appendix 1). Similar NLP-based data augmentation techniques were applied on the external dataset for cohort identification. The proposed progression extraction workflow was applied on the external dataset for extraction of progression events during the defined study period (60 months) of the patients. For validation, similar to the primary cohort, a 2-tier approach was applied, including sentence-level and patient-level validation. For patient-level validation, the time of the first progression event in the patient cohort was manually abstracted and validated. Performance metrics such as precision, recall, accuracy, and F_1-score were calculated to evaluate the alignment of automated progression captures with manual annotations.

Sensitivity Analysis on rwPFS Estimates

Two sensitivity analyses were conducted to assess the robustness of rwPFS estimates: (1) to evaluate the effect of data incompleteness, 10%, 20%, and 30% of rows capturing progression events (missingness at random) were systematically removed from the Kaplan-Meier source data. Median rwPFS and survival probabilities were descriptively analyzed to quantify variations introduced by missing records. (2) The impact of treating death as a progression event versus censoring was assessed by generating Kaplan-Meier curves under both scenarios. Differences were evaluated using the log-rank test, with comparisons of median rwPFS and survival probabilities at predefined time points. These analyses ensured the robustness of rwPFS estimates by addressing potential biases from data structure and event definitions.

Patient-Reported Outcomes and Clinical Progression: Analysis Using the Patient Health Questionnaire-8

Patient-reported outcomes (PROs) were integrated by analyzing Patient Health Questionnaire-8 (PHQ-8) scores to complement clinician-documented progression events. Cumulative declines and peak reductions in PHQ-8 scores were compared between progressed and nonprogressed patients using a *t* test. Peak reduction was determined as the largest decrease observed between any 2 recorded scores during the study period. Cumulative decline, representing the total improvement over time, was calculated as the sum of all positive reductions (decline in PHQ-8 value) in scores across all pairwise comparisons during the study period. This approach aimed to provide a holistic perspective by linking patient-reported mental health outcomes to clinical progression.

Outcomes Assessment

The primary outcome was rwPFS, calculated as the time between the start of the intervention of interest and the first rwP event captured or a change in the line of therapy for the patient. The secondary outcome was real-world overall survival (rwOS), calculated as the time between the start of the intervention of interest and the date of death. A subgroup analysis was performed to assess the impact of prior and other concomitant medications on rwPFS and rwOS in the metastatic setting. The validation metrics of the primary outcome were reported as sensitivity, specificity, accuracy, precision, and F_1-scores.

Median follow-up time was computed from the date of the start of therapy till their last encounter in the EHR system. Time to treatment after the diagnosis of advanced disease (first evidence of metastasis) and follow-up after the start of therapy were imputed using *the date of the first evidence of metastasis* (identified by structured disease code or NLP-positive confirmation) and *the date of the first structured drug order for the combination therapy* (palbociclib and letrozole) as the anchor dates, respectively.

Ethical Considerations

This study analyzed deidentified primary patient-level data extracted from the Nference's, nSights electronic health record database under a data-use agreement that obviates the need for additional institutional review board review. Nference, in collaboration with the AMC data partner that provided the deidentified data for this study, has established a secure data environment, hosted by and within the AMC, that houses the AMC's deidentified patient data. The provisioning of and access to this data are governed by an expert determination that satisfies the Health Insurance Portability and Accountability Act Privacy Rule requirements for the deidentification of PHI. Each AMC's deidentified data environment is specifically designed and operated to enable access to and analysis of deidentified data without the need for institutional review board oversight, approval, or an exemption confirmation. Participants retain the right to opt out at any time. The data are accessible only to authorized users subject to a robust credentialing and authentication process. Data shown and reported in the manuscript have been extracted from this environment using an established protocol for data extraction, aimed at preserving patient privacy. The data have been deidentified pursuant to an expert determination in accordance with the Health Insurance Portability and Accountability Act Privacy Rule. No compensation was provided to individuals whose deidentified records were included.

Statistical Analysis

Data hosted on Nference's nSights environment were imported on demand into the secure code workspaces deployed with Python (version 3.10.6). Missing data imputation was not undertaken. The analysis workflow uses proprietary Python packages with APIs for database querying and data standardization. The descriptive statistics were reported as n (%) and median, IQR. Loss to follow-up was considered as a censoring event for survival estimates. The Kaplan-Meier estimator from the lifelines package 0.27.7 was used in this analysis. The median rwPFS and rwOS were reported, with a 95% CI.

Results

Workflow Configuration

We applied 3 selection conditions to ensure that the progression captures from the clinical NLP engine are relevant and up-to-date with the patient's current status with respect to the clinical note or report. First, entities labeled with "YES," "PATIENT," and "CURRENT," each having a sentiment prediction confidence of ≥0.9, were deemed relevant. Second, to address the issue of "copy-forwarding" in clinical notes, only the first chronological instance of each extracted sentence was retained. Finally, sentences from "Past History" sections were excluded, as they are unlikely to reflect events occurring at the time of documentation. The workflow configuration is illustrated in Figure 1.

Performance Evaluation

The accuracy of the progression capture was evaluated at 2 levels (Table 1): in level 1, sentence-level progression capture validation yielded an accuracy of 98.2% for the relevant progression captures, and in level 2, patient-level validation yielded an accuracy of 88.0%. Ablation analysis revealed the essentiality of the individual components of the clinical NLP engine for progression capture and selection conditions. All steps except the subject sentiment model labels substantially contribute to the overall model performance. This model can be disabled and the workflow performance remains the same. The patient-level workflow performance at each ablation step is outlined in Table 2.

Table . Manual validation was performed for progression at 2 levels.

	Values
Progression capture analysis in level 1[a] (manual validation of sampled raw progression captures [n=1000])	
Sensitivity	99.8%
Specificity	96.7%
Precision	96.6%
Accuracy	98.2%
F_1-score	98.2
First progression capture analysis in level 2[b] (manual validation of first progression [n=100])	
Sensitivity	92.5%
Specificity	83.0%
Precision	86.0%
Accuracy	88.0%
F_1-score	89.1

[a]Level 1: sentence-level review to validate the capture of progression sentiments at the sentence level. At this level, we reviewed the extracted sentences to ascertain the validity of the progression capture at a sentence level.

[b]Level 2: patient-level review to identify the first progression date. Here, we undertook a full review of patient records to ascertain the first progression capture of the metastatic disease.

Table . Output of the ablation analysis showcasing the performance metrics at each step.

	Validation against manually abstracted patient-level dataset (N=100)			
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Median PFS[a] (months), value (95% CI)
Overall workflow	88.0	92.5	83.0	20 (18 - 26)
Temporal model ablation	79.0	91.5	67.9	19 (15 - 23)
Subject model ablation	88.0	92.5	83.0	20 (18 - 26)
Certainty model ablation	42.0	100	20.5	7 (6-8)
All 3 sentiment models ablated	35	100	15.6	6 (5-7)
Postprocessing ablation	87.0	96.2	76.6	19 (16 - 23)

[a]PFS: progression-free survival.

Cohort Description

The baseline characteristics are detailed in Table 3. The median age at metastasis was 59 (IQR 50.5 - 69) years. The median follow-up time after metastasis diagnosis was 43.3 (IQR 28.1 - 61.2) months and the median follow-up time after the start of the therapy was 39.8 (IQR 25.5 - 57.9) months. The starting dose of palbociclib and letrozole was available for 53.2% and 61%, respectively. The median number of drug orders for palbociclib and letrozole was 6 (IQR 3 - 12) and 4 (IQR 2 - 7), respectively. The treatment characteristics, including prior and concomitant exposure to other chemotherapy agents and a history of prior radiotherapy and breast surgery, are also detailed in Table 3. The breakdown of outcomes and censoring events that contributed to the Kaplan-Meier survival estimates are further detailed in Table 4.

Table . Study cohort characteristics.

Category and variable		Overall mBC[a] cohort (N=316)	mBC validation set (n=100)
Demographics			
	Age at metastasis (y), median (IQR)	59 (50.5 - 68.2)	59.1 (47.8 - 69.2)
	Female gender, n (%)	316 (100)	100 (100)
Ethnicity, n (%)			
	Not Hispanic or Latino	298 (94.3)	94 (94)
	Hispanic or Latino	11 (3.5)	3 (3)
	Unknown or choose not to disclose	7 (2.2)	3 (3)
Race, n (%)			
	Caucasian	293 (92.8)	95 (95)
	Asian	6 (1.8)	1 (1)
	African American or other	17 (5.4)	4 (4)
Tumor markers, n (%)			
	HR-positive[b]	316 (100)	100 (100)
	ER-positive[c] and PR-positive[d]	223 (70.6)	68 (68)
	ER-positive and PR-negative	27 (8.5)	11 (11)
	PR-positive and ER-negative	50 (15.8)	15 (15)
	ER, PR status unknown	16 (5.1)	6 (6)
	HER-2/neu-negative[e]	316 (100)	100 (100)
Disease severity, n (%)			
	Patients with confirmed stage IV [−30,+30] within 30 d of primary diagnosis[f]	196 (61.1)	62 (62)
	ECOG[g] performance score <3	316 (100)	100 (100)
Disease-related follow-up, median (IQR)			
	Follow-up after metastasis (months)	43.3 (28.1 - 61.2)	50.8 (38.6 - 64.2)
Treatment, n (%)			
	Prior systemic therapy	22 (7.5)	6 (6)
	Prior radiotherapy	125 (39.5)	38 (38)
	Prior surgical resection	128 (40.5)	37 (37)
	Other concomitant systemic therapy	36 (11)	6 (6)
Treatment follow-up, median (IQR)			
	Follow-up after start of treatment in months	39.8 (25.5 - 57.9)	48.6 (37.8 - 61.4)
	Time to treatment after advanced disease diagnosis in months	0.5 (0.2 - 1.7)	0.6 (0.2 - 1.7)

[a]mBC: metastatic breast cancer.

[b]HR: hormone receptor

[c]ER: estrogen receptor.

[d]PR: progesterone receptor.

[e]HER-2: human epidermal growth factor receptor-2.

[f]All patients included in the study are stage 4 cancer. The provided numbers represent those diagnosed within the stated period.

[g]ECOG: Eastern Cooperative Oncology Group.

Table . Breakdown analysis of outcomes and censoring events in the mBC[a] cohort.

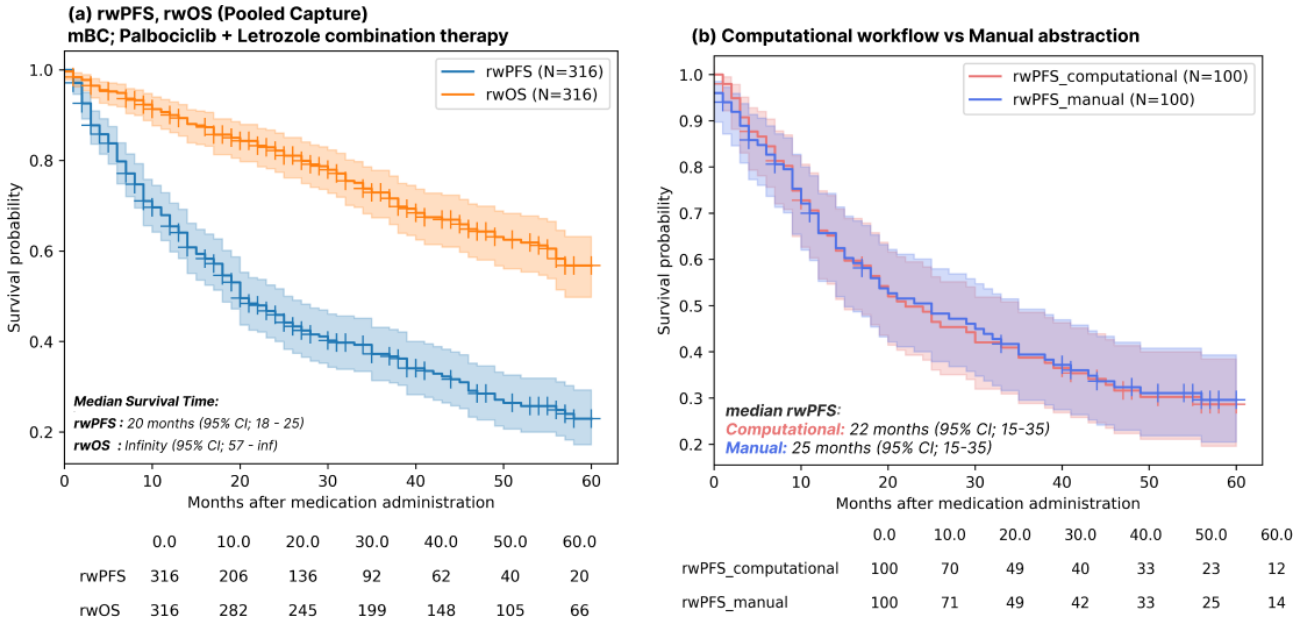
Source of capture		Events, n (%)
Breakdown of progression events (n=199)		
The first event is a progression capture from the pooled sources (n=152)		
	Radiology Reports	78 (51.3)
	Clinical documents	74 (48.7)
The first event is the start of a second-line drug (n=47)		
	Capecitabine	13 (27.7)
	Everolimus	11 (23.4)
	Abemaciclib	9 (19.1)
	Ribociclib	9 (19.1)
	5-Fluorouracil	3 (6.4)
	Olaparib	1 (2.1)
	Cyclophosphamide	1 (2.1)
Breakdown of censoring events (n=117)		
Last encounter date		76 (65)
Patient death date		22 (18.8)
End of study period		19 (16.2)

[a]mBC: metastatic breast cancer.

Outcomes

In the study cohort (N=316), 199 (62.9%) patients progressed during the observation period (60 mo starting from Jan 1, 2015). Out of the progressed patients, 152 (48.1%) were based on progression captures from unstructured data, and 47 (14.8%) were based on changes in the line of therapy from structured data. The median rwPFS for the overall cohort was 20 months (95% CI 18.0 - 25.0; Figure 3A). The median rwOS was not reached during the study period (95% CI 57- not reached [NR]).

Figure 3. Kaplan-Meier survival plots for the overall study cohort and validation sets: (**A**) Kaplan-Meier survival plots indicating the real-world progression-free survival (rwPFS) and real-world overall survival (rwOS) in the study cohort of patients with metastatic breast cancer using pooled note sources. (**B**) Patient-level validation of first progression capture and comparing outcomes estimated by computational workflow with manual curation. mBC: metastatic breast cancer.



In the mBC validation set of 100 patients, the median rwPFS was determined to be 25 (95% CI 15-35) months by manual curation and 22 (95%CI 15-35) months by the computational workflow outlined in Figure 3B. Based on the data source for progression capture, the rwPFS estimated exclusively from radiology reports was 30 (95% CI 24.0-39.0) months, compared to 23 (95% CI 19.0-28.0) months when estimated exclusively from CDs, as represented in Figure 4. Subgroup analysis on the

rwPFS based on prior or concomitant therapies is detailed in Figure 5.

Figure 4. Kaplan-Meier survival plots for real-world progression-free survival (rwPFS) based on the patient note source. Survival plots indicating the real-world rwPFS with progressions captured from solitary sources of radiology reports (RR) and routine clinical documents (CD). mBC: metastatic breast cancer.

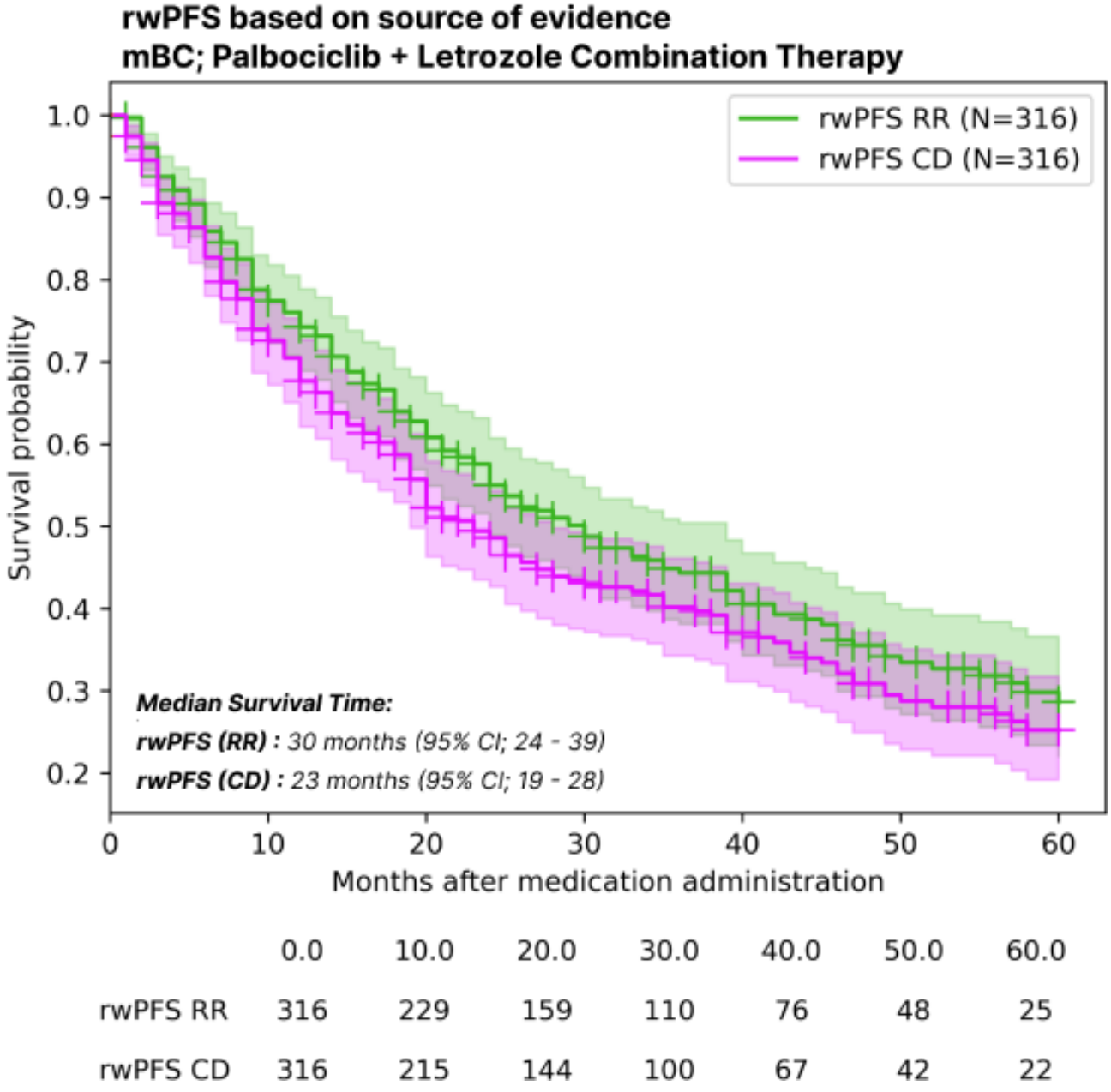
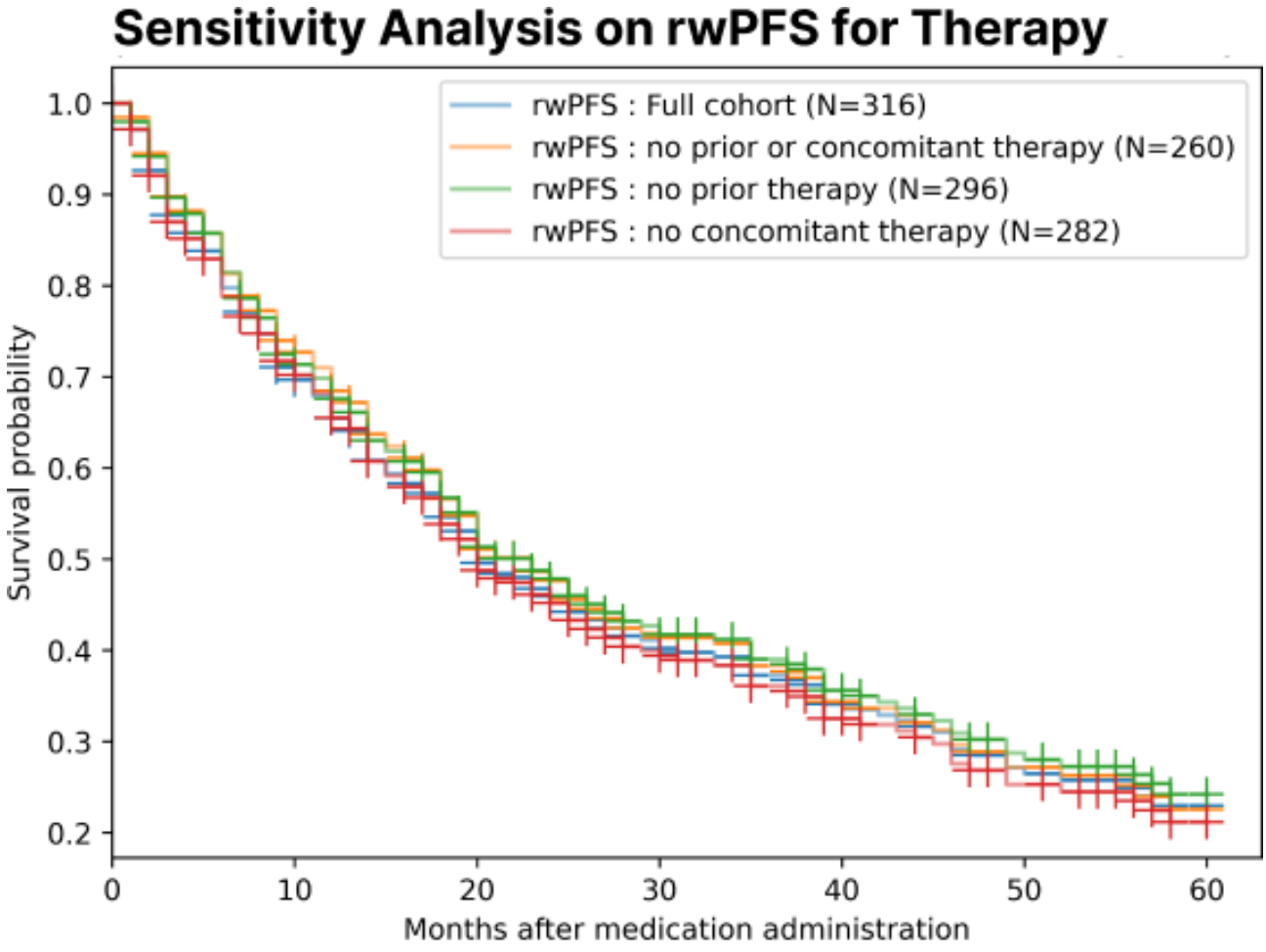


Figure 5. Kaplan-Meier survival curves for subgroup analysis. Each of the subgroups account for different variations in treatment patterns. The survival curves and risks table showcase the effect of other prior or concomitant systemic therapies on the median real-world progression-free survival (rwPFS).



Time	Median rwPFS estimate	Time(months)						
		0(N)	10	20	30	40	50	60
rwPFS : Full cohort	20.00 months (95% CI 18.0-25.0)	316	206	136	92	62	40	20
rwPFS : no prior or concomitant therapy	23.00 months (95% CI 19.0-27.0)	260	176	118	78	50	34	14
rwPFS : no prior therapy	23.00 months (95% CI 19.0-28.0)	296	196	132	89	60	39	19
rwPFS : no concomitant therapy	20.00 months (95% CI 18.0-25.0)	282	186	122	81	52	35	15

External Dataset Validation

External validation was conducted using data from a partner AMC, representing a health system distinct from the main study, to assess the generalizability and robustness of the progression capture pipeline. At Level 1, manual validation of 200 sampled raw progression captures achieved an accuracy of 92.5% and an F_1-score of 92.8%, while Level 2 validation of the first progression in 61 patients reported an accuracy of 90.2% and an F_1-score of 92.5%. Two patients were excluded from performance metrics because the textual evidence identified during manual abstraction of the first progression event was unavailable to the automated extraction pipeline. Comprehensive performance metrics, including sensitivity, specificity, and F_1-scores, are detailed in Table 5.

Table . Manual validation of the progression capture workflow on the external dataset.

Validation step	Sensitivity	Specificity	Accuracy	Precision	F_1-score
Level 1					
Manual validation of sampled raw progression captures (n=200)	89.7%	95.7%	92.5%	96%	92.8%
Level 2					
Manual validation of first progression (n=61)	97.4%	78.3%	90.2%	88.1%	92.5%

Sensitivity Analysis on rwPFS Estimates

The sensitivity analysis demonstrated that rwPFS estimates were robust under varying conditions. Systematic removal of 10%, 20%, and 30% of progression events resulted in median rwPFS values of 20 (95% CI 18-26) months, 20 (95% CI 18-27) months, 22 (95% CI 18-28) months, and 23 (95% CI 19-29) months for the complete, 10%, 20%, and 30% datasets, respectively, with widening CIs indicating increased uncertainty. Similarly, median rwPFS estimates were comparable when death was treated as censorship versus a progression event, with values of 20 months (95% CI 18-25) and 18 months (95% CI 15-21) in the main dataset. These findings, along with overlapping CIs, indicate that the rwPFS estimates were not meaningfully affected by missing data or event definitions. See Figure S5 in Multimedia Appendix 1 for Kaplan-Meier curves and event tables.

PHQ-8 Outcomes and Disease Progression

Of the 316 patients, 94 had at least 2 PHQ-8 scores recorded during the study period, including 30 nonprogressed and 64 progressed patients. Nonprogressed patients showed greater mean peak reduction (5.57, SD 5.90 vs 2.95, SD 4.30; t_{92}=−2.397, P=.02) and cumulative decline (mean 8.00, SD 12.68 vs 3.66, SD 6.40; t_{92}=−2.201, P=.03) in PHQ-8 scores compared to progressed patients. See Table S3 in Multimedia Appendix 1 for details.

Discussion

Overview

The study showcases the development and validation of a novel semiautomated workflow for estimating rwPFS in patients with mBC using deidentified EHRs. One of its key strengths lies in the integration of NLP techniques to extract clinician-reported progression events from unstructured data sources such as clinical notes and radiology reports, combined with structured patient data like drug orders and clinical encounters. This approach enhances the accuracy and comprehensiveness of capturing progression events, as evidenced by the high sensitivity (99.8%) and specificity (96.7%) at the sentence level with good patient-level accuracy (88%).

While our initial goal was to develop a fully automated workflow for capturing disease progression, we have successfully implemented a semiautomated approach. This is advantageous because the semiautomated method allows for disease-specific adjustments to clinical concept recognition configurations, ensuring relevance across various cancer types. This flexibility underscores the potential for broader applicability beyond mBC, making it a valuable tool for oncological research and RWE generation.

Principal Findings

The median rwPFS of 20 months (95% CI 18 - 25) reported in this study is comparable to those reported in previous real-world studies and clinical trial results, validating the workflow's reliability and accuracy [20,21]. Subgroup analysis revealed the impact of prior and other concomitant medications on median rwPFS. For instance, the PALOMA-2 trial reported a median progression-free survival (PFS) of 24.8 months (95% CI 22.1-inf), which is comparable to the real-world observation in the study subcohort (N=260) of patients who received palbociclib and letrozole in the first-line metastatic setting (patients with no other prior or concomitant drugs), which was 23.00 months [22]. A matched comparison between the study cohort and other real-world cohorts or clinical trials could further establish the concordance between the survival estimates. Integrating progression events from both radiology reports and routine clinical or oncology notes standardizes the identification of disease progression, mitigating biases and overestimation that can arise from relying exclusively on a single data source. Ablation analysis also revealed the futility of using the subject sentiment analysis model in the workflow, as physicians are unlikely to describe the progression status of a family member or a blood relative in the patient's notes. While this model is useful for extracting other concepts using the clinical NLP engine, it has shown no benefit in its usage for progression capture.

The external validation further demonstrated the robustness and generalizability of the progression capture workflow across health systems. Manual validation on this dataset also achieves high accuracy at the sentence level (92.5%) and at capturing the first progression event (90.2%). Sensitivity analysis confirmed that rwPFS estimates were stable, regardless of whether death was treated as censorship or an event, with overlapping CIs observed across both scenarios. Sensitivity analysis confirmed that rwPFS estimates were stable across varying levels of missing data and event definitions, with slight increases in median rwPFS and wider CIs under data incompleteness and overlapping intervals when treating death as censorship or an event. Furthermore, integration of PHQ-8 outcomes revealed significant associations between patient-reported mental health and progression status,

highlighting the potential of PROs to provide complementary insights.

Comparison to Prior Work

The study also highlights the importance of source data used for determining rwPFS. Relying solely on radiology reports overestimated the median rwPFS compared to estimates derived from both clinical notes and radiology reports combined. The median rwPFS from the pool of free-text data excluding radiology reports (23 months) was closer to the median survival of the overall study cohort with all available pooled free-text data (20 months) when compared to the median rwPFS computed from free-text data exclusively from radiology reports (30 months). This discrepancy can be explained by the observation that patients can undergo radiological evaluations outside the EHR data network, with their findings being documented by treating physicians within the EHR network in patients' routine clinical notes. Similar findings were observed in a previous study that analyzed the impact of source data on real-world survival estimates [6]. Additionally, relying solely on structured data like drug records (time to discontinuation or time to next treatment) as a surrogate for rwPFS has been shown to underestimate the median rwPFS substantially in a prior study [23]. PROs provide direct insight into a patient's symptoms and quality of life and have been linked to progression-free and overall survival in prior studies. Although direct comparisons with alternative workflows were not performed, our method demonstrates performance metrics that are in line with those reported in previous studies, warranting further comparative analyses in future work.

Among other computational techniques for characterizing cancer response in real-world data (RWD), the use of LLMs has also shown promise. A prior study evaluating this has shown GatorTron to be the best-performing model, achieving an accuracy of 89% at the radiology report level upon fine-tuning [24]. However, applying LLMs across a broader patient corpus needs further investigation to fully ascertain their validity and generalizability. PROs provide direct insight into a patient's symptoms and quality of life and have been linked to progression-free and overall survival in prior studies [25-27]. We have observed significant associations between the decline in PHQ-8 scores and the patient's progression status.

Limitations

There are, however, limitations to this study. First, the reliance on clinician-reported events means that the accuracy of the workflow is reliant on the quality and completeness of clinical documentation. Incomplete or inconsistent documentation could lead to underestimation or overestimation of progression events. To mitigate this, careful validation of extraction patterns and data completeness checks were implemented. Second, although the semiautomated workflow reduces the resource-intensive nature of manual abstraction, it requires initial manual rule definition and configuration, which could introduce biases based on the selected rules and criteria. Representative evaluation samples were curated across the breast cancer cohorts to reduce the biases. Third, sensitivity analyses were limited primarily to variations in clinical text sources and censoring definitions. Expanding sensitivity analyses to include demographic factors, alternative definitions of progression, and data-source reliability could further strengthen the robustness of the findings. Fourth, augmenting the mortality data with commercial and federal death registries could enhance the accuracy of survival estimates. This was not feasible in the present analysis but represents an important area for future improvement. Fifth, integration of PROs could provide a more comprehensive understanding of patient well-being in relation to progression events. However, demonstrating this in RWD was challenging due to the limited availability of patient-reported records. Finally, the ensemble deep learning engine's performance was evaluated within a specific cohort of mBC patients; thus, further validation across more diverse external datasets and different cancer types is necessary to truly establish the generalizability of the workflow.

Future Directions

These findings align with the growing body of research advocating for integrating artificial intelligence and machine learning in health care data analysis, as these technologies can substantially enhance the speed, accuracy, and breadth of data processing capabilities. Future work will explore more advanced text data analysis and extraction methods, such as adaptive machine learning techniques and LLMs, to minimize manual import and enhance scalability. Furthermore, by using federated learning, insights and patterns from diverse populations across various institutions can be pooled securely, enriching the model's generalizability and performance across different health care settings. The successful implementation of this automated workflow demonstrates its potential to streamline the data extraction process from EHRs from various health systems. It also paves the way for its application in other oncological studies, where similar challenges in data abstraction exist.

Conclusions

Developing a practical and scalable method for capturing real-world progression from EHR data is crucial to improving oncological research and patient care. Overall, this technology represents a step forward in realizing the full potential of EHR data in oncology. Our findings establish a workflow for automated data capture to provide a more efficient and scalable method than traditional manual processes, particularly in handling complex, unstructured EHR data. Although the principles of progression capture remain the same across other cancer types, further research across other types of solid tumors is needed to ascertain the generalizability of the workflow.

Acknowledgments

We thank the peer reviewers for their thoughtful reviews and feedback, which greatly improved the quality of the final manuscript.

Data Availability

The deidentified patient-level EHR data that underpin the findings of this study are not publicly available due to the expert-determination process required to preserve Health Insurance Portability and Accountability Act–compliant deidentification and the terms of our data use agreement. Aggregate results and metadata supporting the conclusions are available from the corresponding author on reasonable request and subject to a data-sharing agreement. Qualified researchers may also request access via the Nference nSights Analytics Platform [15], under existing privacy and security controls. The analytic code and any training or validation datasets remain proprietary and are not publicly available.

Authors' Contributions

GV led data curation, formal analysis, visualization, writing of the original draft, review, editing, and contributed to the methodology. RKY contributed to the methodology, data curation, formal analysis, investigation, and visualization. PK was responsible for conceptualization, methodology, lead investigation, funding acquisition, lead supervision, and manuscript review and editing. BSA, SC, SKR, VK, and MM contributed to data curation and validation, with MM also involved in investigation. KP contributed to data curation, methodology, and investigation. AvA developed the clinical NLP engine software for NLP model inference. SJ developed the NLP models and software. AkA contributed to software development and provided feedback. RB and VT provided feedback, project administration, with VT also providing resources. PL and SAS contributed to validation, feedback, and writing review and editing; SAS also provided resources. VS provided supervision. All authors read and approved the final manuscript.

Conflicts of Interest

GV, PK, KP, SC, AvA, MM, SJ, AkA, RB, VT, PL, and VS are current employees of Nference, inc and hold a minority stake in the company. RKY, BSA, and VK are past employees of Nference. SAS is affiliated with Mayo Clinic, Rochester. The authors declare no further competing interests in the findings of the study.

Multimedia Appendix 1
Additional tables and figures.
[DOCX File, 3415 KB - cancer_v11i1e64697_app1.docx]

References

1. Klonoff DC. The new FDA real-world evidence program to support development of drugs and biologics. J Diabetes Sci Technol 2020 Mar;14(2):345-349. [doi: 10.1177/1932296819832661] [Medline: 30862182]
2. Feinberg BA, Gajra A, Zettler ME, Phillips TD, Phillips EG Jr, Kish JK. Use of real-world evidence to support FDA approval of oncology drugs. Value Health 2020 Oct;23(10):1358-1365. [doi: 10.1016/j.jval.2020.06.006] [Medline: 33032780]
3. Khozin S, Blumenthal GM, Pazdur R. Real-world data for clinical evidence generation in oncology. J Natl Cancer Inst 2017 Nov 1;109(11). [doi: 10.1093/jnci/djx187] [Medline: 29059439]
4. Grimaldi S, Terroir M, Caramella C. Advances in oncological treatment: limitations of RECIST 1.1 criteria. Q J Nucl Med Mol Imaging 2018 Jun;62(2):129-139. [doi: 10.23736/S1824-4785.17.03038-2] [Medline: 29166754]
5. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009 Jan;45(2):228-247. [doi: 10.1016/j.ejca.2008.10.026] [Medline: 19097774]
6. Griffith SD, Tucker M, Bowser B, et al. Generating real-world tumor burden endpoints from electronic health record data: comparison of RECIST, radiology-anchored, and clinician-anchored approaches for abstracting real-world progression in non-small cell lung cancer. Adv Ther 2019 Aug;36(8):2122-2136. [doi: 10.1007/s12325-019-00970-1] [Medline: 31140124]
7. Meyers DE, Stukalin I, Vallerand IA, et al. The lung immune prognostic index discriminates survival outcomes in patients with solid tumors treated with immune checkpoint inhibitors. Cancers (Basel) 2019 Nov 2;11(11):1713. [doi: 10.3390/cancers11111713] [Medline: 31684111]
8. Griffith SD, Miksad RA, Calkins G, et al. Characterizing the feasibility and performance of real-world tumor progression end points and their association with overall survival in a large advanced non-small-cell lung cancer data set. JCO Clin Cancer Inform 2019 Aug;3:1-13. [doi: 10.1200/CCI.19.00013] [Medline: 31403818]
9. Amorrortu R, Garcia M, Zhao Y, et al. Overview of approaches to estimate real-world disease progression in lung cancer. JNCI Cancer Spectr 2023 Oct 31;7(6):pkad074. [doi: 10.1093/jncics/pkad074] [Medline: 37738580]
10. Torres AZ, Nussbaum NC, Parrinello CM, et al. Analysis of a real-world progression variable and related endpoints for patients with five different cancer types. Adv Ther 2022 Jun;39(6):2831-2849. [doi: 10.1007/s12325-022-02091-8] [Medline: 35430670]
11. Suárez-Paniagua V, Dong H, Casey A. A multi-BERT hybrid system for named entity recognition in spanish radiology reports. Presented at: CLEF 2021 – Conference and Labs of the Evaluation Forum 2021; Sep 21-24, 2021; Bucharest, Romania p. 846-856.

12. Bai L, Mulvenna MD, Wang Z, Bond R. Clinical entity extraction: comparison between metamap, ctakes, CLAMP and amazon comprehend medical. 2021 Presented at: 2021 32nd Irish Signals and Systems Conference (ISSC); Jun 10-11, 2021; Athlone, Ireland p. 1-6. [doi: 10.1109/ISSC52156.2021.9467856]
13. Shah-Mohammadi F, Finkelstein J. A comparison of automated named entity recognition tools applied to clinical text. 2022 Presented at: 2022 7th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS); Nov 24-26, 2022; Nara, Japan p. 227-231. [doi: 10.1109/ICIIBMS55689.2022.9971532]
14. Yang X, Chen A, PourNejatian N, et al. GatorTron: a large clinical language model to unlock patient information from unstructured electronic health records. arXiv [csCL]. Preprint posted online on 2022 URL: http://arxiv.org/abs/2203.03540
15. Lenehan P, Venkatakrishnan AJ, Murugadoss K, et al. Clinical nSights: a software platform to accelerate real world oncology analyses. J Clin Oncol 2024 Jun 1;42(16_suppl):e23316-e23316. [doi: 10.1200/JCO.2024.42.16_suppl.e23316]
16. Murugadoss K, Rajasekharan A, Malin B, et al. Building a best-in-class automated de-identification tool for electronic health records through ensemble learning. Patterns (N Y) 2021 Jun 11;2(6):100255. [doi: 10.1016/j.patter.2021.100255] [Medline: 34179842]
17. Wagner T, Shweta F, Murugadoss K, et al. Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis. Elife 2020 Jul 7;9:32633720. [doi: 10.7554/eLife.58227]
18. Devlin J, Chang MW, Lee K, Toutanova K. B. Pre-training of deep bidirectional transformers for language understanding. 2019 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); Minneapolis, Minnesota p. 4171-4186.
19. Beltagy I, Lo K, Cohan A. SciBERT: a pretrained language model for scientific text. 2019 Presented at: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); Hong Kong, China. [doi: 10.18653/v1/D19-1371]
20. DeMichele A, Cristofanilli M, Brufsky A, et al. Comparative effectiveness of first-line palbociclib plus letrozole versus letrozole alone for HR+/HER2- metastatic breast cancer in US real-world clinical practice. Breast Cancer Res 2021 Mar 24;23(1):37. [doi: 10.1186/s13058-021-01409-8] [Medline: 33761995]
21. Wilkie J, Schickli MA, Berger MJ, et al. Progression-free survival for real-world use of palbociclib in hormone receptor-positive metastatic breast cancer. Clin Breast Cancer 2020 Feb;20(1):33-40. [doi: 10.1016/j.clbc.2019.06.010] [Medline: 31451366]
22. Finn RS, Martin M, Rugo HS, et al. Palbociclib and Letrozole in advanced breast cancer. N Engl J Med 2016 Nov 17;375(20):1925-1936. [doi: 10.1056/NEJMoa1607303] [Medline: 27959613]
23. Walker MS, Herms L, Miller PJE. Performance of time to discontinuation and time to next treatment as proxy measures of progression-free survival, overall and by treatment group. JCO 2020 May 20;38(15_suppl):e19135-e19135. [doi: 10.1200/JCO.2020.38.15_suppl.e19135]
24. Tan R, Lin Q, Low GH, et al. Inferring cancer disease response from radiology reports using large language models with data augmentation and prompting. J Am Med Inform Assoc 2023 Sep 25;30(10):1657-1664. [doi: 10.1093/jamia/ocad133] [Medline: 37451682]
25. Bhatt AS, Schabath MB, Hoogland AI, Jim HSL, Brady-Nicholls R. Patient-reported outcomes as interradiographic predictors of response in non-small cell lung cancer. Clin Cancer Res 2023 Aug 15;29(16):3142-3150. [doi: 10.1158/1078-0432.CCR-23-0396] [Medline: 37233986]
26. Balitsky AK, Rayner D, Britto J, et al. Patient-reported outcome measures in cancer care: an updated systematic review and meta-analysis. JAMA Netw Open 2024 Aug 1;7(8):e2424793. [doi: 10.1001/jamanetworkopen.2024.24793] [Medline: 39136947]
27. Badaoui S, Kichenadasse G, Rowland A, Sorich MJ, Hopkins AM. Patient-reported outcomes predict progression-free survival of patients with advanced breast cancer treated with Abemaciclib. Oncologist 2021 Jul;26(7):562-568. [doi: 10.1002/onco.13806] [Medline: 33914991]

Abbreviations

AMC: academic medical center
API: application programming interface
BERT: Bidirectional Encoder Representations from Transformers
CD: clinical document
EHR: electronic health record
HER-2: human epidermal growth factor receptor-2
***ICD-10*:** *International Statistical Classification of Diseases, Tenth Revision*
***ICD-9*:** *International Classification of Diseases, Ninth Revision*
LLM: large language model
mBC: metastatic breast cancer
NLP: natural language processing
PFS: progression-free survival

PHI: protected health information
PHQ-8: Patient Health Questionnaire-8
PRO: patient-reported outcome
RECIST: response evaluation criteria in solid tumors
RWD: real-world data
RWE: real-world evidence
rwOS: real-world overall survival
rwP: real-world progression
rwPFS: real-world progression-free survival

Edited by N Cahill; submitted 24.07.24; peer-reviewed by GK Gupta, M Harao; revised version received 09.03.25; accepted 21.03.25; published 15.05.25.

