Contents

Reviews

Digital Health Psychosocial Intervention in Adult Patients With Cancer and Their Families: Systematic Review and Meta-Analysis (e46116)
Yingzi Zhang, Marie Flannery, Zhihong Zhang, Meghan Underhill-Blazey, Melanie Bobry, Natalie Leblanc, Darcey Rodriguez, Chen Zhang. . . 3

Machine Learning Approaches to Predict Symptoms in People With Cancer: Systematic Review (e52322)
Nahid Zeinali, Nayung Youn, Alaa Albashayreh, Weiguo Fan, Stéphanie Gilbertson White. ................................................................. 23

Viewpoint

Need for Culturally Competent and Responsive Cancer Education for African Immigrant Families and Youth Living in the United States (e53956)
Olufunmilola Abraham, Adeola Agoke, Kazeem Sanuth, Abimbola Fapohunda, Motolani Ogunsanya, Megan Piper, Amy Trentham-Dietz. . . 40

Original Papers

Iterative Patient Testing of a Stimuli-Responsive Swallowing Activity Sensor to Promote Extended User Engagement During the First Year After Radiation: Multiphase Remote and In-Person Observational Cohort Study (e47359)
Eileen Shinn, Adam Garden, Susan Peterson, Dylan Leupi, Minxing Chen, Rachel Blau, Laura Becerra, Tarek Rafieidi, Julian Ramirez, Daniel Rodriguez, Finley VanFossen, Sydney Zehner, Patrick Mercier, Joseph Wang, Kate Hutcheson, Ehab Hanna, Darren Lipomi. .......................... 52

Toxic Relationships Described by People With Breast Cancer on Reddit: Topic Modeling Study (e48860)
Cara Davidson, Richard Booth, Kimberley Jackson, Tara Mantler. .................................................................................................. 69

Improving Concordance Between Clinicians With Australian Guidelines for Bowel Cancer Prevention Using a Digital Application: Randomized Controlled Crossover Study (e46625)
Tsai-Wing Ow, Olga Sukhocheva, Peter Bampton, Guruparan Iyngkaran, Christopher Rayner, Edmund Tse. ........................................... 82

Assessing the Quality, Privacy, and Security of Breast Cancer Apps for Arabic Speakers: Systematic Search and Review of Smartphone Apps (e48428)
Dari Alhuwail, Aisha Alhouti, Latifah Alsarhan. ..................................................................................................................... 96
Research Letter

Web-Based Scaffolds: The Feasibility of a Constructivist Approach to Oncology Fellow Learning (e52501)
Sam Brondfield, Matthew Schwede, Tyler Johnson, Shagun Arora.
Digital Health Psychosocial Intervention in Adult Patients With Cancer and Their Families: Systematic Review and Meta-Analysis

Yingzi Zhang1*, PhD; Marie Flannery2, PhD; Zhihong Zhang2, MSc; Meghan Underhill-Blazey2, PhD; Melanie Bobry2, MSc; Natalie Leblanc2, MPH, PhD; Darcey Rodriguez3, MLiS; Chen Zhang2*, MPH, PhD

1Magnet Program and Nursing Research Department, UT Southwestern Medical Center, Dallas, TX, United States
2School of Nursing, University of Rochester Medical Center, Rochester, NY, United States
3Edward G Miner Library, University of Rochester Medical Center, Rochester, NY, United States
*these authors contributed equally

Corresponding Author:
Yingzi Zhang, PhD
Magnet Program and Nursing Research Department
UT Southwestern Medical Center
8200 Brookriver Dr
Dallas, TX, 75247
United States
Phone: 1 469 291 4808
Email: yingzi.zhang@utsouthwestern.edu

Abstract

Background: Patients with cancer and their families often experience significant distress and deterioration in their quality of life. Psychosocial interventions were used to address patients’ and families’ psychosocial needs. Digital technology is increasingly being used to deliver psychosocial interventions to patients with cancer and their families.

Objective: A systematic review and meta-analysis were conducted to review the characteristics and effectiveness of digital health interventions on psychosocial outcomes in adult patients with cancer and their family members.

Methods: Databases (PubMed, Cochrane Library, Web of Science, Embase, CINAHL, PsycINFO, ProQuest Dissertations and Theses Global, and ClinicalTrials.gov) were searched for randomized controlled trials (RCTs) or quasi-experimental studies that tested the effects of a digital intervention on psychosocial outcomes. The Joanna Briggs Institute’s critical appraisal checklists for RCTs and quasi-experimental studies were used to assess quality. Standardized mean differences (ie, Hedges g) were calculated to compare intervention effectiveness. Subgroup analysis was planned to examine the effect of delivery mode, duration of the intervention, type of control, and dosage on outcomes using a random-effects modeling approach.

Results: A total of 65 studies involving 10,361 patients (mean 159, SD 166; range 9-803 patients per study) and 1045 caregivers or partners (mean 16, SD 54; range 9-244 caregivers or partners per study) were included in the systematic review. Of these, 32 studies were included in a meta-analysis of the effects of digital health interventions on quality of life, anxiety, depression, distress, and self-efficacy. Overall, the RCT studies’ general quality was mixed (applicable scores: mean 0.61, SD 0.12; range 0.38-0.91). Quasi-experimental studies were generally of moderate to high quality (applicable scores: mean 0.75, SD 0.08; range 0.63-0.89). Psychoeducation and cognitive-behavioral strategies were commonly used. More than half (n=38, 59%) did not identify a conceptual or theoretical framework. Most interventions were delivered through the internet (n=40, 62%). The median number of intervention sessions was 6 (range 1-56). The frequency of the intervention was highly variable, with self-paced (n=26, 40%) being the most common. The median duration was 8 weeks. The meta-analysis results showed that digital psychosocial interventions were effective in improving patients’ quality of life with a small effect size (Hedges g=0.05, 95% CI –0.01 to 0.10; I²=42.7%; P=.01). The interventions effectively reduced anxiety and depression symptoms in patients, as shown by moderate effect sizes on Hospital Anxiety and Depression Scale total scores (Hedges g=–0.72, 95% CI –1.89 to 0.46; I²=97.6%; P<.001).

Conclusions: This study demonstrated the effectiveness of digital health interventions on quality of life, anxiety, and depression in patients. Future research with a clear description of the methodology to enhance the ability to perform meta-analysis is needed. Moreover, this study provides preliminary evidence to support the integration of existing digital health psychosocial interventions in clinical practice.

Trial Registration: PROSPERO CRD42020189698; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=189698
cancer; anxiety; decision-making; depression; digital health; distress; family; mental health; mortality; psychosocial intervention; quality of life

**Introduction**

Cancer is often associated with psychological distress in patients and their family members. Emerging evidence shows that psychological distress contributes to cancer mortality [1,2]. Given that over 2 million new cancer cases are expected to be diagnosed in 2024 in the United States, psychosocial distress is a significant public health problem [3]. Psychosocial distress can be triggered by many challenges, such as decision-making regarding treatment, self-care challenges due to side effects from cancer treatment, maintaining work-life balance, and financial burden. A large body of research documents the negative influence of a cancer diagnosis and treatment on a patient’s experience, including depression, anxiety, and decreased quality of life [4,5]. Cancer not only affects the patient but also imposes changes on the family [6]. Family members, who often assume caregiving roles to complement the roles of the health care team, often experience deteriorating quality of life and significant psychological distress [7,8]. For many years, researchers have examined psychosocial interventions addressing patients’ and family members’ needs to help maintain psychosocial well-being and quality of life during the cancer experience [9-12].

Increasingly, studies have used digital technology to deliver psychosocial interventions. In this report, we refer to digital health intervention as the use of digital, mobile, and wireless technologies to deliver an intervention. Digital health interventions have gained popularity due to their geographic accessibility, self-paced nature, user-friendly design, up-to-date information provision, and time-sensitive interaction with health care providers [13,14]. Further, digital interventions have significant potential for reaching people, mainly in rural areas or people with limited mobility [15]. There are various delivery modes for digital interventions, such as smartphone apps, websites, the internet, and virtual reality. There are also drawbacks, including concerns related to security and privacy and inaccessibility for people without smart device ownership. Psychosocial interventions may incorporate various components, such as communication skills training, cognitive behavioral therapy, patient education, peer support, and problem-solving training [16].

Despite the plethora of individual research studies, a synthesis of digital psychosocial interventions for patients with cancer and their families is needed to provide a summary of existing evidence regarding the effects of interventions and provide directions for future research and clinical practice. A range of systematic reviews have examined digital health psychosocial interventions for patients with cancer [17-22] and their family members [23,24]. However, these reviews have limitations. For example, some reviews primarily focused on a specific population, such as individuals with breast [17] or prostate cancer [19,20]; a particular delivery mode, such as internet-based [17,23,24]; or a specific psychosocial outcome, such as quality of life or psychological distress [21,22]. In addition, Slev et al [25] synthesized evidence from systematic reviews of interventions delivered through computers or the internet for patients with cancer and their caregivers; however, the authors failed to quantify the effectiveness of interventions across studies using advanced statistical techniques, such as a meta-analysis. To date, no studies have used meta-analytical strategies to quantify the impact of digital health interventions on psychosocial outcomes in patients with cancer and family members. To fill these gaps, we conducted a systematic review and meta-analysis to comprehensively review the characteristics and effectiveness of digital psychosocial interventions on psychosocial outcomes across different available delivery modes in adult patients with cancer and their family members.

The specific aims were to answer the following questions:

1. What are the characteristics of digital psychosocial interventions for adult patients and families living with cancer? (ie, intervention component, theoretical or conceptual framework, tailored or standardized, mode of delivery, prescribed dosage, duration of the intervention, and actual dosage)?

2. What is the efficacy of interventions on psychosocial outcomes for adult individuals diagnosed with cancer and their family members and associated factors (ie, delivery mode, control condition, and dosage, including the number of sessions, frequency, and duration)?