Please cite as:
Varma G, Yenukoti RK, Kumar M P, Ashrit BS, Purushotham K, Subash C, Ravi SK, Kurien V, Aman A, Manoharan M, Jaiswal S, Anand A, Barve R, Thiagarajan V, Lenehan P, Soefje SA, Soundararajan V
A Deep Learning–Enabled Workflow to Estimate Real-World Progression-Free Survival in Patients With Metastatic Breast Cancer: Study Using Deidentified Electronic Health Records
JMIR Cancer 2025;11:e64697
URL: https://cancer.jmir.org/2025/1/e64697
doi:10.2196/64697

Interpretable Machine Learning to Predict the Malignancy Risk of Follicular Thyroid Neoplasms in Extremely Unbalanced Data: Retrospective Cohort Study and Literature Review

Rui Shan[1], MM; Xin Li[2*], MD; Jing Chen[1], MM; Zheng Chen[3], MD; Yuan-Jia Cheng[4], MD; Bo Han[5,6], PhD; Run-Ze Hu[1], BMed; Jiu-Ping Huang[7], MM; Gui-Lan Kong[8,9], PhD; Hui Liu[10], PhD; Fang Mei[11,12], PhD; Shi-Bing Song[2], MD; Bang-Kai Sun[13], MS; Hui Tian[3], MM; Yang Wang[14], MD, PhD; Wu-Cai Xiao[1], MM; Xiang-Yun Yao[7], MD; Jing-Ming Ye[4], MD; Bo Yu[7], BMed; Chun-Hui Yuan[2], MD; Fan Zhang[7], MD; Zheng Liu[1*], PhD

[1]Department of Maternal and Child Health, School of Public Health, Peking University, Beijing, China

[10]Institute of Advanced Clinical Medicine, Peking University, Beijing, China

[11]Department of Pathology, Peking University Third Hospital, Beijing, China

[12]School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China

[13]Information Management and Big Data Center, Peking University Third Hospital, Beijing, China

[14]Department of Cardiovascular Medicine, First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

[2]Department of General Surgery, Peking University Third Hospital, Beijing, China

[3]Department of Ultrasound, Peking University People's Hospital, Beijing, China

[4]Department of Thyroid and Breast Surgery, Peking University First Hospital, Beijing, China

[5]Department of Pathology, Peking University People's Hospital, Beijing, China

[6]The Key Laboratory of Experimental Teratology, Ministry of Education and Department of Pathology, School of Basic Medical Sciences, Jinan, China

[7]Department of Ultrasound, Peking University Third Hospital, Beijing, China

[8]National Institute of Health Data Science, Peking University, Beijing, China

[9]Advanced Institute of Information Technology, Peking University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Zheng Liu, PhD
Department of Maternal and Child Health, School of Public Health, Peking University, Beijing, China

Abstract

Background: Diagnosing and managing follicular thyroid neoplasms (FTNs) remains a significant challenge, as the malignancy risk cannot be determined until after diagnostic surgery.

Objective: We aimed to use interpretable machine learning to predict the malignancy risk of FTNs preoperatively in a real-world setting.

Methods: We conducted a retrospective cohort study at the Peking University Third Hospital in Beijing, China. Patients with postoperative pathological diagnoses of follicular thyroid adenoma (FTA) or follicular thyroid carcinoma (FTC) were included, excluding those without preoperative thyroid ultrasonography. We used 22 predictors involving demographic characteristics, thyroid sonography, and hormones to train 5 machine learning models: logistic regression, least absolute shrinkage and selection operator regression, random forest, extreme gradient boosting, and support vector machine. The optimal model was selected based on discrimination, calibration, interpretability, and parsimony. To address the highly imbalanced data (FTA:FTC ratio>5:1), model discrimination was assessed using both the area under the receiver operating characteristic curve and the area under the precision-recall curve (AUPRC). To interpret the model, we used Shapley Additive Explanations values and partial dependence and individual conditional expectation plots. Additionally, a systematic review was performed to synthesize existing evidence and validate the discrimination ability of the previously developed Thyroid Imaging Reporting and Data System for Follicular Neoplasm scoring criteria to differentiate between benign and malignant FTNs using our data.

Results: The cohort included 1539 patients (mean age 47.98, SD 14.15 years; female: n=1126, 73.16%) with 1672 FTN tumors (FTA: n=1414; FTC: n=258; FTA:FTC ratio=5.5). The random forest model emerged as optimal, identifying mean thyroid-stimulating hormone (TSH) score, mean tumor diameter, mean TSH, TSH instability, and TSH measurement levels as the top 5 predictors in discriminating FTA from FTC, with the area under the receiver operating characteristic curve of 0.79 (95% CI 0.77 - 0.81) and AUPRC of 0.40 (95% CI 0.37-0.44). Malignancy risk increased nonlinearly with larger tumor diameters and

higher TSH instability but decreased nonlinearly with higher mean TSH scores or mean TSH levels. FTCs with small sizes (mean diameter 2.88, SD 1.38 cm) were more likely to be misclassified as FTAs compared to larger ones (mean diameter 3.71, SD 1.36 cm). The systematic review of the 7 included studies revealed that (1) the FTA:FTC ratio varied from 0.6 to 4.0, lower than the natural distribution of 5.0; (2) no studies assessed prediction performance using AUPRC in unbalanced datasets; and (3) external validations of Thyroid Imaging Reporting and Data System for Follicular Neoplasm scoring criteria underperformed relative to the original study.

Conclusions: Tumor size and TSH measurements were important in screening FTN malignancy risk preoperatively, but accurately predicting the risk of small-sized FTNs remains challenging. Future research should address the limitations posed by the extreme imbalance in FTA and FTC distributions in real-world data.

(JMIR Cancer 2025;11:e66269) doi:10.2196/66269

KEYWORDS

follicular thyroid neoplasm; machine learning; prediction model; malignancy; unbalanced data; literature review

Introduction

Globally, thyroid neoplasms are becoming increasingly prevalent [1]. Among them, follicular thyroid neoplasms (FTNs) represent a major type but have garnered significantly less attention compared to papillary thyroid carcinoma. A key challenge is that over 95% of FTN cases cannot be reliably distinguished as benign (follicular thyroid adenoma [FTA]) or malignant (follicular thyroid carcinoma [FTC]) until diagnostic surgery [2]. This uncertainty often leads to both over- and undertreatment of patients with FTN. On one hand, it is estimated that over 80% of patients who undergo thyroidectomy might ultimately be diagnosed as benign FTN based on postoperative pathology [3]. On the other hand, those with malignant FTN may have already developed distant metastases to the lungs, bones, or other organs by the time they receive surgical treatment.

Several guidelines advocate for enhanced screening, accurate diagnosis, and appropriate treatment for patients with FTN [4,5]. One crucial solution is to develop prediction models to aid clinical decision-making for these patients. To date, machine learning has been proven effective in constructing predictive models for various cancers such as oral, gastrointestinal, and breast cancers [6-8]. Our literature review also indicated that machine learning technology excels at capturing complex, nonlinear relationships and high-dimensional intercorrelations among predictors [9-12].

However, our literature review revealed several limitations among most of the existing studies, mainly including (1) small sample sizes ranging from 18 to 888 participants [13-19], (2) the ratio of FTA to FTC deviating from the real population distributions, (3) reliance on simple linear models unable to capture the complex nonlinearity or interactions underlying predictor-outcome relationships [14,18], (4) using inappropriate metrics to evaluate model performance for the unbalanced data [13-19], (5) lack of assessing the extent to which a predictor influences the model's prediction (ie, model interpretability) [13,16,19], (6) not evaluating whether the predicted probabilities were consistent with actual outcomes (ie, model calibration) in the development and validation of clinical prediction models [20], and (7) predictors are predominantly confined to sonographic features with limited consideration of other factors such as the presence of Hashimoto thyroiditis (an autoimmune disease that may increase the risk for differentiated thyroid cancer [21]).

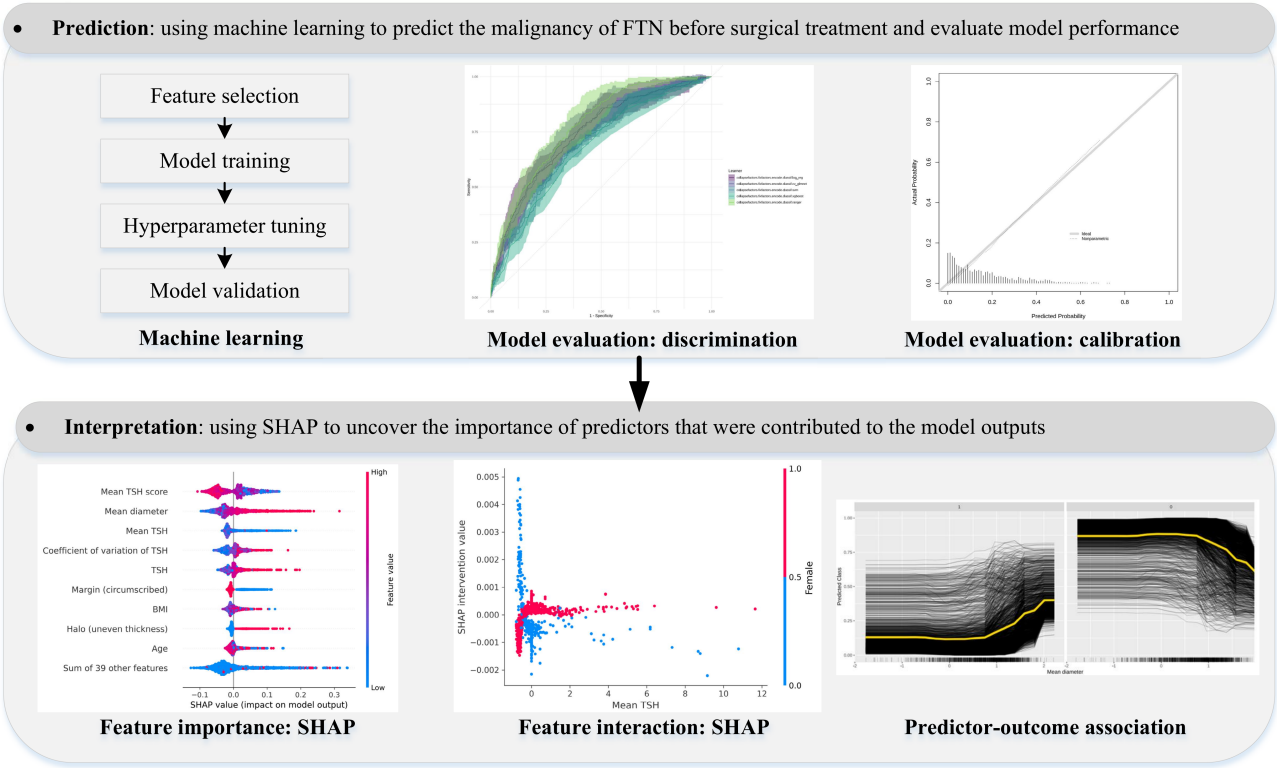
To address these limitations, our study has united a multidisciplinary treatment team for thyroid neoplasms and accumulated a cohort of over 1500 patients with FTN over the past decade [22]. This provided us a unique opportunity to develop and validate clinical prediction models to bridge the current research gaps in the field of FTN. Specifically, we aimed to individualize the clinical decision-making for patients with FTN by using interpretable machine learning to not only predict the malignancy risk of FTN but also identify the important predictors that might contribute to the prediction.

Methods

Study Design

This retrospective cohort study followed the suggestions of the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) statement [23]. Figure 1 shows the framework of our study.

Figure 1. Study framework of the machine learning–based modeling to facilitate the clinical decision-making for patients of FTN. FTN: follicular thyroid neoplasm; SHAP: Shapley Additive Explanations; TSH: thyroid-stimulating hormone.



Study Population

Our multidisciplinary research team included experts in the fields of epidemiology, surgery, pathology, ultrasound, and endocrinology. We conducted a retrospective cohort study at Peking University Third Hospital in Beijing, China, from January 2012 to September 2023. Eligible patients were those who underwent surgery and were pathologically diagnosed with FTA or FTC following the procedure. Patients were excluded if they did not undergo ultrasound examinations prior to surgery or if they had nodules classified as follicular tumors of uncertain malignant potential (UMPs). This exclusion was based on two considerations: (1) accurate diagnosis of FTNs required both an experienced pathologist and a complete biopsy sample. The key to distinguishing between benign and malignant FTNs was determining whether the tumor invaded the capsule. Tumors that invaded the capsule or blood vessels were classified as FTC, while those that did not were considered FTA. If the pathologist struggled to assess capsule invasion due to inexperience, or if the sample was inadequately collected during surgery, leading to capsule damage, the tumor might not have been accurately classified as either FTA or FTC. In such cases, it could have been labeled as a UMP. (2) Through a literature review, we found that all previous research had excluded UMPs [13-19]. Therefore, our study also excluded UMPs, enhancing comparability with prior research. We paid close attention to the accuracy of the pathological diagnosis of FTN due to its high professional requirements, which include not only complete sampling but also a thorough examination of all areas of the tumor margin. To ensure this, we invited pathologists with expertise in thyroid tumors to double-check all the postoperational pathological diagnoses in the study population, based on the most recent 2022, 5th edition WHO Classification of Thyroid Neoplasms [24].

It is important to note that our study population reflected the natural distribution of FTNs (ie, the ratio of FTA and FTC), resulting in imbalanced data, with 84.57% (n=1414) of cases was FTA. Specifically, we did not restrict the ratio of FTA to FTC to 1:1 or any other fixed ratio in the main analyses. This approach allowed the results from the developed prediction model to be more readily applicable to external populations with a similar natural distribution.

Data Sources and Processing

The data for this study were sourced from the electronic health records, extracted by professional information management personnel from the hospital's electronic information system. For critical data sources like thyroid pathology and neck ultrasound reports, a tailored data extraction form was designed using EpiData (EpiData Association), aligning with the study's research questions. The form was iteratively refined through discussions among researchers, surgeons, ultrasound specialists, and pathologists, followed by trial entries and revisions until finalized. Trained clinical doctors and medical students performed manual data entry, with researchers conducting 2 rounds of random checks to ensure accuracy and consistency. Senior doctors performed a final review to verify data quality. Missing values were imputed using the mean for continuous variables and the mode for categorical variables.

Predictors

We selected the predictors to develop the machine learning–based model based on our systematic review [17,18], domain knowledge [25-27], and data available. The predictors

included sonographic features, patients' age, sex, BMI, whether or not diagnosed as Hashimoto thyroiditis (an autoimmune disease that destroys thyroid cells by cell and antibody-mediated immune processes [28]), and measurements of thyroid hormones. Specifically, the sonographic features included mean diameter, composition (solid, predominantly solid, predominantly cystic, or cystic), echogenicity (hyperechoic, isoechoic, hypoechoic, or anechoic), taller-than-wide (the length in the vertical direction is greater than the width in the horizontal direction: absent or present), margin (circumscribed, ill-defined, irregular, or lobulated), calcifications (microcalcifications, macrocalcifications, peripheral calcifications, punctate echogenic foci of undetermined significance, microcalcifications with comet-tail artifacts, or no echogenic foci), halo (absent halo, even thickness halo, uneven thickness halo, or present halo without evenness of thickness reported), internal blood flow (absent or present), vascularity (mainly central vascularity, mainly peripheral vascularity, mixed vascularity, or avascularity), trabecular formation (typically appears as elongated, band-like, or fibrous echogenicity, arranged in a reticular or cord-like pattern: absent or present), and nodule-in-nodule appearance (a smaller nodule or an area with different echogenic characteristics is present within a larger thyroid nodule: absent or present); the measurements at the latest examination of thyroid hormones included thyroid-stimulating hormone (TSH), free triiodothyronine, and free thyroxine; additionally, TSH-related features derived from all examinations included mean TSH score (interval-adjusted detailed TSH score) [29], time-adjusted root mean square of successive differences of TSH [30], mean TSH (mean value of preoperative TSH), and coefficient of variation of TSH (the ratio of SD of preoperative TSH to mean value of preoperative TSH), and detailed definitions were introduced in previous publication [31]. All selected predictors were carefully checked by both clinicians and researchers to ensure the accuracy and reliability of the study results.

Development and Validation of Machine Learning–Based Models

We established the machine learning–based model as shown in Figure 1. We selected features, trained models, tuned hyperparameters, and validated models, as briefly described below. We used the mlr3 [32] ecosystem in R (version 4.3.3; R Foundation for Statistical Computing), scikit-learn [33], and Shapley Additive Explanations (SHAP) [34] in Python (version 3.11.1; Python Software Foundation) to conduct machine learning.

Feature selection, which aims to reduce the number of features, offers several benefits including minimizing overfitting, enhancing model robustness, and accelerating predictions. Notably, it is particularly advantageous for datasets with a high feature-to-sample ratio, where the number of features exceeds the limited size of data points. To identify a core set of predictors that could effectively predict the outcome without redundancy, we used a novel information-gain approach for feature selection [35]. To finalize the optimal model, we also compared model performance between that with full predictors and that with selected predictors.

We trained 5 classification models including logistic regression, least absolute shrinkage and selection operator (LASSO) regression, random forest, extreme gradient boosting, and support vector machine. We comprehensively considered and weighed (trade-off) the performance, calibration, parsimony, and interpretability of models and selected the most appropriate one as our prediction model.

The random search and cross-validation were combined to select model hyperparameters when training the machine learning model. We performed a random search over more than 45,000 hyperparameter combinations to select the best hyperparameter combination and trained the final classifiers. Additionally, to address the issue of imbalance, both oversampling (increasing the amount of minority class samples with producing new samples or repeating some samples) and undersampling (decreasing the amount of majority class samples) techniques were applied [36].

Evaluation of Model Performance

We evaluated the performance of the developed model in terms of discrimination (the ability of the model to distinguish between those with and without the outcome) and calibration (the consistency or agreement between the observed outcomes and predicted risks from the model). For discrimination, we first showed the confusion matrix including the numbers and percentages of true positive, true negative, false positive, and false negative. We then calculated both the threshold-free and threshold-sensitive metrics. Threshold-free metrics included the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC). While AUROC was a common performance metric for discrimination, AUPRC was considered more useful and informative for handling the unbalanced data in this study [37]. Threshold-sensitive metrics included sensitivity, precision, specificity, and accuracy. For calibration, we first plotted predicted risks (x-axis) against observed outcomes (y-axis) using a smoothed flexible calibration curve based on individual data. We also quantitatively assessed calibration using the calibration slope and calibration-in-the-large.

Interpretation of Model Prediction Results

First, we evaluated the feature importance (ie, the extent of the model depended on the feature) and the feature interaction by using the SHAP summary plot and SHAP interaction value dependence plot (see details in Multimedia Appendix 1) [34]. Second, we figured the partial dependence plots to visualize the direction of predictor-outcome associations, illustrate whether the risk of the outcome increased with a rise or decline in the predictor values, and assess whether this relationship is linear. Third, we plotted individual conditional expectations curves to explore potential modifiers that could influence predictor-outcome associations. Finally, we separated FTC into 2 groups based on whether they were correctly predicted and compared their characteristics. The significance of differences between the groups was tested using the Mann-Whitney *U* test, as the data did not follow to a normal distribution.

A Systematic Review of the Previous Studies

We conducted a systematic review of previous studies addressing similar topics. According to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [38] (see details in Checklist 1), we searched PubMed, Web of Science, Embase, and IEEE Xplore using the terms “follicular thyroid cancer” and “predict” for papers published up to October 1, 2023 (see details in Multimedia Appendix 2). Eligible studies included those that established prediction models to distinguish FTC from FTA before operation with various preoperative predictors. We included studies with either deep learning models, machine learning models, traditional statistical models, or other relevant methodologies. Studies were excluded if fewer than 50% of patients had FTN or if the papers were not written in English.

We evaluated the Thyroid Imaging Reporting and Data System for Follicular Neoplasm (F-TIRADS) scoring criteria developed by Li et al [18] to differentiate between benign and malignant cases in our dataset. These criteria are based on 6 key features: mean diameter, composition, echogenicity, margin, calcifications, and trabecular formation. Each specific characteristic of these features is assigned a corresponding point value, and the total points across the 6 features indicate the risk level of FTC. For instance, a total score of 12 points or higher suggests an FTC risk exceeding 90% (refer to Figure S1 in Multimedia Appendix 1).

Ethical Considerations

This study was classified as human participant research and was reviewed and approved by the medical research ethics committee of Peking University Third Hospital (IRB00006761-M2023168). As a retrospective analysis, the study was granted a waiver for additional informed consent. During the data extraction process, strict confidentiality measures were implemented to ensure patient privacy and data security. All extracted data were anonymized, with any information that could directly identify patients being removed.

Results

Characteristics of the Study Population

Altogether, we included 1539 patients, 1409 of whom had solitary tumors, and 130 had more than 1 tumor. Thus, a total of 1672 tumors were included and divided into 2 pathological types: FTA (n=1414) and FTC (n=258). The characteristics of the included tumors are listed in Table 1, and the characteristics of the study population are listed in Table 2. The age of the included population was 47.98 (SD 14.15) years (n=1530; missing value=9), the mean BMI was 24.18 (SD 3.66) kg/m^2 (n=1475; missing value=64), and the female population made up 73.16% (n=1126; male: n=342; missing value=71) of all patients.

Table . Characteristics of the tumors.

Characteristics		FTA[a]	FTC[b]
Number of tumors, n (%)		1414 (84.57)	258 (15.43)
Composition, n (%)			
	Solid	650 (45.97)	140 (54.26)
	Predominantly solid	445 (31.47)	75 (29.07)
	Predominantly cystic	161 (11.39)	16 (6.20)
	Cystic	15 (1.06)	1 (0.39)
	N/A[c]	143 (10.11)	26 (10.08)
Echogenicity, n (%)			
	Anechoic	7 (0.50)	0 (0)
	Hyperechoic	31 (2.19)	7 (2.71)
	Isoechoic	660 (46.68)	105 (40.70)
	Hypoechoic	547 (38.68)	130 (50.39)
	N/A	169 (11.95)	16 (6.20)
Margin, n (%)			
	Circumscribed	1073 (75.88)	166 (64.34)
	Ill-defined	38 (2.69)	6 (2.33)
	Irregular	116 (8.20)	39 (15.12)
	Lobulated	69 (4.88)	35 (13.57)
	N/A	118 (8.35)	12 (4.65)
Halo, n (%)			
	Uneven thickness halo	149 (10.54)	55 (21.32)
	Even thickness halo	444 (31.40)	61 (23.64)
	Absent halo	602 (42.57)	108 (41.86)
	Present halo without evenness of thickness reported	83 (5.87)	11 (4.26)
	N/A	136 (9.62)	23 (8.91)
Taller-than-wide, n (%)			
	Absent	1174 (83.03)	214 (82.95)
	Present	73 (5.16)	18 (6.98)
	N/A	167 (11.81)	26 (10.08)
Calcifications, n (%)			
	No echogenic foci	1134 (80.20)	179 (69.38)
	Microcalcifications	106 (7.50)	24 (9.30)
	Macrocalcifications	117 (8.27)	40 (15.50)
	Peripheral calcifications	15 (1.06)	10 (3.88)
	Microcalcifications with comet-tail artifacts	22 (1.56)	5 (1.94)
	Punctate echogenic foci of undetermined significance	20 (1.41)	0 (0)
	N/A	0 (0)	0 (0)
Internal blood flow, n (%)			
	Absent	163 (11.53)	21 (8.14)
	Present	1183 (83.66)	227 (87.98)

Characteristics		FTA[a]	FTC[b]
	N/A	68 (4.81)	10 (3.88)
Vascularity, n (%)			
	Mainly central vascularity	69 (4.88)	17 (6.59)
	Mainly peripheral vascularity	420 (29.70)	64 (24.81)
	Mixed vascularity	578 (40.88)	143 (55.43)
	Avascularity	4 (0.28)	0 (0)
	N/A	343 (24.26)	34 (13.18)
Trabecular formation, n (%)			
	Absent	1224 (86.56)	227 (87.98)
	Present	30 (2.12)	15 (5.81)
	N/A	160 (11.32)	16 (6.20)
Nodule-in-nodule appearance, n (%)			
	Absent	1231 (87.06)	227 (87.98)
	Present	23 (1.63)	15 (5.81)
	N/A	160 (11.32)	16 (6.20)
Mean diameter			
	Mean (SD) (cm)	2.30 (1.17)	2.94 (1.39)
	N/A, n (%)	0 (0)	0 (0)

[a]FTA: follicular thyroid adenoma.

[b]FTC: follicular thyroid carcinoma.

[c]N/A: not available data.

Table . Characteristics of the study population.

Characteristics		FTA[a]	FTC[b]
Hashimoto thyroiditis, n (%)			
	Absent	854 (66.30)	129 (51.39)
	Present	357 (27.72)	83 (33.07)
	N/A[c]	77 (5.98)	39 (15.54)
Sex, n (%)			
	Male	286 (22.20)	56 (22.31)
	Female	943 (73.21)	183 (72.91)
	N/A	59 (4.58)	12 (4.78)
Age (years)			
	Mean (SD)	47.89 (14.05)	48.47 (14.68)
	N/A, n (%)	1 (0.08)	8 (3.19)
BMI (kg/m ^{2})			
	Mean (SD)	24.08 (3.64)	24.67 (3.70)
	N/A, n (%)	59 (4.58)	5 (1.99)
Thyroid-stimulating hormone (μIU/mL)			
	Mean (SD)	1.76 (1.83)	1.99 (1.55)
	N/A, n (%)	325 (25.23)	90 (35.86)
Free triiodothyronine (pg/mL)			
	Mean (SD)	3.27 (0.66)	3.32 (0.67)
	N/A, n (%)	326 (25.31)	85 (33.86)
Free thyroxine (ng/dL)			
	Mean (SD)	1.27 (0.20)	1.26 (0.27)
	N/A, n (%)	325 (25.23)	85 (33.86)

[a]FTA: follicular thyroid adenoma.

[b]FTC: follicular thyroid carcinoma.

[c]N/A: not available data.

Model Performance in Discrimination and Calibration

We compared performance among 5 models (logistic regression, LASSO regression, random forest, extreme gradient boosting, and support vector machine) using the AUROC and AUPRC. As shown in Multimedia Appendices 3 and 4, the random forest model performed better in both AUROC and AUPRC than the other 4 models. With comprehensive consideration of the discrimination, calibration, parsimony, and interpretability of models, we selected the random forest model as the optimal.

We developed a random forest model with a total of 22 features: age, sex, BMI, Hashimoto thyroiditis, thyroid hormones (TSH, free triiodothyronine, and free thyroxine), ultrasonic predictors (mean diameter, composition, echogenicity, taller-than-wide, margin, calcifications, halo, internal blood flow, vascularity, trabecular formation, and nodule-in-nodule appearance), and TSH-related variables (mean TSH score, time-adjusted root mean square of successive differences of TSH, mean TSH, and coefficient of variation of TSH). After 5-fold cross-validation, the AUROC of the prediction model was 0.79 (95% CI 0.77-0.81) and the AUPRC was 0.40 (95% CI 0.37-0.44). When the threshold is gradually lowered from 50%, 40%, 30%, 20%, and finally to 10%, the accuracy, specificity, and precision decreased step by step while the sensitivity increased progressively (Multimedia Appendix 5). The calibration slope and calibration-in-the-large were 1.16 and 0.13, respectively (Multimedia Appendix 6).

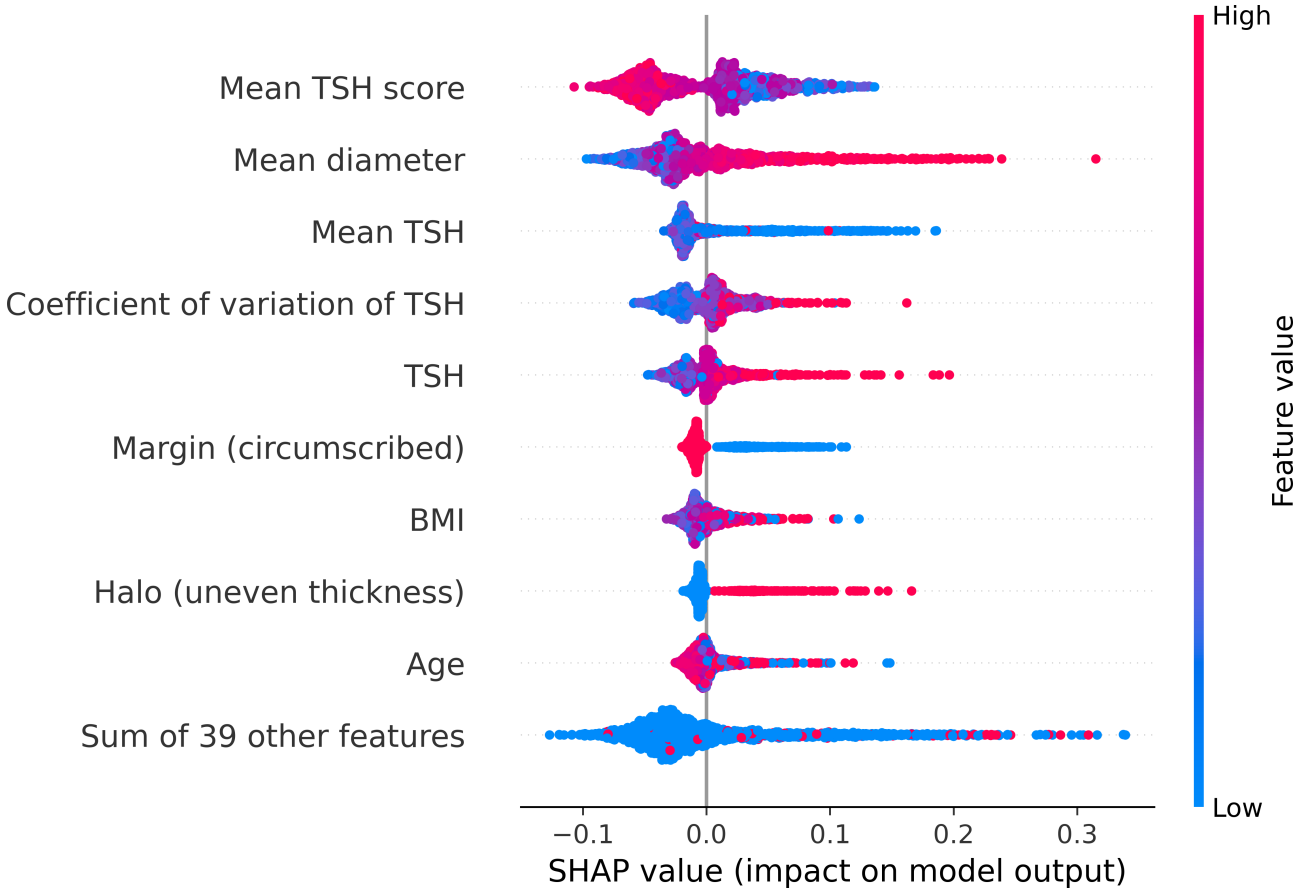
In addition, we implemented both oversampling and undersampling techniques to handle the imbalance in our models. However, following oversampling, the AUROC and AUPRC were 0.76 and 0.37, respectively, while after undersampling, the AUROC and AUPRC were 0.77 and 0.39, respectively. Notably, the model performed better before applying these sampling methods, with an AUROC of 0.79 and an AUPRC of 0.40.

Model Performance in Interpretation

The top 5 predictors were the mean TSH score, mean tumor diameter, mean TSH, coefficient of variation of TSH, and TSH level (Figure 2). The 5 top predictors did not show explicit

interactions with the sex or other predictors (Multimedia Appendix 7). Due to the strong correlation between TSH level and mean TSH score (correlation coefficient>0.6), we only plotted the partial dependence and individual conditional expectation plots for the top 4 features, excluding the TSH level. The associations between those 4 top continuous features and prediction probability were nonlinear (Multimedia Appendix 8). In general, the risk of malignancy tended to rise as the mean tumor diameter or the coefficient of variation of TSH increased, and the risk of malignancy tended to decrease as the mean TSH score or the mean TSH increased.

Figure 2. SHAP summary plot. SHAP: Shapley Additive Explanations; TSH: thyroid-stimulating hormone.



Moreover, we compared the characteristics of FTC groups with incorrect and correct predictions. FTC predicted as FTA by the model was classified into the incorrect-predicted group, while FTC predicted as malignant correctly by the model was then classified into the correct group. The mean diameter of the tumor was smaller in the incorrect-predicted group compared to the correct-predicted group (incorrect vs correct mean 2.88, SD cm 1.38 vs mean 3.71, SD 1.36 cm; mean diameter W=1474.5; P=.02).

A Systematic Review of the Previous Studies

After screening citations, we eventually included 7 studies in this systematic review (refer to Figure 3). The characteristics of the included studies are presented in Table 3. The sample sizes of the studies ranged from 18 to 888 patients [13-19]. The ratio of FTA to FTC in previous studies varied from 0.64 to 4.00 [13-19], which was much smaller than the ratio observed in our study (5.50) and in the real population, where the ratio of FTA to FTC can be as high as 5:1 [3]. In total, 3 studies even set the ratio close to 1 to address the imbalance [15,17,19]. As for the model selection, 4 studies developed deep learning models [13,16,17,19], 1 study used a random forest model [15], and the other 2 studies only established linear regression models [14,18], without concerning nonlinear associations or complicated interactions. Previous studies did not use gene mutations and other biomarkers as predictive variables. Except for 1 study from South Korea, which reported an AUROC of just 0.612 [13], the AUROC of the models in the other studies ranged from 0.75 to 0.96. However, none of them used AUPRC as a metric to assess discrimination. As for the interpretation of the models, Lin et al [15] assessed the feature importance, Tang et al [14] drew a nomogram, Li et al [18] developed F-TIRADS scoring criteria, and Yang et al [17] drew a heat map to visualize the importance of pixel regions, but the other 3 studies did not further explore interpretability, including feature importance, or the linear and nonlinear associations and interactions between features and targets.

Figure 3. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) study flow diagram. FTN: follicular thyroid neoplasm.

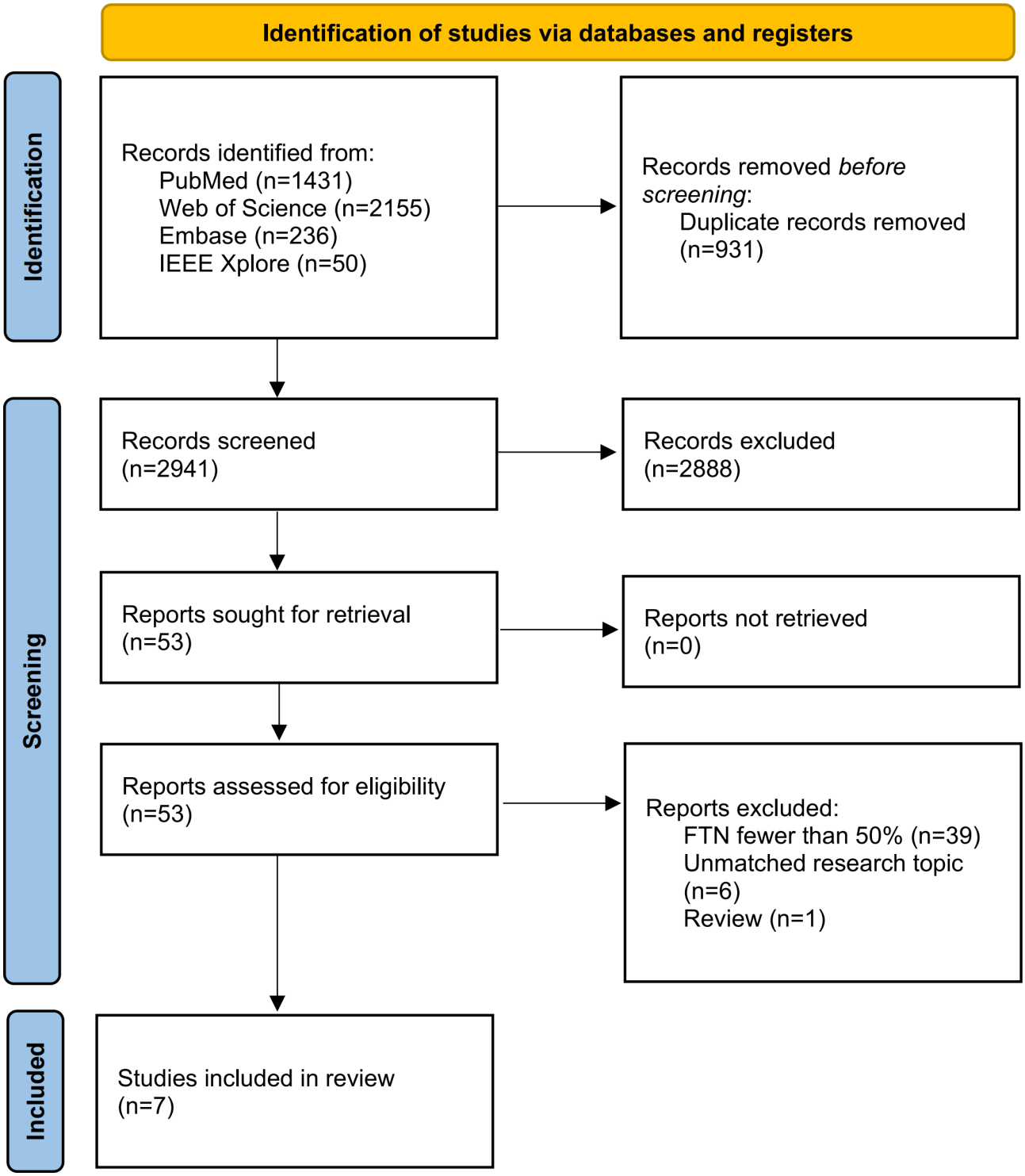


Table . Characteristics of the included studies.

Study	Country	Pathological ratio (FTA[a]/FTC[b]~ratio)	Features	Model	External test set	Sample size (n, patients)	Discrimination	Interpretation
Seo et al (2017) [16]	Korea	250/83~3.01	Ultrasound image	CNN[c]	No	Training set: 78; validation set: 229	Sensitivity: 71.05%; specificity: 93.19%; precision: 89.52%; AUROC[d]: 0.8088	No
Shin et al (2020) [13]	Korea	252/96~2.63	Ultrasound image	ANN[e] and SVM[f]	No	Training set: 340; validation set: leave-one-out cross-validation	ANN: precision: 74.1%; sensitivity: 32.3%; specificity: 90.1%; AUROC: 0.612; SVM: precision: 69%; sensitivity: 41.7%; specificity: 79.4%; AUROC: 0.605	No
Yang et al (2020)[g] [19]	China	Number of images: training set: 340/324~1.05; validation set: 85/81~1.05; additional test set: 154/146~1.05	Ultrasound image	CNN	No	Training set: 664 images; validation set: 166 images; test set: 300 images	Sensitivity: 95.89%; specificity: 96.10%; precision: 96%; AUROC:0.96	No
Tang et al (2021) [14]	China	112/28~4.00	Computed tomography features and clinical features and hormone level	LASSO[h] regression	No	Training set: 140; validation set: 60	Sensitivity: 92.9%; specificity: 77.7%; precision: 80%; AUROC: 0.913 (95% CI 0.850 - 0.975)	Nomogram
Li et al (2023) [18]	China	Training set: 515/188~2.74; validation set: 122/33~3.70	Ultrasound features	LASSO regression and logistic regression	No	Pretraining set: 30; training set: 703; validation set: 155	LASSO regression: sensitivity: 66%; specificity: 72%; precision: 71%; AUROC: 0.76 (95% CI 0.72 - 0.79); Logistic regression: sensitivity: 64%; specificity: 75%; precision: 72%; AUROC: 0.75 (95% CI 0.71 - 0.79)	F-TIRADS[i] scoring criteria

Study	Country	Pathological ratio (FTA[a]/FTC[b]~ratio)	Features	Model	External test set	Sample size (n, patients)	Discrimination	Interpretation
Yang et al (2023) [17]	China	Training set: 705/687~1.02; validation set: 177/172~1.03; external test set: 150/159~0.94	Ultrasound image	CNN	Yes	Training set: 352; validation set: 80; external test set: 71	Sensitivity: 66.7%; specificity: 79.6%; precision: 73%; AUROC: 0.81 (95% CI, 0.76 - 0.86)	Heat map
Lin et al (2024) [15]	United States	7/11~0.64	Ultrasound image features and clinical features	Random forest	No	Training set: 18; validation set: leave-one-out cross-validation	Sensitivity: 100%; specificity: 43%; AUROC: 0.792	Feature importance

[a]FTA: follicular thyroid adenoma.