**Methods**

The review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [26].

**Study Identification**

The medical librarian (DR) and first author (YZ) worked together to identify search terms to build a comprehensive search strategy (Multimedia Appendix 1). Using controlled vocabulary and keywords when available, the search strategy was executed in the following databases: PubMed, Cochrane Library, Web of Science, Embase, CINAHL, PsyNFO, ProQuest Dissertations and Theses Global, and ClinicalTrials.gov. The results were limited to the English language and those published from each resource’s inception until March 2019, when the search was completed. An initial limited search of PubMed and CINAHL was undertaken, followed by an analysis of the text words in the abstract and the index terms used to describe the article. Relevancy was determined by the first author (YZ) and medical librarian (DR). A second search was undertaken across all included databases using all identified keywords and index terms.
Study Selection
The inclusion criteria were studies that (1) included adult patients (≥18 years of age with any cancer diagnosis) or their adult family members (eg, partner, caregiver, adult children, parent, or relative); (2) tested a digital health psychosocial intervention, which was defined as any nonpharmacological therapeutic intervention that addressed the psychological, social, personal, or relational adjustment needs associated with cancer through a digital health mechanism (eg, application and website); (3) measured at least 1 psychosocial outcome; and (4) used an experimental (randomized controlled trial [RCT]) or a quasi-experimental design. Studies were excluded if they enrolled pediatric patients with cancer; were review articles, letters to the editor, editorial reports, case reports, or commentaries; were published as abstracts only; and were not published in English. For meta-analysis, we excluded articles that did not provide data or when only a single study included the outcome measure.

After removing duplicates, the first author (YZ) read all titles and abstracts to identify articles based on inclusion and exclusion criteria. The full texts of all included articles were then screened independently by 2 reviewers (master’s-level or above), and final decisions were made based on consensus. Finally, articles identified in the search were imported to Endnote X8 (Clarivate Analytics).

Data Extraction and Management
A Microsoft Excel (Microsoft Corporation) spreadsheet was used to record information [27], including the description of the interventions (eg, theory basis, mode of delivery, content, actual dosage, planned dosage, standardized, or tailored), study sample (eg, age, sex, education, race, ethnicity, and cancer diagnosis), study characteristics (eg, design, randomization method, and control condition), intervention outcome variables and measurements, follow-ups, and quantitative data (ie, mean, SD, and sample size). Dosage was described as the number of intervention sessions, frequency, and duration of access to intervention. A standardized intervention was defined as all participants receiving the same intervention, while a tailored intervention involved customization of the intervention based on individual characteristics or needs [28]. We defined the prescribed dosage as the intended treatment dose, including the number of intervention sessions, frequency, and total length according to the study protocol. A codebook was created for data extraction, and the team’s decisions were tracked and recorded. All authors extracted data from 3 articles to pilot-test the spreadsheet. The research team discussed any ambiguity, resolved differences in interpretation, and modified the data extraction spreadsheet. Subsequently, each article underwent independent data extraction by YZ and another author (6 trained reviewers). The research team met throughout the study period every other week to resolve any discrepancies. A total of 15 original study authors were contacted to request missing information (eg, mean, SD, and sample size), and no additional data were received.

Assessment of Methodological Quality
The reviewers assessed the included studies for methodological rigor using standardized critical appraisal instruments from the 13-item Joanna Briggs Institute (JBI) Critical Appraisal Checklist for RCT and the 9-item JBI Critical Appraisal Checklist for quasi-experimental studies [29]. Reviewers answered each risk of bias item as “yes” (score=1), “no” (score=0), “unclear” (score=0), or “not applicable.” Possible composite scores ranged from 0 to 9 for quasi-experimental studies and 0-13 for RCTs, with higher scores indicating less risk of bias and better study quality. The applicable score (range 0-1) was calculated by dividing the composite score by the maximum score possible after subtracting any “not applicable” responses [30]. All studies were double-coded, and any disagreements were resolved through discussion with the research team [26].

Data Synthesis and Meta-Analysis

Data Synthesis
Data synthesis was completed on all articles that met the inclusion criteria. Only primary study results were included if multiple articles were published from the same intervention study. Simple descriptive statistics (ie, mean, SD, frequency, and percentage) were used to summarize study characteristics (eg, study design and participant characteristics) and key features of interventions (ie, theory, mode of delivery, number of sessions, frequency, and total length). Intervention content was grouped and narratively summarized according to the description of the intervention components.

Meta-Analytical Procedure
An a priori decision was made to only include studies in the meta-analysis if at least 2 studies used the same instrument to assess the same psychosocial outcome [31]. Standardized mean differences (ie, Hedges g) were calculated to compare intervention effectiveness across studies that used different scales or measurements. Mean differences between the scores before the intervention and the follow-up assessment after the intervention were calculated for pre-post interventions. Similarly, for the RCT studies, the results from follow-up in each study were selected and analyzed using difference scores from before and after the intervention for both intervention and control groups, with the pooled SDs. We computed the overall effect size across different time points for studies with multiple follow-ups. By doing so, we captured the time-varying effect on intervention effectiveness [31]. The overall effect (including all information across all time points) and time-varying effects, including the interim effect (during the intervention period), immediate effect (after the intervention), short-term effect (follow-up ≤8 weeks after completion of the intervention), and long-term effect (follow-up >8 weeks after completion of the intervention), were calculated. A cutoff of 8 weeks was chosen because it was the median length of the follow-up period across the included studies.

To assess study heterogeneity, the F statistic was examined. The F-statistic quantifies the proportion of total variance across studies caused by a fundamental difference between trials rather than chance. An F statistic of <25% indicates low heterogeneity.
between 25% and 75% indicates moderate heterogeneity, and >75% indicates high heterogeneity [32]. Lower heterogeneity is better. Funnel plots (ie, to visually assess the asymmetry) and Egger test (ie, to test the asymmetry statistically) assessed publication bias [33]. In funnel plots, if points are distributed equally between positive and negative effects, bias is lacking; variability is expected to be greater near the bottom of the chart among smaller sample size studies. For the analysis of data from studies with more than 1 digital psychosocial intervention group, we compared each digital psychosocial intervention group to the control group separately. Additionally, subgroup analysis was planned based on the review’s focus on examining the effect of delivery mode, type of control condition, and dosage on outcomes. Furthermore, we performed sensitivity analyses by including and excluding studies with extreme weights in the analyses. We used the DerSimonian-Laird random-effects model to weight and pool the individual estimates to capture variance across different studies, as all included studies were conducted in heterogeneous populations across various settings [34]. We performed all statistical and meta-analyses using STATA (version 17; StataCorp LLC).

Results

Search Results

After removing duplicates, a total of 2108 studies were identified. Figure 1 shows a flow diagram of studies identified, screened, included, and excluded from this systematic review and meta-analysis. After screening titles and abstracts and applying inclusion and exclusion criteria, a total of 70 records with 65 unique studies (for multiple manuscripts published from the same intervention study, only primary manuscripts were included) were included in the systematic review [35-99] and 32 studies [35,36,38,40,41,43,44,47,50,53-55,57,64,66,68-72, 74,77,80,82,83,86,88,91,93,95,96,99] with available data were included in the meta-analysis. A total of 33 studies were excluded from the meta-analysis because either data were unavailable to calculate the effect size (n=14) [42,46,51,62,73,76,78,79,81,84,87,89,94,98] or no other study used the same measure (n=19) [37,39,45,48,49,52, 56,58-61,63,65,67,75,85,90,92,97].

Study Characteristics

Overview

Of the 65 studies, 48 (74%) were RCTs [35-37, 39-50,52-57,59-65,72,73,76-82,85,89,91,93-99], and 17 (26%) were quasi-experimental [38,51,58,66-71,74,75,83,84,86-88,92]. More than half (n=37, 58%) of the studies were conducted in the United States [36-40,44,48-50,54,56-61,63, 65,67-69,71,72,74,77,79,80,83-85,87-90,94,97,98], and the rest were from the Netherlands (n=9, 14%) [35,41,43,47,64,70,91,92,96], Australia (n=5, 8%) [45,66,75,86,95], and other countries (eg, Denmark and Ireland). In total, 10,361 patients (mean 159, SD 166; range 9-803

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.
patients per study) were included: 7098 female patients and 3263 male patients; 1045 caregivers or partners were enrolled (mean 16, SD 54; range 9-244 caregivers or partners per study), including 781 female individuals and 264 male individuals. The average age of patients ranged from 39.9 to 72 years, and the average age of caregivers or partners ranged from 51.5 to 58.8 years. In the 33 studies that provided information about race and ethnicity, most patients (n=3495, 90%) and family members (n=259, 97%) were described as “White” or “Caucasian.” The cancer diagnoses varied across studies, with the most prevalent being breast cancer (n=24, 37%) [35,37,38,42,44,52,54,58,60,61,64,66,70-72,74,76,80,82,84,91,93,95], mixed cancer diagnosis (n=19, 29%) [36,40,45,47,50,51,53,57,62,69,75,77,78,81,83,94,96,97,99], and prostate cancer (n=7, 11%) [39,48,56,67,85,88,99]. The attrition rate ranged from 0% to 76%, with a median of 16.8% (mean 20.4%, SD 13.7%). The recruitment rate ranged from 4.4% to 94.2%, with a median of 59.5% (mean 56%, SD 24.6%). Detailed information about the study and sample characteristics from the included studies is provided in Table S1 in Multimedia Appendix 2 [35-99].

Control Condition

Of the 48 RCTs, 29 (60%) studies included a usual care control group [35,36,39,41-44,46,47,52,54,55,62,64,65,73,76-80,82,85,91,93-96,99], and 19 (40%) included an active control [37,40,45,48-50,53,56,57,59-61,63,72,81,89,90,97,99]. Among the 17 quasi-experimental studies, 11 (65%) did not have a control group [38,51,67-71,74,86-88], and 6 (35%) studies included a usual care control group [58,66,75,83,84,92].

Outcome Assessment

A total of 21 studies had 1 follow-up assessment [22,38,45,47,48,50,51,57,68-70,72,74,76-78,88-90,98,99], 23 had 2 follow-up assessments [35,39,44,46,52-55,63-65,67,71,75,79,83,84,86,87,92,93,95,97], 11 had 3 follow-up assessments [37,40,43,60-62,66,80,91], and 7 had 4 or more follow-up assessments [49,56,58,79,83,85]. The timing of follow-up assessments varied, ranging from immediately to 6 months after the intervention. The commonly reported outcomes and relevant measures are reported below in Aim 2: Effects on Patients’ and Family Members’ Psychosocial Outcomes.

Quality Assessment (Risk of Bias)

The quality assessment scores of the included studies are summarized in Table S1 in Multimedia Appendix 2. Overall, the RCT studies’ general quality was mixed, with applicable scores ranging from 0.38 to 0.91 (mean 0.61, SD 0.12). Quasi-experimental studies were generally of moderate to high quality, with applicable scores ranging from 0.63 to 0.89 (mean 0.75, SD 0.08) on the JBI Critical Appraisal Checklist for quasi-experimental studies. The publication year and applicable scores ranged from 0.63 to 0.89 (mean 0.75, SD 0.08) on the JBI Critical Appraisal Checklist for RCTs. The quality assessment scores of the included studies are summarized in Table S1 in Multimedia Appendix 2. Overall, the RCT studies’ general quality was mixed, with applicable scores ranging from 0.38 to 0.91 (mean 0.61, SD 0.12). Quasi-experimental studies were generally of moderate to high quality, with applicable scores ranging from 0.63 to 0.89 (mean 0.75, SD 0.08) on the JBI Critical Appraisal Checklist for quasi-experimental studies. The publication year and applicable scores ranged from 0.63 to 0.89 (mean 0.75, SD 0.08) on the JBI Critical Appraisal Checklist for RCTs.

Aim 1: Intervention Characteristics

Overview

There was large heterogeneity in intervention components, theoretical or conceptual framework, type of intervention (ie, tailored or standardized), mode of delivery, prescribed dosage (ie, number of sessions, frequency, and length), and received dosage (Table S2 in Multimedia Appendix 2).

Intervention Components

A total of 37 (57%) out of 65 studies included a single intervention component [36,38-40,44,45,47,51,52,55-57,63-66,68,69,71-77,79,80,82-84,86,87,90,91,93,94,98], 13 (20%) studies included 2 intervention components [35,41,43,46,48,50,62,67,78,85,88,95,96], and 15 (23%) studies included 3-5 intervention components [37,42,49,53,54,58-61,70,81,89,92,97,99]. The most common intervention components were information and resources, or psychoeducation (n=29, 45%) [35,37,39-43,46,48,49,52,56-58,61-70,81,82-87,90,92,95,97], and cognitive-behavioral strategies (n=20, 31%) [44,45,47,50,54,57,63,64,67,68,71,74,75,80-85,87,89,91,98].

Theoretical or Conceptual Framework

More than half (n=38, 59%) of the included studies did not identify a conceptual or theoretical framework [37,41,42,45-47,51,52,54,55,57,62,66,69,71,73,75,76,78,81,85,90,92,93,95,97,98].

Standardized or Tailored Intervention

Of the 65 studies, 26 (40%) included both standardized and tailored interventions [37,39,42,43,46,47,49,53,55,59-61,64,68,70-73,75,77-79,81,85,89,92,94], 28 (43%) studies included only standardized interventions [36,38,41,44,50,51,54,56-58,63,65-67,69,71,72,76,80-82,84,87,88,91,93,95,98], and 11 (17%) studies had only tailored interventions [35,40,45,48,52,62,86,90,96,97,99].

Modes of Delivery

The majority of studies conducted interventions through an internet website (n=40, 62%) [35-37,39-50,53,55,59-62,64,67,68,70,72,75,77-81,82,85-88,91,93,95-99] or smart device app (n=8) [43,52,56,57,63,69,80,87]. A total of 7 (11%) studies conducted interventions through virtual reality [51,66,73,76,78,83,84], 3 (5%) studies through telehealth [54,74,79], and 2 (3%) studies through a computer program [38,65]. Electronic health information systems [92], interaction portals [58], and videocounters [86] were each used in 1 study. Overall, 2 studies used multimodal interventions delivered through the combination of either telephone and videocounter [94] or internet and telephone [61].

Dosage

The dosage prescribed and received were highly variable. The number of intervention sessions ranged from 1 to 56, with a median of 6. A total of 27 (42%) studies did not specify the prescribed dosage; 19 (29%) only stated the number of days participants had access to the intervention [35,36,42,49,56,59-62,67,70,80,81,85,87,89,92,95,99] and 8 (12%) did not provide information on the prescribed dosage [37,39,40,52,55,58,72,73]. Frequency was highly variable, with
self-paced (n=26, 40%) as the most common [35,36,42,45,49,50,56,59-63,67,70,80,81,85,87-90,92,95-97,99], meaning no specific intervention frequency was defined and the intervention content was available throughout the study period. The other common frequencies of intervention sessions were weekly (n=17, 26%) [38,41,44,45,47,53,54,64,68,71,74,75,77,86,91,94,98] and 1-time intervention sessions (n=8, 12%) [48,65,66,76,78,82-84]. The median length of the intervention was 8 weeks, with the length ranging from 1 hour (ie, use of the intervention on an iPad for an hour) to 24 months. Received dosage was defined as the uptake of the intervention by the participants. A total of 18 (28%) studies did not report the received dosage [36,37,39,48,52,56,58,65,69,75,76,78,79,82,85,86,93,94]. Various information was reported, including attendance rate, number of times participants used the app, frequency with which participants logged into the website, number of website pages reviewed, skill practice time, and intervention session completion rate. Most of the interventions (n=43, 66%) were self-delivered without an interventionist, with self-paced being most common [35-37,39,40,42,44,45,48-53,55,57-60,62,63,67,69-73,75,76,78,80-84,87,88,91-93,95,96,99].

Aim 2: Effects on Patients’ and Family Members’ Psychosocial Outcomes

Patients’ Outcomes

Overview
A meta-analysis was conducted on 32 studies. Overall, 5 outcomes were examined. A summary of the interventions’ overall effect sizes; time-varying effect sizes for quality of life, anxiety, depression, distress, and self-efficacy; and heterogeneity statistics for each outcome is displayed in Table 1. The forest plots for overall effect sizes and time-varying effects are displayed in Multimedia Appendix 3. The funnel plots for overall effect sizes and time-varying effects are displayed in Multimedia Appendix 4.
<table>
<thead>
<tr>
<th>Population, outcome, measure, and value</th>
<th>Effect at different time points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
</tr>
<tr>
<td><strong>Patient</strong></td>
<td></td>
</tr>
<tr>
<td><strong>QOL</strong></td>
<td></td>
</tr>
<tr>
<td>FACT-B</td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>0.13 (–0.05 to 0.31)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>62.3</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>10.61 (4)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.03</td>
</tr>
<tr>
<td>FACT-G</td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>–0.04 (–0.17 to 0.09)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>0</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>1.91 (4)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.43</td>
</tr>
<tr>
<td>QLQ-30</td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>0.05 (–0.04 to 0.14)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>58.4</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>19.95 (6)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.03</td>
</tr>
<tr>
<td>SF36</td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>0.03 (–0.10 to 0.15)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>14.4</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>8.41 (8)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.31</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>0.05 (–0.01 to 0.10)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>42.7</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>48.12 (20)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Anxiety and depression</strong></td>
<td></td>
</tr>
<tr>
<td>HADS total score</td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>–0.72 (–1.89 to 0.46)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>97.6</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>165.82 (14)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Population, outcome, measure, and value</td>
<td>Effect at different time points</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td></td>
<td>Overall&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>HADS-depression</strong></td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>-0.13 (-0.23 to -0.02)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>0</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>4.17 (7)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.73</td>
</tr>
<tr>
<td><strong>CESD</strong></td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>0.10 (-0.10 to 0.30)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>0</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>0.99 (4)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.91</td>
</tr>
<tr>
<td><strong>PHQ9</strong></td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>-0.05 (-0.17 to 0.08)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>0</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>0.78 (1)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.38</td>
</tr>
<tr>
<td><strong>Multiple scales</strong></td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>0.32 (-0.35 to 0.99)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>95</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>19.86 (1)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>0.03 (-0.10 to 0.16)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>60.9</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>40.77 (16)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Anxiety</strong></td>
<td></td>
</tr>
<tr>
<td><strong>HADS-anxiety</strong></td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>0.32 (-0.20 to 0.84)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>94.3</td>
</tr>
<tr>
<td>Heterogeneity, $\chi^2 (df)$</td>
<td>123.33 (7)</td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>SATI</strong></td>
<td></td>
</tr>
<tr>
<td>Pooled ES, Hedges g (95% CI)</td>
<td>-0.19 (-0.41 to 0.04)</td>
</tr>
<tr>
<td>$I^2$</td>
<td>26.8</td>
</tr>
<tr>
<td>Population, outcome, measure, and value</td>
<td>Effect at different time points</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td></td>
<td>Overall(^a)</td>
</tr>
<tr>
<td>Heterogeneity, (\chi^2 (df))</td>
<td>5.46 (4)</td>
</tr>
<tr>
<td>(P) value</td>
<td>.24</td>
</tr>
</tbody>
</table>

**Overall**

| Pooled ES, Hedges g (95% CI)             | 0.12 (–0.19 to 0.43) | −0.10 (–0.19 to 0) | −0.04 (–0.19 to 0.12) | −0.13 (–0.43 to 0.17) | —      |
| \(I^2\)                                | 90.2 | 6.7 | 35.1 | 10.5 | —      |
| Heterogeneity, \(\chi^2 (df)\)          | 132.99 (13) | 13.94 (13) | 6.16 (4) | 1.12 (1) | —      |
| \(P\) value                           | <.001 | .38 | .19 | .29 | —      |