[b]FTC: follicular thyroid carcinoma.

[c]CNN: convolutional neural network.

[d]AUROC: area under the receiver operating characteristic curve.

[e]ANN: artificial neural network.

[f]SVM: support vector machine.

[g]Yang et al (2020) [19] did not report the numbers of patients or nodules.

[h]LASSO: least absolute shrinkage and selection operator.

[i]F-TIRADS: Thyroid Imaging Reporting and Data System for Follicular Neoplasm.

Additionally, we tested the F-TIRADS scoring criteria developed by Li et al [18] with our dataset. The criteria specify 6 key features for scoring. After filtering our data to include only cases with complete information for these 6 features, we selected 1025 tumors from 993 patients as an external test set. When applying the F-TIRADS scoring criteria, the predictive performance was suboptimal. With a threshold for FTC risk set at >90%, the model achieved an accuracy of 0.82, sensitivity of 0.04, specificity of 0.99, and precision of 0.53. When using a >50% FTC risk threshold, the sensitivity increased to 0.27, while accuracy, specificity, and precision decreased to 0.79, 0.91, and 0.39, respectively. For threshold-independent metrics, the AUROC and AUPRC for the F-TIRADS scoring criteria were 0.47 and 0.59, respectively, in our external test set.

Discussion

Principal Findings

This study systematically established the interpretable machine learning–based model to address the challenge of clinical decision-making for the FTN before surgery. We developed a model using 22 readily available predictors with a preferable AUROC (0.79, 95% CI 0.77 - 0.81). Additionally, the model demonstrated excellent interpretability, identifying the mean TSH score, mean tumor diameter, mean TSH, coefficient of variation of TSH, and TSH level as the most important predictors. After comparing groups of incorrect-predicted and correct-predicted FTC, we found that smaller FTCs were more likely to be misclassified as FTA.

Comparison to Prior Work

It is crucial to evaluate the performance of clinical prediction models comprehensively, that is, the models should be well performed in discrimination and calibration. Concerning the discrimination, our developed model was comparable to that of previous studies aimed at predicting the malignancy risk of FTN before surgical treatment. For example, according to Li et al [18], the AUROC reached 0.76 in the LASSO regression model consisting of ultrasound features (the ratio of FTA to FTC: training set: 2.74 and validation set: 3.70); also, in the LASSO regression model, the AUROC reached 0.913 in discriminating FTA from FTC on selected clinical parameters, computed tomography signs, and radiomic features referring to Tang et al [14] (the ratio of FTA to FTC: 4.00). However, neither study used AUPRC as the evaluation metric. We also acknowledge that the model was derived from extremely imbalanced data (the ratio of FTA to FTC: 5.50), and in this context, the AUPRC metric for assessing model performance is more informative and intuitive than the AUROC [37]. For example, the prediction model might perform relatively well when measured by AUROC but may perform unsatisfactorily when measured by AUPRC, in the scenario of imbalanced data. Furthermore, to address the imbalance in our models, we applied oversampling and undersampling techniques, but both failed to improve performance. Our findings were in line with the previous research, which found that oversampling and undersampling generally did not enhance prediction models in large observational health datasets [39]. The possible reasons might be as follows: (1) oversampling and undersampling would modify the outcome proportions in the training data, leading to miscalibration, such as overestimated risks [39]; (2) the synthetic data generated by oversampling may not accurately represent

the original distribution of minority class, potentially affecting classification performance [40]; and (3) undersampling reduced the number of majority class samples, limiting the model's ability to fully use the features of the majority class during training [41].

As one previous systematic review indicated, calibration was commonly overlooked during the development and validation of clinical prediction models [20]. However, calibration metrics are also important to assess the size of the gap between the predicted risk probability and the true risk probability. For instance, grouping can be manipulated to obscure the evaluation of miscalibration in a particular range without a calibration curve and its numerical quantification [42]. In general, our model had relatively good calibration, as the calibration slope was close to 1 and the calibration-in-the-large close to 0.

Based on the results of our systematic review, most of the previous studies show relatively satisfying AUROC, but none of them reported AUPRC. Although Li et al [18] reported a handy score-risking tool for clinicians to assess the malignancy risk of FTN at the diagnosis stage, this tool seemed to not perform ideally in the practice of our data (AUROC 0.47; AUPRC 0.59; sensitivity 0.04 [threshold 90%] and 0.27 [threshold 50%]).

Limitations and Strengths

We should interpret the study findings cautiously. As with other single-center studies, the results from this study were limited in generalizability to patients and clinical settings with distinct characteristics. However, the pathological diagnosis of FTC was highly heterogeneous across different clinical settings due to its challenge in sufficient sampling and accurate diagnosis. Therefore, we advocated for the standardization of FTC diagnosis before the conduction of a multicenter study soon. Additionally, the prediction performance of models, comparable to the previous work with similar predictors, had room to further improve. The clinical utility of the screening stage was also less than ideal. Building on the experiences and lessons learned from this study, we are conducting a prospective cohort study to further optimize the model performance through collecting other costly multidimensional predictors including genomics, ultrasound images, and videos. Besides, our study was retrospective in nature, which may introduce selection bias. Furthermore, we excluded patients with nodules of UMPs, potentially limiting the model's accuracy in identifying borderline follicular tumors. Moreover, our models did not incorporate other potential predictors, such as genetic markers (eg, BRAF, TRET, and RAS mutations), computed tomography or magnetic resonance imaging characteristics, or family history, due to constraints in data availability.

Our study had several strengths. Our models were advantageous in the large sample size for the present topic, the clinically easy-accessible and clinician-validated predictors, and the comprehensive evaluation with the metrics appropriate for the nature of the data (imbalanced data) [43]. Furthermore, the disease distribution of FTA and FTC in the study population was fully consistent with that of patients with FTN in real-world settings, that is, we did not deliberately over- or undersample patients with any type of disease in the model development, as commonly seen in previous studies [17,18]. As such, findings from our study had theoretically better fidelity and generalizability in real-world settings. In addition, we conducted a systematic review to synthesize findings from previous studies, comprehensively integrate the evidence, and identify research gaps.

Future Directions

Our study paved the way for future research in terms of predictors, models, and targets. Concerning predictors and models, further studies might consider taking advantage of the rapidly developing deep learning models and fully using high-dimension predictors such as ultrasound images and genomics. In terms of targets, it is important to standardize the pathological diagnosis of FTC across multiple centers before conducting a future multicenter study.

Our study is also important for future clinical practice. First, findings from the interpretation of our models indicate that clinicians should comprehensively consider patients' variables such as thyroid hormones in addition to the ultrasound results. Second, in a natural distribution population with severely unbalanced data (FTA is far more than FTC), preoperative prediction of FTA and FTC by thyroid hormone and ultrasound features alone may face challenges, especially for relatively small-sized FTCs, which are easy to miss detection.

Conclusions

In clinical practice, it remained challenging to sensitively screen, precisely diagnose, and appropriately treat patients with FTN. Interpretation of our developed machine learning–based model suggests that clinicians should also pay attention to patients' variables such as TSH along with tumor size. However, it may be hard to correctly predict FTNs preoperatively with thyroid hormone and ultrasound features alone, especially for FTCs with small sizes. The findings of our study bridged the gaps of previous work and paved the way for connecting machine learning to interpretation in the field of FTN research. We call for subsequent studies to further examine the generalizability to other contexts.

Acknowledgments

This work used the data of the Peking University Third Hospital. The authors sincerely thank the research team of Thyroid Tumor Multi-Disciplinary Treatment. We appreciate the health professionals at Peking University Third Hospital for data collection and management. Funding support was received from the National Natural Science Foundation of China (82373694), Young Elite Scientists Sponsorship Program by CAST (2023QNRC001), Beijing Education Sciences Planning Program during the 14th

Five-Year Plan (BECA23111), Proof of Concept Program of Zhongguancun Science City and Peking University Third Hospital (HDCXZHKC2022210), and the Fundamental Research Funds for the Central Universities (BMU2021YJ030).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ZL and XL contributed to the conception. Both XL and ZL contributed equally to this work as corresponding authors. RS, XL, and ZL were responsible for data analysis and interpretation. RS contributed to manuscript writing, and ZL contributed to the critical revision of the manuscript. ZC, YJC, BH, JPH, GLK, HL, FM, SBS, BKS, HT, YW, XYY, JMY, BY, CHY, and FZ provided study materials or patients. All authors participated in the collection and assembly of data. All authors reviewed and approved the final manuscript. All authors are accountable for all aspects of the work. Coauthors, from JC to FZ, are listed alphabetically by last name.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information for methods, definitions, and scoring criteria.
[DOCX File, 476 KB - cancer_v11i1e66269_app1.docx]

Multimedia Appendix 2
Search strategy and eligibility criteria.
[DOCX File, 17 KB - cancer_v11i1e66269_app2.docx]

Multimedia Appendix 3
Comparison between 5 different models.
[PDF File, 731 KB - cancer_v11i1e66269_app3.pdf]

Multimedia Appendix 4
Performance of 5 different models.
[DOCX File, 23 KB - cancer_v11i1e66269_app4.docx]

Multimedia Appendix 5
Area under the curve and confusion matrix (true positive, true negative, false positive, and false negative).
[DOCX File, 20 KB - cancer_v11i1e66269_app5.docx]

Multimedia Appendix 6
Calibration plot.
[PDF File, 94 KB - cancer_v11i1e66269_app6.pdf]

Multimedia Appendix 7
Shapley Additive Explanations interaction dependence plot.
[PDF File, 809 KB - cancer_v11i1e66269_app7.pdf]

Multimedia Appendix 8
Partial dependence plot and individual conditional expectation plot.
[PDF File, 1331 KB - cancer_v11i1e66269_app8.pdf]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 32 KB - cancer_v11i1e66269_app9.docx]

References

1. Boucai L, Zafereo M, Cabanillas ME. Thyroid cancer: a review. JAMA 2024 Feb 6;331(5):425-435. [doi: 10.1001/jama.2023.26348] [Medline: 38319329]
2. Grani G, Lamartina L, Durante C, Filetti S, Cooper DS. Follicular thyroid cancer and Hürthle cell carcinoma: challenges in diagnosis, treatment, and clinical management. Lancet Diabetes Endocrinol 2018 Jun;6(6):500-514. [doi: 10.1016/S2213-8587(17)30325-X] [Medline: 29102432]
3. McHenry CR, Phitayakorn R. Follicular adenoma and carcinoma of the thyroid gland. Oncologist 2011;16(5):585-593. [doi: 10.1634/theoncologist.2010-0405] [Medline: 21482585]
4. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid cancer. Thyroid 2016 Jan;26(1):1-133. [doi: 10.1089/thy.2015.0020]
5. Durante C, Hegedüs L, Czarniecka A, et al. 2023 European Thyroid Association Clinical Practice Guidelines for thyroid nodule management. Eur Thyroid J 2023 Oct 1;12(5):37358008. [doi: 10.1530/ETJ-23-0067] [Medline: 37358008]
6. Zachariah FJ, Rossi LA, Roberts LM, Bosserman LD. Prospective comparison of medical oncologists and a machine learning model to predict 3-month mortality in patients with metastatic solid tumors. JAMA Netw Open 2022 May 2;5(5):e2214514. [doi: 10.1001/jamanetworkopen.2022.14514] [Medline: 35639380]
7. Farrokhian N, Holcomb AJ, Dimon E, et al. Development and validation of machine learning models for predicting occult nodal metastasis in early-stage oral cavity squamous cell carcinoma. JAMA Netw Open 2022 Apr 1;5(4):e227226. [doi: 10.1001/jamanetworkopen.2022.7226] [Medline: 35416990]
8. Tseng YJ, Wang HY, Lin TW, Lu JJ, Hsieh CH, Liao CT. Development of a machine learning model for survival risk stratification of patients with advanced oral cancer. JAMA Netw Open 2020 Aug 3;3(8):e2011768. [doi: 10.1001/jamanetworkopen.2020.11768] [Medline: 32821921]
9. Bennett M, Hayes K, Kleczyk EJ, Mehta R. Similarities and differences between machine learning and traditional advanced statistical modeling in healthcare analytics. arXiv. Preprint posted online on Jan 7, 2022. [doi: 10.48550/arXiv.2201.02469]
10. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina (Kaunas) 2020 Sep 8;56(9):455. [doi: 10.3390/medicina56090455] [Medline: 32911665]
11. Bzdok D, Krzywinski M, Altman N. Points of significance: machine learning: a primer. Nat Methods 2017 Nov 30;14(12):1119-1120. [doi: 10.1038/nmeth.4526] [Medline: 29664466]
12. Carmichael I, Marron JS. Data science vs. statistics: two cultures? Jpn J Stat Data Sci 2018 Jun;1(1):117-138. [doi: 10.1007/s42081-018-0009-3]
13. Shin I, Kim YJ, Han K, et al. Application of machine learning to ultrasound images to differentiate follicular neoplasms of the thyroid gland. Ultrasonography 2020 Jul;39(3):257-265. [doi: 10.14366/usg.19069] [Medline: 32299197]
14. Tang P, Ren C, Shen L, Zhou Z. Development and validation of a diagnostic nomogram for the preoperative differentiation between follicular thyroid carcinoma and follicular thyroid adenomas. J Comput Assist Tomogr 2021;45(1):128-134. [doi: 10.1097/RCT.0000000000001078] [Medline: 33475318]
15. Lin AC, Liu Z, Lee J, et al. Generating a multimodal artificial intelligence model to differentiate benign and malignant follicular neoplasms of the thyroid: a proof-of-concept study. Surgery 2024 Jan;175(1):121-127. [doi: 10.1016/j.surg.2023.06.053] [Medline: 37925261]
16. Seo JK, Kim YJ, Kim KG, Shin I, Shin JH, Kwak JY. Differentiation of the follicular neoplasm on the gray-scale US by image selection subsampling along with the marginal outline using convolutional neural network. Biomed Res Int 2017;2017:3098293. [doi: 10.1155/2017/3098293] [Medline: 29527533]
17. Yang Z, Yao S, Heng Y, et al. Automated diagnosis and management of follicular thyroid nodules based on the devised small-dataset interpretable foreground optimization network deep learning: a multicenter diagnostic study. Int J Surg 2023 Sep 1;109(9):2732-2741. [doi: 10.1097/JS9.0000000000000506] [Medline: 37204464]
18. Li J, Li C, Zhou X, et al. US risk stratification system for follicular thyroid neoplasms. Radiology 2023 Nov;309(2):e230949. [doi: 10.1148/radiol.230949] [Medline: 37987664]
19. Yang B, Yan M, Yan Z, Zhu C, Xu D, Dong F. Segmentation and classification of thyroid follicular neoplasm using cascaded convolutional neural network. Phys Med Biol 2020 Dec 22;65(24):33137800. [doi: 10.1088/1361-6560/abc6f2] [Medline: 33137800]
20. Yang C, Kors JA, Ioannou S, et al. Trends in the conduct and reporting of clinical prediction model development and validation: a systematic review. J Am Med Inform Assoc 2022 Apr 13;29(5):983-989. [doi: 10.1093/jamia/ocac002] [Medline: 35045179]
21. Paparodis R, Imam S, Todorova-Koteva K, Staii A, Jaume JC. Hashimoto's thyroiditis pathology and risk for thyroid cancer. Thyroid 2014 Jul;24(7):1107-1114. [doi: 10.1089/thy.2013.0588] [Medline: 24708347]
22. MDT team of thyroid and parathyroid tumor [Article in Chinese]. Peking University Third Hospital. URL: https://www.puh3.net.cn/info/2891/118191.htm [accessed 2024-04-15]
23. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 7;350:g7594. [doi: 10.1136/bmj.g7594] [Medline: 25569120]

24. Baloch ZW, Asa SL, Barletta JA, et al. Overview of the 2022 WHO classification of thyroid neoplasms. Endocr Pathol 2022 Mar;33(1):27-63. [doi: 10.1007/s12022-022-09707-3] [Medline: 35288841]
25. Li X, Xiao WC, Mei F, et al. The association of pregnancy with disease progression in patients previously treated for differentiated thyroid cancer: a propensity score-matched retrospective cohort study. J Womens Health (Larchmt) 2023 Nov;32(11):1174-1181. [doi: 10.1089/jwh.2023.0172] [Medline: 37638829]
26. Xiao WC, Li X, Shan R, et al. Pregnancy and progression of differentiated thyroid cancer: a propensity score-matched retrospective cohort study. J Clin Endocrinol Metab 2024 Feb 20;109(3):837-843. [doi: 10.1210/clinem/dgad557] [Medline: 37738427]
27. Shan R, Li X, Tao M, et al. Pregnancy and the disease recurrence of patients previously treated for differentiated thyroid cancer: a systematic review and meta analysis. Chin Med J (Engl) 2024 Mar 5;137(5):547-555. [doi: 10.1097/CM9.0000000000003008] [Medline: 38311812]
28. Mincer DL, StatPearls J. Treasure Island (FL) Ineligible Companies Disclosure: Ishwarlal Jialal Declares No Relevant Financial Relationships with Ineligible Companies: StatPearls Publishing LLC; 2024.
29. Ito Y, Miyauchi A, Fujishima M, et al. Thyroid-stimulating hormone, age, and tumor size are risk factors for progression during active surveillance of low-risk papillary thyroid microcarcinoma in adults. World J Surg 2023 Feb;47(2):392-401. [doi: 10.1007/s00268-022-06770-z] [Medline: 36182976]
30. Taquet M, Griffiths K, Palmer EOC, et al. Early trajectory of clinical global impression as a transdiagnostic predictor of psychiatric hospitalisation: a retrospective cohort study. Lancet Psychiatry 2023 May;10(5):334-341. [doi: 10.1016/S2215-0366(23)00066-4] [Medline: 36966787]
31. Li X, Fu P, Xiao WC, et al. Associations of gestational thyrotropin levels with disease progression among pregnant women with differentiated thyroid cancer: a retrospective cohort study. Front Endocrinol (Lausanne) 2024;15:1369344. [doi: 10.3389/fendo.2024.1369344] [Medline: 39493771]
32. Kotthoff L, Sonabend R, Foss N, Bischl B. Introduction and overview. In: Bischl B, Sonabend R, Kotthoff L, Lang M, editors. Applied Machine Learning Using mlr3 in R: CRC Press; 2024. [doi: 10.1201/9781003402848-1]
33. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
34. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020 Jan;2(1):56-67. [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
35. Mukhopadhyay S. InfoGram and admissible machine learning. Mach Learn 2022 Jan;111(1):205-242. [doi: 10.1007/s10994-021-06121-4]
36. Mohammed R, Rawashdeh J, Abdullah M. Machine learning with oversampling and undersampling techniques: overview study and experimental results. Presented at: 2020 11th International Conference on Information and Communication Systems (ICICS); Apr 7-9, 2020; Irbid, Jordan. [doi: 10.1109/ICICS49469.2020.239556]
37. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015;10(3):e0118432. [doi: 10.1371/journal.pone.0118432] [Medline: 25738806]
38. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339:b2700. [doi: 10.1136/bmj.b2700] [Medline: 19622552]
39. Yang C, Fridgeirsson EA, Kors JA, Reps JM, Rijnbeek PR. Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data. J Big Data 2024;11(1):7. [doi: 10.1186/s40537-023-00857-7]
40. Elreedy D, Atiya AF, Kamalov F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. Mach Learn 2024 Jul;113(7):4903-4923. [doi: 10.1007/s10994-022-06296-4]
41. Liu XY, Wu J, Zhou ZH. Exploratory undersampling for class-imbalance learning. IEEE Trans Syst Man Cybern B Cybern 2009 Apr;39(2):539-550. [doi: 10.1109/TSMCB.2008.2007853] [Medline: 19095540]
42. Riley RD, Archer L, Snell KIE, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. BMJ 2024 Jan 15;384:e074820. [doi: 10.1136/bmj-2023-074820] [Medline: 38224968]
43. Hashimoto DA, Varas J, Schwartz TA. Practical guide to machine learning and artificial intelligence in surgical education research. JAMA Surg 2024 Apr 1;159(4):455-456. [doi: 10.1001/jamasurg.2023.6687] [Medline: 38170510]

Abbreviations:

AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
F-TIRADS: Thyroid Imaging Reporting and Data System for Follicular Neoplasm
FTA: follicular thyroid adenoma
FTC: follicular thyroid carcinoma
FTN: follicular thyroid neoplasm
LASSO: least absolute shrinkage and selection operator

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SHAP: Shapley Additive Explanations
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
TSH: thyroid-stimulating hormone
UMP: follicular tumor of uncertain malignant potential

Edited by N Cahill; submitted 09.09.24; peer-reviewed by J Lai, K Matthew; revised version received 03.12.24; accepted 04.12.24; published 10.02.25.

Please cite as:
Shan R, Li X, Chen J, Chen Z, Cheng YJ, Han B, Hu RZ, Huang JP, Kong GL, Liu H, Mei F, Song SB, Sun BK, Tian H, Wang Y, Xiao WC, Yao XY, Ye JM, Yu B, Yuan CH, Zhang F, Liu Z
Interpretable Machine Learning to Predict the Malignancy Risk of Follicular Thyroid Neoplasms in Extremely Unbalanced Data: Retrospective Cohort Study and Literature Review
JMIR Cancer 2025;11:e66269
URL: https://cancer.jmir.org/2025/1/e66269
doi:10.2196/66269

Original Paper

Next-Generation Sequencing–Based Testing Among Patients With Advanced or Metastatic Nonsquamous Non–Small Cell Lung Cancer in the United States: Predictive Modeling Using Machine Learning Methods

Alan James Michael Brnabic[1], MS; Ilya Lipkovich[2], PhD; Zbigniew Kadziola[2], MS; Dan He[3], MS; Peter M Krein[2], PhD; Lisa M Hess[2], PhD

[1]Eli Lilly and Company, Sydney, Australia

[2]Eli Lilly and Company, Indianapolis, IN, United States

[3]Syneos Health, Morrisville, NC, United States

Corresponding Author:
Lisa M Hess, PhD
Eli Lilly and Company
LCC Corporate Center
Indianapolis, IN, 46285
United States
Phone: 1 317 908 1872
Email: hess_lisa_m@lilly.com

Abstract

Background: Next-generation sequencing (NGS) has become a cornerstone of treatment for lung cancer and is recommended in current treatment guidelines for patients with advanced or metastatic disease.

Objective: This study was designed to use machine learning methods to determine demographic and clinical characteristics of patients with advanced or metastatic non–small cell lung cancer (NSCLC) that may predict likelihood of receiving NGS-based testing (ever vs never NGS-tested) as well as likelihood of timing of testing (early vs late NGS-tested).

Methods: Deidentified patient-level data were analyzed in this study from a real-world cohort of patients with advanced or metastatic NSCLC in the United States. Patients with nonsquamous disease, who received systemic therapy for NSCLC, and had at least 3 months of follow-up data for analysis were included in this study. Three strategies, logistic regression models, penalized logistic regression using least absolute shrinkage and selection operator penalty, and extreme gradient boosting with classification trees as base learners, were used to identify predictors of ever versus never and early versus late NGS testing. Data were split into D1 (training+validation; 80%) and D2 (testing; 20%) sets; the 3 strategies were evaluated by comparing their performance on multiple m=1000 splits in the training (70%) and validation data (30%) within the D1 set. The final model was selected by evaluating performance using the area under the receiver operating curve while taking into account considerations of simplicity and clinical interpretability. Performance was re-estimated using the test data D2.

Results: A total of 13,425 met the criteria for the ever NGS-tested, and 17,982 were included in the never NGS-tested group. Performance metrics showed the area under the receiver operating curve evaluated from validation data was similar across all models (77%-84%). Among those in the ever NGS-tested group, 84.08% (n=11,289) were early NGS-tested, and 15.91% (n=2136) late NGS-tested. Factors associated with both ever having NGS testing as well as early NGS testing included later year of NSCLC diagnosis, no smoking history, and evidence of programmed death ligand 1 testing (all P<.05). Factors associated with a greater chance of never receiving NGS testing included older age, lower performance status, Black race, higher number of single-gene tests, public insurance, and treatment in a geography with Molecular Diagnostics Services Program adoption (all P<.05).

Conclusions: Predictors of ever versus never as well as early versus late NGS testing in the setting of advanced or metastatic NSCLC were consistent across machine learning methods in this study, demonstrating the ability of these models to identify factors that may predict NGS-based testing. There is a need to ensure that patients regardless of age, race, insurance status, and geography (factors associated with lower odds of receiving NGS testing in this study) are provided with equitable access to NGS-based testing.

(JMIR Cancer 2025;11:e64399) doi:10.2196/64399

KEYWORDS

lung cancer; NGS testing; next-generation sequencing; real-world data; machine learning; biomarkers; predictive modeling; artificial intelligence; treatment guidelines; tumor biomarker; oncology

Introduction

The care of patients with non–small cell lung cancer (NSCLC) has changed dramatically since the early 2010s, from a chemotherapy-based approach that was tailored only to the disease histology (squamous or nonsquamous tumors) to becoming a disease with multiple actionable biomarkers that can identify targeted therapies associated with superior outcomes based on individual patient genomic characteristics [1,2]. This has led to the adoption of next-generation sequencing (NGS) recommendations included in treatment guidelines for patients with NSCLC [3].

Unfortunately, despite these recommendations, multiple studies have shown that NGS-based testing is not being used for all patients with advanced or metastatic NSCLC, and only about half of all patients in some studies receive comprehensive biomarker testing [4-6]. The reasons for the lack of testing are unclear but may include barriers to ordering tests, insufficient tissue, clinical deterioration, or a crisis that requires immediate care [6]. More recent studies have also demonstrated a racial disparity in receipt of biomarker testing; patients who are Black are significantly less likely than those who are White to receive NGS-based testing in the United States [7].

Studies evaluating the barriers to testing have typically taken a specific hypothesis-driven a priori categorization of potential barriers to investigate the lack of testing [6,7]. While certainly this approach is critical to investigate specific issues such as racial disparities, this falls short when trying to evaluate the complexity of care and the multiple and potentially interacting factors. Clinical prediction models are an alternative approach to using patient-level evidence to help inform health care decision makers about patient care. These models have been used for decades by health care professionals [8]. Traditionally, prediction models combine patient demographic, clinical, and treatment characteristics in the form of a statistical or mathematical model, usually regression, classification, or neural networks, but deal with a limited number of predictor variables (usually below 25). Flexible machine learning methods can be used, by which the researcher does not force the model to evaluate a limited set of covariates, but rather the models themselves learn by trial and error from the data to make predictions, without having a predefined set of rules for decision-making. Simply, machine learning can be better understood as “learning from data” [9]. The setting of biomarker testing provides an opportunity to apply these methods to more thoroughly explore the factors that are associated with the lack of recommended biomarker testing.

While machine learning methods have been more commonly used for biomarker identification and treatment selection, there is little evidence of these methods applied to the prediction of biomarker testing itself. To date, the investigations surrounding the gaps in biomarker testing have remained largely limited to descriptive research and opinion pieces [10-13]. Therefore, this study was designed to fill this gap in evidence by applying machine learning methods to the question of biomarker testing for patients with advanced or metastatic nonsquamous NSCLC to determine demographic and clinical characteristics that may predict receipt of NGS-based testing. A second objective was to further determine the characteristics that predict receipt of NGS-based testing (early testing) in accordance with clinical guidelines that can inform first-line therapy (vs those who receive NGS-based testing after the first-line therapy is underway). These objectives were pursued to better understand factors associated with experiencing barriers to recommended testing and the timing of such testing to inform future intervention strategies.

Methods

Data Source

This study used the Advanced NSCLC Analytic Cohort from the nationwide Flatiron Health electronic health record–derived longitudinal database, comprising deidentified patient-level structured and unstructured data, curated via technology-enabled abstraction [14,15]. The data are deidentified and subject to obligations to prevent reidentification and protect patient confidentiality and are not considered human participants in accordance with the US Code of Federal Regulations [16]. These deidentified data originate from approximately 280 cancer clinics (~800 sites of care) in the United States. Patients in this database are those who have lung cancer *ICD* (*International Classification of Diseases*) codes 162.x (*ICD-9* [*International Classification of Diseases, Ninth Revision*]), C34x, or C39.9 (*ICD-10* [*International Statistical Classification of Diseases, Tenth Revision*]) on at least 2 documented clinical visits on different days occurring on or after January 1, 2011. Longitudinal patient-level data were available through November 2021. Patients must further have had pathology consistent with NSCLC and have advanced or metastatic disease (diagnosed with stage IIIB, IIIC, IVA, or IVB disease or diagnosed with early-stage NSCLC and subsequently developed recurrent or progressive disease).

Definitions of NGS Testing Cohorts

Patients were included in this analysis if they were in the Flatiron Health Advanced NSCLC Analytic Cohort, had nonsquamous NSCLC, evidence of receipt of systemic therapy, and at least 3 months of follow-up in the database. Receipt of testing by NGS is a field recorded in the electronic medical record database by the health care provider that was used for testing identification in this study. The method of NGS testing (tissue or circulating tumor) is not specified. Patients were excluded who had evidence of NGS-based testing more than 20 days prior to initial NSCLC diagnosis. Patients meeting the

inclusion criteria for this study were categorized into 2 groups. The ever NGS-tested group included patients with at least 1 NGS test recorded in the database. All remaining patients were included in the never NGS-tested group, as this group was comprised of patients with no evidence of any NGS test recorded in the database. Among those in the ever NGS-tested group, individuals were further subgrouped by the timing of NGS-based testing. Each patient in the ever NGS-tested group was either included in the early NGS-tested subgroup, including patients whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy, or the late NGS-tested subgroup, all remaining patients whose first NGS-based test occurred 8 days or later after the start of first-line therapy. The date of advanced or metastatic diagnosis was considered the index diagnosis date.

Candidate Predictors

Candidate predictors for receipt and timing of NGS-based testing were prespecified based on published literature, analyses of real-world data, and expert input from the field of cancer diagnostics [4,7,17]. These variables included patient age at advanced or metastatic diagnosis date (years), sex (male or female), race (Asian, Black, White, and other), insurance type (public, private, or other), Eastern Cooperative Oncology Group (ECOG) performance status (0-4), smoking history (ever vs never smoker), body weight (kilograms), BMI (kg/m^2), practice setting (academic or community), practice volume (the average number of those with NSCLC receiving care at the site where the included patient received care by index year over the period 2011 to 2021), biomarker result (positive, not positive, and not tested) by each available biomarker (anaplastic lymphoma kinase [ALK]; epidermal growth factor receptor [EGFR]; V-Raf murine sarcoma viral oncogene homolog B [BRAF]; Kirsten rat sarcoma virus [KRAS]; c-ros oncogene 1 [ROS1]; mesenchymal epithelial transition [MET]; neurotrophic tyrosine receptor kinase [NTRK]; rearranged during transfection [RET]; and programmed death ligand 1 [PD-L1]), stage of disease at initial diagnosis (0-IV), laboratory value (low, normal, high, or not tested) by blood test (alkaline phosphatase, alanine transaminase, aspartate transferase, bilirubin, creatinine, lymphocyte count, red blood cell count, hematocrit, platelet count, white blood cell count, and hemoglobin), number of non-NGS biomarker tests received (total number of fluorescence in situ hybridization, immunohistochemistry, polymerase chain reaction, or other non-NGS–based tests), as well as 2 variables to identify periods of environmental changes. The first of these variables categorized the status of National Comprehensive Cancer Network (NCCN) Clinical Guidelines: prior to 2016, before NGS was recommended in the guidelines; 2016-2019, when broad-based testing was recommended; and 2020 and later, when NGS-based testing was recommended [18]. The second variable evaluated the timing of US Food and Drug Administration approval of drugs that targeted the available biomarkers: period (1) January 1, 2011-August 25, 2011 (EGFR drugs only); period (2) August 26, 2011-March 10, 2016 (EGFR+ALK); period (3) March 11, 2016-June 21, 2017 (EGFR+ALK+ROS1); period (4) June 22, 2017-November 25, 2018 (EGFR+ALK+ROS1+BRAF); period (5) November 26, 2018-May 5, 2020 (EGFR+ALK+ROS1+BRAF+NTRK); period (6) May 6, 2020-May 26, 2021 (EGFR+ALK+ROS1+BRAF+NTRK+MET+RET); and period (7) May 27, 2021, and later (EGFR+ALK+ROS1+BRAF+MET+NTRK+RET+KRAS) [19]. Additionally, candidate predictors of Medicare Administrative Contractor (MAC) region [20] and Molecular Diagnostics Services (MolDX) Program adoption (yes or no) [21] were included. These variables explored the policies in place at the geography in which the patient received care. MACs are private companies that process claims for Medicare beneficiaries. These companies are geographically distinct and identifiable by unique alphanumeric designations (eg, J8=jurisdiction 8) and by private company names (eg, Noridian and Palmetto) [22]. The MolDX Program determines the coverage of diagnostic testing in 4 MACs across 28 states [20,21]. Importantly, all candidate predictor variables were required to be recorded prior to the end of the early NGS testing period to ensure that no covariates were recorded after the measurement of the NGS testing outcome.

The following interactions were deemed to be clinically relevant and forced into the models for evaluation: smoking and sex, smoking and NCCN guideline periods, race and insurance type, age and ECOG performance status, MAC region and public insurance, and MolDX region and public insurance. The estimates of the expected direction of these relationships were defined in the study protocol and are summarized in Table 1.

Table 1. Expected direction of candidate predictors for next-generation sequencing (NGS) testing.

Candidate predictor variable	Expected direction
Year of advanced or metastatic diagnosis	As year increases, NGS testing is more likely.
Smoking status (yes vs no)	Smoking=no, NGS testing is more likely.
Sex (male vs female)	Sex=female, NGS testing is more likely.
Race (Asian, Black, White, other)	Race=Asian or White, NGS testing is more likely.
Practice volume (continuous)	As practice volume increases, NGS testing is more likely.
BMI (using WHO[a] categories)	BMI=underweight, NGS testing is less likely.
ECOG[b] performance status (0, 1, 2, 3, or 4)	As ECOG performance status increases, NGS testing is less likely.
Body weight (continuous, in kilograms)	As weight increases, NGS testing is more likely.
Stage at initial diagnosis (0-I, II, III, or IV)	Stage 0-II=NGS is more likely than stage III; stage IV=NGS is more likely than stage III.
EGFR[c] (not tested, positive, not positive) by non-NGS test	EGFR=positive, NGS less likely.
ROS1[d] (not tested, positive, not positive) by non-NGS test	ROS1=positive, NGS less likely.
ALK[e] (not tested, positive, not positive) by non-NGS test	ALK=positive, NGS less likely.
BRAF[f] (not tested, positive, not positive) by non-NGS test	BRAF=positive, NGS less likely.
KRAS[g] (not tested, positive, not positive) by non-NGS test	KRAS=positive, NGS less likely.
PD-L1[h] (not tested, positive, not positive)	PD-L1=positive, NGS less likely.
Number of single-gene tests (continuous)	As the number of single-gene tests increase, NGS less likely.
Practice setting (academic, community)	Practice setting=academic, NGS more likely.
Insurance status (public, private, other)	This relationship is unknown. It is possible that insurance status=public, NGS less likely; however, it is possible that in some cases, insurance status=private only, NGS could be less likely.
MAC[i] region	No direction is known.
MolDX[j]	While this only applies to Medicare, states may adopt broader policies, and the relationship is uncertain. MolDX may make NGS more likely, but it is largely unknown.
NCCN[k] guidelines (pre, broad, or NGS)	NCCN guidelines=NGS, NGS more likely.
Drug approval periods (1, 2, 3, 4, 5, 6, 7)	As drug approval periods increase, NGS more likely.
Laboratory values (high, normal, low, not tested) for alkaline phosphatase, alanine transaminase, aspartate transferase, bilirubin, creatine, lymphocyte count, red blood cell count, hematocrit, platelet count, white blood cell count, hemoglobin	The direction of a single laboratory value is unknown. However, generally one would expect multiple out-of-range values to reflect poor health and may make NGS less likely, but the a priori assumed direction is unknown.

[a]WHO: World Health Organization.

[b]ECOG: Eastern Cooperative Oncology Group.

[c]EGFR: epidermal growth factor receptor.

[d]ROS1: c-ros oncogene 1.

[e]ALK: anaplastic lymphoma kinase.

[f]BRAF: V-Raf murine sarcoma viral oncogene homolog B.

[g]KRAS: Kirsten rat sarcoma virus.

[h]PD-L1: programmed death ligand 1.

[i]MAC: Medicare Administrative Contractor.

[j]MolDX: Molecular Diagnostics Services.

[k]NCCN: National Comprehensive Cancer Network.

Statistical Analysis

Descriptive analyses were conducted to summarize available data and to understand the extent of missingness in the database. Categorical variables were assessed using a 1-sided chi-square test or Fisher exact test and continuous variables using a 2-sided *t* test. Missing values were imputed using the random forest

missing data algorithm (impute.rfsrc function in R package *randomForestSRC*) [23].

Three modeling strategies were used to identify potential predictors of NGS-based testing with 2 sets of outcomes for ever versus never NGS-tested (model 1) and early versus late NGS-tested (model 2). The 3 modeling strategies included logistic regression (LR) models, penalized logistic regression (PLR) using least absolute shrinkage and selection operator (LASSO) penalty, and extreme gradient boosting (XGBoost) with trees as base learners. LR was implemented using forward selection on the main effects and predefined interactions (listed earlier), starting with the predefined variables and adding the most significant terms to the model. PLR was implemented using sparse group LASSO on the main effects and predefined interactions, forcing some predefined variables into the model with the penalty selected using 5-fold cross-validation. XGBoost is a decision tree–based machine learning algorithm [24]. The model matrix for XGBoost was built using main effects and predefined interactions. Hyperparameters were selected based on 5-fold cross-validation over a grid search, and hyperparameters included the shrinkage (learning rate), the number of trees, and tree depth. Table S1 in Multimedia Appendix 1 contains the full list of hyperparameters used in this study. The data extraction approach and modeling process is summarized in Figure 1.

In step 1, data were extracted based on the prespecified inclusion and exclusion criteria. Step 2 involved variable recoding, which included transforming all categorical variables with missing information by creating an additional level to represent missing data. Step 3 was a data quality method used to identify any unusual observations that needed to be excluded or recoded in addition to any imputation that was required. Steps 4 to 6 outline the implementation of models, evaluation of the performance of these models, and interpretation of the final features selected using LR. Figure 2 provides an overview of the model strategy evaluation process for the 2 outcomes mentioned in step 4 of Figure 1.

First, the data were split into D1 (training+validation; 80%) and D2 (testing; 20%) sets. Then, the 3 strategies were evaluated by comparing their performance on multiple m=1000 splits in the training (70%) and validation data (30%) within the D1 set. Specifically, for each split, all 3 strategies were fit to training data, and performance measures (eg, area under the receiver operating curve) were computed on the validation data. Modeling was done using R packages, sparsegl was used for LASSO, XGBoost for gradient boosting, and PRROC, which computes the areas under the precision-recall and ROC curve, for performance measures. PLR and XGBoost involved hyperparameters that were fine-tuned using 5-fold cross-validation nested within training datasets. Prediction models were developed on 2 different groups: ever versus never and early versus late NGS-tested groups. In total, 146 features (including all levels of all variables) were entered into both the XGBoost and LASSO models, with only 36 features (main effects and interactions) being used in the LR model. Preselection of features consisted of excluding variables that have little to no association with the outcomes of interest.

The final model was selected by evaluating performance as described earlier (area under the receiver operating curve from validation data) and by considering the simplicity and clinical interpretability. Model performance was re-estimated using the test data D2. For the final model choice, the features with nonzero coefficients selected by PLR were run on the D1 data. These variables were fitted to an LR model within the test data D2 to calculate model estimates (odds ratios, 95% CIs, and P values). Odds ratios for main effects in the presence of interaction terms were calculated using the analytical formula presented in Multimedia Appendix 2. All analyses were conducted using SAS (version 9.4; SAS Institute Inc) and R (version 4.0.3; R Foundation for Statistical Computing).

Figure 1. Data extraction and modeling flow. AUC: area under the curve; CI: confidence interval; NSCLC: non–small cell lung cancer; OR: odds ratio; ROC: receiver operating curve.

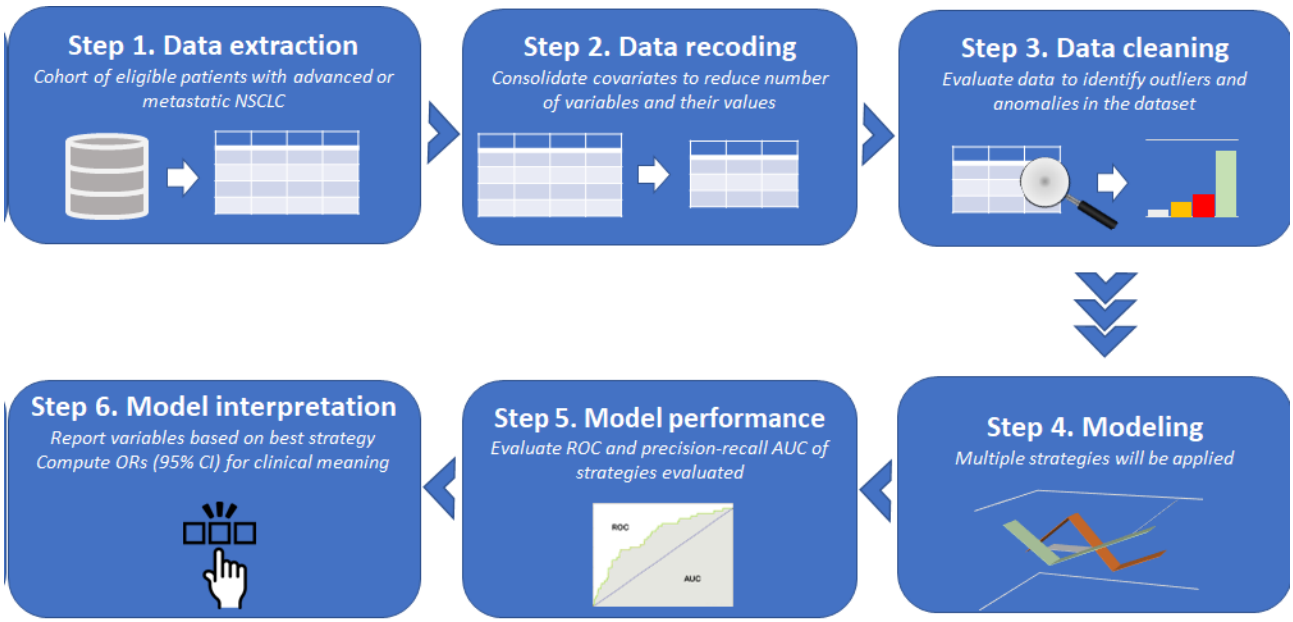
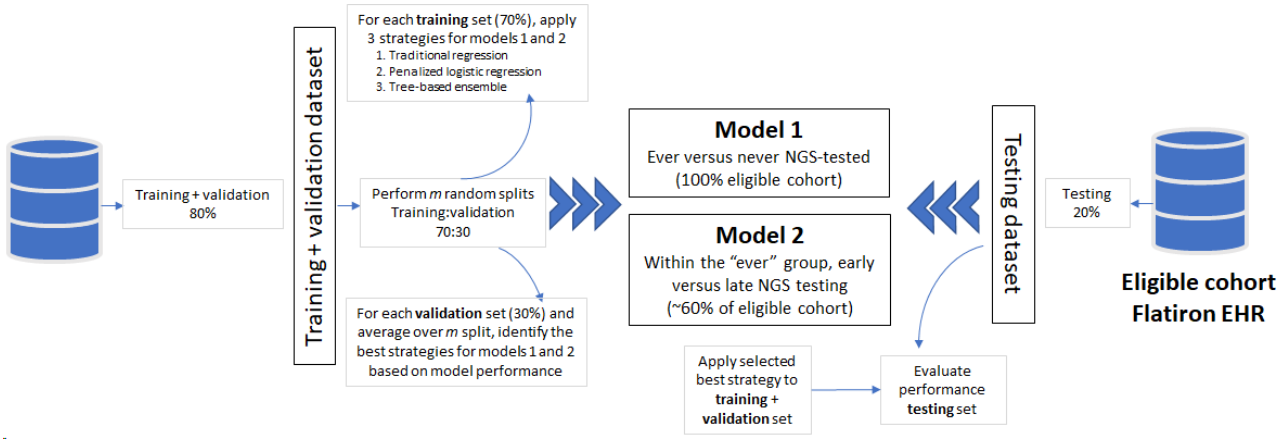


Figure 2. Modeling evaluation flow. EHR: electronic health record; NGS: next-generation sequencing.



Ethical Considerations

The data used for this study are deidentified and subject to obligations to prevent reidentification and protect patient confidentiality, and as such are not considered human subjects research and are exempt from review in accordance with the US Code of Federal Regulations [16].

Results

A total of 74,211 patient records were available in the Flatiron Health NSCLC dataset for this analysis. After applying eligibility criteria, a total of 31,407 patients were included in this analysis. Of all patients, 42.75% (n=13,425) were included in the ever NGS-tested group and 57.25% (n=17,982) were included in the never NGS-tested group. Among those in the ever NGS-tested group, 84.08% (n=11,289) were early NGS-tested, and 15.91% (n=2136) late NGS-tested. Characteristics of these groups and subgroups used as features in the machine learning models are listed in Tables 2-11.

Most features were significantly different between both the ever and never NGS-tested as well as the early NGS versus late NGS-tested groups. Of note, smoking rates and testing conducted during the NCCN prerecommendations period were lower for the ever NGS-tested group (n=10,589, 78.88% vs n=14,987, 83.34% and n=2663, 19.84% vs n=10,734, 59.69%, respectively), and ECOG status of 0 (n=4410, 32.85% vs n=4665, 25.94%) was higher for the ever NGS-tested group versus those who were never tested. Similarly, for the early versus late NGS-tested groups, there was a higher proportion of patients with a history of smoking (n=9025, 79.95% vs n=1564, 73.22%) and a lower proportion of testing conducted during the NCCN prerecommendations period (n=1746, 15.47% vs n=917, 42.93%) as well as a lower proportion of ECOG status of 0 (n=3606, 31.94% vs n=804, 37.64%) for the early tested group.

Comparison of performance metrics for each model showed that the percent AUC was similar across models (80%-84% and 77%-80%) and marginally better when the models were fit on the ever versus never NGS-tested groups. In addition, other metrics were also comparable (Table S2 in Multimedia Appendix 1). The final model chosen was the LASSO model, as it was able to identify important features including interactions (those with nonzero coefficients after shrinkage) and the metrics for each model were highly comparable (Table S2 in Multimedia Appendix 1). Figures S1 and S2 in Multimedia Appendix 1 show the feature importance plots for both groups. The most important factors associated with ever versus never testing included year of diagnosis, observation of a PD-L1 test, Black or African American race, and number of single-gene tests observed. The most important factors associated with early versus late testing included the observation of a PD-L1 test, a positive single-gene test result, the year of diagnosis, and the geographical region of care. Later year of diagnosis, evidence of PD-L1 testing, patient race, positive single-gene test results, and region were among the top 5 predictors of NGS testing for both ever versus never as well as early versus late NGS testing.

Table 2. Demographic characteristics of the overall, ever, and never NGS[a]-tested study cohorts prior to imputation.

Characteristic	Overall (N=31,407)	Ever NGS-tested[b] (n=13,425)	Never NGS-tested[c] (n=17,982)	Ever NGS-tested versus never NGS-tested, *P* value[d]
Age at initial diagnosis (years), mean (SD)	67.2 (9.8)	67.2 (10.1)	67.3 (9.5)	.66
Sex, n (%)				.0007
Female	16,680 (53.11)	7281 (54.23)	9399 (52.27)	
Male	14,726 (46.89)	6144 (45.77)	8582 (47.73)	
Unknown or missing	1 (0)	0 (0)	1 (0.01)	
Race, n (%)				<.0001
Asian	1050 (3.34)	552 (4.11)	498 (2.77)	
Black or African American	2845 (9.06)	1089 (8.11)	1756 (9.77)	
White	21,248 (67.65)	9109 (67.85)	12,139 (67.51)	
Other	3269 (10.41)	1392 (10.37)	1877 (10.44)	
Unknown or missing	2995 (9.54)	1283 (9.56)	1712 (9.52)	
Smoking status, n (%)				<.0001
History of smoking	25,576 (81.43)	10,589 (78.88)	14,987 (83.34)	
No history of smoking	5657 (18.01)	2826 (21.05)	2831 (15.74)	
Unknown or missing	174 (0.55)	10 (0.07)	164 (0.91)	
ECOG[e] performance status, n (%)				<.0001
0	9075 (28.89)	4410 (32.85)	4665 (25.94)	
1	11,215 (35.71)	5275 (39.29)	5940 (33.03)	
2	3401 (10.83)	1393 (10.38)	2008 (11.17)	
3	762 (2.43)	306 (2.28)	456 (2.54)	
4	51 (0.16)	17 (0.13)	34 (0.19)	
Unknown or missing	6903 (21.98)	2024 (15.08)	4879 (27.13)	

[a]NGS: next-generation sequencing.

[b]Patients in the overall study cohort with evidence of NGS-based biomarker testing in the database.

[c]Patients in the overall study cohort with no evidence of NGS-based biomarker testing.

[d]Two-sided *t* test for continuous variables; chi-square or Fisher exact test (where expected cell size <5) for categorical variables.

[e]ECOG: Eastern Cooperative Oncology Group.

Table 3. Biomarker status of the overall, ever, and never NGS[a]-tested study cohorts prior to imputation.

Characteristic	Overall (N=31,407)	Ever NGS-tested[b] (n=13,425)	Never NGS-tested[c] (n=17,982)	Ever NGS-tested vs never NGS-tested, *P* value[d]
Non-NGS–based (single gene) ALK[e] status, n (%)				<.0001
Positive	617 (1.96)	253 (1.88)	364 (2.02)	
Not positive	15,626 (49.75)	6278 (46.76)	9348 (51.99)	
Not tested	15,164 (48.28)	6894 (51.35)	8270 (45.99)	
Non-NGS–based (single gene) BRAF[f] status, n (%)				<.0001
Positive	94 (0.30)	32 (0.24)	62 (0.34)	
Not positive	3775 (12.02)	1729 (12.88)	2046 (11.38)	
Not tested	27,538 (87.68)	11,664 (86.88)	15,874 (88.28)	
Non-NGS–based (single gene) EGFR[g] status, n (%)				<.0001
Positive	2822 (8.99)	928 (6.91)	1894 (10.53)	
Not positive	12,312 (39.20)	3427 (25.53)	8885 (49.41)	
Not tested	16,273 (51.81)	9070 (67.56)	7203 (40.06)	
Non-NGS–based (single gene) KRAS[h] status, n (%)				<.0001
Positive	1141 (3.63)	298 (2.22)	843 (4.69)	
Not positive	2958 (9.42)	1082 (8.06)	1876 (10.43)	
Not tested	27,308 (86.95)	12,045 (89.72)	15,263 (84.88)	
Non-NGS–based (single gene) ROS1[i] status, n (%)				<.0001
Positive	128 (0.41)	58 (0.43)	70 (0.39)	
Not positive	9383 (29.88)	5011 (37.33)	4372 (24.31)	
Not tested	21,896 (69.72)	8356 (62.24)	13,540 (75.30)	
Non-NGS–based (single gene) MET[j] status, n (%)				<.0001
Positive	7 (0.02)	3 (0.02)	4 (0.02)	
Not positive	1965 (6.26)	1517 (11.30)	448 (2.49)	
Not tested	29,435 (93.72)	11,905 (88.68)	17,530 (97.49)	
Non-NGS–based (single gene) RET[k] status, n (%)				<.0001
Positive	34 (0.11)	27 (0.20)	7 (0.04)	
Not positive	2381 (7.58)	1679 (12.51)	702 (3.90)	
Not tested	28,992 (92.31)	11,719 (87.29)	17,273 (96.06)	
Non-NGS–based (single gene) NTRK[l] status, n (%)				<.0001
Positive	2 (0.01)	1 (0.01)	1 (0.01)	
Not positive	747 (2.38)	617 (4.60)	130 (0.72)	
Not tested	30,658 (97.62)	12,807 (95.40)	17,851 (99.27)	
Non-NGS–based (single gene) testing[m], n (%)				<.0001
Any positive result observed	4795 (15.27)	1576 (11.74)	3219 (17.90)	
Never tested	11,968 (38.11)	5661 (42.17)	6307 (35.07)	
Tested, but no positive results observed	14,644 (46.63)	6188 (46.09)	8456 (47.02)	
PD-L1[n] status, n (%)				<.0001
Positive	1826 (5.81)	1289 (9.60)	537 (2.99)	
Not positive	9988 (31.80)	6354 (47.33)	3634 (20.21)	

Characteristic	Overall (N=31,407)	Ever NGS-tested[b] (n=13,425)	Never NGS-tested[c] (n=17,982)	Ever NGS-tested vs never NGS-tested, *P* value[d]
Not tested	19,593 (62.38)	5782 (43.07)	13,811 (76.80)	
Single-gene tests received[m], mean (SD)	2.1 (2.0)	2.3 (2.0)	2.0 (1.9)	<.0001

[a]NGS: next-generation sequencing.

[b]Patients in the overall study cohort with evidence of NGS-based biomarker testing in the database.

[c]Patients in the overall study cohort with no evidence of NGS-based biomarker testing.

[d]Two-sided *t* test for continuous variables; chi-square or Fisher exact test (where expected cell size <5) for categorical variables.

[e]ALK: anaplastic lymphoma kinase.

[f]BRAF: V-Raf Murine Sarcoma Viral Oncogene Homolog B.

[g]EGFR: epidermal growth factor receptor.

[h]KRAS: Kirsten rat sarcoma virus.

[i]ROS1: c-ros oncogene 1.

[j]MET: mesenchymal epithelial transition.

[k]RET: rearranged during transfection.

[l]NTRK: neurotrophic tyrosine receptor kinase.

[m]Results are based on biomarkers ALK, BRAF, EGFR, KRAS, ROS1, MET, RET, and NTRK.

[n]PD-L1: programmed death ligand 1.

Table 4. Geographic and time characteristics of the overall, ever, and never NGS[a]-tested study cohorts prior to imputation.

Characteristic	Overall (N=31,407), n (%)	Ever NGS-tested[b] (n=13,425), n (%)	Never NGS-tested[c] (n=17,982), n (%)	Ever NGS-tested vs never NGS-tested, *P* value[d]
MAC[e] region				<.0001
JE Noridian	2097 (6.68)	814 (6.06)	1283 (7.13)	
JF Noridian	2476 (7.88)	1111 (8.28)	1365 (7.59)	
J6 NGS	856 (2.73)	335 (2.50)	521 (2.90)	
J5 WPS	603 (1.92)	235 (1.75)	368 (2.05)	
J8 WPS	2025 (6.45)	1051 (7.83)	974 (5.42)	
JK NGS	2459 (7.83)	1102 (8.21)	1357 (7.55)	
JL Novitas	2817 (8.97)	1283 (9.56)	1534 (8.53)	
JM Palmetto	2218 (7.06)	858 (6.39)	1360 (7.56)	
J15 CGS	924 (2.94)	397 (2.96)	527 (2.93)	
JJ Cahaba	4194 (13.35)	2049 (15.26)	2145 (11.93)	
JH Novitas	6093 (19.40)	2176 (16.21)	3917 (21.78)	
Unknown or missing	4645 (14.79)	2014 (15)	2631 (14.63)	
MolDX[f] Program				<.0001
Yes	14,294 (45.51)	6399 (47.66)	7895 (43.91)	
No	12,468 (39.70)	5012 (37.33)	7456 (41.46)	
Unknown or missing	4645 (14.79)	2014 (15)	2631 (14.63)	
NCCN[g] guideline period				<.0001
Prerecommendations	13,397 (42.66)	2663 (19.84)	10,734 (59.69)	
Broad-based testing recommended	13,552 (43.15)	7339 (54.67)	6213 (34.55)	
NGS-based testing recommended	4458 (14.19)	3423 (25.50)	1035 (5.76)	
Timing of diagnosis by drug approval period				<.0001
Period 1	1223 (3.89)	96 (0.72)	1127 (6.27)	
Period 2	12,850 (40.91)	2823 (21.03)	10,027 (55.76)	
Period 3	4396 (14)	1868 (13.91)	2528 (14.06)	
Period 4	4877 (15.53)	2724 (20.29)	2153 (11.97)	
Period 5	4613 (14.69)	3224 (24.01)	1389 (7.72)	
Period 6	2858 (9.10)	2216 (16.51)	642 (3.57)	
Period 7	590 (1.88)	474 (3.53)	116 (0.65)	

[a]NGS: next-generation sequencing.

[b]Patients in the overall study cohort with evidence of NGS-based biomarker testing in the database.

[c]Patients in the overall study cohort with no evidence of NGS-based biomarker testing.

[d]Two-sided *t* test for continuous variables; chi-square or Fisher exact test (where expected cell size <5) for categorical variables.

[e]MAC: Medicare Administration Contractor.

[f]MolDX: Molecular Diagnostics Services.

[g]NCCN: National Comprehensive Cancer Network.

Table 5. Clinical care characteristics of the overall, ever, and never NGS[a]-tested study cohorts prior to imputation.

Characteristic	Overall (N=31,407)	Ever NGS-tested[b] (n=13,425)	Never NGS-tested[c] (n=17,982)	Ever NGS-tested vs never NGS-tested, *P* value[d]
Practice setting, n (%)				<.0001
Academic	3626 (11.55)	1783 (13.28)	1843 (10.25)	
Community	27,781 (88.45)	11,642 (86.72)	16,139 (89.75)	
Insurance type, n (%)				<.0001
Private+public	4301 (13.69)	1940 (14.45)	2361 (13.13)	
Private only	7083 (22.55)	3601 (26.82)	3482 (19.36)	
Public only	4037 (12.85)	1560 (11.62)	2477 (13.77)	
Multiple types	8997 (28.65)	4066 (30.29)	4931 (27.42)	
Unknown or missing	6989 (22.25)	2258 (16.82)	4731 (26.31)	
Stage at initial diagnosis, n (%)				<.0001
0-I	2736 (8.71)	1208 (9)	1528 (8.50)	
II	1453 (4.63)	671 (5)	782 (4.35)	
III	5621 (17.90)	2227 (16.59)	3394 (18.87)	
IV	20,929 (66.64)	9096 (67.75)	11,833 (65.80)	
Unknown or missing	668 (2.13)	223 (1.66)	445 (2.47)	
Year of index diagnosis, n (%)				<.0001
2011	1896 (6.04)	158 (1.18)	1738 (9.67)	
2012	2402 (7.65)	229 (1.71)	2173 (12.08)	
2013	2699 (8.59)	476 (3.55)	2223 (12.36)	
2014	3054 (9.72)	664 (4.95)	2390 (13.29)	
2015	3346 (10.65)	1136 (8.46)	2210 (12.29)	
2016	3397 (10.82)	1372 (10.22)	2025 (11.26)	
2017	3472 (11.05)	1708 (12.72)	1764 (9.81)	
2018	3401 (10.83)	1966 (14.64)	1435 (7.98)	
2019	3282 (10.45)	2293 (17.08)	989 (5.50)	
2020	2777 (8.84)	2066 (15.39)	711 (3.95)	
2021	1681 (5.35)	1357 (10.11)	324 (1.80)	
Practice volume[e], mean (SD)	154.1 (143.6)	169.2 (156.0)	142.8 (132.5)	<.0001
BMI, n (%)				<.0001
Underweight	1373 (4.37)	597 (4.45)	776 (4.32)	
Normal weight	10,593 (33.73)	4638 (34.55)	5955 (33.12)	
Overweight	8897 (28.33)	4019 (29.94)	4878 (27.13)	
Obese	6492 (20.67)	2920 (21.75)	3572 (19.86)	
Unknown or missing	4052 (12.90)	1251 (9.32)	2801 (15.58)	
Body weight (kg), mean (SD)	75.0 (18.6)	75.3 (18.8)	74.8 (18.4)	.04
Duration of follow-up (days), mean (SD)	704.8 (638.1)	735.1 (636.5)	682.2 (638.3)	<.0001

[a]NGS: next-generation sequencing.

[b]Patients in the overall study cohort with evidence of NGS-based biomarker testing in the database.

[c]Patients in the overall study cohort with no evidence of NGS-based biomarker testing.

[d]Two-sided *t* test for continuous variables; chi-square or Fisher exact test (where expected cell size <5) for categorical variables.

[e]Number of patients with non-small cell lung cancer receiving care at the same practice per year.

Table 6. Laboratory values of the overall, ever, and never NGS[a]-tested study cohorts prior to imputation.

Characteristic	Overall (N=31,407), n (%)	Ever NGS-tested[b] (n=13,425), n (%)	Never NGS-tested[c] (n=17,982), n (%)	Ever NGS-tested vs never NGS-tested, *P* value[d]
ALP[e]				<.0001
High	3550 (11.30)	1583 (11.79)	1967 (10.94)	
Low	146 (0.46)	60 (0.45)	86 (0.48)	
Normal	15,295 (48.70)	6805 (50.69)	8490 (47.21)	
Not tested	12,416 (39.53)	4977 (37.07)	7439 (41.37)	
ALT[f]				<.0001
High	1480 (4.71)	676 (5.04)	804 (4.47)	
Low	850 (2.71)	384 (2.86)	466 (2.59)	
Normal	16,606 (52.87)	7389 (55.04)	9217 (51.26)	
Not tested	12,471 (39.71)	4976 (37.07)	7495 (41.68)	
AST[g]				<.0001
High	1364 (4.34)	579 (4.31)	785 (4.37)	
Low	1018 (3.24)	447 (3.33)	571 (3.18)	
Normal	16,706 (53.19)	7479 (55.71)	9227 (51.31)	
Not tested	12,319 (39.22)	4920 (36.65)	7399 (41.15)	
Bilirubin				<.0001
High	461 (1.47)	212 (1.58)	249 (1.38)	
Low	1200 (3.82)	545 (4.06)	655 (3.64)	
Normal	16,014 (50.99)	7138 (53.17)	8876 (49.36)	
Not tested	13,732 (43.72)	5530 (41.19)	8202 (45.61)	
Creatinine				<.0001
High	2272 (7.23)	950 (7.08)	1322 (7.35)	
Low	2143 (6.82)	965 (7.19)	1178 (6.55)	
Normal	15,512 (49.39)	6917 (51.52)	8595 (47.80)	
Not tested	11,480 (36.55)	4593 (34.21)	6887 (38.30)	
Lymphocyte count				<.0001
High	435 (1.39)	162 (1.21)	273 (1.52)	
Low	7325 (23.32)	3270 (24.36)	4055 (22.55)	
Normal	12,238 (38.97)	5504 (41)	6734 (37.45)	
Not tested	11,409 (36.33)	4489 (33.44)	6920 (38.48)	
Red blood cell count				<.0001
High	371 (1.18)	135 (1.01)	236 (1.31)	
Low	5751 (18.31)	2336 (17.40)	3415 (18.99)	
Normal	12,350 (39.32)	5551 (41.35)	6799 (37.81)	
Not tested	12,935 (41.19)	5403 (40.25)	7532 (41.89)	
Hematocrit				<.0001
High	482 (1.53)	187 (1.39)	295 (1.64)	
Low	6440 (20.50)	2772 (20.65)	3668 (20.40)	
Normal	13,085 (41.66)	6026 (44.89)	7059 (39.26)	
Not tested	11,400 (36.30)	4440 (33.07)	6960 (38.71)	

Characteristic	Overall (N=31,407), n (%)	Ever NGS-tested[b] (n=13,425), n (%)	Never NGS-tested[c] (n=17,982), n (%)	Ever NGS-tested vs never NGS-tested, *P* value[d]
Platelet count				.003
High	2605 (8.29)	1038 (7.73)	1567 (8.71)	
Low	675 (2.15)	271 (2.02)	404 (2.25)	
Normal	14,807 (47.15)	6436 (47.94)	8371 (46.55)	
Not tested	13,320 (42.41)	5680 (42.31)	7640 (42.49)	
White blood cell count				.03
High	5171 (16.46)	2166 (16.13)	3005 (16.71)	
Low	461 (1.47)	195 (1.45)	266 (1.48)	
Normal	13,237 (42.15)	5790 (43.13)	7447 (41.41)	
Not tested	12,538 (39.92)	5274 (39.28)	7264 (40.40)	
Hemoglobin, whole blood				<.0001
High	406 (1.29)	141 (1.05)	265 (1.47)	
Low	6973 (22.20)	2969 (22.12)	4004 (22.27)	
Normal	13,193 (42.01)	5997 (44.67)	7196 (40.02)	
Not tested	10,835 (34.50)	4318 (32.16)	6517 (36.24)	

[a]NGS: next-generation sequencing.

[b]Patients in the overall study cohort with evidence of NGS-based biomarker testing in the database.

[c]Patients in the overall study cohort with no evidence of NGS-based biomarker testing.

[d]Two-sided *t* test for continuous variables; chi-square or Fisher exact test (where expected cell size <5) for categorical variables.

[e]ALP: alkaline phosphatase.

[f]ALT: alanine transaminase.

[g]AST: aspartate aminotransferase.

Table 7. Demographic characteristics of early and late NGS[a]-tested study cohorts prior to imputation.

Characteristic	Early NGS-tested[b] (n=11,289)	Late NGS-tested[c] (n=2136)	Early NGS-tested vs late NGS-tested, *P* value[d]
Age at initial diagnosis (years), mean (SD)	67.5 (10.1)	65.5 (10.0)	<.0001
Sex, n (%)			.02
Female	6073 (53.80)	1208 (56.55)	
Male	5216 (46.20)	928 (43.45)	
Race, n (%)			<.0001
Asian	408 (3.61)	144 (6.74)	
Black or African American	897 (7.95)	192 (8.99)	
White	7655 (67.81)	1454 (68.07)	
Other	1215 (10.76)	177 (8.29)	
Unknown or missing	1114 (9.87)	169 (7.91)	
Smoking status, n (%)			<.0001
History of smoking	9025 (79.95)	1564 (73.22)	
No history of smoking	2256 (19.98)	570 (26.69)	
Unknown or missing	8 (0.07)	2 (0.09)	
ECOG[e] performance status, n (%)			<.0001
0	3606 (31.94)	804 (37.64)	
1	4440 (39.33)	835 (39.09)	
2	1220 (10.81)	173 (8.10)	
3	282 (2.50)	24 (1.12)	
4	15 (0.13)	2 (0.09)	
Unknown or missing	1726 (15.29)	298 (13.95)	

[a]NGS: next-generation sequencing.

[b]Patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy.

[c]Patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy.

[d]Two-sided *t* test for continuous variables; chi-square or Fisher exact test (where expected cell size <5) for categorical variables.

[e]ECOG: Eastern Cooperative Oncology Group.

Table 8. Biomarker status of early and late NGS[a]-tested study cohorts prior to imputation.

Characteristic	Early NGS-tested[b] (n=11,289)	Late NGS-tested[c] (n=2136)	Early NGS-tested vs late NGS-tested, *P* value[d]
Non-NGS–based (single gene) ALK[e] status, n (%)			<.0001
Positive	193 (1.71)	60 (2.81)	
Not positive	5208 (46.13)	1070 (50.09)	
Not tested	5888 (52.16)	1006 (47.10)	
Non-NGS–based (single gene) BRAF[f] status, n (%)			.04
Positive	25 (0.22)	7 (0.33)	
Not positive	1420 (12.58)	309 (14.47)	
Not tested	9844 (87.20)	1820 (85.21)	
Non-NGS–based (single gene) EGFR[g] status, n (%)			<.0001
Positive	435 (3.85)	493 (23.08)	
Not positive	2589 (22.93)	838 (39.23)	
Not tested	8265 (73.21)	805 (37.69)	
Non-NGS–based (single gene) KRAS[h] status, n (%)			<.0001
Positive	221 (1.96)	77 (3.60)	
Not positive	825 (7.31)	257 (12.03)	
Not tested	10,243 (90.73)	1802 (84.36)	
Non-NGS–based (single gene) ROS1[i] status, n (%)			<.0001
Positive	44 (0.39)	14 (0.66)	
Not positive	4376 (38.76)	635 (29.73)	
Not tested	6869 (60.85)	1487 (69.62)	
Non-NGS–based (single gene) MET[j] status, n (%)			<.0001
Positive	2 (0.02)	1 (0.05)	
Not positive	1449 (12.84)	68 (3.18)	
Not tested	9838 (87.15)	2067 (96.77)	
Non-NGS–based (single gene) RET[k] status, n (%)			<.0001
Positive	27 (0.24)	0 (0)	
Not positive	1558 (13.80)	121 (5.66)	
Not tested	9704 (85.96)	2015 (94.34)	
Non-NGS–based (single gene) NTRK[l] status, n (%)			<.0001
Positive	1 (0.01)	0 (0)	
Not positive	596 (5.28)	21 (0.98)	
Not tested	10,692 (94.71)	2115 (99.02)	
Non-NGS–based (single gene) testing[m], n (%)			<.0001
Any positive result observed	931 (8.25)	645 (30.20)	
Never tested	4959 (43.93)	702 (32.87)	
Tested, but no positive results observed	5399 (47.83)	789 (36.94)	
PD-L1[n] status, n (%)			<.0001
Positive	1228 (10.88)	61 (2.86)	
Not positive	5785 (51.24)	569 (26.64)	

Characteristic	Early NGS-tested[b] (n=11,289)	Late NGS-tested[c] (n=2136)	Early NGS-tested vs late NGS-tested, *P* value[d]
Not tested	4276 (37.88)	1506 (70.51)	
Number of single-gene tests received[m], mean (SD)	2.3 (2.1)	2.2 (2.0)	.002

[a]NGS: next-generation sequencing.

[b]Patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy.

[c]Patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy.

[d]Two-sided *t* test for continuous variables; chi-square or Fisher exact test (where expected cell size <5) for categorical variables.

[e]ALK: anaplastic lymphoma kinase.

[f]BRAF: V-Raf murine sarcoma viral oncogene homolog B.

[g]EGFR: epidermal growth factor receptor.

[h]KRAS: Kirsten rat sarcoma virus.

[i]ROS1: c-ros oncogene 1.

[j]MET: mesenchymal epithelial transition.

[k]RET: rearranged during transfection.

[l]NTRK: neurotrophic tyrosine receptor kinase.

[m]Results are based on biomarkers ALK, BRAF, EGFR, KRAS, ROS1, MET, RET, and NTRK.

[n]PD-L1: programmed death ligand 1.

Table 9. Geographic and time characteristics of early and late NGS[a]-tested study cohorts prior to imputation.

Characteristic	Early NGS-tested[b] (n=11,289), n (%)	Late NGS-tested[c] (n=2136), n (%)	Early NGS-tested vs late NGS-tested, *P* value[d]
MAC[e] region			<.0001
JE Noridian	639 (5.66)	175 (8.19)	
JF Noridian	956 (8.47)	155 (7.26)	
J6 NGS	283 (2.51)	52 (2.43)	
J5 WPS	205 (1.82)	30 (1.40)	
J8 WPS	921 (8.16)	130 (6.09)	
JK NGS	924 (8.18)	178 (8.33)	
JL Novitas	1094 (9.69)	189 (8.85)	
JM Palmetto	707 (6.26)	151 (7.07)	
J15 CGS	339 (3)	58 (2.72)	
JJ Cahaba	1734 (15.36)	315 (14.75)	
JH Novitas	1786 (15.82)	390 (18.26)	
Unknown or missing	1701 (15.07)	313 (14.65)	
MolDX[f] Program			.38
Yes	5402 (47.85)	997 (46.68)	
No	4186 (37.08)	826 (38.67)	
Unknown or missing	1701 (15.07)	313 (14.65)	
NCCN[g] guideline period			<.0001
Pre recommendations	1746 (15.47)	917 (42.93)	
Broad-based testing recommended	6286 (55.68)	1053 (49.30)	
NGS-based testing recommended	3257 (28.85)	166 (7.77)	
Timing of diagnosis by drug approval period			<.0001
Period 1	43 (0.38)	53 (2.48)	
Period 2	1902 (16.85)	921 (43.12)	
Period 3	1458 (12.92)	410 (19.19)	
Period 4	2347 (20.79)	377 (17.65)	
Period 5	2955 (26.18)	269 (12.59)	
Period 6	2122 (18.80)	94 (4.40)	
Period 7	462 (4.09)	12 (0.56)	

[a]NGS: next-generation sequencing.

[b]Patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy.

[c]Patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy.

[d]Two-sided *t* test for continuous variables; chi-square or Fisher exact test (where expected cell size <5) for categorical variables.

[e]MAC: Medicare Administration Contractor.

[f]MolDX: Molecular Diagnostics Services.

[g]NCCN: National Comprehensive Cancer Network.

Table 10. Clinical care characteristics of early and late NGS[a]-tested study cohorts prior to imputation.

Characteristic	Early NGS-tested[b] (n=11,289)	Late NGS-tested[c] (n=2136)	Early NGS-tested vs late NGS-tested, *P* value[d]
Practice setting, n (%)			.50
Academic	1509 (13.37)	274 (12.83)	
Community	9780 (86.63)	1862 (87.17)	
Insurance type, n (%)			<.0001
Private+public	1689 (14.96)	251 (11.75)	
Private only	3026 (26.80)	575 (26.92)	
Public only	1317 (11.67)	243 (11.38)	
Multiple types	3491 (30.92)	575 (26.92)	
Unknown or missing	1766 (15.64)	492 (23.03)	
Stage at initial diagnosis, n (%)			.0004
0-I	1051 (9.31)	157 (7.35)	
II	580 (5.14)	91 (4.26)	
III	1822 (16.14)	405 (18.96)	
IV	7655 (67.81)	1441 (67.46)	
Unknown or missing	181 (1.60)	42 (1.97)	
Year of index diagnosis			<.0001
2011	69 (0.61)	89 (4.17)	
2012	121 (1.07)	108 (5.06)	
2013	295 (2.61)	181 (8.47)	
2014	430 (3.81)	234 (10.96)	
2015	831 (7.36)	305 (14.28)	
2016	1038 (9.19)	334 (15.64)	
2017	1425 (12.62)	283 (13.25)	
2018	1712 (15.17)	254 (11.89)	
2019	2111 (18.70)	182 (8.52)	
2020	1939 (17.18)	127 (5.95)	
2021	1318 (11.68)	39 (1.83)	
Practice volume[e], mean (SD)	169.6 (156.7)	166.8 (152.4)	.44
BMI, n (%)			<.0001
Underweight	529 (4.69)	68 (3.18)	
Normal weight	3963 (35.10)	675 (31.60)	
Overweight	3410 (30.21)	609 (28.51)	
Obese	2435 (21.57)	485 (22.71)	
Unknown or missing	952 (8.43)	299 (14)	
Body weight (kg), mean (SD)	75.2 (18.8)	75.9 (18.7)	.11
Duration of follow-up (days), mean (SD)	644.1 (547.5)	1216.2 (829.1)	<.0001

[a]NGS: next-generation sequencing.

[b]Patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy.

[c]Patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy.

[d]Two-sided *t* test for continuous variables; chi-square or Fisher exact test (where expected cell size <5) for categorical variables.

[e]Number of patients with non-small cell lung cancer receiving care at the same practice per year.

Table 11. Laboratory values of early and late NGS[a]-tested study cohorts prior to imputation.

Characteristic	Early NGS-tested[b] (n=11,289)	Late NGS-tested[c] (n=2136)	Early NGS-tested vs late NGS-tested, *P* value[d]
ALP[e]			.79
High	1339 (11.86)	244 (11.42)	
Low	48 (0.43)	12 (0.56)	
Normal	5720 (50.67)	1085 (50.80)	
Not tested	4182 (37.04)	795 (37.22)	
ALT[f]			.01
High	577 (5.11)	99 (4.63)	
Low	345 (3.06)	39 (1.83)	
Normal	6192 (54.85)	1197 (56.04)	
Not tested	4175 (36.98)	801 (37.50)	
AST[g]			.07
High	486 (4.31)	93 (4.35)	
Low	396 (3.51)	51 (2.39)	
Normal	6278 (55.61)	1201 (56.23)	
Not tested	4129 (36.58)	791 (37.03)	
Bilirubin			.47
High	182 (1.61)	30 (1.40)	
Low	470 (4.16)	75 (3.51)	
Normal	5992 (53.08)	1146 (53.65)	
Not tested	4645 (41.15)	885 (41.43)	
Creatinine			.52
High	815 (7.22)	135 (6.32)	
Low	813 (7.20)	152 (7.12)	
Normal	5808 (51.45)	1109 (51.92)	
Not tested	3853 (34.13)	740 (34.64)	
Lymphocyte count			.003
High	122 (1.08)	40 (1.87)	
Low	2792 (24.73)	478 (22.38)	
Normal	4611 (40.85)	893 (41.81)	
Not tested	3764 (33.34)	725 (33.94)	
Red blood cell count			.001
High	108 (0.96)	27 (1.26)	
Low	2004 (17.75)	332 (15.54)	
Normal	4594 (40.69)	957 (44.80)	
Not tested	4583 (40.60)	820 (38.39)	
Hematocrit			.02
High	150 (1.33)	37 (1.73)	
Low	2378 (21.06)	394 (18.45)	
Normal	5031 (44.57)	995 (46.58)	
Not tested	3730 (33.04)	710 (33.24)	

Characteristic	Early NGS-tested[b] (n=11,289)	Late NGS-tested[c] (n=2136)	Early NGS-tested vs late NGS-tested, *P* value[d]
Platelet count			.04
High	855 (7.57)	183 (8.57)	
Low	233 (2.06)	38 (1.78)	
Normal	5372 (47.59)	1064 (49.81)	
Not tested	4829 (42.78)	851 (39.84)	
White blood cell count			.04
High	1837 (16.27)	329 (15.40)	
Low	162 (1.44)	33 (1.54)	
Normal	4811 (42.62)	979 (45.83)	
Not tested	4479 (39.68)	795 (37.22)	
Hemoglobin, whole blood			.0004
High	111 (0.98)	30 (1.40)	
Low	2564 (22.71)	405 (18.96)	
Normal	4987 (44.18)	1010 (47.28)	
Not tested	3627 (32.13)	691 (32.35)	

[a]NGS: next-generation sequencing.

[b]Patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy.

[c]Patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy.

[d]Two-sided *t* test for continuous variables; chi-square or Fisher exact test (where expected cell size <5) for categorical variables.

[e]ALP: alkaline phosphatase.

[f]ALT: alanine transaminase.

[g]AST: aspartate aminotransferase.

Over the 1000 bootstrap samples over the training data D1, an average of 135 and 89 features were identified by the LASSO models for the ever versus never and early versus late NGS-tested groups, respectively. These variables were then entered into an LR model using the testing set. The final model was established after removing any nonsignificant interaction terms, as explained earlier in the study methods. Details of the model fit statistics are shown in Table S3 in Multimedia Appendix 1. All main effects identified from the modeling for each group are shown in Figures 3-9.

There were lower odds of ever receiving NGS testing among patients with later age at initial diagnosis, bilirubin not tested, worse ECOG performance status, treated in geographies under the MolDX Program, a total higher number of genetic tests received, had only public insurance, and who were of Black or African American race as compared with those who were never tested. Patients who were obese, had a later year of initial NSCLC diagnosis, were from larger practices, had evidence of PD-L1 testing, no results for platelet testing, no history of smoking, had stage II disease, and were treated in a MAC region other than JH Novitas or J6 NGS had higher odds of ever receiving NGS-based testing.

For early versus late NGS testing (Figures 10-17), there were greater odds of receiving early NGS-based testing among patients with a later year of initial NSCLC diagnosis, who had no history of smoking, who were in later drug period approval periods, had a PD-L1 test, treated in the MAC J8 WPS, and who had no other biomarker tests or inconclusive testing.

Figure 3. Forest plot of ever versus never NGS-tested: variables determined by a logistic regression model from variables preselected by a LASSO model: clinical care and demographic variables. Ever NGS-tested: patients in the overall study cohort with evidence of NGS-based biomarker testing in the database; never NGS-tested: patients in the overall study cohort with no evidence of NGS-based biomarker testing. Index year: year of index diagnosis; LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio.