**Distress**

**DT\(^m\)**

| Pooled ES, Hedges g (95% CI)             | 0.98 (–0.18 to 2.14) | 0.51 (0.10 to 0.92) | —         | —     | —      |
| \(I^2\)                                | 98.5 | 54.2 | —         | —     | —      |
| Heterogeneity, \(\chi^2 (df)\)          | 332.71 (2) | 4.37 (2) | —         | —     | —      |
| \(P\) value                           | <.001 | .11 | —         | —     | —      |

**Self-efficacy**

**CBI\(^n\)**

| Pooled ES, Hedges g (95% CI)             | −1.41 (–4.02 to 1.20) | 2.56 (−1.22 to 6.35) | —         | —     | —      |
| \(I^2\)                                | 99 | 98.2 | —         | —     | —      |
| Heterogeneity, \(\chi^2 (df)\)          | 1.06 (1) | 55.43 (1) | —         | —     | —      |
| \(P\) value                           | .29 | <.001 | —         | —     | —      |

**Family member**

**Depression**

**HADS-depression**

| Pooled ES, Hedges g (95% CI)             | −0.25 (−0.72 to 0.21) | —         | —     | —     | —      |
| \(I^2\)                                | 0 | —         | —     | —     | —      |
| Heterogeneity, \(\chi^2 (df)\)          | 0.41 (1) | —         | —     | —     | —      |
| \(P\) value                           | .52 | —         | —     | —     | —      |

**Anxiety**

**HADS-anxiety**

| Pooled ES, Hedges g (95% CI)             | −0.23 (−0.70 to 0.23) | —         | —     | —     | —      |
| \(I^2\)                                | 0 | —         | —     | —     | —      |
| Heterogeneity, \(\chi^2 (df)\)          | 0.65 (1) | —         | —     | —     | —      |
The overall effect accounts for time-varying effect across different time points.

bQOL: quality of life.

cFACT-B: Functional Assessment of Cancer Therapy–Breast.

dES: effect size.

eNot applicable.


gQLQ-30: Quality of Life Questionnaire, 30 items.

hSF36: Short Form Survey 36-item.

iHADS: Hospital Anxiety and Depression Scale.

jCESD: Center for Epidemiologic Studies Depression Scale.

kPHQ: Patient Health Questionnaire-9.

lSATI: State-Trait Anxiety Inventory.

mDT: Distress Thermometer.

nCBI: Coping Behaviors Inventory.

Quality of Life

Quality of life was measured by the Functional Assessment of Cancer Therapy–Breast [44,54,80,93,95], Functional Assessment of Cancer Therapy–General [38,53,57,86,88], European Organization for Research and Treatment of Cancer Quality of Life Questionnaire, 30 items [35,40,43,74,91,96,99], and 36-item Short Form Survey [41,54,64,70,71]. Overall, a total of 21 studies with 1847 participants in the intervention groups showed an increase in quality of life, with a mean difference between groups of Hedges $g$ = 0.05 (95% CI –0.01 to 0.10). The impact of heterogeneity within the studies was significant ($I^2$ = 42.7%; $P$ = .01). With respect to publication bias, the funnel plot displayed a greater number of studies toward the top of the mean (Egger test, $P$ = .001). The time-varying effects were as follows: Hedges $g$ = –0.16 (95% CI –0.39 to 0.06) for the interim effect; Hedges $g$ = 0.95 (95% CI –1.99 to 3.89) for the immediate effect; Hedges $g$ = 2.25 (95% CI 0.36-4.14) for the short-term effect; and Hedges $g$ = 0.18 (95% CI 0-0.35) for the long-term effect. The statistical heterogeneity among studies was $I^2$ = 70% ($P$ = .07) for the interim effect; $I^2$ = 100% ($P$ < .001) for the immediate effect; $I^2$ = 98% ($P$ < .001) for the short-term effect; and $I^2$ = 18.3% ($P$ = .29) for the long-term effect.

Anxiety and Depression

Hospital Anxiety and Depression Scale (HADS) total scores (without subscale scores reported) were reported in 5 studies with 338 participants in the intervention groups [43,47,64,86,91]. Overall, participants receiving interventions reported decreased anxiety with a standardized mean difference of Hedges $g$ = –0.72 (95% CI –1.89 to 0.46). The heterogeneity within the studies was significant ($I^2$ = 97.6%; $P$ < .001). The funnel plot was found to be asymmetric, and Egger test was found to be not statistically significant ($P$ = .77). The time-varying effects were as follows: Hedges $g$ = –0.04 (95% CI –0.23 to 0.16) for the immediate effect; Hedges $g$ = –0.22 (95% CI –0.54 to 0.10) for the short-term effect; and Hedges $g$ = 0.14 (95% CI –0.09 to 0.38) for the medium-term effect. The statistical heterogeneity among studies was $I^2$ = 0% across all time-varying effects.

Depression

Depression was assessed by the HADS-depression subscale [50,69,74,86,95,96,99], Center for Epidemiologic Studies Depression Scale [41,68,71,77], Patient Health Questionnaire-9 (PHQ-9) [40,53], and a combination of the PHQ-9 and HADS-anxiety [43,57] in 1509 participants in the intervention groups. Overall, interventions were not more effective than control conditions for reducing depression (Hedges $g$ = 0.03, 95% CI –0.10 to 0.16), with a high heterogeneity of 60.9% ($P$ < .001). With respect to publication bias, the funnel plot displayed a greater number of studies toward the top of the mean (Egger test, $P$ = .25). The time-varying effects were as follows: Hedges $g$ = 0.06 (95% CI –0.10 to 0.22) for the immediate effect and Hedges $g$ = –0.04 (95% CI –0.22 to 0.14) for the interim effect. The statistical heterogeneity among studies was $I^2$ = 69.4% for the immediate effect and $I^2$ = 29.8% for the interim effect.

Anxiety

Anxiety was assessed by the HADS-anxiety subscale [57,64,69,74,86,95,96,99], State-Trait Anxiety Inventory (STAI) [66,71,72,82,83], and a combination of the STAI and HADS-anxiety [43] in 1075 participants in the intervention groups. Overall, interventions were not more effective than control conditions for reducing anxiety (Hedges $g$ = 0.12, 95% CI –0.19 to 0.43), with high heterogeneity of 90.2% ($P$ < .001). The funnel plot displayed a greater number of studies toward the top of the mean (Egger test, $P$ = .46). The interim effect was Hedges $g$ = –0.04 (95% CI –0.19 to 0.12), and the immediate effect was Hedges $g$ = –0.10 (95% CI –0.19 to 0.0). The short-term effect was Hedges $g$ = –0.13 (95% CI –0.43 to 0.17). The statistical heterogeneity among studies was $I^2$ = 35.1% for the interim effect, $I^2$ = 6.7% for the immediate effect, and $I^2$ = 10.5% for the short-term effect.
Distress
Psychological distress was assessed in 182 participants in the intervention groups using the distress thermometer [35,69,91]. Overall, participants in the intervention groups showed no reduction in distress, with a mean difference between groups of Hedges $g=0.98$ (95% CI –0.18 to 2.14). The impact of heterogeneity within the studies was significant ($I^2=98.5\%$; $P<.001$). Regarding publication bias, the funnel plot displayed a symmetric distribution around the mean effect (Egger test, $P=.46$). The immediate effect was Hedges $g=0.51$ (95% CI 0.10-0.92), with statistical heterogeneity $I^2=54.2\%$.

Self-Efficacy
Self-efficacy was measured by the Coping Behaviors Inventory in 174 participants in the intervention groups [44,55]. Overall, participants in the intervention groups did not report improvement in self-efficacy, with a standardized mean difference of Hedges $g=–1.41$ (95% CI –4.02 to 1.20). However, the impact of heterogeneity within studies was significant ($I^2=99\%$; $P<.001$). Regarding publication bias, the funnel plot displayed a symmetric distribution around the mean effect (Egger test, $P=.22$). The immediate effect was Hedges $g=2.56$ (95% CI –1.22 to 6.35) with high heterogeneity ($I^2=98.2\%$; $P<.001$).

Subgroup Analyses
Given the heterogeneity of reporting on dosage information and limited data, the subgroup analysis of dosage on intervention effect was not conducted. Table 2 includes the results of the subgroup analysis on the effect on quality of life, depression and anxiety, and distress. Overall, the associations between delivery mode and control condition with patient outcomes were not statistically significant ($P>.05$).

Table 2. Subgroup analyses on the effect of delivery mode (internet vs noninternet) and control condition (usual care vs active control) on patient outcomes.

<table>
<thead>
<tr>
<th>Outcome and moderators</th>
<th>Effect size, Hedges g (95% CI)</th>
<th>SE</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quality of life (27 studies)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery mode</td>
<td>0.04 (–0.06 to 0.14)</td>
<td>0.05</td>
<td>.45</td>
</tr>
<tr>
<td>Control condition</td>
<td>–0.01 (–0.99 to 0.06)</td>
<td>0.04</td>
<td>.78</td>
</tr>
<tr>
<td><strong>HADS$^a$ total (6 studies)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery mode</td>
<td>0.16 (–0.30 to 0.62)</td>
<td>0.24</td>
<td>.50</td>
</tr>
<tr>
<td>Control condition</td>
<td>N/A$^b$</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Depression (21 studies)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery mode</td>
<td>0.10 (–0.09 to 0.29)</td>
<td>0.10</td>
<td>.31</td>
</tr>
<tr>
<td>Control condition</td>
<td>–0.04 (–0.17 to 0.09)</td>
<td>0.07</td>
<td>.55</td>
</tr>
<tr>
<td><strong>Anxiety (15 studies)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery mode</td>
<td>0.01 (–0.07 to 0.09)</td>
<td>0.04</td>
<td>.79</td>
</tr>
<tr>
<td>Control condition</td>
<td>–0.06 (–0.34 to 0.22)</td>
<td>0.14</td>
<td>.67</td>
</tr>
<tr>
<td><strong>Distress (8 studies)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery mode</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Control condition</td>
<td>–0.03 (–0.34 to 0.28)</td>
<td>0.15</td>
<td>.86</td>
</tr>
</tbody>
</table>

$a$HADS: Hospital Anxiety and Depression Scale.