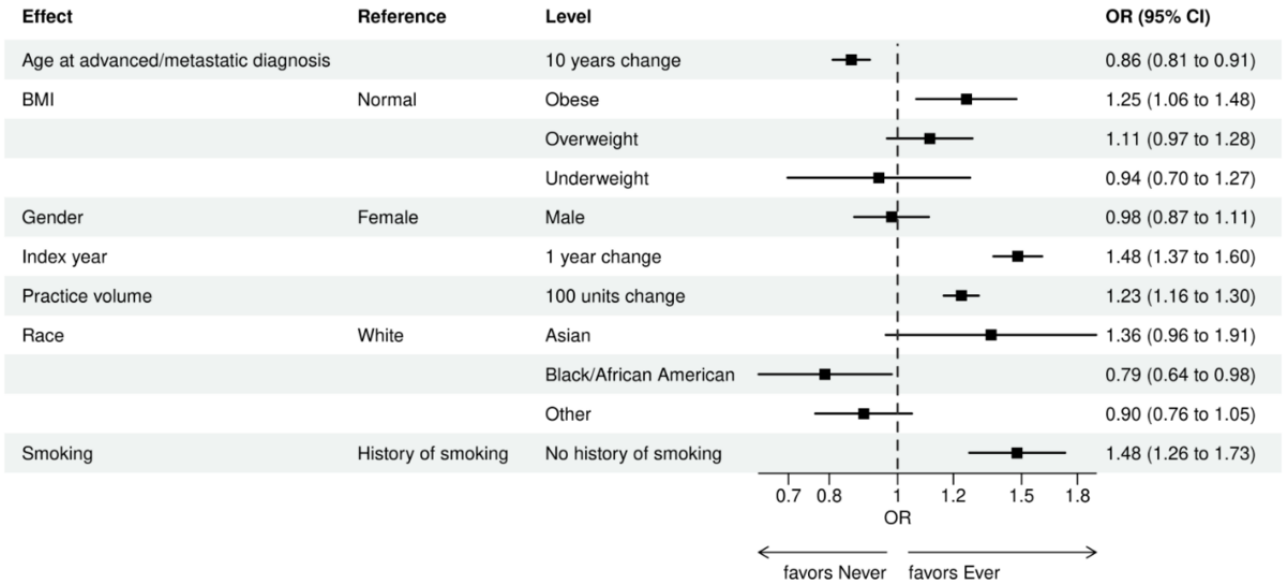


Figure 4. Forest plot of ever versus never NGS-tested: variables determined by a logistic regression model from variables preselected by a LASSO model: ECOG performance status and stage. Ever NGS-tested: patients in the overall study cohort with evidence of NGS-based biomarker testing in the database; never NGS-tested: patients in the overall study cohort with no evidence of NGS-based biomarker testing. ECOG: Eastern Cooperative Oncology Group; LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio.

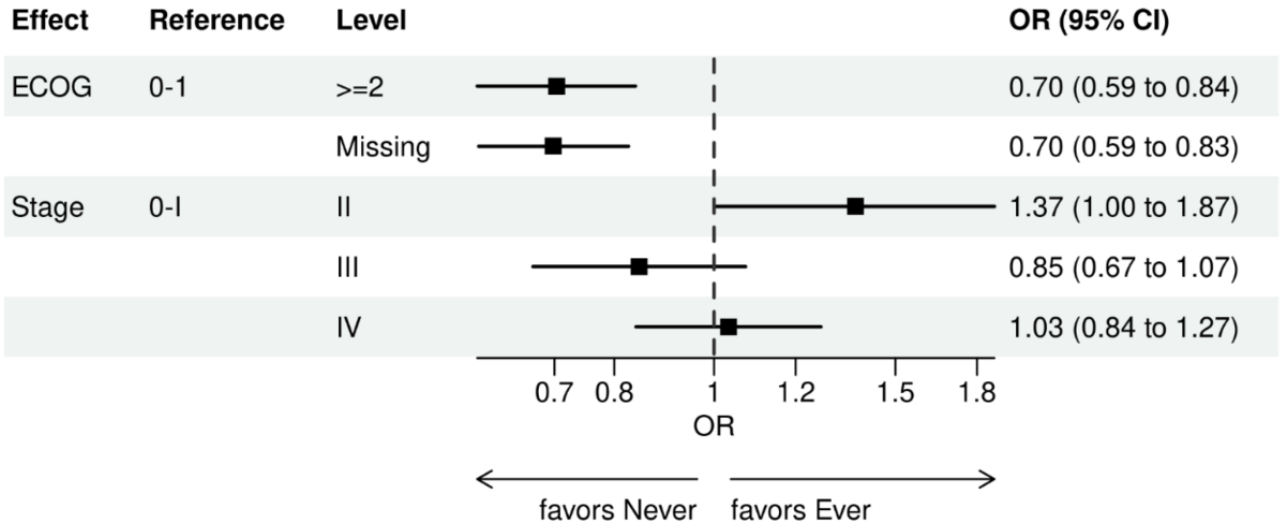


Figure 5. Forest plot of ever versus never NGS-tested: variables determined by a logistic regression model from variables preselected by a LASSO model: biomarkers and MolDX region. Ever NGS-tested: patients in the overall study cohort with evidence of NGS-based biomarker testing in the database; never NGS-tested: patients in the overall study cohort with no evidence of NGS-based biomarker testing. LASSO: least absolute shrinkage and selection operator; MolDX: Molecular Diagnostics Services; NGS: next-generation sequencing; OR: odds ratio; PDL1: programmed death ligand 1.

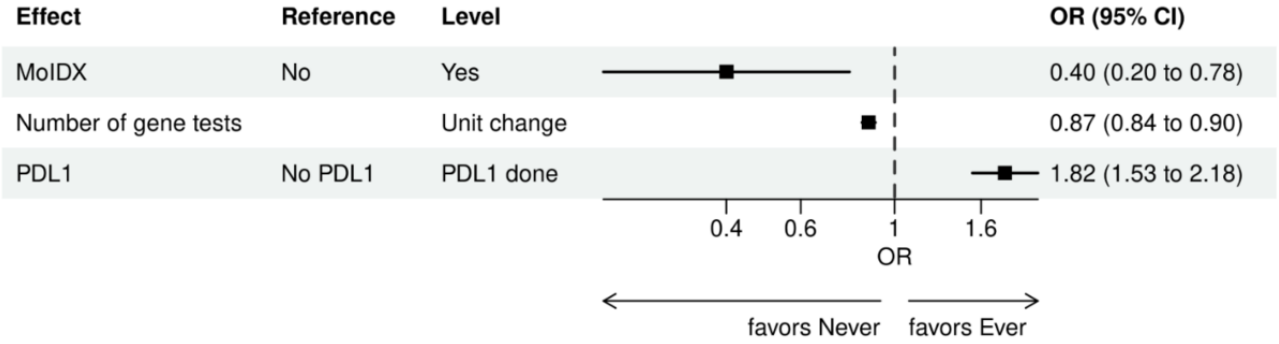


Figure 6. Forest plot of ever versus never NGS-tested: variables determined by a logistic regression model from variables preselected by a LASSO model: insurance. Ever NGS-tested: patients in the overall study cohort with evidence of NGS-based biomarker testing in the database; never NGS-tested: patients in the overall study cohort with no evidence of NGS-based biomarker testing. LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio.

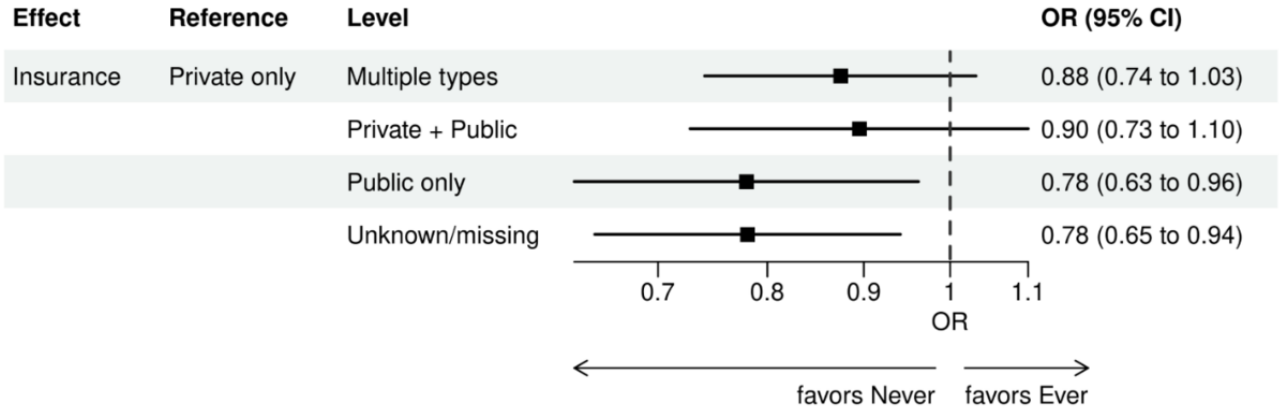


Figure 7. Forest plot of ever versus never NGS-tested: variables determined by a logistic regression model from variables preselected by a LASSO model: NCCN guidelines. Ever NGS-tested: patients in the overall study cohort with evidence of NGS-based biomarker testing in the database; never NGS-tested: patients in the overall study cohort with no evidence of NGS-based biomarker testing. LASSO: least absolute shrinkage and selection operator; NCCN: National Comprehensive Cancer Network; NGS: next-generation sequencing; OR: odds ratio.

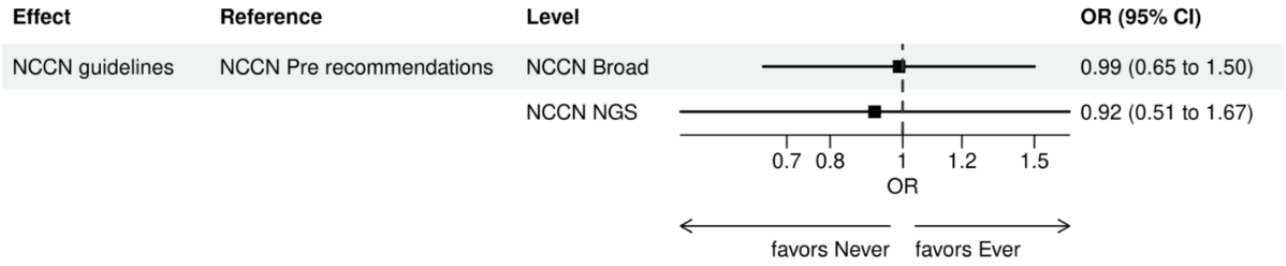


Figure 8. Forest plot of ever versus never NGS-tested: variables determined by a logistic regression model from variables preselected by a LASSO model: laboratory values. Ever NGS-tested: patients in the overall study cohort with evidence of NGS-based biomarker testing in the database; never NGS-tested: patients in the overall study cohort with no evidence of NGS-based biomarker testing. ALP: alkaline phosphatase; ALT: alanine transaminase; AST: aspartate aminotransferase; LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio.

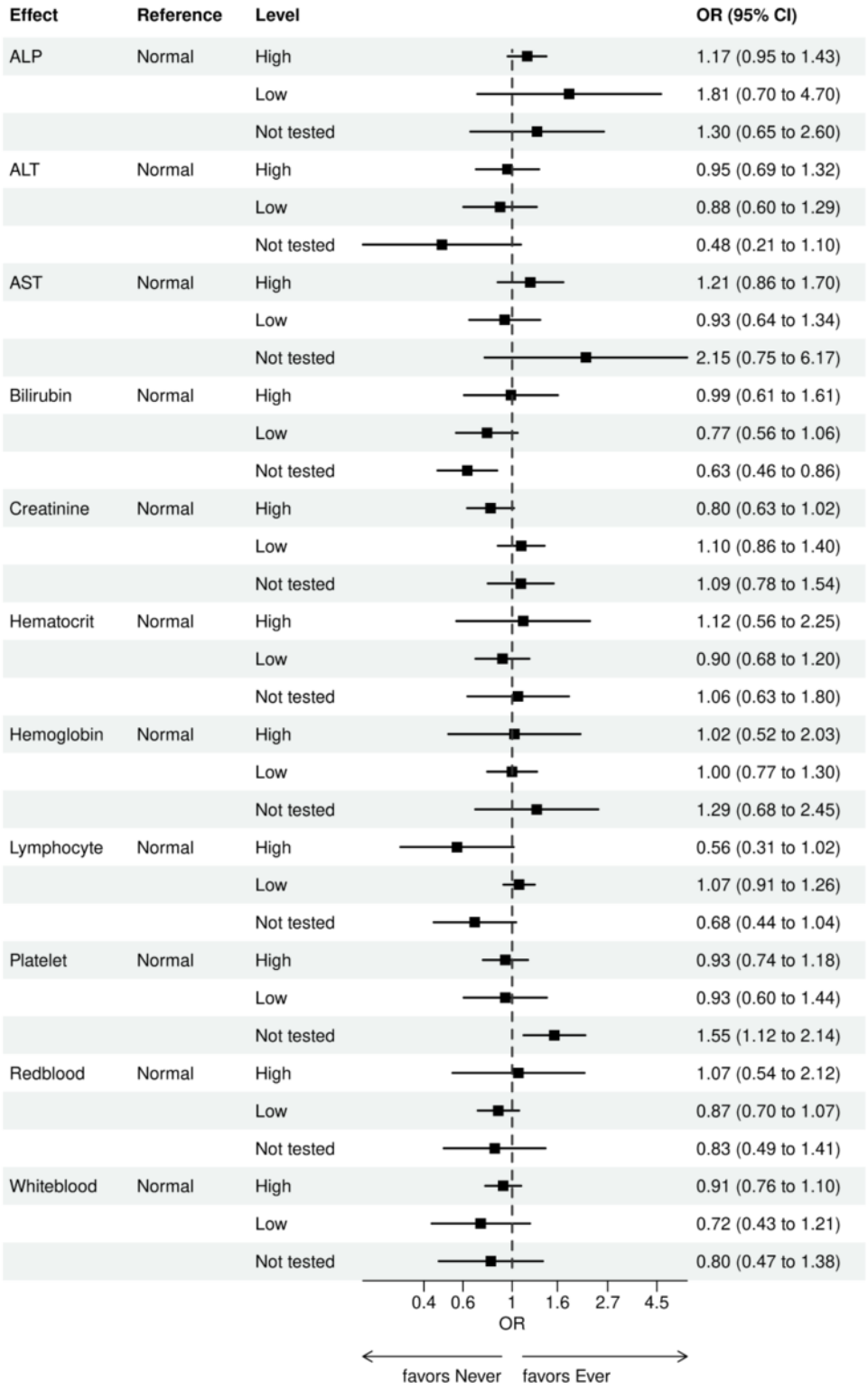


Figure 9. Forest plot of ever versus never NGS-tested: variables determined by a logistic regression model from variables preselected by a LASSO model: geographic region. Geographic regions reflect Medicare Administration Contractors. Ever NGS-tested: patients in the overall study cohort with evidence of NGS-based biomarker testing in the database; never NGS-tested: patients in the overall study cohort with no evidence of NGS-based biomarker testing. LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio.

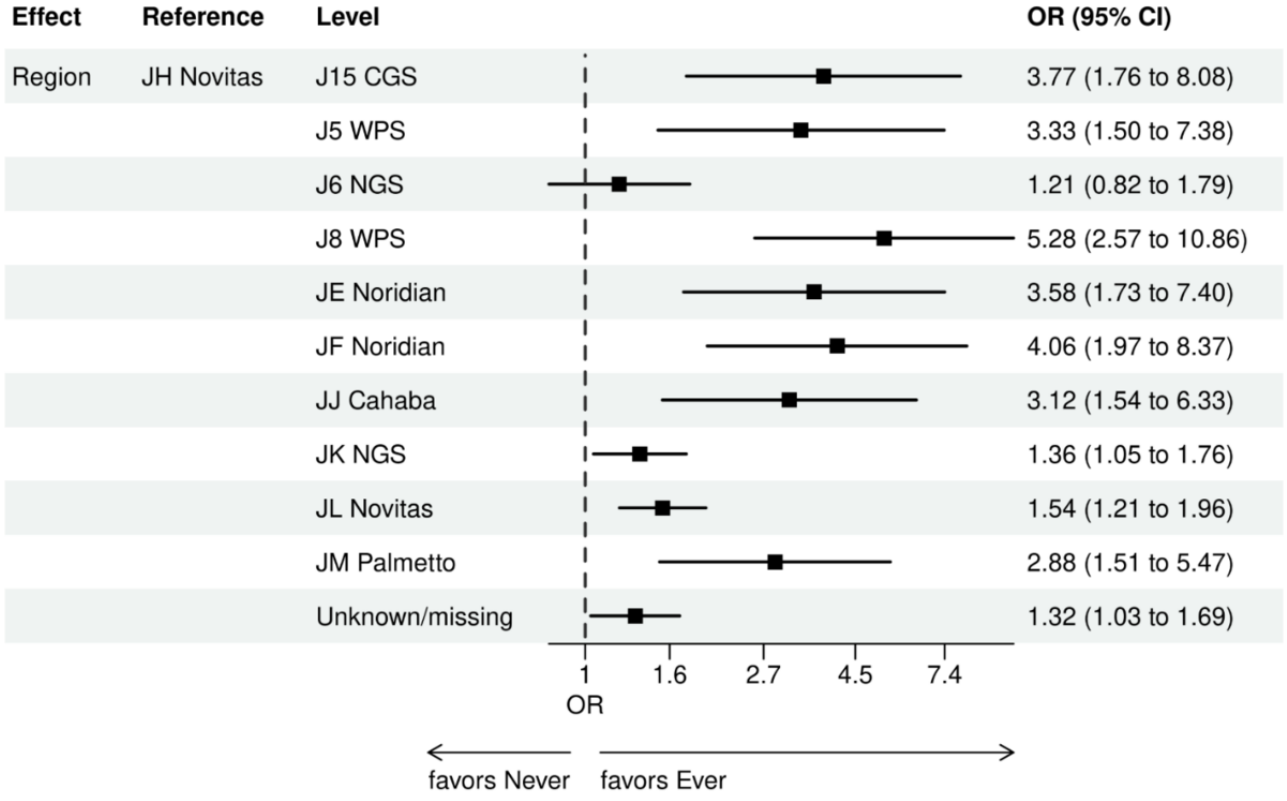


Figure 10. Forest plots early versus late NGS-tested: variables determined by logistic regression from variables preselected by a LASSO model: clinical care and demographic variables. Early NGS-tested: patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy; late NGS-tested: patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy. Index year: year of index diagnosis; LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio.

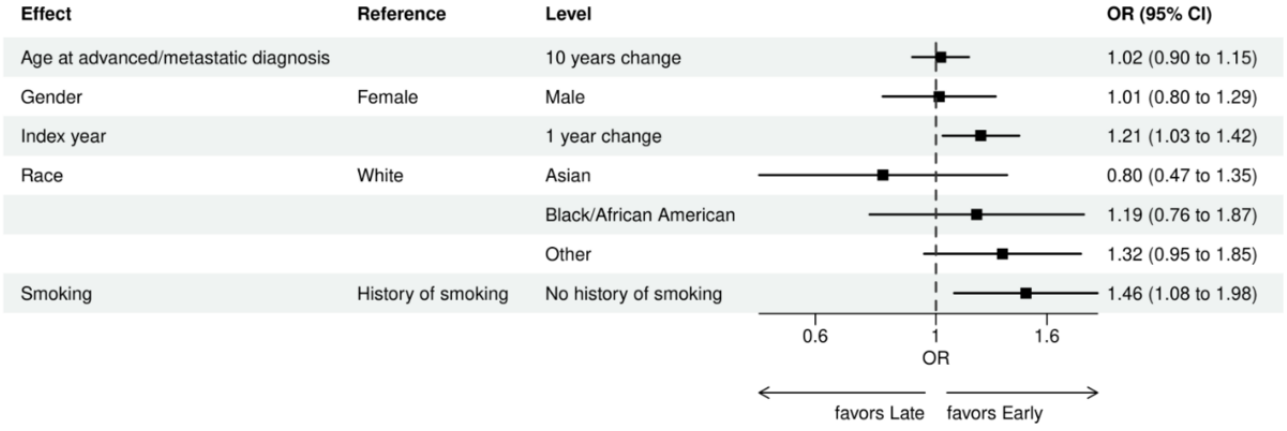


Figure 11. Forest plots early versus late NGS-tested: variables determined by logistic regression from variables preselected by a LASSO model: ECOG performance status and stage. Early NGS-tested: patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy; late NGS-tested: patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy. ECOG: Eastern Cooperative Oncology Group; LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio.

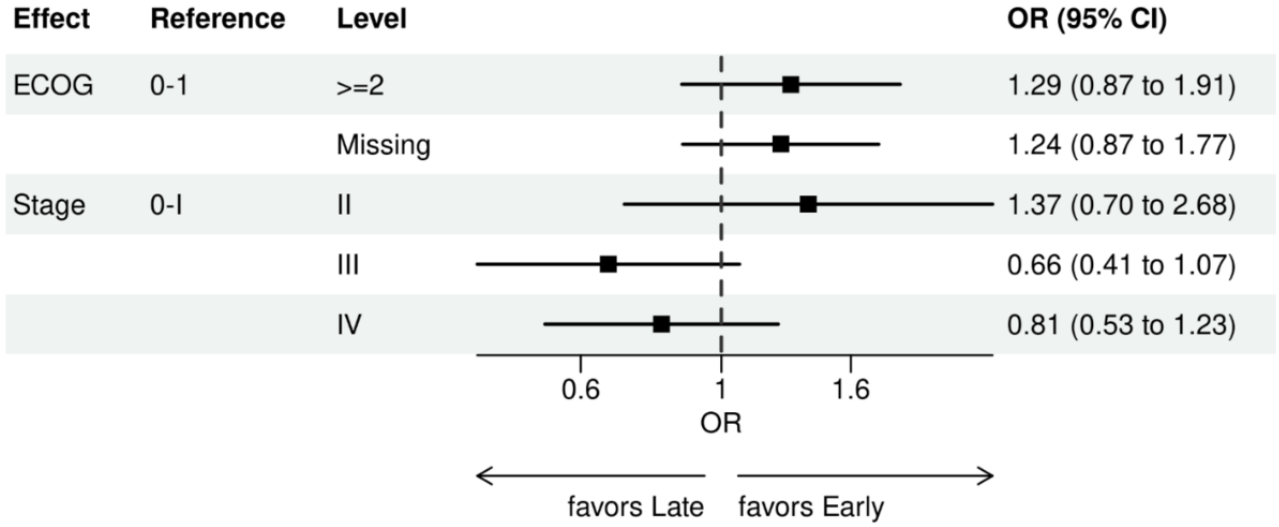


Figure 12. Forest plots early versus late NGS-tested: variables determined by logistic regression from variables preselected by a LASSO model: biomarkers. Early NGS-tested: patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy; late NGS-tested: patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy. LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio; PDL1: programmed death ligand 1.

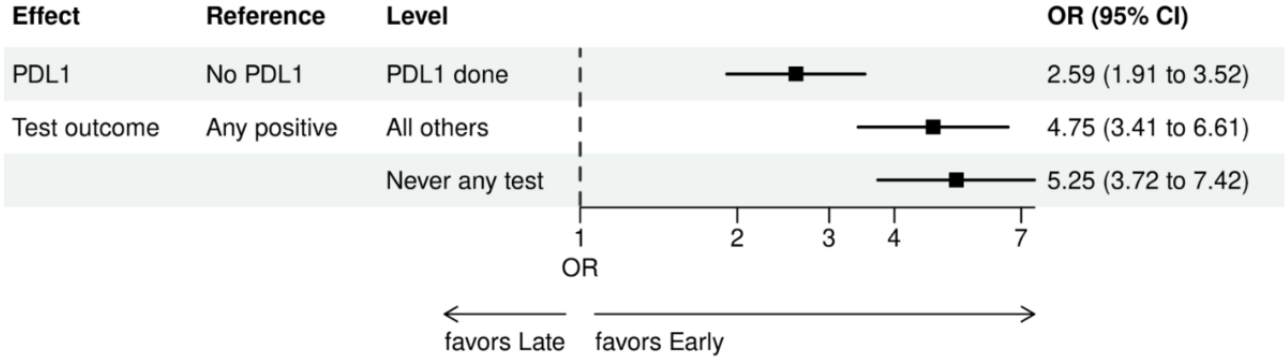


Figure 13. Forest plots early versus late NGS-tested: variables determined by logistic regression from variables preselected by a LASSO model: insurance. Early NGS-tested: patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy; late NGS-tested: patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy. LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio.

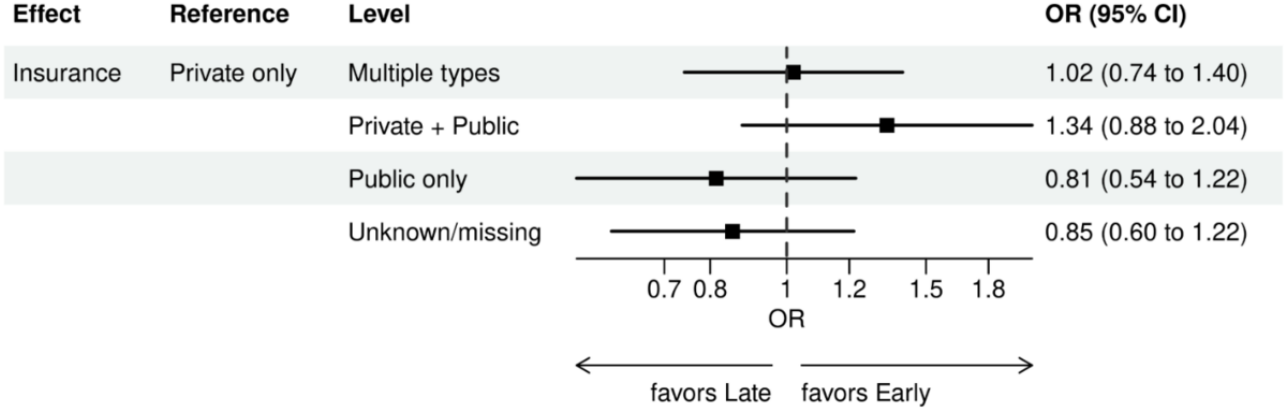


Figure 14. Forest plots early versus late NGS-tested: variables determined by logistic regression from variables preselected by a LASSO model: NCCN guidelines. Early NGS-tested: patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy; late NGS-tested: patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy. LASSO: least absolute shrinkage and selection operator; NCCN: National Comprehensive Cancer Network; NGS: next-generation sequencing; OR: odds ratio.

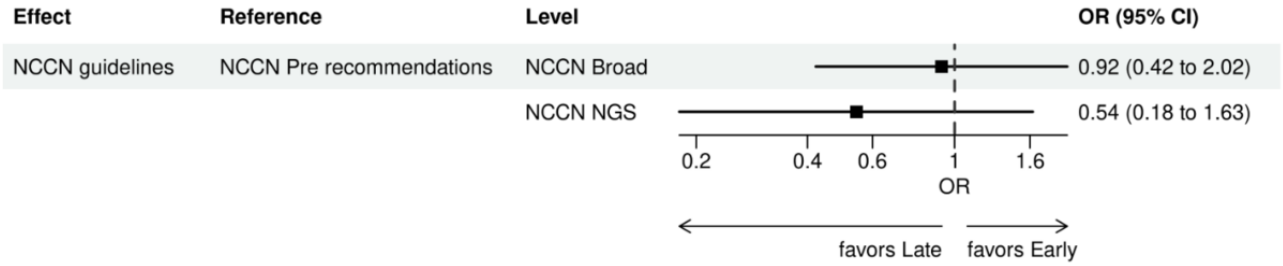


Figure 15. Forest plots early versus late NGS-tested: variables determined by logistic regression from variables preselected by a LASSO model: laboratory values. Early NGS-tested: patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy; late NGS-tested: patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy. ALT: alanine transaminase; LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio.

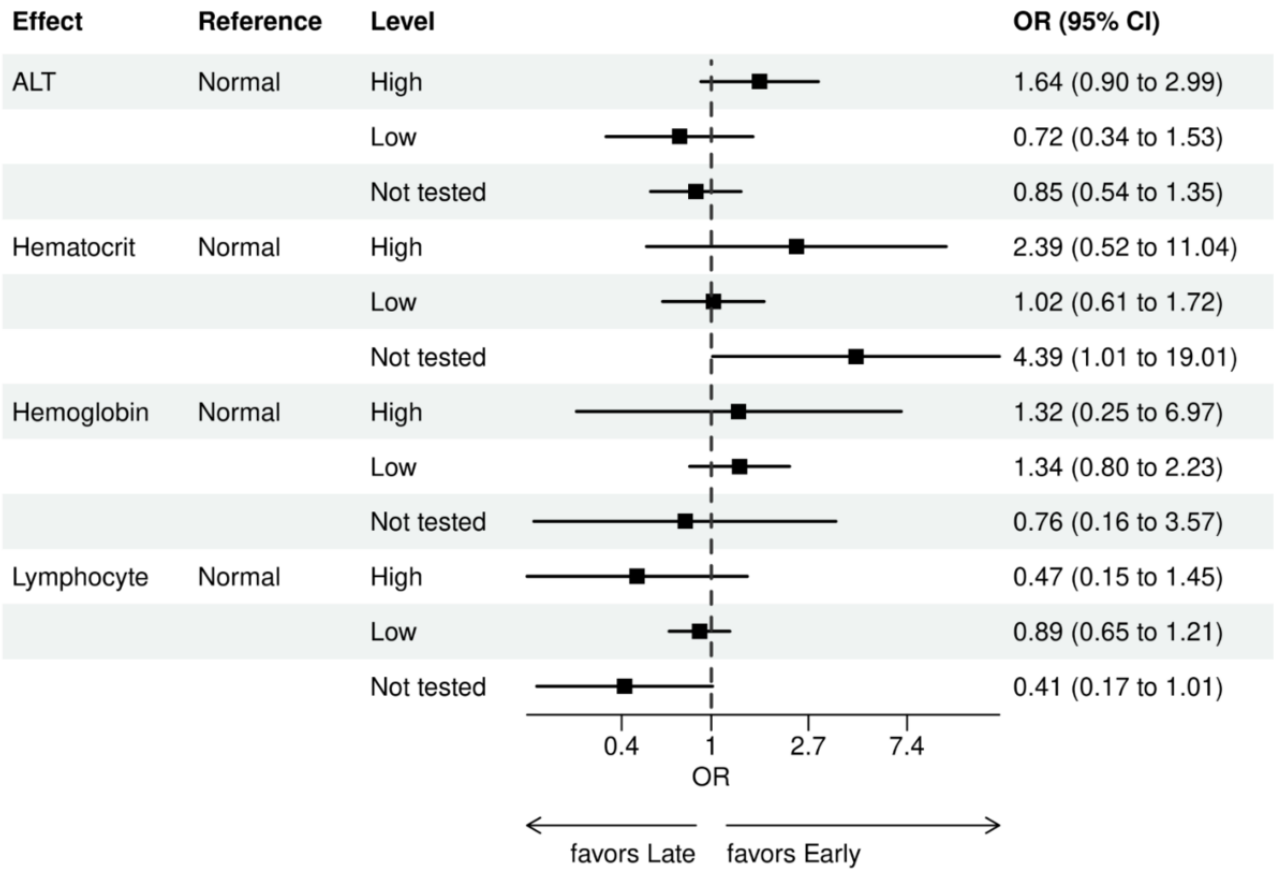


Figure 16. Forest plots early versus late NGS-tested: variables determined by logistic regression from variables preselected by a LASSO model: geographic region. Geographic regions reflect Medicare Administration Contractors. Early NGS-tested: patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy; late NGS-tested: patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy. LASSO: least absolute shrinkage and selection operator; NGS: next-generation sequencing; OR: odds ratio.

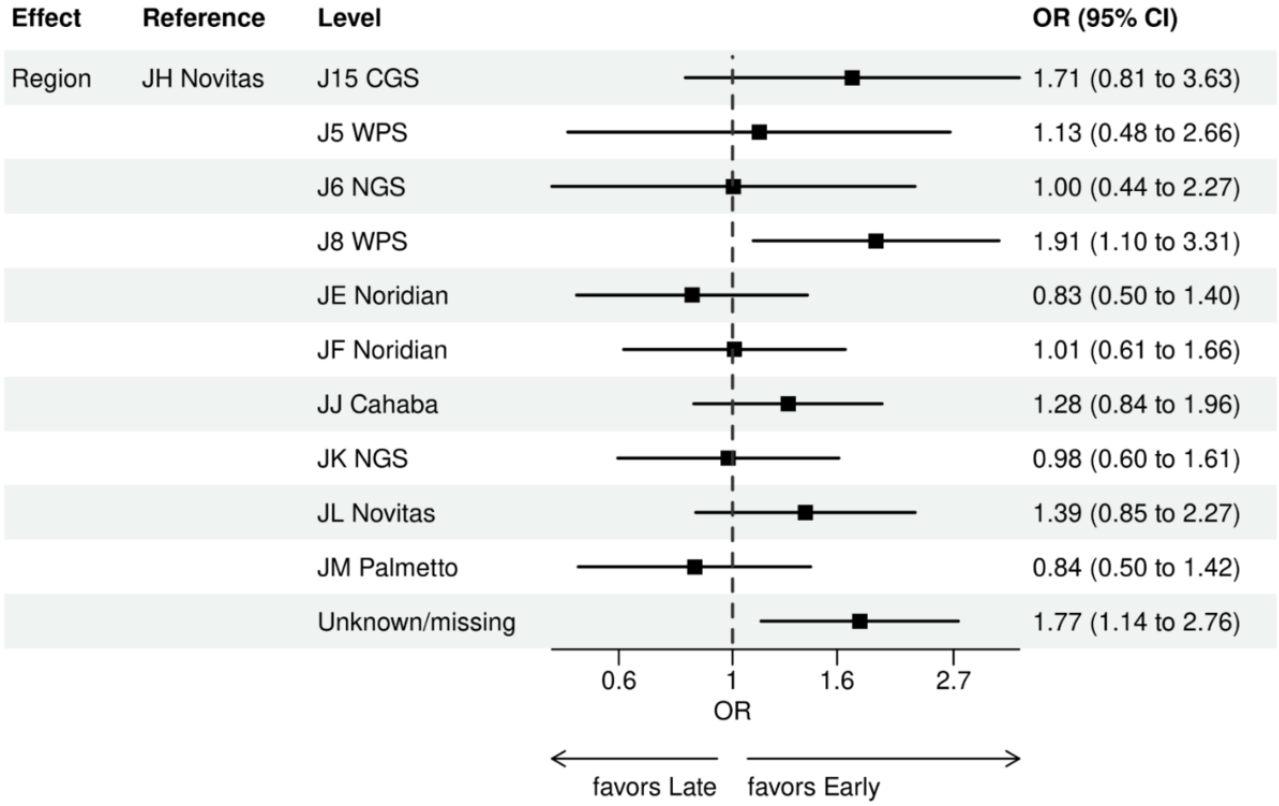
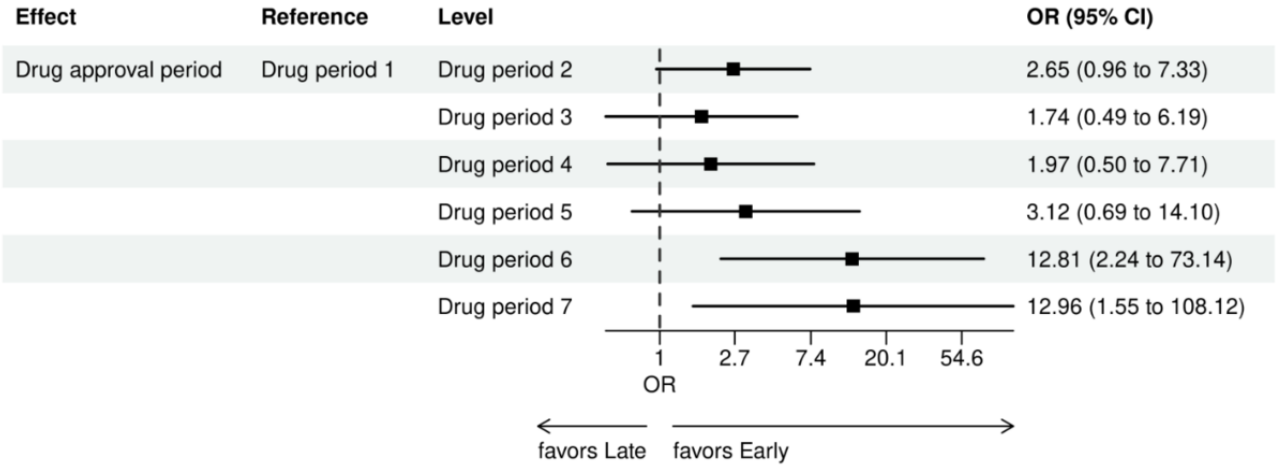


Figure 17. Forest plots early versus late NGS-tested: variables determined by logistic regression from variables preselected by a LASSO model: time period variables. Early NGS-tested: patients in the ever NGS-tested group whose first or only NGS-based test occurred prior to the start of first-line therapy through day 7 of first-line therapy; late NGS-tested: patients in the ever NGS-tested group whose first NGS-based test occurred 8 days or later after the start of first-line therapy. ALK: anaplastic lymphoma kinase; BRAF: V-Raf murine sarcoma viral oncogene homolog B; EGFR: epidermal growth factor receptor; KRAS: Kirsten rat sarcoma virus; LASSO: least absolute shrinkage and selection operator; MET: mesenchymal epithelial transition; NGS: next-generation sequencing; NTRK: neurotrophic tyrosine receptor kinase; OR: odds ratio; RET: rearranged during transfection; ROS1: c-ros oncogene 1; period 1: January 1-August 25, 2011 (EGFR drugs only); period 2: August 26, 2011-March 10, 2016 (EGFR+ALK); period 3: March 11, 2016-June 21, 2017 (EGFR+ALK+ROS1); period 4: June 22, 2017-November 25, 2018 (EGFR+ALK+ROS1+BRAF); period 5: November 26, 2018-May 5, 2020 (EGFR+ALK+ROS1+BRAF+NTRK); period 6: May 6, 2020-May 26, 2021 (EGFR+ALK+ROS1+BRAF+NTRK+MET+RET); period 7: May 27, 2021 and later (EGFR+ALK+ROS1+BRAF+MET+NTRK+RET+KRAS).



Discussion

Overview

This study applied machine learning methods and traditional statistical tools that identified several factors that were significantly associated with not only receiving NGS-based testing but also receiving the testing early when there is a potential for early intervention with targeted therapies. Factors associated with both ever having NGS testing as well as early NGS testing included later year of NSCLC diagnosis, no history of smoking, and evidence of PD-L1 testing. These factors were consistent with the hypothesized direction of candidate variables, as NGS-based testing has been increasing over time, and it was not unexpected that the rate of testing has increased in recent years [4,25]. In addition, consistent with the hypothesized direction of these relationships, patients without a smoking history were more likely to undergo NGS-based testing. The lack of environmental causal factors would lead one to seek other explanations for the onset of lung cancer, including certain genomic abnormalities, which are frequently observed among nonsmokers with lung cancer [26]. PD-L1 testing is generally conducted alongside the NGS test and was only available in later years, so the observation of these relationships was also not unexpected.

Principal Findings

Factors associated with a greater chance of never receiving NGS testing included older age, lower ECOG performance status, Black race, higher number of single-gene tests, public insurance, and treatment in a geography associated with MolDX Program adoption. Patient age and public insurance are factors that are closely related. Patients aged 65 years and older generally have Medicare coverage, whereas younger patients will have private insurance. The median age of lung cancer diagnosis is 71 years [27], and it is highly likely that a younger patient presenting with NSCLC could raise questions about the genomic aspects of the disease that should be investigated as a result be associated with a higher likelihood of receiving early NGS-based testing as noted in the published literature [28]. Importantly, patient race, similar to prior research [7], remains a significant factor that continues to demonstrate the lack of equity in receipt of NGS-based testing. Of all factors evaluated in this study, racial inequity cannot be explained by any reasonable clinical factors and requires immediate attention by the health care community.

Several factors that did not have a clear association with NGS-based testing were those that also did not have a hypothesized direction associated with a potential relationship. While blood test results may have captured some aspect of well-being, there was no consistent relationship identified. Similarly, while patients with better performance status were more likely to receive NGS-based testing, this relationship was not strong, and the factor was not among those with the highest importance scores observed in this study. Therefore, this study suggests that these factors are likely not largely factored into a decision to receive NGS-based testing and could be why little data were observed in the published literature related to these factors.

The roles of the MolDX program and the MAC region are unclear. The emergence of MAC region J8 WPS as a predictive factor for greater odds of receiving early NGS testing and both JH Novitas and J6 NGS at lower odds of receiving any NGS testing could be an artifact of a large dataset with multiple subgroups or could reflect underlying factors related to this region that could not be explored, given the available data in the electronic data used for this study. Additionally, the timing of MolDX program adoption was not taken into account, so the patients in these regions could have had the decision made at a time that was unrelated to this variable (“yes” or “no”). Other geographic factors such as distance to a clinic, access to testing resources, and site of care could certainly have played a role as well; therefore, the relationship with MolDX should not be overinterpreted. Additionally, not all patients in these regions had Medicare coverage, so there is a great deal of uncertainty in these variables. A study with more comprehensive variables related to patient care in these regions would be needed to come to any clear conclusion about these relationships.

Limitations

First, this study is based on real-world data. The Flatiron Health deidentified data, as with most other electronic health record–based datasets, do not contain all potentially relevant variables to investigate all aspects of the complex question of NGS-based testing. Factors such as tissue availability, tissue quality, a patient crisis requiring immediate care, and other health care system–related factors were not recorded and may be additional factors that could impact access and receipt of NGS-based testing. The availability of these data, however, would not invalidate the factors that were observed in this study. Second, there were some patients who could have received NGS-based testing at an early stage diagnosis who were not included in this study due to our eligibility criteria, requiring testing within the time frame of advanced or metastatic diagnosis. Therefore, this study may not be generalizable to those diagnosed and tested at earlier stages of the disease. Third, as with all real-world data sources, missingness is a potential issue. However, the rates of NGS-based testing in this study are very similar to other estimates from different data sources, which provides confidence in the outcome variable assessed within the database used for this study [10]. Finally, when evaluating predictive models, a cutoff of 0.5 was applied to the predicted probability of events. While this may result in a suboptimal trade-off of specificity versus sensitivity for certain models (eg, for modeling “early vs late” NGS testing, it resulted in low specificities of ~20% and very high sensitivity of ~98%; Table S2 in Multimedia Appendix 1), the objective of this study was to identify predictors of NGS testing rather than optimizing predictive rules. The probability cutoffs could be further calibrated to strike a desired balance between false positives and negatives.

Conclusions

Despite the limitations of these data, this study reinforces the need to assure equity in access to NGS-based testing that has been observed in prior research. Black race is consistently associated with lower biomarker testing rates [7]. Other factors may be more associated with disease trajectory (eg, age, lower

ECOG performance status, and single-gene tests), emphasizing the flexibility needed in testing for those patients who may not be well enough for systemic therapy or who have an actionable biomarker previously identified. While efforts must be made to ensure all patients diagnosed with NSCLC have equal access to NGS-based testing early in the trajectory of the disease, there may be consideration for the specific patient needs in these cases.

Acknowledgments

This was an unfunded research project with employee time and materials provided by Eli Lilly and Company.

Data Availability

The data that support the findings of this study have been originated by Flatiron Health, Inc. Requests for data sharing by license or by permission for the specific purpose of replicating results in this manuscript can be submitted to dataaccess@flatiron.com.

Authors' Contributions

AJMB conceptualized the study, planned the methodology, investigated the study, and revised and edited the final manuscript. LMH conceptualized the study, planned the methodology, investigated, wrote the original manuscript, and revised and edited the final manuscript. PMK conceptualized the study, investigated the study, and revised and edited the final manuscript. IL planned the methodology, conducted the final analysis, validated and investigated the study, and revised and edited the final manuscript. ZK planned the methodology, investigated the study, conducted final analysis and validation, produced visualization, and revised and edited the final manuscript. DH investigated the study, conducted final analysis, validation, and data curation, created visualization, and revised and edited the final manuscript.

Conflicts of Interest

AJMB, IL, ZK, and LMH are employees of Eli Lilly and Company, and PMK was an employee of Loxo@Lilly, subsequently Eli Lilly and Company, when this work was conducted. DH is an employee of Syneos Health, an organization that receives funding from Eli Lilly and Company for analytic support services.

Multimedia Appendix 1
Details of study methods, model performance, and variable importance plots.
[DOCX File , 406 KB - cancer_v11i1e64399_app1.docx]

Multimedia Appendix 2
Computing partial effects from covariates from logistic regression models.
[DOCX File , 16 KB - cancer_v11i1e64399_app2.docx]

References

1. Mitchell CL, Zhang AL, Bruno DS, Almeida FA. NSCLC in the era of targeted and immunotherapy: what every pulmonologist must know. Diagnostics (Basel) 2023;13(6):1117 [FREE Full text] [doi: 10.3390/diagnostics13061117] [Medline: 36980426]
2. VanderLaan PA, Rangachari D, Majid A, Parikh MS, Gangadharan SP, Kent MS, et al. Tumor biomarker testing in non-small-cell lung cancer: a decade of change. Lung Cancer 2018;116:90-95 [FREE Full text] [doi: 10.1016/j.lungcan.2018.01.002] [Medline: 29413057]
3. NCCN Clinical Practice Guidelines in Oncology: non-small cell lung cancer version 3.2023. National Comprehensive Cancer Network. 2023. URL: https://www.nccn.org/guidelines/category_1 [accessed 2023-07-02]
4. Hess LM, Krein PM, Haldane D, Han Y, Sireci AN. Biomarker testing for patients with advanced/metastatic nonsquamous NSCLC in the United States of America, 2015 to 2021. JTO Clin Res Rep 2022;3(6):100336 [FREE Full text] [doi: 10.1016/j.jtocrr.2022.100336] [Medline: 35677681]
5. Sireci AN, Krein PM, Hess LM, Khan T, Willey J, Ayars M, et al. Real-world biomarker testing patterns in patients with metastatic non-squamous non-small cell lung cancer (NSCLC) in a US community-based oncology practice setting. Clin Lung Cancer 2023;24(5):429-436. [doi: 10.1016/j.cllc.2023.03.002] [Medline: 37080814]
6. Evangelist MC, Butrynski JE, Paschold JC, Ward PJ, Spira AI, Ali K, et al. Contemporary biomarker testing rates in both early and advanced NSCLC: results from the MYLUNG pragmatic study. J Clin Oncol 2023;41(16 Suppl):9109-9109. [doi: 10.1200/jco.2023.41.16_suppl.9109]
7. Bruno DS, Hess LM, Li X, Su EW, Patel M. Disparities in biomarker testing and clinical trial enrollment among patients with lung, breast, or colorectal cancers in the United States. JCO Precis Oncol 2022;6:e2100427. [doi: 10.1200/po.21.00427]
8. Steyerberg EW. Clinical Prediction Models. Washington DC: Springer; 2019.

9. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 2019;18(6):463-477 [FREE Full text] [doi: 10.1038/s41573-019-0024-5] [Medline: 30976107]
10. Robert NJ, Espirito JL, Chen L, Nwokeji E, Karhade M, Evangelist M, et al. Biomarker testing and tissue journey among patients with metastatic non-small cell lung cancer receiving first-line therapy in The US Oncology Network. Lung Cancer 2022;166:197-204 [FREE Full text] [doi: 10.1016/j.lungcan.2022.03.004] [Medline: 35313244]
11. Boehmer L, Basu Roy UK, Schrag J, Martin NA, Salinas GD, Coleman B, et al. Identifying barriers to equitable biomarker testing in underserved patients with NSCLC: A mixed-methods study to inform quality improvement opportunities. J Clin Oncol 2021;39(28 Suppl 123). [doi: 10.1200/jco.2020.39.28_suppl.123]
12. Mileham KF, Schenkel C, Bruinooge SS, Freeman-Daily J, Basu Roy U, Moore A, et al. Defining comprehensive biomarker-related testing and treatment practices for advanced non-small-cell lung cancer: results of a survey of U.S. oncologists. Cancer Med 2022;11(2):530-538 [FREE Full text] [doi: 10.1002/cam4.4459] [Medline: 34921524]
13. Burns L, Jani C, Radwan A, Omari OA, Patel M, Oxnard GR, et al. Implementation challenges and disparities in molecular testing for patients with stage IV NSCLC: perspectives from an urban safety-net hospital. Clin Lung Cancer 2023;24(2):e69-e77. [doi: 10.1016/j.cllc.2022.11.002] [Medline: 36464575]
14. Ma X, Long L, Moon S, Adamson BJS, Baxi SS. Comparison of population characteristics in real-world clinical oncology databases in the US: Flatiron Health, SEER, and NPCR. medRxiv Preprint posted online January 5, 2023. [doi: 10.1101/2020.03.16.20037143]
15. Birnbaum B, Nussbaum N, Seidl-Rathkopf K, Agrawal M, Estevez M, Estola E, et al. Model-assisted cohort selection with bias analysis for generating large-scale cohorts from the EHR for oncology research. ArXiv Preprint posted online on January 13, 2020 [FREE Full text]
16. Definition of human subjects research. National Institutes of Health. 2024. URL: https://grants.nih.gov/policy/humansubjects/research.htm [accessed 2024-03-07]
17. Bruno DS, Li X, Hess LM. Biomarker testing, targeted therapy and clinical trial participation by race among patients with lung cancer: a real-world Medicaid database study. JTO Clin Res Rep 2024;5(3):100643. [doi: 10.1016/j.jtocrr.2024.100643] [Medline: 38496377]
18. NCCN Clinical Practice Guidelines in Oncology. National Comprehensive Cancer Network. 2022. URL: https://www.nccn.org/guidelines/category_1 [accessed 2022-07-18]
19. Oncology (cancer)/hematologic malignancies approval notification. Food and Drug Administration. 2022. URL: https://www.fda.gov/drugs/resources-information-approved-drugs/oncology-cancer-hematologic-malignancies-approval-notifications [accessed 2022-07-01]
20. Experian. MAC Jurisdiction Map. 2016. URL: https://www.experian.com/blogs/healthcare/wp-content/uploads/2016/08/MAC-A-B-Jurisdiction-Map.jpg [accessed 2016-06-30]
21. MolDX® Program (Administered by Palmetto GBA). Palmetto-GBA. 2022. URL: https://palmettogba.com/palmetto/moldxv2.nsf [accessed 2025-05-30]
22. CMS. Medicare Administrative Contractors (MACs). 2023. URL: https://www.cms.gov/medicare/coding-billing/medicare-administrative-contractors-macs/whats-mac [accessed 2025-04-10]
23. Tang F, Ishwaran H. Random forest missing data algorithms. Stat Anal Data Min 2017;10(6):363-377. [doi: 10.1002/sam.11348] [Medline: 29403567]
24. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13, 2016; San Francisco, CA, United States. [doi: 10.1145/2939672.2939785]
25. Swaminathan A, Stergiopoulos S, Snider J, Fisher V, Castellanos E, Snow T. Changes over time in real-world next-generation sequencing (NGS) test use in patients (pts) with advanced non-small cell lung cancer (aNSCLC). Cancer Res 2021;81(13_Suppl):864. [doi: 10.1158/1538-7445.am2021-864]
26. Mack PC, Klein MI, Ayers KL, Zhou X, Guin S, Fink M, et al. Targeted next-generation sequencing reveals exceptionally high rates of molecular driver mutations in never-smokers with lung adenocarcinoma. Oncologist 2022;27(6):476-486. [doi: 10.1093/oncolo/oyac035] [Medline: 35298662]
27. Cancer stat facts: lung and bronchus cancer. National Cancer Institute. SEER. URL: https://seer.cancer.gov/statfacts/html/lungb.html [accessed 2024-03-20]
28. Wu X, Zhao J, Yang L, Nie X, Wang Z, Zhang P, et al. Next-generation sequencing reveals age-dependent genetic underpinnings in lung adenocarcinoma. J Cancer 2022;13(5):1565-1572. [doi: 10.7150/jca.65370] [Medline: 35371328]

Abbreviations

ALK: anaplastic lymphoma kinase
BRAF: V-Raf murine sarcoma viral oncogene homolog B
ECOG: Eastern Cooperative Oncology Group
EGFR: epidermal growth factor receptor

ICD: International Classification of Diseases
ICD-9: International Classification of Diseases, Ninth Revision
ICD-10: International Statistical Classification of Diseases, Tenth Revision
KRAS: Kirsten rat sarcoma virus
LASSO: least absolute shrinkage and selection operator
LR: logistic regression
MAC: Medicare Administrative Contractor
MET: mesenchymal epithelial transition
MolDX: Molecular Diagnostics Services
NCCN: National Comprehensive Cancer Network
NGS: next-generation sequencing
NSCLC: non–small cell lung cancer
NTRK: neurotrophic tyrosine receptor kinase
PD-L1: programmed death ligand 1
PLR: penalized logistic regression
RET: rearranged during transfection
ROS1: c-ros oncogene 1
XGBoost: extreme gradient boosting

Edited by N Cahill; submitted 16.07.24; peer-reviewed by P Wu; comments to author 17.12.24; revised version received 26.02.25; accepted 26.03.25; published 11.06.25.

Please cite as:
Brnabic AJM, Lipkovich I, Kadziola Z, He D, Krein PM, Hess LM
Next-Generation Sequencing–Based Testing Among Patients With Advanced or Metastatic Nonsquamous Non–Small Cell Lung Cancer in the United States: Predictive Modeling Using Machine Learning Methods
JMIR Cancer 2025;11:e64399
URL: https://cancer.jmir.org/2025/1/e64399
doi:10.2196/64399
PMID:

Leveraging Patient-Reported Outcome Measures for Optimal Dose Selection in Early Phase Cancer Trials

Bill Byrom[1], PhD; Anthony Everhart[1], MD; Paul Cordero[2], PhD; Chris Garratt[3], MBChB; Tim Meyer[4], MBBS, PhD

[1]Science and Medicine, Signant Health, 785 Arbor Way, Blue Bell, PA, United States

[2]Patient Informed Development & Health Value Translation, Sanofi Pharmaceuticals, Reading, United Kingdom

[3]New Medicines Development, Orion Pharma, Nottingham, United Kingdom

[4]Research Department of Oncology, University College London, London, United Kingdom

Corresponding Author:
Bill Byrom, PhD
Science and Medicine, Signant Health, 785 Arbor Way, Blue Bell, PA, United States

Abstract

While patient-reported outcome measures are regularly incorporated into phase 3 clinical trials, they have been infrequently used in early phase trials. However, the patient's perspective is vital to fully understanding dose toxicity and selecting an optimal dose. This viewpoint paper reviews the rationale for and practical approach to collecting patient-reported outcome data in early phase oncology drug development and the rationale for electronic collection.

(JMIR Cancer 2025;11:e64611) doi:10.2196/64611

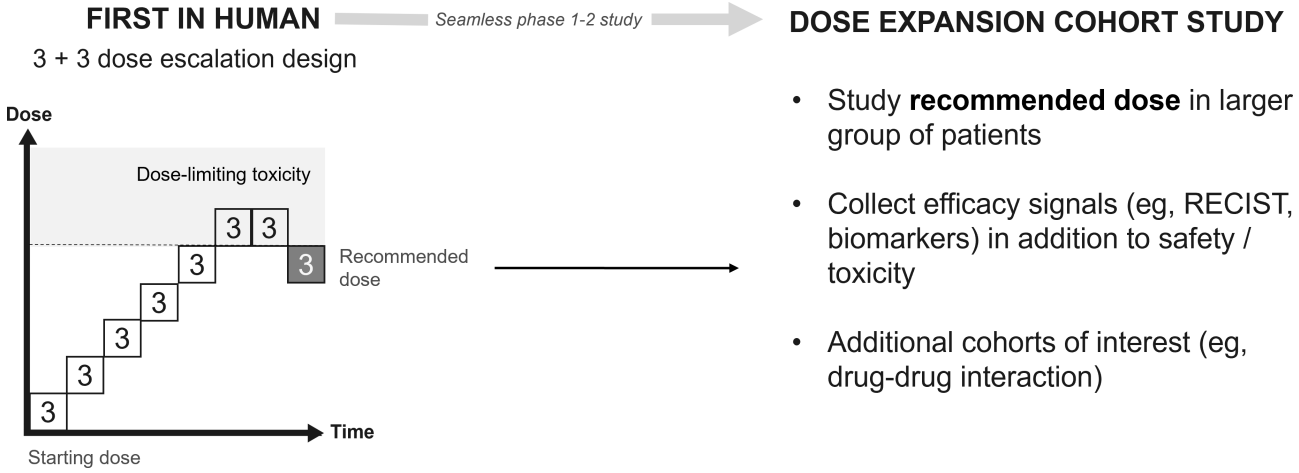
KEYWORDS

clinical trials; early phase; dose finding; patient-reported outcome; PRO; electronic patient-reported outcome; ePRO; PRO-CTCAE; adverse events; tolerability; optimal dose; cancer trials; dose toxicity; oncology; drug development; electronic collection; dose level; pharmacodynamic; cytotoxic chemotherapy drugs; cytotoxic; chemotherapy drug; life-threatening disease; Common Terminology Criteria for Adverse Events

Dose Finding in Oncology Drug Development

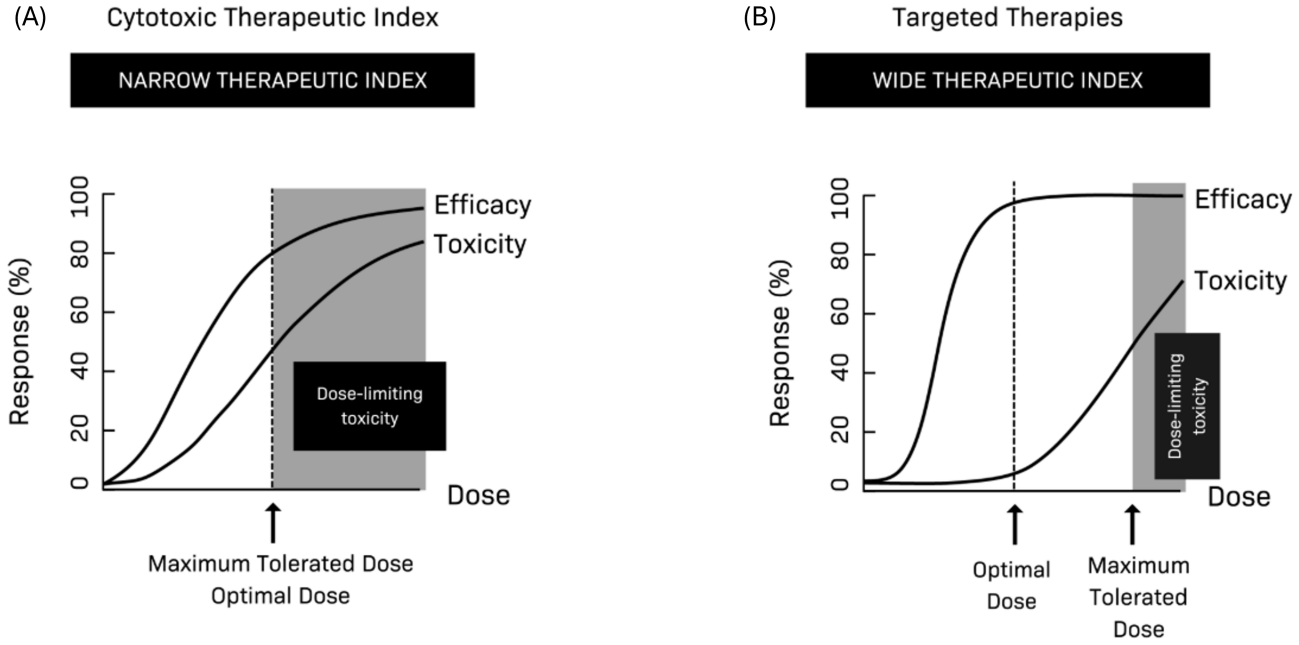
Traditional dose findings in oncology clinical drug development have focused on determining the maximum tolerated dose, assuming that the efficacy-dose relationship follows a steep monotonic increasing curve. By this assumption, the recommended dose is defined as the highest dose in which dose-limiting toxicity is not observed. Typical early phase programs (Figure 1) often use a 3 + 3 dose escalation design in which subsequent cohorts of 3 patients are studied, each receiving a higher dose than the last. Dose levels often follow a modified Fibonacci sequence whereby dose increments become smaller as the dose increases [1]. When dose-limiting toxicity is observed in at least one patient, the dose level is repeated in a further cohort of 3 patients, and if dose-limiting toxicity is observed again, further escalation stops, identifying the previous dose as the recommended maximum tolerated dose to take forward. Further study of the recommended dose is achieved, often using a seamless phase 1-2 design, by recruiting an additional larger group of patients into a dose expansion study. The primary end point in an expansion cohort is usually to determine efficacy, most frequently according to the radiological response rate. Additionally, further safety data is gathered, and pharmacodynamic markers may also be developed.

Figure 1. Traditional early phase oncology dose-finding studies. RECIST: Response Evaluation Criteria in Solid Tumors.



This approach has been acceptable for cytotoxic chemotherapy drugs due to their steep dose-response relationships, their limited drug target specificity, and the willingness of patients to trade off substantial toxicity to treat serious, life-threatening diseases [2]. However, it may lead to the recommendation of higher doses and a suboptimal tolerability profile when used in dose finding for modern, more targeted oncology drugs such as kinase inhibitors and monoclonal antibodies (Figure 2). In these cases, the wider therapeutic index means that a range of doses may show relevant efficacy, and doses below the maximum tolerated dose may have similar efficacy with reduced toxicity [3]. This can be particularly important because targeted therapies are often taken for much longer periods, during which lower grade, persistent symptomatic toxicities can present a greater challenge to patients [2]. Dose finding limitations have been illustrated in 26% of Food and Drug Administration (FDA)–approved kinase inhibitors (2001‐2015) requiring postmarketing requirements/commitments to study alternative doses [4].

Figure 2. Dose-response relationships and optimal dose selection for cytotoxic chemotherapy drugs and targeted therapies.



Better characterizing dose response to optimize dose finding has been underlined by the FDA's Project Optimus, which aims to reform how doses are selected in oncology clinical trials, with a particular focus on maximizing efficacy and optimizing safety and tolerability [5]. This led to their subsequent guidance on dose optimization for new cancer treatments [2]. Studying more dose levels in the dose expansion study may be one approach to enable this and may better enable characterization of the dose-response relationship, albeit qualitatively given the likely small cohort sizes.

Understanding Tolerability

Tolerability is defined in good clinical practice as "the degree to which overt adverse effects can be tolerated by the subject" [6]. In oncology, assessment of tolerability typically comprises clinician-reported treatment-related adverse events (AEs) using, for example, the Common Terminology Criteria for Adverse Events (CTCAE) [7], along with other data such as dose modifications, discontinuations and interruptions, safety biomarkers, hospitalization, and death [8]. However, these tools and data fail to fully account for the patient's perspective or to

fully measure the impact of AEs on the patient's activities and quality of life. Many studies comparing physician and patient reports of treatment-related AEs have consistently shown underreporting and reduced severity rating in physician interpretations compared to patient reports [9-16]. This represents a challenge for drug developers in accurately quantifying the dose-toxicity relationship and limits the ability to define optimal doses, leading to a greater risk of exposing greater numbers of patients to doses that are too high, potentially resulting in increased discontinuation and a less favorable safety profile.

For example, measuring the frequency of individual AEs reported by early phase patients using the full Patient-Reported Outcomes version of the CTCAE (PRO-CTCAE) item bank [17], Veitch and colleagues [9] evaluated the associated frequency of reporting of the same AEs by physicians using the CTCAE. They found that all 50 AEs reported by at least 10% of patients undergoing cancer treatment (n=243) were consistently underreported by physicians using the CTCAE, in some cases markedly. For example, 9 AEs were identified at least 50 times less frequently by physicians: decreased libido (31.4% vs 0.1%), palpitations (14.7% vs 0.1%), wheezing (14.5% vs 0.2%), voice alteration (14.1% vs 0.2%), hiccups (13.9% vs 0.1%), hyperhidrosis (23.9% vs 0.4%), vaginal dryness (11.0% vs 0.1%), nail ridging (10.0% vs 0.2%), and urinary incontinence (10.0% vs 0.2%). Further, 19 CTCAE items were reported 1% or less of the time by physicians, compared to 10%-31.4% by patients.

A further study in 1933 patients with a variety of oncology conditions treated in various routine care settings reported an underestimation of AE severity by clinicians in comparison to patient reports [10]. The frequency of symptoms assessed as moderate or severe by patients and physicians, respectively, were pain (67% vs 47%), fatigue (71% vs 54%), generalized weakness (65% vs 47%), anorexia (47% vs 25%), depression (31% vs 17%), constipation (45% vs 30%), poor sleep (32% vs 21%), dyspnea (30% vs 16%), nausea (27% vs 14%), vomiting (14% vs 6%), and diarrhea (14 vs 6%). While this study was not conducted in an early phase setting, it is likely that the discordance in clinician and patient assessments, consistent with Veitch et al [9], would be similarly reflected in early phase research.

These examples demonstrate that physician assessment of patient AEs may be both incomplete and underestimated in comparison to the patient perspective. Reasons for this may include patient difficulties in spontaneously raising or describing AEs, patient fears of delay or discontinuation of treatment options in which they have high expectations of positive results, introduction of clinician subjectivity, and time constraints and practical limitations with current physician tools.

While valuable in addressing the underreporting and lower scoring of AE severity by physicians, the PRO-CTCAE alone fails to assess the cumulative impact of the AE profile and the effects on functioning and quality of life. The cumulative impact may be especially important in newer treatments taken for sustained periods, where multiple, concurrent, low-grade but persistent AEs may together represent an intolerable burden for the patient. As we describe later, supplementing the rating of individual AEs with an overall single-item measure of the cumulative impact of AEs (eg, using the Functional Assessment of Cancer Therapy–Item GP5 [FACT GP5] [18] or item 168 from the European Organisation for Research and Treatment of Cancer [EORTC] item library [19]) and the adverse impact on patient physical function and role function (eg, measured using the associated subscales of the EORTC Quality of Life Questionnaire—Core Questionnaire [EORTC QLQ-C30] [20]) provides a valuable assessment of the impact of the AE profile experienced.

Using Patient-Reported Outcomes in Early Phase Oncology Trials

Overview

While the use of patient-reported outcome measures (PROMs) is increasingly incorporated into phase 3 clinical trials, and regulatory recommendations on measurement strategy have been recently published by the FDA [21], there is little use of PROMs in early phase trials [22]. Barriers to adoption in early phase oncology studies include a lack of guidance regarding PROM selection, concerns relating to dealing with missing patient-reported outcome (PRO) data, overburdening site staff and patients, handling patient and data queries [23], and low power associated with small sample sizes. Nonetheless, the patient perspective is a vital element of fully understanding dose toxicity and selecting an optimal dose for later phase development.

Adverse Events

In later phase trials, there is typically enough understanding of the AE profile of the investigational treatment to enable a reliable preselection of items for measurement (eg, using a small subset of PRO-CTCAE items). The same is not true for first in human and other early phase trials. Preclinical data may provide some signals to drive thinking, but these are unlikely to be robust and comprehensive, and while the AE evidence from other drugs with the same mechanism of action may be available and relevant, this is not always the case. The full PRO-CTCAE instrument contains 124 items across 78 distinct terms [17], and this is impractical to use in a full list format for regular ongoing measurement.

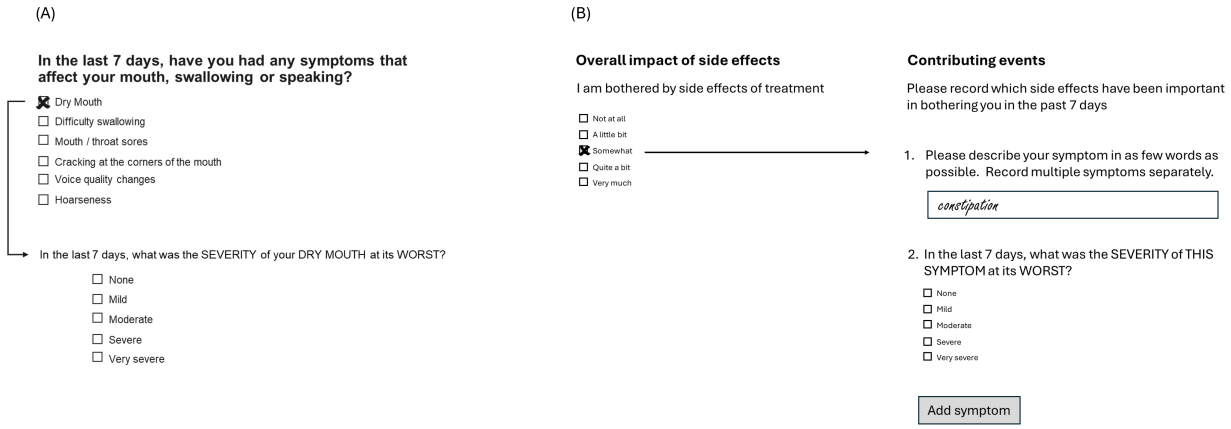
Janse van Rensburg and colleagues [24] used a statistical approach to develop a reduced list of PRO-CTCAE items considered most likely to occur in a phase 1 population using the same dataset reported by Veitch et al [9]. Using that dataset, they eliminated AEs recorded less than 5% of the time; those recorded as "mild" severity by at least 75% of patients; and AEs associated with interference scores of "not at all," frequency scores of "never," or amount scores of "not at all" by at least 80% of patients. Finally, terms with the lowest internal reliability within each organ system domain, as measured using Cronbach α, were also eliminated. With further refinements from physician perspectives, this led to a tailored PRO-CTCAE survey consisting of 58 items assessing 30 terms. While a useful and interesting approach, the generalizability of this reduced survey may be limited by the relatively small sample size, the limited

set of treatments and tumor types represented, and the risk that using historical data may miss important aspects of tolerability for new targeted therapies.

When selecting a reduced set of predefined items, it is helpful, as PRO-CTCAE recommends and allows, to include a free-text item to capture other important AEs not listed [25] and to use this information to allow the item list to adapt with the emerging understanding gained through continued study.

Because the PRO-CTCAE items are grouped by organ system domains, it is possible to optimize the completion of the full item list compared to an individual symptom-by-symptom approach (Figure 3A). Alternatively, a free-text approach asking patients to list and rate the AEs that contribute most to their overall impact rating (scored using the FACT GP5 item, for example) might provide a less burdensome approach (Figure 3B). Leveraging an electronic PRO (ePRO) solution, using a smartphone app, for example, would enable free-text symptom text to be resurfaced as a list of existing AEs to be easily rescored at future time points.

Figure 3. Approaches to simplify adverse event capture and scoring (A) using organ system grouping and (B) collecting the most bothersome adverse events associated with overall impact score.

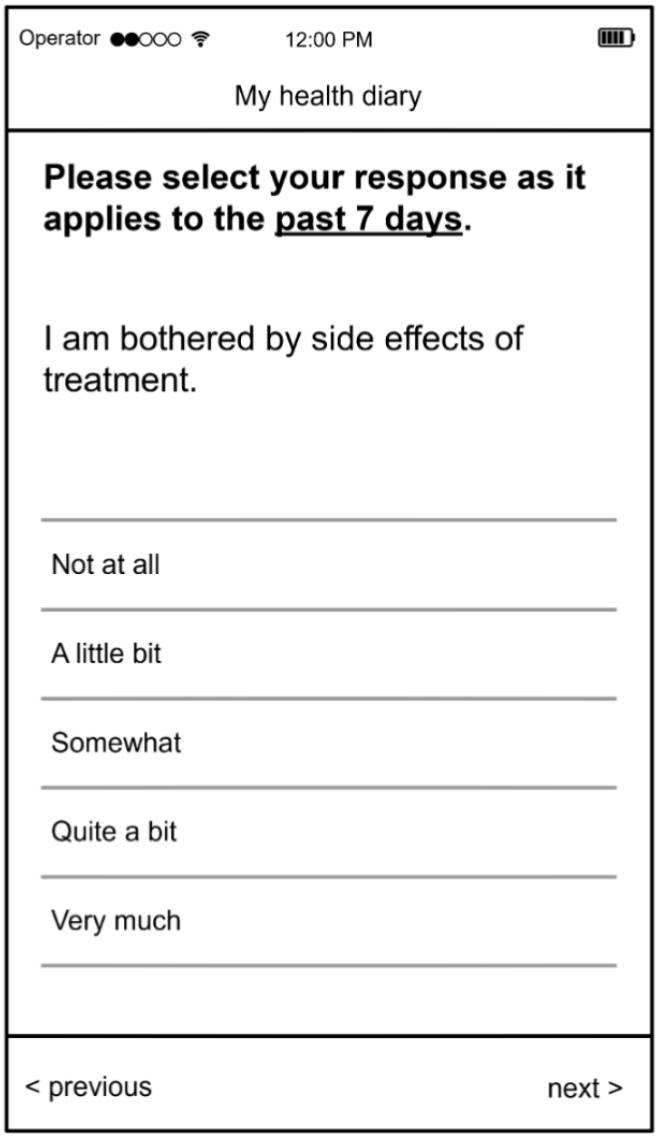


Collecting the most bothersome AEs (Figure 3B) has similarities to some existing PROMs that measure the most bothersome symptoms (MBSs). For example, MBS has been shown to be a useful patient-centric measure of migraine symptoms [26,27] and is referred to in the patient-focused drug development guidance published by the FDA [28]. Challenges with collecting an MBS include how to pool data in the statistical analysis and different symptoms becoming the most bothersome over time. However, these challenges may be less relevant when collecting the set of most bothersome AEs to understand the dose-toxicity relationship in early phase cancer trials.

Other Recommended PROMs

The Friends of Cancer white paper [8] and FDA guidance on PROs in cancer trials [21] both recommend that in addition to the collection and scoring of individual AEs, an overall measure of the AE impact is included, along with measures to assess the impact of treatment on physical function and role function, although the FDA guidance is more focused toward use in confirmatory trials. A single item to score the overall impact of AEs, such as the FACT GP5 [18] (Figure 4) or Q168 of the EORTC item bank [19], is important as it enables the patient to account for the impact of any AEs not covered by PRO-CTCAE administration and to attach greater importance to the combined impact of multiple low-level AEs. This understanding may be particularly important in assessing the impact of newer treatments taken over longer periods. The FDA has identified measures of physical function and role function they consider suitable, including the EORTC QLQ-C30 physical function and role function subscales [20] and the Patient-Reported Outcomes Measurement Information System (PROMIS) physical function scale [29].

Figure 4. Functional Assessment of Cancer Therapy–Item GP5 (FACT GP5). An example of a single-item measure of the impact of adverse events. FACT GP5 is reproduced with permission, copyright of Dr David Cella, and licensed by FACIT.org [18].



Mitigating Barriers to PROM Adoption in Early Phase Trials

As described earlier, barriers to the adoption of PROMs in early phase cancer trials include a lack of guidance regarding PROM selection, concerns relating to dealing with missing PRO data, overburdening site staff and patients, handling patient and data queries [23], and low power associated with small cohorts. We address PROM selection in the discussion above. Dealing with missing data is always an important consideration in clinical research, as different missing data approaches rely on assumptions that, if violated, can lead to biased estimates. Although researchers in early phase trials may use less formal approaches to interpreting the data and determining the optimal dose, it will remain important to consider the impact of missing data using a variety of sensitivity evaluations.

In terms of patient burden, the UK National Cancer Research Institute Consumer Forum survey indicated that most patients and carers affected by cancer and involved in research activities (n=57) were willing to spend up to 15 minutes per day completing PROMs [23]. This time duration seems high for frequent collection but perhaps reflects the value that patients see in communicating this data to their treating physician. The measures we have discussed above typically use a 7-day recall period, and it is, therefore, most likely that a weekly completion schedule would be recommended. Recall bias using the weekly recall periods associated with these validated measures is unlikely to be a concern, but ensuring completion times do not overburden patients with the debilitating effects of the disease and treatment is an important consideration. Median per-item completion rates of PROMs commonly used in oncology trials have been reported as 6 - 14 seconds [30], which suggests that a weekly PROM assessment of, for example, the PRO-CTCAE implemented using the approach outlined in Figure 3A, an overall AE impact item, and the physical function and role function subscales of the QLQ-C30 (items 5 and 2, respectively) might translate to an average completion time of less than 5 minutes per week. This seems to be a feasible assessment strategy, and ensuring a flexible completion window across more than 1 day may drive higher completion rates.

The remaining barriers cited may be mitigated by electronic collection of PRO data, for example, by using an ePROs smartphone app. The burden on site staff and patients during busy clinic visits can be mitigated by enabling at-home completion, and electronic tools can eliminate data queries by prohibiting ambiguous or invalid entries. The easy implementation of longer lists of items using branching logic to speed completion is only practical using an electronic approach. Further, the use of ePRO solutions can also lead to reduced missing data through alarms, reminders, and remote monitoring to drive on-time completion.

Electronic collection of PRO data may be perceived as a significant additional cost relative to the smaller numbers of patients involved in early phase studies, but this should be considered in the context of the value of the data. The more frequent assessment schedules and the nature of the measures implemented drive the use of electronic solutions. In the context of the increased expense of studying more patients in early phase trials due to the need to better characterize dose response, the use of ePROs to drive more accurate, reliable data may lead to accurate decision-making using relatively smaller sample sizes and offset the cost of ePROs many times over.

Smaller sample sizes associated with early phase studies may limit the robust characterization of the dose-response relationship, but this limitation is not unique to PRO data and also applies to other measures of efficacy and tolerability that inform dose selection. A thoughtful approach is required to balance the cost of increased sample size with the statistical robustness of dose-response characterization. Supplementary qualitative data collected using in-trial patient interviews may be a valuable addition to understanding the AEs experienced by the patients and aid the interpretation of PROM data when limited by small sample sizes.

Conclusion

There is growing interest in more completely quantifying the dose-response relationship to inform optimal dose determination for new oncological treatments. PROs play a vital role in understanding dose tolerability profiles, especially as treatment-related AEs tend to be underreported and underscored by physicians. While AE profiles are less understood in early phase drug development, this should not prevent the capturing of this data to inform dose selection decisions as early as the first in human study using some of the approaches discussed in this paper.

Of course, we lack experience in interpreting PRO tolerability data from such early studies and need to remember that we may not have adequately defined the patient population at this early stage, and so the clinical interpretation of dose-response relationships associated with the PROs and other efficacy and tolerability data needs to be interpreted with this in mind.

With newer targeted therapeutics, there is a need to learn much more about safety and tolerability across a wide range of doses, and the current dose-finding models focusing on a single “optimal” dose may no longer work. A fundamental element of the decision-making process for determining safety and tolerability currently missing is the patient experience. It has been assumed that the physician, through patient interaction and AE reporting, can provide a sufficient reflection of the patient experience, but the evidence demonstrates that this is not reliable. Some important symptoms for patients are missed completely. The severity of other symptoms is underreported. Further, with newer targeted agents, AEs may accumulate over time, and the chronic nature or combination of events may make a dose become intolerable for the patient later. If the patient’s perspective is not considered, there is a risk of selecting dose groups that are too high, leading to reduced compliance. It is, therefore, necessary to build a PROM assessment strategy for early phase trials that combines elements of well-established scales to assess safety and tolerability in a package that is practical and not burdensome, yielding vital data to support decision-making as the trial progresses. Maximizing the value of the early patient experience is ethically appropriate and feasible, and drives efficiency in development programs and patient exposure.

Authors' Contributions

BB and AE wrote and edited the manuscript. PC, CG, and TM edited the manuscript.

Conflicts of Interest

BB and AE are employees of Signant Health and may own stock or stock options. PC is an employee of Sanofi and may own stock or stock options. CG is an employee and shareholder of Orion Corporation. TM reports consultancy to Roche, Astra Zeneca, Signant Health, GreyWolf, Guerbet, Geneos, Eisai, Beigene, and MSD, and research funding from MSD, Bayer, and Boston Scientific.