$b$N/A: not applicable.

Family Members’ Outcomes
The forest plots for overall effect sizes are displayed in Multimedia Appendix 5. The funnel plots for overall effect sizes are displayed in Multimedia Appendix 6. For family members’ data, we pooled 2 studies [36,69] on anxiety and depression for the meta-analysis with 68 participants in the intervention groups. Due to the small sample size, no time-varying effect or subgroup analysis was calculated. The overall effect on anxiety was Hedges $g=–0.23$ (95% CI –0.70 to 0.23), with heterogeneity of $I^2=0\%$ ($P=.42$). The overall effect on depression was Hedges $g=–0.25$ (95% CI –0.72 to 0.21), with heterogeneity of $I^2=0\%$ ($P=.52$). Regarding publication bias, the funnel plot displayed asymmetrical scattered points with statistical significance (Egger test, $P<.001$).

Discussion
Overview
This systematic review and meta-analysis of 65 unique digital psychosocial intervention studies for patients with cancer and their family members provides strong evidence that psychosocial interventions delivered through digital health significantly improve psychosocial outcomes. There were 3 major findings. First, this review included a large group of participants with various cancer diagnoses; however, underrepresented populations affected by cancer were not included, and the results
predominantly focused on White patients. Second, we found that various intervention modes and components were used. There is a lack of specificity with respect to the description of interventions or theoretical basis for interventions, which may hinder future replication or refinement of the interventions and understanding of underlying mechanisms. Third, despite high heterogeneity across studies, the available data suggest that digital psychosocial interventions effectively improve some psychosocial outcomes, including patients’ quality of life, anxiety, and depression.

**Principal Findings**

First, the majority of participants in the included studies were White and female, which does not reflect the broader patient population with cancer, including non-White ethno-racial groups (ie, African American or Black, American Indian and Alaska Native, Asian, Native Hawaiian or other Pacific Islander, and Hispanic or Latino populations). It is well documented in the literature that the impact of cancer on psychological distress and quality of life is worse for racial and ethnic minority groups [100-102]. Therefore, future trials should include more participants from underrepresented groups to reduce health care disparities and improve generalizability in diverse populations [103]. Family members and caregivers were rarely included in the studies reviewed. However, there is ample evidence that family members and caregivers experience significant caregiver burden, worsening quality of life, and difficulty with psychological adjustment, therefore needing support [104,105]. Previous systematic reviews suggest that interventions targeting problem-solving and communication skills may ease the burden related to patient care and improve caregivers’ quality of life [106]. Many reviews focus on the evaluation of nondigital interventions targeting the psychosocial experience in family members and caregivers, including several reviews of caregiver interventions [9,107-109]. Therefore, with growing technology usage, more digital interventions are needed to address family members’ or caregivers’ needs.

Few RCTs met all quality criteria, including blinding, analysis by treatment assignment, and standardized outcome assessment [110]. While concealing assignments from participants and those delivering interventions is not always possible, single blinding of assessors should occur in well-designed research. Few studies used power calculations for sample size, making it difficult to determine whether sample sizes were adequate [111]. Generally, results from group sizes <20 are questionable. There are several effective strategies known to increase the retention rate, such as adding monetary incentives and using an open trial design [112,113]. The critical appraisal also depends on comprehensive reporting of study details, which were limited in the identified studies. Although attempts have been made to improve reporting using the CONSORT (Consolidated Standards of Reporting Trials) statement for RCTs and the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) statement for nonrandomized intervention studies in the early 2000s [114,115], we did not see improvement in appraisal scores over time. The main limitations of the results include underpowered and methodologically weaker studies. These highlight the need for improved methodologies in future research, as the overall methodological quality was moderate.

Second, this study identified various intervention modes and components, which is consistent with a previous systematic review of psychosocial interventions for patients with advanced cancer, which identified similar intervention components, including psychoeducation and CBT-based intervention, as 2 of the most common [11]. However, more than half of the included studies did not use theoretical or conceptual frameworks to guide the development of intervention components or selection of outcomes. The lack of a theoretical framework leads to a lack of clarity about the mechanisms through which intervention components impact psychosocial outcomes [116]. In future research, theories or conceptual frameworks need to be incorporated to help us better understand the mechanisms that explain the changes in psychosocial outcomes when using digital health interventions.

In addition, the prescribed dosage information (ie, the number of sessions, duration, and frequency) was inconsistently reported, making it difficult to estimate an efficacious intervention dose. Most of the interventions were self-paced, without the involvement of an interventionist, which gives the patient autonomy to choose which intervention component or module they would like to focus on and how much time to allocate. However, there is a lack of information on intervention uptake, which may have influenced the effectiveness of interventions. Approaches that tackle barriers to adherence at various levels (eg, individual, family, clinician, agency, and environment) and improve engagement should be implemented [117]. For example, a scoping review about engagement strategies in digital interventions for mental health promotion recommended personalized feedback, e-coaching to guide content and individual progress, social platforms and interaction with peers, content gamification, reminders, and ease of use [118].

Third, we found some significant improvement in the patient’s quality of life. Some studies with a smaller number of participants or with a focus on internet-based interventions reported an improvement, but the results from these studies were not consistent [21,119]. This meta-analysis, including 21 studies, revealed a small effect size for overall effectiveness of digital health interventions in improving patients’ quality of life (Hedges g=0.05, 95% CI –0.01 to 0.10), with time-varying effects shown as promising. Another meta-analysis that pooled 16 studies demonstrated a larger positive effect of mHealth interventions on the quality of life of patients with cancer (standardized mean difference 0.28, 95% CI 0.03-0.53) [21]. Another meta-analysis that included 6 internet-based psychoeducational interventions for patients with cancer showed no significant improvement in quality of life (mean difference 1.10, 95% CI –4.42 to 6.63) [119]. Importantly, our analyses found the largest improvements in quality of life occurred from post intervention to 8 weeks (Hedges g=2.25, 95% CI 0.36-4.14). The effect of psychosocial interventions decreased after 8 weeks of follow-up, suggesting that interventions may need booster sessions or tailoring to time-sensitive needs in order to maintain effectiveness in the long term. This result was limited by substantial inconsistency across studies in all evaluation periods except the medium-term effect [32]. In addition, given the heterogeneity of follow-up periods in selected...
studies, the time-varying effect was only tested with a small number of studies, not in the 21 studies we used to calculate the overall effect.

This meta-analysis was able to demonstrate the effectiveness of digital health interventions on both anxiety and depression (measured by HADS: Hedges $g = -0.72$, 95% CI $-1.89$ to $0.46$). Our finding was partially consistent with the other meta-analyses. One study showed that internet-based psychoeducational interventions had a significant effect on decreasing depression (standardized mean difference $-0.58$, 95% CI $-1.12$ to $-0.03$), but found no evidence for effects on distress (standardized mean difference $-1.03$, 95% CI $-2.63$ to $0.57$) [119]. However, there was considerable heterogeneity in measurements among the studies included in the review by Wang et al [119]; it is difficult to determine how meaningful it is to make direct comparisons between the studies included in this meta-analysis and past reports. Possible ways to address this problem could be using similar outcome measures and a standardized study report.

Our meta-analysis demonstrated that the interventions were effective in reducing anxiety and depression in family caregivers. However, the effect size was small, perhaps due to the limited number of studies. This is partially consistent with findings from another meta-analysis, which found depressive symptoms decreased from baseline to post intervention (Hedges $g = -0.44$, 95% CI $-1.03$ to $0.15$) [120], while anxiety remained relatively stable when comparing intervention to control either at postintervention (Hedges $g = 0.12$, 95% CI $0.16$ to $0.44$) or during follow-up (Hedges $g = -0.08$, 95% CI $-0.34$ to $0.19$).

Strengths and Limitations
This review’s strength lies in its rigorous design, sophisticated data synthesis, and enduring empirical contributions. We acknowledge that our literature search was conducted 4 years before manuscript submission. The findings and contributions from our research remain pertinent and enduring. This is because, as digital psychosocial interventions continue to evolve, the core intervention content and outcomes have remained relatively consistent over the past 4 years. It would be valuable to conduct a reassessment of the evolving body of evidence concerning digital psychosocial interventions that have emerged since the onset of the COVID-19 pandemic. Moreover, this study was conducted in line with best practices by double-coding and following the PRISMA guidelines. The meta-analysis, including subgroup analysis, was conducted using appropriate methods for combining studies across various follow-up periods. Although we did an extensive search at the start of this review, we may have missed some critical studies, unreported, or unfinished studies. If all data were available, the meta-analysis could have reduced the chances of inflated type-1 error for both observed and unobserved effects that were available for assessment [121]. There was not enough data to perform post hoc analyses to examine the effect of factors such as intervention components and length of intervention on outcomes due to insufficient data.

Conclusions
Patients with cancer and their family members need high-quality psychosocial interventions throughout the cancer trajectory. Digital technologies provide a platform to deliver evidence-based psychosocial interventions from a distance, without the heightened risk of contracting viruses, especially for patients with cancer whose immune systems are compromised. This study comprehensively synthesized the effects of digital psychosocial interventions for people affected by cancer. Our findings suggest that digital health interventions are effective for adult patients with cancer and their family members. Further research development in this area needs to include large, high-quality studies with a clear description of the methodology, theoretical foundations, and standardized tools to permit inclusion in meta-analyses to inform the effectiveness of interventions for a better understanding of the mechanisms.