References

1. Le Tourneau C, Lee JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. J Natl Cancer Inst 2009 May 20;101(10):708-720. [doi: 10.1093/jnci/djp079] [Medline: 19436029]
2. Optimizing the dosage of human prescription drugs and biological products for the treatment of oncologic diseases: guidance for industry; availability. Food and Drug Administration. 2024 Aug. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/optimizing-dosage-human-prescription-drugs-and-biological-products-treatment-oncologic-diseases [accessed 2024-12-13]

3. Murphy R, Halford S, Symeonides SN. Project Optimus, an FDA initiative: considerations for cancer drug development internationally, from an academic perspective. Front Oncol 2023 Mar 3;13:1144056. [doi: 10.3389/fonc.2023.1144056] [Medline: 36937434]
4. Jänne PA, Kim G, Shaw AT, Sridhara R, Pazdur R, McKee AE. Dose finding of small-molecule oncology drugs: optimization throughout the development life cycle. Clin Cancer Res 2016 Jun 1;22(11):2613-2617. [doi: 10.1158/1078-0432.CCR-15-2643] [Medline: 27250931]
5. Project Optimus: reforming the dose optimization and dose selection paradigm in oncology. Food and Drug Administration. 2024. URL: https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus [accessed 2024-06-17]
6. ICH E9 statistical principles for clinical trials - scientific guideline. European Medicines Agency. 1998 Sep 1. URL: https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline [accessed 2024-06-17]
7. Common Terminology Criteria for Adverse Events (CTCAE): version 4.0. National Cancer Institute. 2009 May 28. URL: https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03/Archive/CTCAE_4.0_2009-05-29_QuickReference_8.5x11.pdf [accessed 2024-06-17]
8. Basch E, Campbell A, Hudgens S, Jones L, King-Kallimanis B, Kluetz P, et al. Broadening the definition of tolerability in cancer clinical trials to better measure the patient experience. Friends of Cancer Research. 2018. URL: https://friendsofcancerresearch.org/wp-content/uploads/Comparative-Tolerability-Whitepaper_FINAL.pdf [accessed 2024-06-17]
9. Veitch ZW, Shepshelovich D, Gallagher C, et al. Underreporting of symptomatic adverse events in phase I clinical trials. J Natl Cancer Inst 2021 Aug 2;113(8):980-988. [doi: 10.1093/jnci/djab015] [Medline: 33616650]
10. Laugsand EA, Sprangers MAG, Bjordal K, Skorpen F, Kaasa S, Klepstad P. Health care providers underestimate symptom intensities of cancer patients: a multicenter European study. Health Qual Life Outcomes 2010 Sep 21;8:104. [doi: 10.1186/1477-7525-8-104] [Medline: 20858248]
11. Atkinson TM, Ryan SJ, Bennett AV, et al. The association between clinician-based Common Terminology Criteria for Adverse Events (CTCAE) and patient-reported outcomes (PRO): a systematic review. Support Care Cancer 2016 Aug;24(8):3669-3676. [doi: 10.1007/s00520-016-3297-9] [Medline: 27260018]
12. Pavlović Mavić M, Šeparović R, Tečić Vuger A, Vazdar L. Difference in estimation of side effects of chemotherapy between physicians and patients with early-stage breast cancer: the use of patient reported outcomes (PROs) in the evaluation of toxicity in everyday clinical practice. Cancers (Basel) 2021 Nov 25;13(23):5922. [doi: 10.3390/cancers13235922] [Medline: 34885032]
13. Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst 2009 Dec 2;101(23):1624-1632. [doi: 10.1093/jnci/djp386] [Medline: 19920223]
14. Basch E, Iasonos A, McDonough T, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. Lancet Oncol 2006 Nov;7(11):903-909. [doi: 10.1016/S1470-2045(06)70910-X] [Medline: 17081915]
15. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care 2008 Aug;14(8):530-539. [Medline: 18690769]
16. Dueck AC, Mendoza TR, Mitchell SA, et al. Validity and reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncol 2015 Nov;1(8):1051-1059. [doi: 10.1001/jamaoncol.2015.2639] [Medline: 26270597]
17. Basch E, Reeve BB, Mitchell SA, et al. Development of the National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). J Natl Cancer Inst 2014 Sep 29;106(9):dju244. [doi: 10.1093/jnci/dju244] [Medline: 25265940]
18. FACIT Item GP5: Functional Assessment of Chronic Illness Therapy - Item GP5. FACIT. 2024. URL: https://www.facit.org/measures/facit-item-gp5 [accessed 2024-06-17]
19. Item library. EORTC Quality of Life. 2024. URL: https://qol.eortc.org/item-library/ [accessed 2024-06-17]
20. EORTC QLQ-C30 scoring manual. European Organisation for Research and Treatment of Cancer. 2001. URL: https://www.eortc.org/app/uploads/sites/2/2018/02/SCmanual.pdf [accessed 2025-02-18]
21. Core patient-reported outcomes in cancer clinical trials: guidance for industry. Food and Drug Administration. 2024 Oct. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/core-patient-reported-outcomes-cancer-clinical-trials [accessed 2024-12-13]
22. Coleman RL, Beck JT, Baranda JC, et al. The use of patient-reported outcome measures in phase I oncology clinical trials. Oncology (Williston Park, NY) 2021;99(7):444-453. [doi: 10.1159/000514874] [Medline: 33823518]
23. Lai-Kwon J, Vanderbeek AM, Minchom A, et al. Using patient-reported outcomes in dose-finding oncology trials: surveys of key stakeholders and the National Cancer Research Institute Consumer Forum. Oncologist 2022 Sep 2;27(9):768-777. [doi: 10.1093/oncolo/oyac117] [Medline: 35762393]
24. Janse van Rensburg HJ, Liu Z, Watson GA, et al. A tailored phase I-specific patient-reported outcome (PRO) survey to capture the patient experience of symptomatic adverse events. Br J Cancer 2023 Sep;129(4):612-619. [doi: 10.1038/s41416-023-02307-w] [Medline: 37419999]

25. Bhatnagar V, Dutcus C, Ghiorghiu S, Kluetz P, Jones L, McJunkins KR, et al. Supporting a patient-centric approach to dose optimization in oncology: the essential role of patient-reported outcomes (PROs). Friends of Cancer Research 2022 [FREE Full text]
26. Lipton RB, Goadsby PJ, Dodick DW, et al. Evaluating the clinical utility of the patient-identified most bothersome symptom measure from PROMISE-2 for research in migraine prevention. Headache 2022 Jun;62(6):690-699. [doi: 10.1111/head.14295] [Medline: 35466430]
27. Dodick DW, Tepper SJ, Friedman DI, Gelfand AA, Kellerman DJ, Schmidt PC. Use of most bothersome symptom as a coprimary endpoint in migraine clinical trials: a post-hoc analysis of the pivotal ZOTRIP randomized, controlled trial. Headache 2018 Jul;58(7):986-992. [doi: 10.1111/head.13327] [Medline: 29782049]
28. Patient-focused drug development: incorporating clinical outcome assessments into endpoints for regulatory decision-making – draft guidance for industry. Food and Drug Administration. 2023 Apr. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-incorporating-clinical-outcome-assessments-endpoints-regulatory [accessed 2024-06-17]
29. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). J Clin Epidemiol 2008 Jan;61(1):17-33. [doi: 10.1016/j.jclinepi.2006.06.025] [Medline: 18083459]
30. Kalpadakis-Smith A, Kusanto BT, Byrom B, Everhart A. PCR75 assessment of completion time to optimize the administration of common patient-reported outcome measures (PROMS) in oncology clinical trials. Value Health 2023 Jun;26(6):S325. [doi: 10.1016/j.jval.2023.03.1853]

Abbreviations

AE: adverse event
CTCAE: Common Terminology Criteria for Adverse Events
EORTC: European Organisation for Research and Treatment of Cancer
EORTC QLQ-C30: EORTC Quality of Life Questionnaire–Core Questionnaire
ePRO: electronic patient-reported outcome
FACT GP5: Functional Assessment of Cancer Therapy–Item GP5
FDA: Food and Drug Administration
MBS: most bothersome symptom
PRO: patient-reported outcome
PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events
PROM: patient-reported outcome measure
PROMIS: Patient-Reported Outcomes Measurement Information System

Edited by N Cahill; submitted 24.07.24; peer-reviewed by A Degboe, J Tanna; revised version received 17.12.24; accepted 27.01.25; published 28.02.25.

Please cite as:
Byrom B, Everhart A, Cordero P, Garratt C, Meyer T
Leveraging Patient-Reported Outcome Measures for Optimal Dose Selection in Early Phase Cancer Trials
JMIR Cancer 2025;11:e64611
URL: https://cancer.jmir.org/2025/1/e64611
doi:10.2196/64611

Design and Use of Patient-Facing Electronic Patient-Reported Outcomes and Sensor Data Visualizations During Outpatient Chemotherapy

Christianna Bartel, BS; Leeann Chen, MPS; Weiyu Huang, BSc; Qichang Li, MS; Qingyang Li, MS; Jennifer Fedor, MS; Krina C Durica, MA; Carissa A Low, PhD

Department of Medicine, University of Pittsburgh, Suite 5002, 5051 Centre Avenue, Pittsburgh, PA, United States

Corresponding Author:
Carissa A Low, PhD
Department of Medicine, University of Pittsburgh, Suite 5002, 5051 Centre Avenue, Pittsburgh, PA, United States

Abstract

This study describes patients' interaction with a personalized web-based visualization displaying daily electronic patient-reported outcomes and wearable device data during outpatient chemotherapy.

(JMIR Cancer 2025;11:e62711) doi:10.2196/62711

KEYWORDS

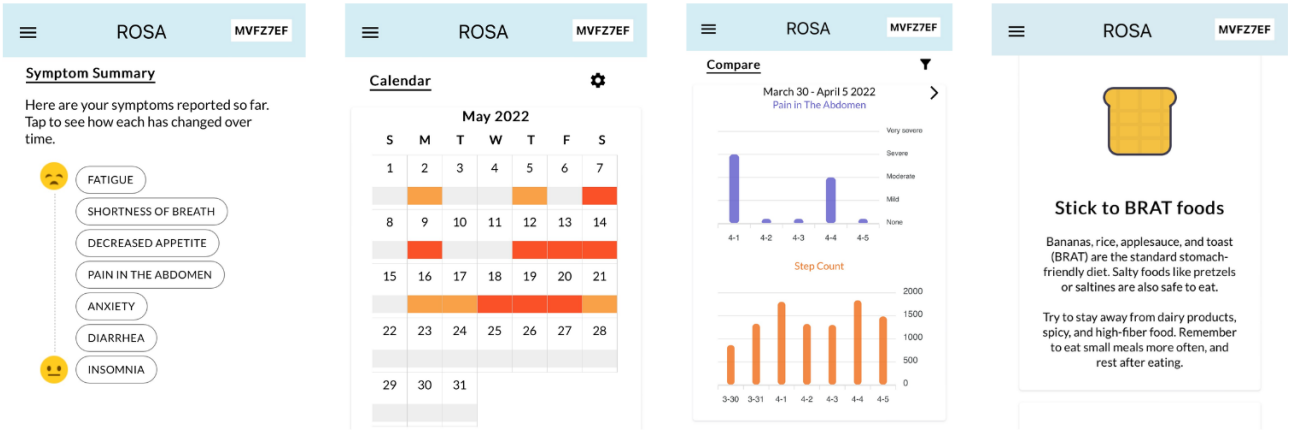
oncology; cancer; data visualization; remote monitoring; mobile technology; patients; outpatient; chemotherapy; symptoms; side effects; cancer treatment; electronic patient-reported outcome; online; monitoring; self-management

Introduction

Chemotherapy can cause significant symptoms that impact the quality of life [1]. Although electronic patient-reported outcome (ePRO) systems for collecting symptom ratings from patients have become increasingly common in cancer care, most of these are designed for clinicians, and fewer than half share data visualizations with the patients [2,3]. Visualization of ePRO and other data (eg, wearable device data) may help patients undergoing cancer treatment find patterns that help them to prepare for future treatment cycles, manage side effects, and have productive conversations with clinicians [4,5].

As part of a prospective longitudinal National Cancer Institute–funded study to develop a remote symptom monitoring system during chemotherapy [6], we created mobile-friendly web visualizations of each patient's daily symptom ratings and wearable device data (Figure 1). The aim of this paper is to describe patterns of use of these novel visualizations.

Figure 1. Visualizations of daily symptom ratings and wearable data with self-care resources.



Methods

Recruitment and Study Design

Participants undergoing chemotherapy for any solid tumor and who owned smartphones were recruited from oncology clinics at a single academic center in Pittsburgh, Pennsylvania. During the study, the participants wore a Fitbit device (Inspire; Google LLC) and reported 16 symptoms commonly experienced during chemotherapy (eg, nausea, fatigue) daily, using the patient-reported version of the Common Terminology Criteria for Adverse Events [7]. The website also included

evidence-based symptom self-management resources as described by Donovan et al [8].

At the time of enrollment, we provided each patient with a personalized link to their real-time visualizations; however, no instructions about viewing frequency for the visualizations or usage reminders were given during the study. At the end of the 3-month study, 141 patients completed an electronic- or paper-based 11-question survey (mean completion time was 5 min) to assess whether they used the visualizations, frequency of use, helpful information, and suggestions for improvement. Data were collected between February 2022 and April 2024.

Ethical Considerations

The University of Pittsburgh's Institutional Review Board approved all study activities (19070011). At the time of enrollment, all participants provided informed written consent. All data were stored in secure locations and identified only by anonymized study ID numbers. Participants received US $100 and could keep the Fitbit device (estimated value $100).

Results

Characteristics of the participants can be found in Table 1. Survey respondents were heterogeneous in age (mean 61, SD 12; range 29-92 years), race (113/141, 80% White; 28/141, 20% other races), and cancer stage (75/135, 56% stage IV). Approximately half (76/141, 54%) of the participants accessed the link to their data visualizations. Participants with non-binary gender (n=1, 0.7%) and unknown cancer stage (n=6, 4.3%) were excluded from χ^2 analysis while comparing participants who accessed their visualizations. There were no significant differences between the participants who clicked on the link and those who did not during the study in terms of mean age (P=.74), gender (P=.66), race (P=.50), or cancer stage (P=.31). Of those who accessed the platform, most (54%, 41/76) viewed it a few times (ie, less than monthly), while 13% (10/76) used it daily. The 10 daily users were within 3 months of starting chemotherapy for the first time. Most participants (58/75, 77%) found the visualizations *somewhat* or *very helpful/informative*. Few participants shared their data with family members or friends (11/141, 8%) and with others (2/141, 1%); none shared data with their providers or other patients. Participant–suggested improvements included reminders to view graphs and the ability to enter treatment and surgery dates.

Table . Participant characteristics.

Characteristics	Overall (N=141)	Did not click the link (n=65)	Clicked the link (n=76)	Test statistic (*df*)[a]	*P* value
Age (years), mean (SD)	61 (12)	61 (13)	60 (11)	t (130.1)=0.68	.74
Gender, n (%)				$\chi^2(1)=0.2$	.66[b]
Female	94 (66.7)	45 (69.2)	49 (64.5)		
Male	45 (31.9)	19 (29.2)	26 (34.2)		
Non-binary	1 (0.7)	0 (0)	1 (1.3)		
Unknown	1 (0.7)	1 (1.5)	0 (0)		
Race, n (%)				$\chi^2(1)=0.45$	.5
White	113 (80.1)	50 (76.9)	63 (82.9)		
Others[c]	28 (19.8)	15 (23.1)	13 (17.1)		
Stage IV cancer, n (%)				$\chi^2(1)=1.04$	.31[d]
Yes	75 (53.2)	39 (60.0)	36 (47.4)		
No	60 (42.5)	25 (38.5)	35 (46.0)		
Unknown	6 (4.3)	1 (1.5)	5 (6.6)		

[a]Welch two sample *t*-test and Pearson's χ^2 test were used, as appropriate; degrees of freedom (*df*) are provided in parentheses.

[b]Participants with non-binary or unknown gender were excluded from this test.

[c]Other race category included Black or African-American (n=24), Asian (n=2), and more than one race (n=2).

[d]Participants with unknown cancer stage were excluded from this test.

Discussion

Overview

Providing real-time visualizations of ePRO and activity data throughout chemotherapy may help patients anticipate and manage symptoms effectively and potentially identify patterns between activity or other sensor data and symptoms. These preliminary findings suggest that patients are motivated to view their data, and these visualizations were accessible to patients of different ages, races, and cancer stages. Daily users, who were mostly new to chemotherapy, may have higher levels of

anxiety and a greater need for health information [9]. Future studies should investigate the potential benefits of patient-facing visualizations for patients beginning chemotherapy.

Limitations

The visualizations and other website content were developed as part of an ancillary project in an ongoing study, and participants received no instructions or reminders regarding website usage. Survey respondents represented a subset (141/158, 89%) of participants, who received personalized visualizations, and the results may be influenced by selection, response, and recall biases.

Conclusion

This study describes initial efforts to share real-time ePRO and wearable device data visualizations with patients undergoing chemotherapy. Further research is needed to improve the usability of data visualizations and evaluate their impact on symptom management, self-efficacy, and other outcomes.

Conflicts of Interest

None declared.

References

1. Reilly CM, Bruner DW, Mitchell SA, et al. A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment. Support Care Cancer 2013 Jun;21(6):1525-1550. [doi: 10.1007/s00520-012-1688-0] [Medline: 23314601]
2. Jensen RE, Gummerson SP, Chung AE. Overview of patient-facing systems in patient-reported outcomes collection: focus and design in cancer care. J Oncol Pract 2016 Oct;12(10):873-875. [doi: 10.1200/JOP.2016.015685] [Medline: 27601515]
3. Warrington L, Absolom K, Conner M, et al. Electronic systems for patients to report and manage side effects of cancer treatment: systematic review. J Med Internet Res 2019 Jan 24;21(1):e10875. [doi: 10.2196/10875] [Medline: 30679145]
4. Chen L, Bartel C, Cai X, et al. Patient and provider perspectives on symptom monitoring during outpatient chemotherapy: interview study. JMIR Form Res 2023 Apr 17;7:e46001. [doi: 10.2196/46001] [Medline: 37067857]
5. Ruzich E, Ritchie J, Ginchereau Sowell F, et al. A powerful partnership: researchers and patients working together to develop a patient-facing summary of clinical trial outcome data. J Am Med Inform Assoc 2024 Jan 18;31(2):363-374. [doi: 10.1093/jamia/ocad099] [Medline: 37341698]
6. McClaine S, Fedor J, Bartel C, Chen L, Durica KC, Low CA. Engagement with daily symptom reporting, passive smartphone sensing, and wearable device data collection during chemotherapy: longitudinal observational study. JMIR Cancer 2024 Dec 10;10:e57347. [doi: 10.2196/57347] [Medline: 39656513]
7. Dueck AC, Mendoza TR, Mitchell SA, et al. Validity and reliability of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncol 2015 Nov;1(8):1051-1059. [doi: 10.1001/jamaoncol.2015.2639] [Medline: 26270597]
8. Donovan HS, Sereika SM, Wenzel LB, et al. Effects of the WRITE symptoms interventions on symptoms and quality of life among patients with recurrent ovarian cancers: an NRG oncology/GOG Study (GOG-0259). J Clin Oncol 2022 May 1;40(13):1464-1473. [doi: 10.1200/JCO.21.00656] [Medline: 35130043]
9. Lim CC, Devi MK, Ang E. Anxiety in women with breast cancer undergoing treatment: a systematic review. Int J Evid Based Healthc 2011 Sep;9(3):215-235. [doi: 10.1111/j.1744-1609.2011.00221.x] [Medline: 21884450]

Abbreviations

ePRO: electronic patient-reported outcomes

Edited by N Cahill; submitted 31.05.24; peer-reviewed by KL Du, LM Valen Schougaard; revised version received 03.12.24; accepted 04.12.24; published 10.01.25.

Please cite as:
Bartel C, Chen L, Huang W, Li Q, Li Q, Fedor J, Durica KC, Low CA
Design and Use of Patient-Facing Electronic Patient-Reported Outcomes and Sensor Data Visualizations During Outpatient Chemotherapy
JMIR Cancer 2025;11:e62711
URL: https://cancer.jmir.org/2025/1/e62711
doi:10.2196/62711

Monthly Variations in Colorectal Cancer Screening Tests Among Federally Qualified Health Center Patients in Missouri: Quality Improvement Project

Jane A McElroy*, PhD; Jamie B Smith*, MA; Kevin D Everett*, PhD

Family and Community Medicine Department, University of Missouri, 1 Hospital Dr DC032.00, Columbia, MO, United States

*all authors contributed equally

Corresponding Author:
Jane A McElroy, PhD
Family and Community Medicine Department, University of Missouri, 1 Hospital Dr DC032.00, Columbia, MO, United States

Abstract

Background: Cancer is the second leading cause of death in the United States. Compelling evidence shows screening detects colorectal cancer (CRC) at earlier stages and prevents the development of CRC through the removal of precancerous polyps. The Healthy People 2030 goal for CRC screening is 68.3%, but only 36.5% of Missouri federally qualified health center patients aged 50 - 75 years are up-to-date on CRC screening. For average risk patients, there are three commonly used screening tests in the United States—two types of stool tests collected at home (fecal immunochemical test [FIT]–immunochemical fecal occult blood test [FOBT] and FIT-DNA, such as Cologuard) and colonoscopies completed at procedural centers.

Objective: This study aims to examine variation by month for the three types of CRC testing to evaluate consistent patient care by clinical staff.

Methods: Data from 31 federally qualified health center clinics in Missouri from 2011 to 2023 were analyzed. A sample of 34,124 unique eligible "average risk" patients defined as persons not having a personal history of CRC or certain types of polyps, family history of CRC, personal history of inflammatory bowel disease, and personal history of receiving radiation to the abdomen or pelvic to treat a previous cancer or confirmed or suspected hereditary CRC syndrome. Another eligibility criterion is that patients need to be seen at least once at the clinic to be included in the denominator for the screening rate calculation. Descriptive statistics characterize the sample, while bivariate analyses assess differences in screening types by month.

Results: Completion of CRC screening yielded statistically significant differences for patients completing the different types of CRC screening by month. October-January had the highest proportions of patients (644-680 per month, 8.5% - 10.2%) receiving a colonoscopy, while February-April had the lowest (509-578 per month, 6.9% - 7.8%), with 614 being the average monthly number of colonoscopies. For FIT-FOBT, June-August had the higher proportions of patients receiving this test (563-613 per month, 8.9% - 9.6%), whereas December-February had the lowest (453-495 per month, 7.1% - 8%), with 541 being the average monthly number of FIT-FOBT kits used. For FIT-DNA, March was the most popular month with 11.3% (n=261 per month) of patients using the Cologuard test, followed by April, May, and November (207-220 per month, 8.7% - 9.4%), and January and June (168-171 per month, 7.2%-7.3%) had the lowest proportion of patients using Cologuard, with 193 being the average monthly number of FIT-DNA kits used. Combining all tests, February had the fewest CRC tests completed (1153/16,173, 7.1%).

Conclusions: Home-based tests are becoming popular, replacing the gold standard colonoscopy, but need to be repeated more frequently. Monthly variation of screening over the course of a year suggests that CRC screening efforts and patient care may be less than ideal. Months with lower rates of screening for each type of CRC test represent opportunities for improving CRC screening.

(JMIR Cancer 2025;11:e64809) doi:10.2196/64809

KEYWORDS

colorectal cancer screening; federally qualified health center; FQHC; fecal immunochemical test; FIT; FIT-DNA; colorectal cancer; CRC; cancer; cancer screening; colonoscopy; United States; health center; quality improvement

Introduction

Colorectal cancer (CRC) is the third most common cancer in the United States and the second leading cause of cancer deaths [1]. Evidence shows that screening detects CRC at earlier stages, and its development can be prevented by removing precancerous polyps. For average risk patients, there are three common screening tests—two types of stool tests collected at home (fecal immunochemical test [FIT]–immunochemical fecal occult blood test [FOBT] and FIT-DNA, like Cologuard) and colonoscopies

completed at procedural centers. The revised Healthy People 2030 goal for CRC screening among people aged 45 - 75 years changed from 74.4% to 68.3% [2]. Federally qualified health centers (FQHCs) provide low-cost care for approximately 30 million people, and 90% of FQHCs' patient population (n=17,562,189) have an income less than 200% of the federal poverty level [3,4]. The CRC screening rate of patients using FQHCs in Missouri (n=95,191) is 36.5% compared to 74.1% for patients not using FQHCs (n=1,657,026) [5].

Colonoscopy is considered the gold standard of CRC screening since precancerous polyps can be removed at the time of the test, preventing cancer. However, numerous patient and health system barriers to colonoscopies have been identified [6]. Home-based testing is becoming more common, and FIT-DNA use has increased post COVID-19 [7]. The increased FIT-DNA use may reflect patient preference for home-based testing that does not incur being wait-listed for months to get a colonoscopy [8]. Additionally, the manufacturer of FIT-DNA provides a full service in facilitating patients' completion of the test. This service includes a patient follow-up to encourage returning the kits and results sent directly to the patient's electronic medical record. For the FIT tests, a clinic is responsible for patient follow-ups regarding stool collection and sending the kit in for analysis [9].

Since screening opportunities take place at patients' routine visits to health centers, determining screening variation by month can assist health care systems adjust outreach efforts, targeting low use months to establish consistently high CRC screening opportunities throughout the year.

Objective

This quality improvement project aims to determine if there is variation in the 3 types of CRC testing by month. Identifying variations by month can support targeted attention. The global aim of the quality improvement project was to support FQHCs' in providing CRC screening opportunities with consistent screening rates each month.

Methods

Overview

Starting in 2020 as part of a 5-year Centers for Disease Control and Prevention–funded quality improvement program, our project supported eight health care systems' initiation or enhancement of four evidence-based interventions to increase CRC screening rates of age-eligible patients using a practice facilitator model. As part of this quality improvement program, up to 4 years of annual data on CRC screening by type and date of completed CRC test for the eligible patient population in the selected health care system were available. Patient characteristics including age, race/ethnicity, primary language, and sex were gathered. Screening compliance was defined as a colonoscopy recommended every 10 years, FIT-FOBT every year, and FIT-DNA every 3 years. Screened for CRC was defined as having a medical record of being up-to-date on one of the three types of tests. For this analysis, eligible patients were aged 50 - 75 years with no prior diagnosis of CRC, adenomatous polyps, or inflammatory bowel disease, and no personal diagnosis or family history of known genetic disorders that predispose them to a high lifetime risk of CRC such as Lynch syndrome or familial adenomatous polyposis [10]. Descriptive statistics characterize the sample, while bivariate analyses assess differences in screening types by month. While examining monthly CRC screening rates, data were limited to exclude years where fewer than 10 screenings occurred for any given month. Monthly totals were first calculated, and the average number of tests across all months was used to calculate the average percentage change (increase or decrease) month to month. A χ^2 test for equal proportions of the CRC screening tests by month among the 3 types of CRC tests was then examined. Month was chosen as the unit of analysis since it is easily understood, helping plan and implement activities. A weekly analysis has fewer observations leading to less stable numbers, and holidays influence the days in any week. SAS 9.4 (SAS Institute) was used for the analysis.

Ethical Considerations

This project was approved by University of Missouri's Institutional Review Board (IRB 2034264), which allowed analysis of clinical data extracted from electronic medical records without additional consent for the secondary analysis. The data were deidentified for the analysis. All data were transmitted and stored in a Health Insurance Portability and Accountability Act (HIPAA)–compliant secure system (REDCap) [11].

Results

A total of 31 clinics servicing predominately rural residents yielded 34,124 unique eligible patients from 2011 to 2023. Among these, 6238 (18.3%) were up to date on their CRC screening, another 5170 (15.2%) had received a CRC screening at some time in the past but were not up to date, and the remaining 22,716 (66.6%) patients had no record of being screened for CRC. Most participants were 50 - 64 years old (n=24,014, 70.4%), were female (n=19,229, 56.4%), used English as their primary language (n=31,686, 92.9%), and were White (n=27,677, 81.1%; Table S1 in Multimedia Appendix 1). Fewer participants younger than 65 years were up to date on their CRC screening than those 65 years and older. Patients with the highest proportion of ever being screened were Hispanic (837/2032, 41.2%), compared to White (9391/27,677, 33.9%) and Black (533/1385, 38.5%), but fewer Hispanic participants (n=260, 12.8%) were up to date compared to White (n=5386, 19.5%) and Black (n=268, 19.4%) participants (Table S1 in Multimedia Appendix 1). The FQHC systems in this analysis served 87% of patients who were at or below 200% of the federal poverty guidelines. Most clinics (n=28, 90.3%) were located in rural areas of Missouri. Among the clinics, the 2023 annual CRC screening rates ranged from 13.7% to 63.1% (62/451 and 238/377 eligible patients, respectively).

Table S2 in Multimedia Appendix 1 breaks down the descriptive statistics on monthly CRC screenings. There were 7368 patients who were up to date on CRC screening by colonoscopy with an average of 614 screenings per month from 2014 to 2023. A χ^2 test for equal proportions found significant differences across

monthly colonoscopy screenings (χ^2_{11}=38.9; *P*<.001). January was the highest month for colonoscopy screenings (n=680, 11% higher than the average), while February was the lowest (n=509, 17% lower than the average; Figure 1). For FIT-FOBT (n=6486), there were an average of 540.5 screenings per month from 2017 to 2023. A χ^2 test for equal proportions found significant differences across monthly FIT-FOBT screenings (χ^2_{11}=51.7; *P*<.001). August was the highest month for FIT-FOBT screenings (n=613, 13% higher than the average) compared to January (n=468, 14% lower than the average) and February (n=453, 16% lower than the average; Figure 2). There were 2319 FIT-DNA screenings, with an average of 193.3 per month from 2020 to 2023. A χ^2 test for equal proportions found significant differences across monthly FIT-FOBT screenings (χ^2_{11}=49.2; *P*<.001). March was the highest month (n=261, 35% higher than the average) while January (n=168, 13% lower than the average) and August (n=153, 21% lower than the average) were the lowest months for FIT-DNA testing (Figure 3).

Figure 1. Colonoscopy by month (2014 - 2023).

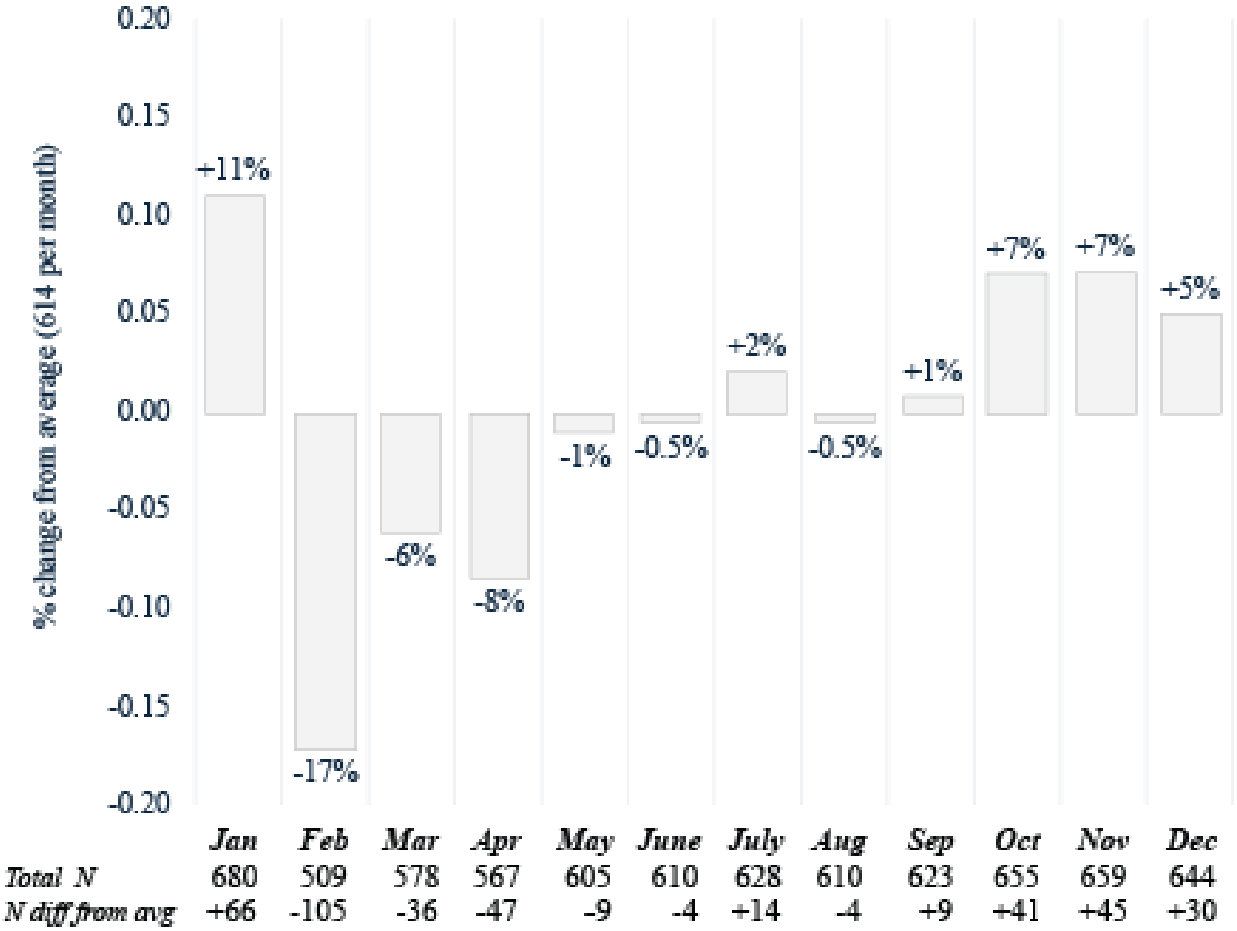


Figure 2. Fecal immunochemistry test–immunochemical fecal occult blood test by month (2017 - 2023).

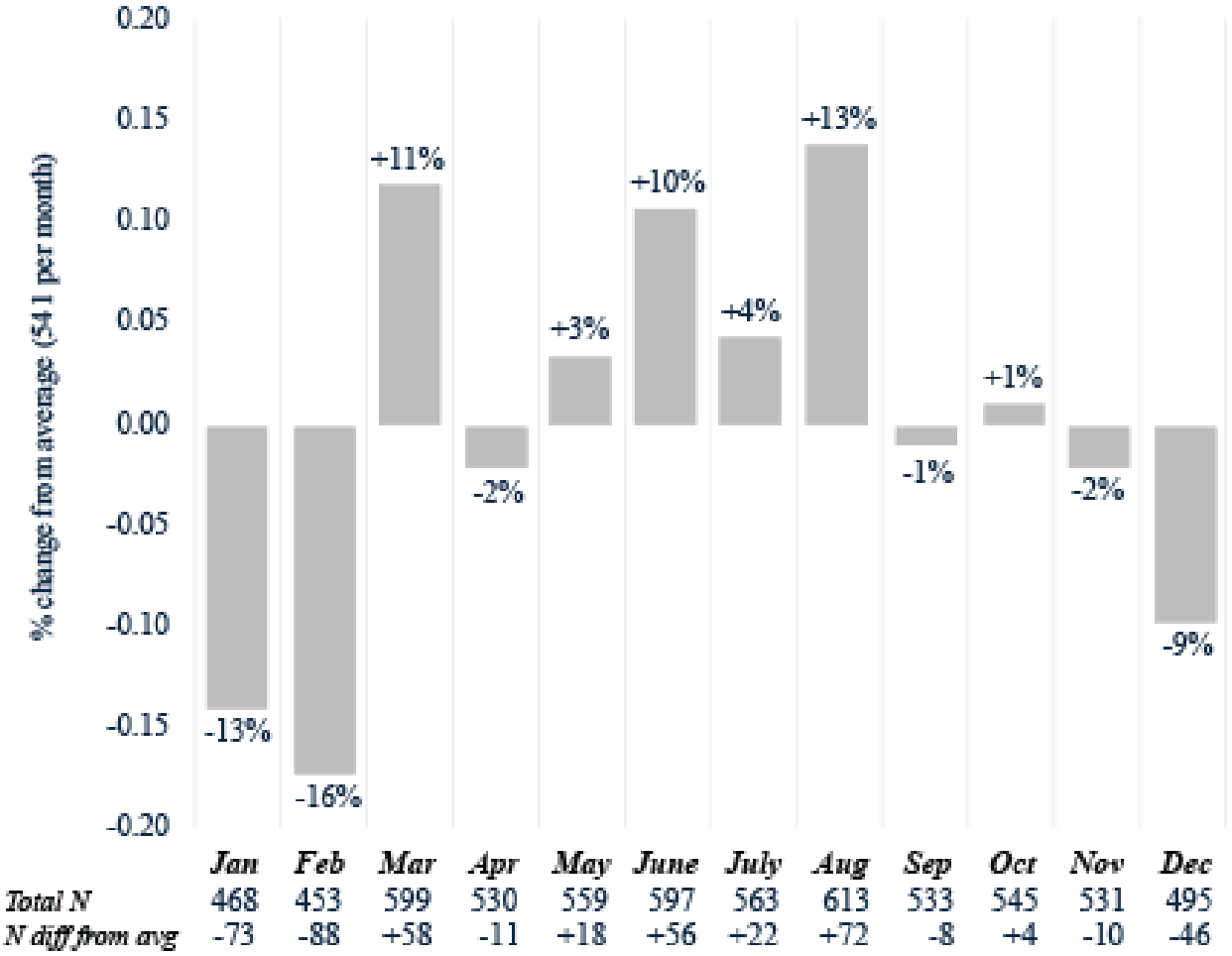
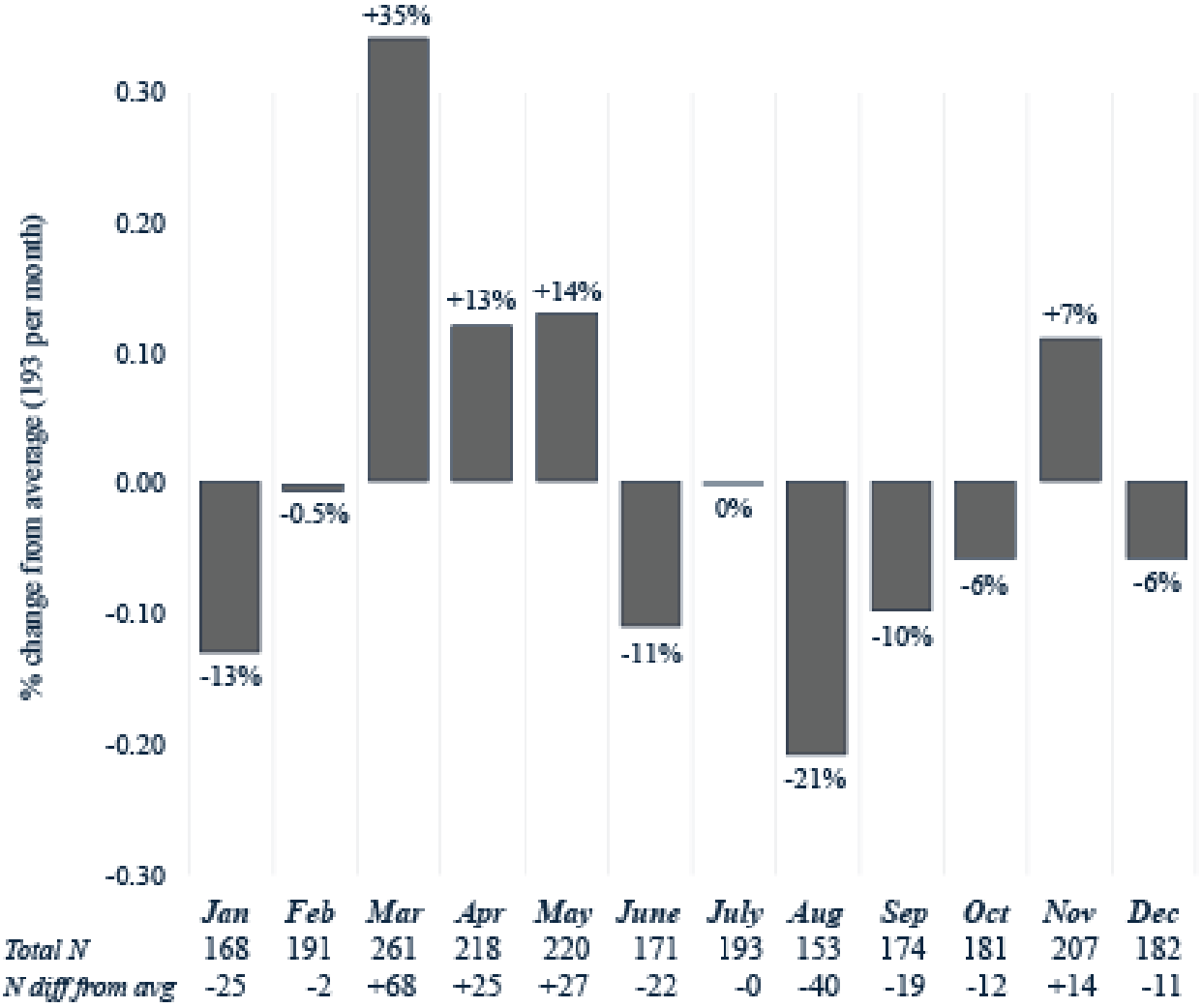


Figure 3. Fecal immunochemistry test–DNA by month (2020 - 2023).



Discussion

Principal Findings

Among the 3 types of CRC screening for average risk patients seen at our FQHCs in the United States, no test was completed consistently by month, and each test had different peak months of completion. We were not able to find any research that compared variation by month in CRC screening test types of colonoscopy, FIT-FOBT, and FIT-DNA. To our knowledge, this is the first study that provides results of CRC screening type by month.

As reflected in our screening choices by patients seen at FQHC clinics, home-based CRC screening increased during the COVID-19 pandemic's closures of specialty care including elective procedures (eg, colonoscopies) [7]. This change in CRC screening options allowed for testing at the discretion of the patient rather than appointment availability.

Strengths and Limitation

One strength of this study was evaluating patients over 12 years from several FQHCs. These data were snapshots of each year's CRC screening behavior by the health care systems. This also captured screening behavior before and after the pandemic.

One limitation of this study was our inability to explain the variability by month of the different screening tests. For example, FIT-DNA and FIT-FOBT tests peaked in CRC awareness month in March but not colonoscopies. Additionally, while the results are informative, only a simple analysis of screening variability was performed, which excluded an examination of temporal changes over time.

The preferences of clinicians on which CRC screening test is recommended and their patient care style were not captured. For example, some clinicians only recommend colonoscopy [12-14]; however, some patients who decline a colonoscopy [15] would be willing to complete a home-based CRC screening test if offered. Further reasons for CRC screening refusal of any test were also not captured. These could be a factor in the CRC test variation by month.

Future Direction

Among the selected participant characteristics, attention is needed on those younger than 65 years to encourage CRC screening. Similarly, while 41.2% of Hispanic participants showed a positive attitude toward CRC screening, only 12.8% were up to date with their screening. This suggests that tailored campaigns and outreach programs could encourage greater participation in CRC screening. For all populations, screening matters since the variance in testing over a year can impact the

health care system's capacity for timely preventive patient care. Gastroenterologist availability to complete colonoscopies may be limited in some regions of the country, but home-based tests can be completed each month [8]. Undoubtedly, individual-level barriers influence CRC screening rates, such as transportation, medical mistrust, financial issues, and low health literacy [16]. However, organizational factors, including monitoring and feedback, have been identified as implementation facilitators [16]. Rockwell and colleagues [6] described health system barriers, especially for colonoscopies, as sludge, "frictions or administrative burdens that make it difficult for people to attain what they want or need." Providing clinical staff information on completed CRC screening rates by month for each test type may facilitate addressing these "sludge" issues and increase CRC screening [8,17].

Acknowledgments

This work was supported by the Centers for Disease Prevention and Control (CDC; award NU58DP006765). The contents are those of the authors and do not represent official views of, nor an endorsement by, the CDC, Health and Human Services, or the US Government.

Data Availability

The deidentified dataset is available from the corresponding author upon reasonable request.

Authors' Contributions

JAM was a multiple principal investigator for the project in which data were available, conceptualized the research aims, worked with the analyst on the analytical plan, and wrote the original draft. JBS curated the data and applied the statistical techniques to analyze study data. KDE was a multiple principal investigator for the project in which data were available and reviewed and edited the manuscript. No generative artificial intelligence was used in writing this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient characteristics and descriptive statistics for monthly screenings.
[DOCX File, 29 KB - cancer_v11i1e64809_app1.docx]

References

1. Cancer facts & figures. American Cancer Society. 2024. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2024/2024-cancer-facts-and-figures-acs.pdf [accessed 2025-03-13]
2. Increase the proportion of adults who get screened for colorectal cancer — C-07: data methodology and measurement. Office of Disease Prevention and Health Promotion. 2024. URL: https://health.gov/healthypeople/objectives-and-data/browse-objectives/cancer/increase-proportion-adults-who-get-screened-colorectal-cancer-c-07/data-methodology [accessed 2025-03-13]
3. National health center program uniform data system (UDS) awardee data - archive. Health Resources & Services Administration. 2024. URL: https://web.archive.org/web/20241220121332/https://data.hrsa.gov/tools/data-reporting/program-data/national [accessed 2025-03-18]
4. What is a health center? Health Resources & Services Administration. 2024. URL: https://bphc.hrsa.gov/about-health-center-program/what-health-center [accessed 2025-03-13]
5. Amboree TL, Montealegre JR, Parker SL, et al. National breast, cervical, and colorectal cancer screening use in federally qualified health centers. JAMA Intern Med 2024 Jun 1;184(6):671-679. [doi: 10.1001/jamainternmed.2024.0693] [Medline: 38683574]
6. Rockwell MS, Frazier MC, Stein JS, et al. A "sludge audit" for health system colorectal cancer screening services. Am J Manag Care 2023 Jul 1;29(7):e222-e228. [doi: 10.37765/ajmc.2023.89402] [Medline: 37523455]
7. Liu PH, Singal AG, Murphy CC. Stool-based tests mitigate impacts of COVID-19 on colorectal cancer screening. Clin Gastroenterol Hepatol 2023 Jun;21(6):1667-1669. [doi: 10.1016/j.cgh.2022.12.014] [Medline: 36566814]
8. Webber C, Flemming JA, Birtwhistle R, Rosenberg M, Groome PA. Regional variations and associations between colonoscopy resource availability and colonoscopy utilisation: a population-based descriptive study in Ontario, Canada. BMJ Open Gastroenterol 2022 Jun;9(1):e000929. [doi: 10.1136/bmjgast-2022-000929] [Medline: 35680174]
9. Brodney S, Bhat RS, Tuan JJ, et al. Barriers and proposed solutions to at-home colorectal cancer screening tests in medically underserved health centers across three US regions to inform a randomized trial. Cancer Med 2024 Aug;13(15):e70040. [doi: 10.1002/cam4.70040] [Medline: 39118261]

10. US Preventive Services Task Force, Davidson KW, Barry MJ, et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. JAMA 2021 May 18;325(19):1965-1977. [doi: 10.1001/jama.2021.6238] [Medline: 34003218]
11. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
12. Wolf MS, Baker DW, Makoul G. Physician-patient communication about colorectal cancer screening. J Gen Intern Med 2007 Nov;22(11):1493-1499. [doi: 10.1007/s11606-007-0289-y] [Medline: 17851721]
13. McQueen A, Bartholomew LK, Greisinger AJ, et al. Behind closed doors: physician-patient discussions about colorectal cancer screening. J Gen Intern Med 2009 Nov;24(11):1228-1235. [doi: 10.1007/s11606-009-1108-4] [Medline: 19763699]
14. Selby K, Cornuz J, Gachoud D, et al. Training primary care physicians to offer their patients faecal occult blood testing and colonoscopy for colorectal cancer screening on an equal basis: a pilot intervention with before-after and parallel group surveys. BMJ Open 2016 May 13;6(5):e011086. [doi: 10.1136/bmjopen-2016-011086] [Medline: 27178977]
15. Martin Y, Braun AL, Biller-Andorno N, et al. Screening refusal associated with choice of colorectal cancer screening methods. a cross-sectional study among Swiss primary care physicians. J Gen Intern Med 2019 Aug;34(8):1409-1411. [doi: 10.1007/s11606-019-05096-2] [Medline: 31190255]
16. Glaser KM, Crabtree-Ide CR, McNulty AD, et al. Improving guideline-recommended colorectal cancer screening in a federally qualified health center (FQHC): implementing a patient navigation and practice facilitation intervention to promote health equity. Int J Environ Res Public Health 2024 Jan 24;21(2):126. [doi: 10.3390/ijerph21020126] [Medline: 38397617]
17. Shim HG, Gupta A, Fu A, et al. A quality improvement study on colonoscopy wait times in underinsured patients following the COVID-19 pandemic. Clin Transl Gastroenterol 2024 Sep 1;15(9):e1. [doi: 10.14309/ctg.0000000000000730] [Medline: 38916225]

Abbreviations

CRC: colorectal cancer
FIT: fecal immunochemical test
FOBT: immunochemical fecal occult blood test
FQHC: federally qualified health center
HIPAA: Health Insurance Portability and Accountability Act

Edited by N Cahill; submitted 26.07.24; peer-reviewed by K Matthew, L Guo; revised version received 19.02.25; accepted 19.02.25; published 19.03.25.

Please cite as:
McElroy JA, Smith JB, Everett KD
Monthly Variations in Colorectal Cancer Screening Tests Among Federally Qualified Health Center Patients in Missouri: Quality Improvement Project
JMIR Cancer 2025;11:e64809
URL: https://cancer.jmir.org/2025/1/e64809
doi:10.2196/64809

Evaluating an AI Chatbot "Prostate Cancer Info" for Providing Quality Prostate Cancer Screening Information: Cross-Sectional Study

Otis L Owens, PhD, MPH; Michael S Leonard, MS, MEd, MEng

College of Social Work, University of South Carolina, 1512 Pendelton Street, Columbia, SC, United States

Corresponding Author:
Otis L Owens, PhD, MPH
College of Social Work, University of South Carolina, 1512 Pendelton Street, Columbia, SC, United States

Abstract

Background: Generative artificial intelligence (AI) chatbots may be useful tools for supporting shared prostate cancer (PrCA) screening decisions, but the information produced by these tools sometimes lack quality or credibility. "Prostate Cancer Info" is a custom GPT chatbot developed to provide plain-language PrCA information only from websites of key authorities on cancer and peer-reviewed literature.

Objective: The objective of this paper was to evaluate the accuracy, completeness, and readability of Prostate Cancer Info's responses to frequently asked PrCA screening questions.

Methods: A total of 23 frequently asked PrCA questions were individually input into Prostate Cancer Info. Responses were recorded in Microsoft Word and reviewed by 2 raters for their accuracy and completeness. Readability of content was determined by pasting responses into a web-based Flesch Kincaid Reading Ease Scores calculator.

Results: Responses to all questions were accurate and culturally appropriate. In total, 17 of the 23 questions (74%) had complete responses. The average readability of responses was 64.5 (SD 8.7; written at an 8th-grade level).

Conclusions: Generative AI chatbots, such as Prostate Cancer Info, are great starting places for learning about PrCA screening and preparing men to engage in shared decision-making but should not be used as independent sources of PrCA information because key information may be omitted. Men are encouraged to use these tools to complement information received from a health care provider.

(JMIR Cancer 2025;11:e72522) doi:10.2196/72522

KEYWORDS

generative artificial intelligence; chatbot; chatGPT; prostate cancer; cancer screening; shared decision making; artificial intelligence

Introduction

Generative artificial intelligence (AI) chatbots such as ChatGPT, Google Gemini, and Microsoft Copilot have become highly publicized for enhancing work efficiency and effectively responding to diverse queries. These sophisticated programs leverage large language models, machine learning, and natural language processing to understand and respond to a query with publicly available or third-party information [1]. Over the past 2 years, researchers have demonstrated a growing interest in evaluating generative AI chatbots for providing quality health and cancer information [2-4]. While the performance of generative AI chatbots has varied depending on the disease queried, complexity of the query, and brand of chatbot used, these tools show promise for being reliable health information resources in the future [3,5,6].

In terms of prostate cancer (PrCA), the second leading cause of cancer mortality among men in the United States [7], the American Cancer Society (ACS) [8], American Urological Association (AUA) [9,10], and the United States Preventive Services Task Force (USPSTF) [11] recommend that men make shared PrCA screening decisions with their health care providers. To prepare for this important decision, men need access to credible, readable, and culturally-appropriate (eg, African Americans have a higher mortality risk [12]) PrCA screening information [13]. Multiple studies have investigated the quality of PrCA information generated by AI chatbots [14-22]. Overall, these studies show that PrCA information produced by chatbots can be accurate, reliable, and moderately comprehensive, but readability and credibility are often compromised. In a recent study by the authors, Owens and Leonard [23] discovered that soliciting plain-language responses from chatbots to PrCA screening inquiries significantly enhanced the response's readability. Conversely, credibility was difficult to ascertain because generative AI chatbots do not consistently reference authoritative information sources [23]. To create a reliable and credible plain-language resource for

PrCA screening information, we have developed "Prostate Cancer Info" (PCI), a generative AI chatbot using Open AI's custom GPT platform [24]. PCI is unique because it only responds to inquiries from credible, PrCA expert-curated websites (like the ACS). This method is different from current generative AI chatbots, which search the entire web and produce responses from a variety of expert-vetted and non-vetted sources. In addition, we have programmed PCI to always provide a source for responses, which is uncommon for current generative AI chatbots. Finally, we have programmed PCI to provide responses that do not exceed 6th to 8th grade readability as recommended by the American Medical Association [25]. The study's purpose is to evaluate the accuracy, completeness, and readability of PCI responses to 23 frequently asked PrCA screening questions. The study will contribute insight into the safety and efficacy of using AI chatbots for shared PrCA screening decision-making and the usefulness of developing customized AI chatbots for PrCA decision-making.

Methods

Intervention Development

Author MSL developed PCI using a multistep process. Websites published by the ACS, AUA, USPSTF, and Centers for Disease Control and Prevention (CDC) were programmed into the GPT builder [24]. The rationale for limiting our search to these websites is that these organizations are globally recognized for providing timely, evidence-based PrCA screening education and recommendations. In particular, the PrCA screening recommendations from the ACS, AUA, and USPSTF are the most widely recognized in US PrCA research and clinical practice. PCI was then directed to draw responses exclusively from these websites in the order listed. Therefore, PCI relied on the ACS website as a primary source unless the information was unavailable or was requested from a non-ACS source. Strict directives were given to PCI to (1) only retrieve information from websites provided, (2) respond with language at or below 8th grade readability, (3) ignore non-PrCA queries, and (4) provide sources for responses. PCI was pretested to confirm its adherence to these directives.

PCI answers user questions through a well-defined process: first, it limits itself to information from preapproved websites. Then, it indexes these sites, reading and organizing their content. When a user asks a question, PCI searches its indexed data for relevant details, analyzes the information to understand the context, and creates a concise, accurate answer from approved sources. This ensures consistent and trustworthy answers.

Study Protocol

A total of 23 frequently asked PrCA questions were adopted from previous studies by Zhu et al [15] and Owens and Leonard [23]. One author entered questions into PCI. Responses were saved in a document for rating by both authors. The authors used a coding form containing questions with key points and answers from ACS and CDC education resources[26,27], along with screening recommendations from the ACS, AUA, and USPSTF [8-11], and checkboxes to evaluate whether a response was accurate (contained correct statements) and complete (presented all salient facts without significant omissions). For example, a response to "What is the prostate?" would be considered accurate if it stated that the prostate is a gland that is a part of the male reproductive system. However, to be considered complete, the response would also need to include information on the size of the prostate, its location, and its purpose. If any parts of the response were not correct, they were rated as inaccurate. Table 1 shows the key points used to determine accuracy and completeness. Each of these key points is critical to a shared PrCA decision because a patient must consider factors such as the risks, benefits, and uncertainties of screening; their age, race, family history; and their personal values and preferences. Our chatbot responses have been included in Multimedia Appendix 1. Additional space was allotted on the coding form to record details about inaccuracies or omissions. The authors had 100% interrater agreement. The readability of responses was determined via a web-based Flesch Kincaid Reading Ease Scores calculator. Each response was copied and pasted into the calculator, excluding the reference website. The Flesch Kincaid Reading Ease Score uses total words, sentences, and syllables in an excerpt of text to calculate a score between 0 and 100, which corresponds to grade-level readability. Scores of 60 to 100 are considered easy to read by someone possessing an education at or below 8th to 9th grade. Scores of 50 to 60 require a 10th to 12th grade education (ie, fairly difficult) and scores below 50 require a college education (ie, very difficult) to comprehend.

Table . Key points for determining accuracy and completeness[a].

Questions	Key points
Basic Questions	
What is the prostate?	• Male reproductive organ. • The size of the prostate increases with age but is walnut-sized in younger men. • Located below the bladder and in front of the rectum. • Produces some of the fluid in semen.
How common is prostate cancer?	• About 313,780 new cases of prostate cancer (1 in 8 men). • About 35,770 deaths from prostate cancer (1 in 44 men).
What are the risks for prostate cancer?	• Risk increases with age. • More common among African-American men. • More prevalent in North America, northwestern Europe, Australia, and the Caribbean islands. • Risk is doubled if a man has a first-degree relative (eg, father, brother, or son) with prostate cancer gene mutations can increase the risk for prostate cancer. • Less common risk factors are diet, obesity, smoking, chemical exposure, inflammation of the prostate, STIs[b], and vasectomy.
When and how often should a man be screened for prostate cancer?	• ACS[c]: ages 50 years (average risk), 45 years (high risk), and 40 years (very high risk). • AUA[d]: ages 45 to 50 years (average risk) and age 40 years (high risk). • USPSTF[e]: age 55 to 69 years (average risk).
What are the symptoms of prostate cancer?	• Can have no symptoms in early stages. • Urinary problems. • Blood in urine or semen. • Erectile dysfunction. • Pain in hips, back, chest, or other areas. • Weakness or numbness in legs or feet. • Loss of bladder or bowel control.
What are the types of screenings for prostate cancer?	• DRE[f] and gloved finger test are not 100% accurate. • PSA[g], a blood test, is not 100% accurate and can produce false positives and false negatives.
What are the benefits and harms of prostate cancer screening?	• Benefit: can find cancer early. • Harms: tests, especially PSA can produce false positives or negatives, which can lead to unnecessary tests or treatments, which carry risks.
How is prostate cancer diagnosed?	• Biopsy: tissue samples from the prostate.
What are the risks of a prostate biopsy?	• Pain, blood in the semen, or infection.
How long can I live if I have prostate cancer?	• The 5-year relative survival rate is 97% on average, but depends on how far the cancer has spread.
Difficult Questions	
Is the PSA or DRE more effective for finding prostate cancer?	• PSA is more effective.
My father had prostate cancer. Will I have prostate cancer too?	• Having a father or brother with prostate cancer can more than double a man's risk of prostate cancer.
I have a high PSA level. Do I have prostate cancer?	• The probability of having prostate cancer increases with PSA level but there is no set PSA level that can definitively indicate the presence of prostate cancer.
What does a PSA level of 4 mean?	• Men with a PSA level between 4 and 10 have about a 1 in 4 chance of having prostate cancer.

Questions	Key points
What does a PSA level of 10 mean?	• The chance of having prostate cancer is 50% with a PSA of 10 or more.
What does a PSA level of 20 mean?	• The chance of having prostate cancer is more than 50% with a PSA of 20 or more.
What newer tests for prostate cancer may be more accurate than the PSA test?	• The prostate health index (PHI). • 4Kscore test. • IsoPSA test. • Urine-based tests.
If my biopsy sample is positive for cancer, should I receive genetic testing?	• Some men who have a strong family history or certain inherited genes, prior cancer diagnosis, or cancer that has spread to other parts of the body, should speak to their health care provider about this option.
If my biopsy sample is positive for cancer, how soon should I start treatment?	Will depend on the stage and grade of the cancer and their: • Age and expected lifespan. • Other serious health conditions. • Feelings about treatment. • The likelihood of a cure and doctor's opinion. • Feelings about treatment side effects.
Are there any cons to taking an at-home PSA test?	• At home PSA tests do not give a man an opportunity to make a shared decision with their health care provider about the risks, benefits, and uncertainties of the PSA test.
I am an African-American male, aged 40, with a family history of prostate cancer, at what age should I begin receiving prostate cancer screening?	• Screening should begin at age 40 based on both the ACS and AUA screening guidelines.
I am an African-American male, aged 40, with a family history of prostate cancer, can you provide me with all of the information I need to know to make a shared decision about prostate cancer screening?	Response should include all key points such as: • Prostate cancer incidence and mortality statistics. • Prostate cancer risks for African-American men. • Symptoms for prostate cancer. • Screenings for prostate cancer for African-American men. • Risks and uncertainties of prostate cancer screening. • Meaning of PSA results. • Biopsy for diagnosis. • Risk of biopsy. • Steps after a positive biopsy.
What are the differences in screening recommendations between major health organizations?	• ACS: ages 50 years (average risk), 45 years (high risk), and 40 years (very high risk). • AUA: age 45 to 50 years (average risk) and age 40 years (high risk). • USPSTF: age 55 to 69 years (average risk).

[a]Key points developed from web sources produced by ACS, CDC, AUA, and UPSTF.

[b]STI: sexually transmitted infection.

[c]ACS: American Cancer Society.

[d]AUA: American Urological Association.

[e]USPSTF: United States Preventive Services Task Force.

[f]DRE:digital rectal exam.

[g]PSA: prostate specific antigen test.

Data Analysis

Data was transferred from coding forms to Microsoft Excel spreadsheets for analysis. Descriptive statistics were calculated to determine the percentage of questions answered accurately and completely. An average mean readability score was also calculated.

Results

Accuracy and Completeness

Responses to all questions were accurate. In total, 17 of 23 questions (74%) were answered completely. Of the 6 questions with less complete responses, one lacked information about geography as a risk for PrCA and the higher prevalence of PrCA in North America. Of note is that this response recognized that

African Americans may be at greater risk for the disease, but statistics were not provided in any responses that substantiated the burden of incidence and mortality among African-American men. A total of 3 questions related to the meanings of PSAs of 4, 10, and 20 lacked statistics about the probability of PrCA, but did state men's greater chance of being diagnosed with PrCA at PSAs higher than 4. A fifth question about how soon a man should start treatment after a positive biopsy lacked information about how age, expected life span, comorbidities, and patient feelings about side effects factor into treatment decisions. Finally, a sixth question about what information an African-American male, aged 40 years with a family history of PrCA needs to know to make a shared screening decision yielded an answer that lacked information about what PSA results mean or the purpose and risks of a prostate biopsy.

Readability

The average readability was 64.5 (SD 8.7), which indicates most responses were written at an 8th-grade level or below. However, 5 of 23 responses (22%) were written at a 10th to 12th grade reading level and 1 response was written at a college level. In addition, 3 of the 5 responses addressed difficult questions. Scores ranged from 48.6 to 81.3. The lowest readability score (ie, 48.6) was in response to a basic question about symptoms of PrCA (see Table 2).

Table . Accuracy, completeness, and readability of Prostate Cancer Info responses to questions about prostate cance.

Questions	Accurate?	Complete?	Readability score
Basic questions			
What is the prostate?	Yes	Yes	81.3
How common is prostate cancer?	Yes	Yes	79.4
What are the risks for prostate cancer?	Yes	No	65.5
When and how often should a man be screened for prostate cancer?	Yes	Yes	70.3
What are the symptoms of prostate cancer?	Yes	Yes	48.6[a]
What are the types of screenings for prostate cancer?	Yes	Yes	70.8
What are the benefits and harms of prostate cancer screening?	Yes	Yes	64.6
How is prostate cancer diagnosed?	Yes	Yes	71.5
What are the risks of a prostate biopsy?	Yes	Yes	56.1[b]
How long can I live if I have prostate cancer?	Yes	Yes	63
Difficult questions			
Is the PSA[c] or DRE[d] more effective for finding prostate cancer?	Yes	Yes	74.7
My father had prostate cancer. Will I have prostate cancer too?	Yes	Yes	60.4
I have a high PSA level. Do I have prostate cancer?	Yes	Yes	70.7
What does a PSA level of 4 mean?	Yes	No	67.5
What does a PSA level of 10 mean?	Yes	No	67.2
What does a PSA level of 20 mean?	Yes	No	67.5
What newer tests for prostate cancer may be more accurate than the PSA test?	Yes	Yes	63.1
If my biopsy sample is positive for cancer, should I receive genetic testing?	Yes	Yes	52.4[b]
If my biopsy sample is positive for cancer, how soon should I start treatment?	Yes	No	58.8[b]
Are there any cons to taking an at-home PSA test?	Yes	Yes	65
I am an African-American male, aged 40, with a family history of prostate cancer, at what age should I begin receiving prostate cancer screening?	Yes	Yes	51.8[b]

Questions	Accurate?	Complete?	Readability score
I am an African-American male, aged 40, with a family history of prostate cancer, can you provide me with all of the information I need to know to make a shared decision about prostate cancer screening?	Yes	No	51.6[b]
What are the differences in screening recommendations between major health organizations?	Yes	Yes	60.6
Total (yes), %	100	74	—[e]
Readability score, mean (SD)	—	—	64.5 (8.7)
Readability score, median (range)	—	—	65 (48.6-81.3)

[a]Readability was very difficult (requires a college education).

[b]Readability was fairly difficult (requires a 10th to 12th grade education).

[c]PSA: prostate specific antigen test.

[d]DRE: digital rectal exam.

[e]Not applicable.

Discussion

Principal Findings

PCI had pristine accuracy and average completeness and readability. On average, completeness and readability were higher on responses to basic questions as compared to difficult questions. Specifically, 9 of 10 (90%) of the responses to basic and 8 of 13 (62%) of the responses to difficult questions were complete. In addition, 8 of 10 or (80%) and 9 of 13 (69%) of readability scores for basic and difficult questions, respectively, were below an 8th to 9th grade level. Difficult questions often contained longer and more complex responses, which likely affected readability. Furthermore, 4 of the 6 incomplete responses only lacked 1 key point which did not significantly dilute these responses. For example, 3 responses on PSA levels at 4, 10, and 20 did not effectively highlight differences in cancer likelihood (eg, over 50% chance), but each response indicated a greater chance of prostate cancer. Therefore, men would be informed that a PSA over 4 is concerning and warrants counsel from a provider. Key points missed in responses about when to start treatment for any man and about shared decision-making for African-American men are more concerning as the omitted information (eg, biopsy as a diagnosis tool) is focal to a PrCA screening decision. Not possessing this knowledge could lead to a PrCA screening decision that is not ideally informed and may not truly be shared between the patient and their healthcare provider. Specifically, knowing that the biopsy, not the PSA, is the only definitive means to diagnose PrCA may somewhat lessen the fear of an increased PSA score because another diagnostic step exists. Being informed about the biopsy could also prompt shared discussion about the relevance of a biopsy for the patient's circumstance. In addition, providing African-American men with all the information necessary to share a PrCA screening decision based on their demographic profile could be exceptionally useful for those men who may lack access to a question list, and not have time to ask multiple questions to an AI chatbot, or simply want a more tailored answer to their given circumstance. This tailored information can also facilitate a shared PrCA screening decision that is more patient-centered.

Limitations

The 23 questions we used may not reflect the full breadth of inquiries someone may have about PrCA screening. PCI was programmed to seek information from a finite set of websites from key medical authorities, but several equally credible websites were not included (eg, Mayo Clinic), which may have slightly improved PCI's performance. While much of the general information about PrCA on these additional websites (eg, signs, symptoms, and prostate anatomy) would likely be similar, there may be cutting-edge research on new PrCA screenings that may not yet be publicized on ACS or similar websites but could provide additional context for more difficult questions like those related to newer tests for PrCA. Finally, although rigorous research methods, such as interrater reliability, were used to mitigate any study bias, we acknowledge that as the developers of PCI, we may be susceptible to unconscious biases that could have affected our ratings. For transparency, we have included all PCI responses in Multimedia Appendix 1. Future studies to evaluate PCI and similar chatbots should include external raters and user feedback.

Comparison With Previous Work

Similar to our previous research [23] and research by others [15-19,21,22], generative AI chatbots like PCI can be highly accurate when responding to PrCA and PrCA screening inquiries. The completeness and readability of Prostate Info's responses to PrCA screening questions varied. PCI generally performed better than Lombardo et al [20] and comparable with Geantă et al [20-22], both of whom investigated chatbot performance on non-US PrCA guidelines. As compared with studies using US PrCA guidelines, PCI performed better than ChatGPT-3.5 (OpenAI), ChatGPT-4 (OpenAI), Microsoft Co-Pilot, Google Gemini, and Google Gemini Advanced, but equal to Microsoft Copilot on completeness of responses to basic PrCA screening queries posed in our previous comparative study, which solicited both standard and plain-language (ie low

literacy) responses [23]. However, the average readability on basic questions was lower than all, but one (ie, Microsoft Co-Pilot) generative AI chatbot when considering plain language responses only [23]. Otherwise, PCI outperformed all, but one (ie, Google Gemini Advanced) chatbot when we asked it to provide a standard response [23]. Compared to Zhu et al [15], who evaluated multiple generative AI chatbots' performance on a similar combination of basic and difficult questions to our study, PCI did not perform nearly as well as ChatGPT and ChatGPT Plus (earlier versions of ChatGPT) on completeness. PCI outperformed all other chatbots evaluated by Zhu et al [15] including Perplexity (by Perplexity AI), YouChat (by You.com), Chatsonic (by Writesonic), and NeevaAI (by Neeva). However, PCI's average readability may have been (but not definitively) lower than all generative AI chatbots evaluated in Zhu and colleagues' [15] study. It is important to note that Zhu et al [15] used a slightly different method for calculating completeness and readability than this study or the study by Owens and Leonard [23]. Zhu et al [15] determined the percentage of comprehensiveness using a Likert approach as opposed to indicating whether a response was simply complete or not complete. Numbers listed in Table 3 for Zhu et al [15] represent the percentage of questions that were "very comprehensive" (ie, fully complete). Readability was rated by reviewers as opposed to using a validated readability measure. Percentages reported in Table 3 for Zhu et al [15], represent that percentage of total responses that were "very easy to read." SDs for Zhu et al [15] were not reported.

Table . Comparison of completeness and readability of chatbot responses on US prostate cancer screening guidelines.

Study	Chatbot name	Completeness, n/N (%)	Average readability score	
			mean (SD)	%, mean (SD)
This study	PCI[a]	17/23 (74)	64.5 (8.7)	—[b]
Zhu et al [15]	ChatGPT	21/22 (95)[c]	—	100 (NR[d])
Zhu et al [15]	ChatGPT Plus	20.3/22 (92)[c]	—	100 (NR)
Zhu et al [15]	ChatSonic	14.3/22 (65)	—	95 (NR)
Zhu et al [15]	YouChat	10.34/22 (47)	—	98 (NR)
Zhu et al [15]	Neeva AI	8.8/22 (40)	—	84 (NR)
Zhu et al [15]	Perplexity Detailed	6.6/22 (30)	—	95 (NR)
Zhu et al [15]	Perplexity Concise	6.6/22 (30)	—	95 (NR)
Owens et al [23]	ChatGPT 3.5 standard response	6/11 (54)	38.0 (7.6)	—
Owens et al [23]	ChatGPT 3.5 low literacy response	4/11 (34)	70.3 (7.2)[e]	—
Owens et al [23]	ChatGPT 4.0 standard response	7/11 (63)	43.1 (9.2)	—
Owens et al [23]	ChatGPT 4.0 low literacy response	7/11 (63)	74.1 (9.9)[e]	—
Owens et al [23]	Google Gemini standard response	6/11 (54)	55.7 (10.4)	—
Owens et al [23]	Google Gemini low literacy response	5/11 (45)	81.0 (3.6)[e]	—
Owens et al [23]	Google Gemini Advanced standard response	6/11 (54)	66.3 (9.4)[e]	—
Owens et al [23]	Google Gemini Advanced low literacy response	6/11 (54)	79.4 (5.1)[e]	—
Owens et al [23]	Microsoft Copilot standard response	8/11 (72)	50.8 (9.3)	—
Owens et al [23]	Microsoft Copilot low literacy response	6/11 (54)	65.1 (6.6)[e]	—
Owens et al [23]	Microsoft Copilot Pro standard response	7/11 (63)	61.2 (9.5)	—
Owens et al [23]	Microsoft Copilot Pro low literacy response	6/1 (54)	78.8 (4.7)[e]	—

[a]PCI: Prostate Cancer Info.

[b]Not applicable.

[c]Chatbot had a higher completeness score than PCI.

[d]NR: not reported.

[e]Chatbot had definitively higher readability scores than PCI based on the Flesch-Kincaid readability. Other scores may also be higher but were not based on a validated measure.

We expected PCI to outperform most other commercially available generative AI chatbots on completeness because of its development using the latest ChatGPT-4.0 technology and its directive to secure information from specific websites, but PCI underperformed earlier versions of ChatGPT. Therefore, additional training of the large language model that undergirds ChatGPT-4.0 will be needed for this niche area. In addition, unexpected was the lower average readability of responses from PCI, especially compared to our previous work, which solicited plain-language responses from multiple chatbots including ChatGPT-3.5 and ChatGPT-4. Nonetheless, the average readability of PCI is suitable for an audience with a middle school education.

Conclusions and Future Directions

Generative AI chatbots, such as PCI, are great starting places for learning about PrCA screening and preparing for shared decision-making but should not yet be used as sole sources of PrCA information because of their periodic omission of key information. Nevertheless, with further testing and validation,

model training, and refinement of the source selection process, we hope PCI can be a publicly available resource for credible, evidence-based, and culturally appropriate information for PrCA screening decisions. In the future, PCI could be integrated into the decision-making workflow by prompting patients to use PCI before their medical visit via an emailed link or a 1-page hard copy with a QR code. This same email or document could contain multiple frequently asked questions about PrCA. Men should be encouraged to pose as many of these questions as possible, but especially those on our list that are more complex (eg, Is the PSA or DRE more effective for finding prostate cancer?) PCI questions and responses could then be saved on their mobile device or printed, notated to indicate areas of concern or need for clarity, and then taken to their appointment to be used to guide the shared PrCA screening discussion. During this discussion, the health care provider should ensure that men understand their personal PrCA risk; screening options; and risks, benefits, and uncertainties of PrCA screening. The discussion should then shift to focus on men's questions and their screening preferences. Using this method of generative AI integration into the shared decision process could fortify men's PrCA knowledge and identify patient values and preferences.

To improve the overall performance of PCI in the future, it will be necessary to iteratively fine-tune our model which will include expanding the sources from which PCI retrieves data, which could include the most current peer-reviewed journal articles in addition to websites of major research hospitals and international health organizations. All sources will be curated by a PrCA expert who will review each data source to ensure it contains quality information. Equally important will be soliciting routine feedback from health care providers and patients through an embedded satisfaction survey that can enable them to comment on the quality of questions developed by the research team, potential questions that should be added to the database, and the quality of the responses generated by the PCI. Additional model training will come from tracking common follow-up questions from users to incorporate them into the initial responses. Based on these continuous feedback loops, PCI's performance could be improved significantly and always remain up-to-date. Future research should focus on the clinical deployment of PCI and testing to assess its acceptability, ease of use in a clinical workflow, and usefulness in the shared PrCA screening decision process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Generative artificial intelligence (AI) chatbot responses.
[DOCX File, 40 KB - cancer_v11i1e72522_app1.docx]

References

1. How ChatGPT and our foundation models are developed. Open AI. URL: https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-foundation-models-are-developed [accessed 2025-02-11]
2. Milne-Ives M, de Cock C, Lim E, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346. [doi: 10.2196/20346] [Medline: 33090118]
3. Wang L, Wan Z, Ni C, et al. A systematic review of ChatGPT and other conversational large language models in healthcare. medRxiv. 2024 Apr 27. [doi: 10.1101/2024.04.26.24306390] [Medline: 38712148]
4. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer 2021 Nov 29;7(4):e27850. [doi: 10.2196/27850] [Medline: 34847056]
5. Potapenko I, Boberg-Ans LC, Stormly Hansen M, Klefter ON, van Dijk EHC, Subhi Y. Artificial intelligence-based chatbot patient information on common retinal diseases using ChatGPT. Acta Ophthalmol 2023 Nov;101(7):829-831. [doi: 10.1111/aos.15661] [Medline: 36912780]
6. Shuaib A. Transforming healthcare with AI: promises, pitfalls, and pathways forward. Int J Gen Med 2024;17(17):1765-1771. [doi: 10.2147/IJGM.S449598] [Medline: 38706749]
7. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74(1):12-49. [doi: 10.3322/caac.21820] [Medline: 38230766]
8. Wolf AMD, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. CA Cancer J Clin 2010;60(2):70-98. [doi: 10.3322/caac.20066] [Medline: 20200110]
9. Wei JT, Barocas D, Carlsson S, et al. Early detection of prostate cancer: AUA/SUO guideline Part I: prostate cancer screening. J Urol 2023 Jul;210(1):46-53. [doi: 10.1097/JU.0000000000003491] [Medline: 37096582]
10. Wei JT, Barocas D, Carlsson S, et al. Early detection of prostate cancer: AUA/SUO guideline part II: considerations for a prostate biopsy. J Urol 2023 Jul;210(1):54-63. [doi: 10.1097/JU.0000000000003492] [Medline: 37096575]
11. US Preventive Services Task Force, Curry SJ, Owens DK, Domingo KB. Screening for prostate cancer: US Preventive Services Task Force recommendation statement. JAMA 2018;319(18):1901-1913. [doi: 10.1001/jama.2018.3710]
12. Giaquinto AN, Miller KD, Tossas KY, Winn RA, Jemal A, Siegel RL. Cancer statistics for African American/Black People 2022. CA A Cancer J Clinicians 2022 May;72(3):202-229. [doi: 10.3322/caac.21718]

13. Grene M, Cleary Y, Marcus-Quinn A. Use of plain-language guidelines to promote health literacy. IEEE Trans Profess Commun 2017;60(4):384-400. [doi: 10.1109/TPC.2017.2761578]
14. Coskun B, Ocakoglu G, Yetemen M, Kaygisiz O. Can ChatGPT, an artificial intelligence language model, provide accurate and high-quality patient information on prostate cancer? Urology 2023 Oct;180:35-58. [doi: 10.1016/j.urology.2023.05.040] [Medline: 37406864]
15. Zhu L, Mou W, Chen R. Can the ChatGPT and other large language models with internet-connected database solve the questions and concerns of patient with prostate cancer and help democratize medical knowledge? J Transl Med 2023 Apr 19;21(1):269. [doi: 10.1186/s12967-023-04123-5] [Medline: 37076876]
16. Alasker A, Alsalamah S, Alshathri N, et al. Performance of large language models (LLMs) in providing prostate cancer information. BMC Urol 2024 Aug 23;24(1):177. [doi: 10.1186/s12894-024-01570-0] [Medline: 39180045]
17. Musheyev D, Pan A, Loeb S, Kabarriti AE. How well do artificial intelligence chatbots respond to the top search queries about urological malignancies? Eur Urol 2024 Jan;85(1):13-16. [doi: 10.1016/j.eururo.2023.07.004] [Medline: 37567827]
18. Gibson D, Jackson S, Shanmugasundaram R, et al. Evaluating the Efficacy of ChatGPT as a Patient Education Tool in Prostate Cancer: Multimetric Assessment. J Med Internet Res 2024 Aug 14;26(1):e55939. [doi: 10.2196/55939] [Medline: 39141904]
19. Chiarelli G, Stephens A, Finati M, et al. Adequacy of prostate cancer prevention and screening recommendations provided by an artificial intelligence-powered large language model. Int Urol Nephrol 2024 Aug;56(8):2589-2595. [doi: 10.1007/s11255-024-04009-5] [Medline: 38564079]
20. Lombardo R, Gallo G, Stira J, et al. Quality of information and appropriateness of Open AI outputs for prostate cancer. Prostate Cancer Prostatic Dis 2025 Mar;28(1):229-231. [doi: 10.1038/s41391-024-00789-0] [Medline: 38228809]
21. Geantă M, Bădescu D, Chirca N, et al. The emerging role of large language models in improving prostate cancer literacy. Bioengineering (Basel) 2024 Jun 27;11(7):654. [doi: 10.3390/bioengineering11070654] [Medline: 39061736]
22. Geantă M, Bădescu D, Chirca N, et al. The potential impact of large language models on doctor-patient communication: a case study in prostate cancer. Healthcare (Basel) 2024 Aug 5;12(15):1548. [doi: 10.3390/healthcare12151548] [Medline: 39120251]
23. Owens OL, Leonard M. A comparison of prostate cancer screening information quality on standard and advanced versions of chatgpt, google gemini, and microsoft copilot: a cross-sectional study. Am J Health Promot 2025 Jun;39(5):766-776. [doi: 10.1177/08901171251316371] [Medline: 39854615]
24. GPT builder. Open AI. URL: https://help.openai.com/en/articles/8770868-gpt-builder [accessed 2025-02-11]
25. Weiss BD. Health Literacy and Patient Safety: Help Patients Understand Manual for Clinicians, 2nd edition: AMA; 2007.
26. Prostate cancer. American Cancer Society. 2025. URL: https://www.cancer.org/cancer/types/prostate-cancer.html
27. Prostate cancer. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/prostate-cancer/index.html

Abbreviations

ACS: American Cancer Society
AI: artificial intelligence
AUA: American Urological Association
CDC: Centers for Disease Control and Prevention
PCI: Prostate Cancer Info
PrCA: prostate cancer
USPSTF: United States Preventive Services Task Force

Edited by J Lau; submitted 16.02.25; peer-reviewed by J Ogunsakin, M Gasmi, N Mungoli, T David; revised version received 19.04.25; accepted 21.04.25; published 21.05.25.

Please cite as:
Owens OL, Leonard MS
Evaluating an AI Chatbot "Prostate Cancer Info" for Providing Quality Prostate Cancer Screening Information: Cross-Sectional Study
JMIR Cancer 2025;11:e72522
URL: https://cancer.jmir.org/2025/1/e72522
doi:10.2196/72522

Correction: Benefits of Remote-Based Mindfulness on Physical Symptom Outcomes in Cancer Survivors: Systematic Review and Meta-Analysis

Maria Komariah[1*], BSN, RN, MHSc, PhD; Sidik Maulana[1*], BSN, RN; Shakira Amirah[2], BSc; Hesti Platini[1], BSN, RN, MNS, CNS; Laili Rahayuwati[1], MHSc, MSc, Prof Dr; Ah Yusuf[3], BSN, RN, MHSc, Prof Dr; Mohd Khairul Zul Hasymi Firdaus[4], BSN, RN, MNS, PhD

[1]Faculty of Nursing, Padjadjaran University, Jl. Ir. Soekarno KM. 21, Jatinangor, Bandung, Indonesia

[2]Faculty of Medicine, University of Indonesia, Depok, Indonesia

[3]Faculty of Nursing, Airlangga University, Surabaya, Indonesia

[4]Kulliyah of Nursing, International Islamic University Malaysia, Selangor, Malaysia

[*]these authors contributed equally

Corresponding Author:
Maria Komariah, BSN, RN, MHSc, PhD
Faculty of Nursing, Padjadjaran University, Jl. Ir. Soekarno KM. 21, Jatinangor, Bandung, Indonesia

Related Article:

Correction of: https://cancer.jmir.org/2025/1/e54154

(JMIR Cancer 2025;11:e71958) doi:10.2196/71958

In "Benefits of Remote-Based Mindfulness on Physical Symptom Outcomes in Cancer Survivors: Systematic Review and Meta-Analysis" (JMIR Cancer 2025;11:e54154) an error was noted.

Reference 46 was previously a duplicate of reference 26, as follows:

Nissen ER, O'Connor M, Kaldo V, et al. Internet-delivered mindfulness-based cognitive therapy for anxiety and depression in cancer survivors: A randomized controlled trial. Psychooncol. Jan 2020;29(1):68-75.

All in-text citations to reference 46 have been changed to 26, the repeated reference information removed, and all subsequent references renumbered accordingly.

The correction will appear in the online version of the paper on the JMIR Publications website on February 13, 2025, together with the publication of this correction notice. Because this was made after submission to PubMed.

Submitted 30.01.25; this is a non–peer-reviewed article;accepted 31.01.25; published 13.02.25.

Please cite as:
Komariah M, Maulana S, Amirah S, Platini H, Rahayuwati L, Yusuf A, Firdaus MKZH
Correction: Benefits of Remote-Based Mindfulness on Physical Symptom Outcomes in Cancer Survivors: Systematic Review and Meta-Analysis
JMIR Cancer 2025;11:e71958
URL: https://cancer.jmir.org/2025/1/e71958
doi:10.2196/71958

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org