Acknowledgments
Dr Flannery receives salary support from the National Institute of Health National Cancer Institute (grants U01 CA233167 and UG1 CA189961).

Conflicts of Interest
None declared.


https://cancer.jmir.org/2024/1/e46116


76. Mohammad EB, Ahmad M. Virtual reality as a distraction technique for pain and anxiety among patients with breast cancer: a randomized control trial. Palliat Support Care 2019;17(1):29-34. [FREE Full text] [doi: 10.1017/S1478951518000639] [Medline: 30198451]


Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
HADS: Hospital Anxiety and Depression Scale
JBI: Joanna Briggs Institute
PHQ-9: Patient Health Questionnaire-9
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
STAI: State-Trait Anxiety Inventory
TREND: Transparent Reporting of Evaluations with Nonrandomized Designs
Machine Learning Approaches to Predict Symptoms in People With Cancer: Systematic Review

Nahid Zeinali1, MS; Nayung Youn2, RN, MSN; Alaa Albashayreh2, RN, MSHI, PhD; Weiguo Fan3, PhD; Stéphanie Gilbertson White2, PhD

1Department of Computer Science and Informatics, University of Iowa, Iowa City, IA, United States
2College of Nursing, University of Iowa, Iowa City, IA, United States
3Department of Business Analytics, University of Iowa, Iowa City, IA, United States

Abstract

Background: People with cancer frequently experience severe and distressing symptoms associated with cancer and its treatments. Predicting symptoms in patients with cancer continues to be a significant challenge for both clinicians and researchers. The rapid evolution of machine learning (ML) highlights the need for a current systematic review to improve cancer symptom prediction.

Objective: This systematic review aims to synthesize the literature that has used ML algorithms to predict the development of cancer symptoms and to identify the predictors of these symptoms. This is essential for integrating new developments and identifying gaps in existing literature.

Methods: We conducted this systematic review in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist. We conducted a systematic search of CINAHL, Embase, and PubMed for English records published from 1984 to August 11, 2023, using the following search terms: cancer, neoplasm, specific symptoms, neural networks, machine learning, specific algorithm names, and deep learning. All records that met the eligibility criteria were individually reviewed by 2 coauthors, and key findings were extracted and synthesized. We focused on studies using ML algorithms to predict cancer symptoms, excluding nonhuman research, technical reports, reviews, book chapters, conference proceedings, and inaccessible full texts.

Results: A total of 42 studies were included, the majority of which were published after 2017. Most studies were conducted in North America (18/42, 43%) and Asia (16/42, 38%). The sample sizes in most studies (27/42, 64%) typically ranged from 100 to 1000 participants. The most prevalent category of algorithms was supervised ML, accounting for 39 (93%) of the 42 studies. Each of the methods—deep learning, ensemble classifiers, and unsupervised ML—constituted 3 (3%) of the 42 studies. The ML algorithms with the best performance were logistic regression (9/42, 17%), random forest (7/42, 13%), artificial neural networks (5/42, 9%), and decision trees (5/42, 9%). The most commonly included primary cancer sites were the head and neck (9/42, 22%) and breast (8/42, 19%), with 17 (41%) of the 42 studies not specifying the site. The most frequently studied symptoms were xerostomia (9/42, 14%), depression (8/42, 13%), pain (8/42, 13%), and fatigue (6/42, 10%). The significant predictors were age, gender, treatment type, treatment number, cancer site, cancer stage, chemotherapy, radiotherapy, chronic diseases, comorbidities, physical factors, and psychological factors.

Conclusions: This review outlines the algorithms used for predicting symptoms in individuals with cancer. Given the diversity of symptoms people with cancer experience, analytic approaches that can handle complex and nonlinear relationships are critical. This knowledge can pave the way for crafting algorithms tailored to a specific symptom. In addition, to improve prediction precision, future research should compare cutting-edge ML strategies such as deep learning and ensemble methods with traditional statistical models.
machine learning; ML; deep learning; DL; cancer symptoms; prediction model

**Introduction**

**Background**

Cancer poses considerable physical and psychological challenges for those diagnosed with the disease. The Global Cancer Observatory estimated that there were 19.3 million new cancer cases and 43.8 million individuals living with cancer within 5 years of diagnosis globally in 2020 [1]. Symptoms such as fatigue, pain, nausea, vomiting, depression, and anxiety often persist beyond treatment [2-5], detrimentally affecting individuals’ quality of life [6]. Moreover, people with cancer frequently grapple with multiple intertwined symptoms [7], intensifying their distress [8]. Unmanaged cancer symptoms can lead to increased health care use, including emergency department visits and unscheduled hospitalizations to address these symptoms; a decline in the quality of life [9]; and even a reduced life expectancy. Providing precision symptom management tailored to the individual at the right moment has the potential to significantly improve outcomes, which is crucial for both people with cancer and their health care providers. Accurately predicting and addressing these symptoms is fundamental to providing such precision in symptom management.

Artificial intelligence, incorporating machine learning (ML) and deep learning (DL) models, excels in handling complex, high-dimensional, and noisy data. It has demonstrated effectiveness in disease diagnosis, predicting disease recurrence, enhancing quality of life, and symptom management [10-16]. There is a growing interest in ML in the emerging field of predictive analytics for cancer symptoms. ML contributes to the development of robust clinical decision systems, enhancing overall health care delivery [17]. ML algorithms can be broadly categorized into supervised learning, unsupervised learning, semisupervised learning, and reinforcement learning. DL, a subset of ML, addresses complex tasks such as speech recognition, image identification, and natural language processing [18].

**Objectives**

This study seeks to offer a comprehensive and systematic review of the literature on the application of ML algorithms in predicting symptoms for people with cancer. Conducting this review of a rapidly expanding body of literature is imperative to understand the current state of the science for ML models in symptom prediction for cancer and to guide future research. This research aims to provide a comprehensive understanding of the current state of research; identify areas for improvement; and understand the limitations and gaps in the current literature, such as a lack of specific focus on ML models for patients with cancer. By comparing model performances across diverse symptom prediction tasks, we can identify the best practices, highlight areas for improvement, and offer informed recommendations that will propel the field of predictive analytics in cancer symptom research forward.

**Methods**

**Search Strategy and Data Sources**

This study was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) protocol [19] and involved a comprehensive database search spanning from 1984 to August 11, 2023, including the PubMed, Embase, CINAHL, and Google Scholar databases. The search terms encompassed cancer, neoplasm, signs and symptoms, neural networks, machine learning, and specific algorithm names. In our study, we used Boolean expressions, using specific combinations of keywords and phrases, acknowledging the variability in terminology across studies. Search results were compiled using EndNote 20 (Clarivate Analytics). The detailed search strategy, which uses Boolean expressions, and the PRISMA checklist can be found in Multimedia Appendices 1 and 2.

**Inclusion and Exclusion Criteria**

To identify relevant research focusing on the application of ML methods in predicting cancer symptoms, we applied the following inclusion criteria: (1) papers published in English, (2) studies that used ML algorithms, and (3) research specifically aimed at predicting cancer symptoms. The exclusion criteria were as follows: (1) nonhuman studies, (2) technical reports, (3) review papers, (4) book chapters or series, (5) conference proceedings, and (6) studies for which full texts were unavailable. Two authors, NZ and NY, independently screened and cross-checked the candidate records. During the screening process, conducted using EndNote 20, any disagreements were resolved by consulting a third reviewer (SGW). The screening process involved an initial review of titles and abstracts, followed by a full-text examination to determine the study’s eligibility for inclusion in the review.

**Data Extraction and Analysis**

In our study, we implemented a systematic, multistep process for data synthesis. Initially, relevant studies were identified and selected based on the predefined inclusion and exclusion criteria. Two independent researchers, NZ and NY, extracted data from 42 selected studies. They worked independently to mitigate bias and enhance the accuracy of the data extraction process. In cases of discrepancies, these were resolved through discussion or consultation with a third reviewer, SGW. The extracted data were aggregated, involving the collation of study characteristics such as research location, sample size, study design, types of ML algorithms, validation metrics, identified significant predictors, cancer types, and the specific symptoms focused on. This comprehensive approach enabled us to reduce the bias and increase the reliability of our findings. For the analysis, we used both quantitative and qualitative methods. Quantitative data,
such as frequencies and percentages, were compiled and analyzed using Python. This included the creation of insightful plots and heat maps to identify patterns and trends, illustrating relationships among variables and highlighting key findings in an easily digestible format. Qualitative aspects, such as algorithm implementation or study design, were explored through narrative synthesis. This allowed for a deeper understanding of the context and nuances in the application of ML algorithms for cancer symptom prediction.

We conducted a cross-analysis to compare findings from different studies, assessing the effectiveness of various ML algorithms across different cancer types and symptoms and identifying common predictors of success and the challenges faced. Finally, we interpreted the findings in the context of the existing literature. We discussed how our results align with or differ from previous studies and what new insights our synthesis brings to the field of ML in cancer symptom prediction.

**Results**

**Overall Results**

A search across the 3 databases produced 1788 papers. After removing 289 duplicates, we screened the records for titles and abstracts, excluding another 1352 irrelevant records. However, 1 study was not retrieved. We reviewed the full text of the remaining 146 records, omitting 105 due to the absence of ML application in predicting cancer symptoms (69/146, 47.3%), not being a research article (34/146, 23.3%), and not being an English article (1/146, 1%). In the second phase, we intend to include Google Scholar in our research methodology to capture an additional 113 articles not found in our main databases, although 1 study was not retrieved. We reviewed the full text of the remaining 99 records, ultimately excluding all of them for reasons such as the lack of ML applications in cancer symptom prediction (89/99, 90%) and not being a research articles (10/99, 10%). Eventually, 42 studies met the inclusion criteria, as depicted in Figure 1.

Of the 42 studies, 42 (100%) is listed in PubMed, Embase covers 37 (88%) studies, and CINAHL includes 18 (43%) studies. The distribution and overlap of these research articles across the databases are illustrated in Multimedia Appendix 3.

The data extracted from these studies, which include the reference number, research location, year, data type, cancer site, symptoms, significant predictors, ML algorithms, and validation methods, are detailed in Table 1 and in Multimedia Appendix 4.
Table 1. Details of the included studies (n=42).

<table>
<thead>
<tr>
<th>Study</th>
<th>Country, year</th>
<th>Data type; number of data</th>
<th>Population</th>
<th>Cancer symptoms</th>
<th>Significant predictors</th>
<th>Algorithms</th>
<th>Validation methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun et al [20]</td>
<td>China, 2023</td>
<td>Clinical data; 1152</td>
<td>People with breast cancer</td>
<td>Pain</td>
<td>Postmenopausal status, urban medical insurance, history of at least one operation, underwent general anesthesia with fentanyl and sevoflurane, and received axillary lymph node dissection.</td>
<td>LR&lt;sup&gt;a,b&lt;/sup&gt;, RF&lt;sup&gt;c&lt;/sup&gt;, GB-DT&lt;sup&gt;d&lt;/sup&gt;, and XGB&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Random</td>
</tr>
<tr>
<td>Xinran et al [21]</td>
<td>China, 2023</td>
<td>Clinical data; 494</td>
<td>People with advanced cancer</td>
<td>Cognitive impairment</td>
<td>Cancer course, anxiety, and age</td>
<td>LR and ANN&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Random</td>
</tr>
<tr>
<td>Shaikh et al [22]</td>
<td>United States, 2023</td>
<td>Clinical data; 1152</td>
<td>Survivors of cancer with osteoarthritis</td>
<td>Depression</td>
<td>Age, education, care fragmentation, polypharmacy, and zip code–level poverty</td>
<td>XGB</td>
<td>10-fold CV&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>Kober et al [23]</td>
<td>United States, 2023</td>
<td>Clinical data; 1217</td>
<td>People with cancer receiving chemotherapy</td>
<td>Morning fatigue</td>
<td>13 individual Li-Fraumeni syndrome items</td>
<td>EN&lt;sup&gt;h&lt;/sup&gt;, RF, LASSO&lt;sup&gt;i&lt;/sup&gt;, LR (filtered/unfiltered), RPAR&lt;sup&gt;j&lt;/sup&gt;, and SVM&lt;sup&gt;k&lt;/sup&gt;</td>
<td>Random</td>
</tr>
<tr>
<td>Du et al [24]</td>
<td>China, 2023</td>
<td>Clinical data; 565</td>
<td>People with cancer</td>
<td>Fatigue</td>
<td>Pain score, Eastern Cooperative Oncology Group score, platelet distribution width, and continuous erythropoiesis receptor activator</td>
<td>LR, RF, NB&lt;sup&gt;l&lt;/sup&gt;, and XGB</td>
<td>5-fold CV</td>
</tr>
<tr>
<td>Moscato et al [25]</td>
<td>Italy, 2022</td>
<td>Clinical data; 21</td>
<td>People with cancer</td>
<td>Pain</td>
<td>N/A&lt;sup&gt;m&lt;/sup&gt;</td>
<td>SVM, RF, MP&lt;sup&gt;n&lt;/sup&gt;, LR, and AdaBoost&lt;sup&gt;o&lt;/sup&gt;</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>Masukawa et al [26]</td>
<td>Japan, 2022</td>
<td>Clinical data; 808</td>
<td>People with cancer</td>
<td>Social distress, spiritual pain, pain, dyspnea, nausea, and insomnia</td>
<td>N/A</td>
<td>LR, RF, light GBM&lt;sup&gt;p&lt;/sup&gt;, SVM, and ensemble</td>
<td>5-fold CV</td>
</tr>
<tr>
<td>Fanizzi et al [27]</td>
<td>Italy, 2022</td>
<td>CT&lt;sup&gt;q&lt;/sup&gt; image data; 61</td>
<td>People with oropharyngeal cancer receiving radiotherapy</td>
<td>Xerostomia</td>
<td>Weight preradiotherapy, induction chemotherapy, sex, platinum-based chemotherapy, current chemotherapy, alcohol history, age at diagnosis, smoking history, surgery, clinical tumor, and clinical node</td>
<td>SVM and CNN&lt;sup&gt;r&lt;/sup&gt;</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>Ueno et al [28]</td>
<td>Japan, 2022</td>
<td>Clinical data; 284</td>
<td>People with breast cancer</td>
<td>Insomnia</td>
<td>General fatigue, physical fatigue, and cognitive fatigue</td>
<td>L2 penalized LR and XGB</td>
<td>8-fold CV</td>
</tr>
<tr>
<td>On et al [29]</td>
<td>Korea, 2022</td>
<td>Clinical data; 935</td>
<td>People with cancer receiving chemotherapy</td>
<td>Nausea-vomiting, fatigue-anorexia, diarrhea, hypersensitivity, stomatitis, hand-foot syndrome, peripheral neuropathy, and constipation</td>
<td>LR, DT&lt;sup&gt;s&lt;/sup&gt;, and ANN</td>
<td>3-fold CV</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Random sampling of data.
<sup>b</sup> LR model with feature selection.
<sup>c</sup> RF model with feature selection.
<sup>d</sup> GB-DT model with feature selection.
<sup>e</sup> XGB model with feature selection.
<sup>f</sup> ANN model with feature selection.
<sup>g</sup> 10-fold cross-validation.
<sup>h</sup> EN model with feature selection.
<sup>i</sup> LASSO models with feature selection.
<sup>j</sup> RPAR model with feature selection.
<sup>k</sup> SVM model with feature selection.
<sup>l</sup> NB model with feature selection.
<sup>m</sup> N/A: Not applicable.
<sup>n</sup> MP model with feature selection.
<sup)o</sup> AdaBoost model with feature selection.
<sup>p</sup> GBM model with feature selection.
<sup>q</sup> CT image data.
<sup>r</sup> CNN model with feature selection.
<table>
<thead>
<tr>
<th>Study</th>
<th>Country, year</th>
<th>Data type; number of data</th>
<th>Population</th>
<th>Cancer symptoms</th>
<th>Significant predictors</th>
<th>Algorithms</th>
<th>Validation methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li et al [30]</td>
<td>China, 2022</td>
<td>Clinical data and CT image data; 365</td>
<td>People with cancer receiving radiotherapy</td>
<td>Xerostomia</td>
<td>Hypertension, age, total radiotherapy dose, dose at 50% of the left parotid volume, mean dose to right parotid gland, mean dose to oral cavity, and course of induction chemotherapy</td>
<td>RF, DT and XGB</td>
<td>External validation</td>
</tr>
<tr>
<td>Kuriusu et al [31]</td>
<td>Japan, 2022</td>
<td>Clinical data; 668</td>
<td>People with advanced cancer receiving pharmacological interventions</td>
<td>Delirium</td>
<td>The baseline Delirium Rating Scale-R98 severity score (cutoff of 15), hypoxia, and dehydration</td>
<td>DT</td>
<td>5-fold CV</td>
</tr>
<tr>
<td>Guo et al [32]</td>
<td>China, 2022</td>
<td>Clinical data; 80</td>
<td>People with lung cancer receiving chemotherapy</td>
<td>Lung infection</td>
<td>Age ≥60 years, length of stay ≥14 days, surgery history, combined chemotherapy, myelosuppression, diabetes, and hormone application</td>
<td>LR and ANN</td>
<td>Random</td>
</tr>
<tr>
<td>Baglione et al [33]</td>
<td>United States, 2022</td>
<td>Clinical data; 40</td>
<td>People with breast cancer</td>
<td>Depressed mood and anxiety</td>
<td>Connectedness, receive support, frequency and duration use of mobile app, and physical pain</td>
<td>RF and XGB</td>
<td>LOOCV</td>
</tr>
<tr>
<td>Chao et al [34]</td>
<td>United States, 2022</td>
<td>Clinical data and CT image data; 155</td>
<td>People with HNC receiving radiotherapy</td>
<td>Xerostomia</td>
<td>N/A</td>
<td>SVM, KNN, NB, and RF</td>
<td>Nested</td>
</tr>
<tr>
<td>Wakabayashi et al [35]</td>
<td>Japan, 2021</td>
<td>Clinical data and CT image data; 69</td>
<td>People with cancer receiving radiotherapy</td>
<td>Pain</td>
<td>Age, numeric rating scale, and biological effective dose 10</td>
<td>RF</td>
<td>LOOCV</td>
</tr>
<tr>
<td>Zhou et al [36]</td>
<td>China, 2021</td>
<td>Clinical data; 386</td>
<td>People with colorectal cancer after chemotherapy</td>
<td>Cognitive impairment</td>
<td>Age, BMI, colostomy, treatment complications, cancer-related anemia, depression, diabetes, Quality of Life Questionnaire Core 30 score, exercise, hypercholesterolemia, diet, marital status, education level, and pathological stage</td>
<td>RF, LR, and SVM</td>
<td>Random</td>
</tr>
<tr>
<td>Xuyi et al [37]</td>
<td>Canada, 2021</td>
<td>Clinical data; 46,104</td>
<td>Specific cancer site or treatment not mentioned</td>
<td>Pain, depression, and well-being</td>
<td>Lung cancer, late-stage cancer, existing chronic conditions such as osteoarthritis, mood disorder, hypertension, diabetes, and coronary disease</td>
<td>ANN</td>
<td>Random</td>
</tr>
<tr>
<td>Xu et al [38]</td>
<td>China, 2021</td>
<td>Clinical data; 598</td>
<td>People with gastrointestinal tumors after surgery</td>
<td>Postoperative fatigue</td>
<td>Age, higher degree of education, lower personal monthly income, advanced cancer, hypoproteinemia, preoperative anxiety or depression, and limited social support</td>
<td>LR, ANN, CART</td>
<td>Random</td>
</tr>
<tr>
<td>Wei et al [39]</td>
<td>China, 2021</td>
<td>Clinical data; 533</td>
<td>People with breast cancer</td>
<td>Lymphedema</td>
<td>N/A</td>
<td>ANN, LR, C5.0, RF, SVM, CART</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>Wang et al [40]</td>
<td>United States, 2021</td>
<td>Clinical data; 823</td>
<td>People with HNC</td>
<td>Pain, taste, and general activity</td>
<td>N/A</td>
<td>SVM, KNN, and RF; Gaussian NB and MLP; and ARIMA and LSTM</td>
<td>Random</td>
</tr>
<tr>
<td>Study</td>
<td>Country, year</td>
<td>Data type: number of data</td>
<td>Population</td>
<td>Cancer symptoms</td>
<td>Significant predictors</td>
<td>Algorithms</td>
<td>Validation methods</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>---------------------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Wang et al [41]</td>
<td>United States, 2021</td>
<td>Clinical data and CT image data; 138</td>
<td>People with cancer receiving chemotherapy</td>
<td>Nausea-vomiting</td>
<td>Smoking, alcohol status, sex, age, and BMI</td>
<td>NB, LR, ANN, SVR&lt;sup&gt;ab&lt;/sup&gt;, and DT</td>
<td>5-fold CV</td>
</tr>
<tr>
<td>Mosa et al [17]</td>
<td>United States, 2021</td>
<td>Clinical data; 6124</td>
<td>People with cancer receiving chemotherapy</td>
<td>Physical activity bouts, sleep, heart rate, and location</td>
<td>Diarrhea, fatigue, and pain</td>
<td>LR, KNN, SVM, RF, GB&lt;sup&gt;ab&lt;/sup&gt;, XGB, and LightGBM</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>Low et al [42]</td>
<td>United States, 2021</td>
<td>Clinical data; 44</td>
<td>People with pancreatic cancer after surgery</td>
<td>Depression</td>
<td>A set of psychological traits (optimism, perceived ability to cope with trauma, resilience as a trait, and ability to understand the illness) and subjective perceptions of personal functionality (physical, social, and cognitive)</td>
<td>RF, SVM, and GB</td>
<td>3-fold CV and LOOCV</td>
</tr>
<tr>
<td>Kourou et al [43]</td>
<td>Greece, 2021</td>
<td>Clinical data; 609</td>
<td>People with breast cancer</td>
<td>Morning fatigue, lower evening energy, and sleep disturbance</td>
<td></td>
<td></td>
<td>5-fold CV</td>
</tr>
<tr>
<td>Kober et al [44]</td>
<td>United States, 2021</td>
<td>Clinical data; 1217</td>
<td>People with cancer receiving chemotherapy</td>
<td>Education level, sex, age, marital status, medical insurance, per capita monthly household income, pathological stage, Suicide Severity Rating Scale, Pittsburgh Sleep Quality Index, and Quality of Life Questionnaire Core 30</td>
<td></td>
<td>RF, LR (filtered or unfiltered), RPAR, and SVM</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>Hu et al [45]</td>
<td>China, 2021</td>
<td>Clinical data; 238</td>
<td>People with non-Hodgkin lymphoma receiving chemotherapy</td>
<td>Fatigue or weakness, insomnia, and pain appeared</td>
<td></td>
<td>SVM, RF, and LASSO+LR</td>
<td>Random</td>
</tr>
<tr>
<td>Haun et al [46]</td>
<td>Germany, 2021</td>
<td>Clinical data; 496</td>
<td>People with cancer seen in primary care</td>
<td>Anxiety</td>
<td>Fatigue or weakness, insomnia, and pain appeared</td>
<td>OLS&lt;sup&gt;x&lt;/sup&gt;, RF&lt;sup&gt;ad&lt;/sup&gt;, LASSO, ENR&lt;sup&gt;ad&lt;/sup&gt;, RF, and XGB</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>Study</td>
<td>Country, year</td>
<td>Data type; number of data</td>
<td>Population</td>
<td>Cancer symptoms</td>
<td>Significant predictors</td>
<td>Algorithms</td>
<td>Validation methods</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
<td>---------------------------</td>
<td>------------</td>
<td>----------------</td>
<td>------------------------</td>
<td>------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Lee et al [47]</td>
<td>United States, 2020</td>
<td>Clinical data and CT Images data; 388</td>
<td>People with lung cancer after intensity-modulated radiation therapy</td>
<td>Weight loss</td>
<td>Joint Gross tumor volume L1+L2+L3 radiomics, Gross tumor volume, and esophagus L3 dosimetric</td>
<td>SVM, DNN, and ensemble classifier</td>
<td>Nested CV</td>
</tr>
<tr>
<td>Juwara et al [48]</td>
<td>Canada, 2020</td>
<td>Clinical data; 204</td>
<td>People with breast cancer after surgery</td>
<td>Np^4</td>
<td>Anxiety, type of surgery, and acute pain</td>
<td>LS^1, RR, ENR, RF, GB, and ANN</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>Men et al [49]</td>
<td>United States, 2019</td>
<td>Clinical data and CT image data; 784</td>
<td>People with HNC receiving radiotherapy</td>
<td>Xerostomia</td>
<td>Feature map visualization</td>
<td>LR and 3D-RC-NN^a^</td>
<td>Random</td>
</tr>
<tr>
<td>Jiang et al [50]</td>
<td>United States, 2019</td>
<td>Clinical data and CT image data; 427</td>
<td>People with HNC</td>
<td>Xerostomia</td>
<td>The patient has human papillomavirus, completed chemotherapy, their baseline xerostomia grade, tumor site, N stage, and use of feeding tube</td>
<td>RR, LASSO, and RF</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>Sheikh et al [51]</td>
<td>United States, 2019</td>
<td>CT images data; 266</td>
<td>People with HNC</td>
<td>Xerostomia</td>
<td>N/A</td>
<td>Generalized linear model</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>Papachristou et al [52]</td>
<td>United States, 2019</td>
<td>Clinical data; 799</td>
<td>People with cancer receiving chemotherapy</td>
<td>Sleep disturbance, anxiety, and depression</td>
<td>Age, gender, cancer site, the number of prior cancer treatment, and initial diagnosis</td>
<td>SVR (linear, polynomial, and radial Sigma) and n-CCA^a</td>
<td></td>
</tr>
<tr>
<td>Zhang et al [53]</td>
<td>China, 2018</td>
<td>Clinical data; 375</td>
<td>People with cancer receiving radiotherapy</td>
<td>Weight loss</td>
<td>Head and neck tumor location and total radiation dose of ≥70 Gray, and without postsurgery</td>
<td>DT and LR</td>
<td>Random</td>
</tr>
<tr>
<td>Ölling et al [54]</td>
<td>Denmark, 2018</td>
<td>Clinical and CT image; 131</td>
<td>People with lung cancer receiving radiotherapy</td>
<td>Odynophagia (painful swallowing)</td>
<td>N/A</td>
<td>Multivariable LR, Lasso and elastic net regularized generalized linear models, and SVM</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>Gabryś et al [55]</td>
<td>Germany, 2018</td>
<td>Clinical and CT image; 153</td>
<td>People with HNC after radiotherapy</td>
<td>Xerostomia</td>
<td>The parotid gland volume, the spread of the contralateral dose-volume histogram, and the parotid gland eccentricity, and sex</td>
<td>LRL^1, LRL^2, LR-EN^a, KNN, SVM, ET^a, and GTH^a</td>
<td>Single and nested CV</td>
</tr>
<tr>
<td>Lötsc'h et al [56]</td>
<td>Germany, 2018</td>
<td>Clinical data; 1000</td>
<td>People with breast cancer after surgery</td>
<td>Pain</td>
<td>Age, chronic pain of any type, number of previous operations, BMI, preoperative pain in the area to be operated on, smoking and psychological factors</td>
<td>Unsupervised ML^a</td>
<td></td>
</tr>
<tr>
<td>Abdollahi et al [57]</td>
<td>Iran, 2018</td>
<td>Clinical and CT image; 47</td>
<td>People with HNC receiving chemotherapy</td>
<td>Hearing loss</td>
<td>10 of the 490 radiomic features selected as the associated features with significant sensorineural hearing loss status</td>
<td>Decision stump, Hoefding, C4.5, NB, AdaBoost, bootstrap aggregating, and LR</td>
<td>10-fold CV</td>
</tr>
<tr>
<td>van Dijk et al [58]</td>
<td>United States, 2018</td>
<td>Clinical data and CT image; 68</td>
<td>People with HNC</td>
<td>Xerostomia</td>
<td>N/A</td>
<td>LR</td>
<td>External validation</td>
</tr>
<tr>
<td>Cvetković [59]</td>
<td>Serbia, 2017</td>
<td>Clinical data; 84</td>
<td>People with breast cancer</td>
<td>Depression</td>
<td>N/A</td>
<td>ELM^a, ANN, and Fuzzy Genetic Algorithm</td>
<td>Random</td>
</tr>
</tbody>
</table>
A total of 2 individual researchers (NZ and NY) separately extracted data from each study, working independently of each other. This approach is used to reduce bias and increase the accuracy of the data extraction process. If discrepancies arise between the 2 independent authors, they are usually resolved through discussion or by consulting a third reviewer (SGW).

**Primary Database Information**

The studies selected were published between 2017 and 2023 and were conducted in North America (18/42, 43%), Asia (16/42, 38%), and Europe (8/42, 19%). Methods of data collection varied, with studies originating from individual centers (23/42, 55%) and multiple centers (19/42, 45%).
average sample size was 1686, and the studies varied in sample size: <100 participants (8/42, 19%), between 100 and 1000 participants (27/42, 64%), and >1000 participants (7/42, 17%). Most studies relied on clinical data (28/42, 67%), although some integrated clinical data with computed tomography (CT) images (14/42, 33%). The study designs were diverse, including retrospective (18/42, 43%), cross-sectional (15/42, 38%), prospective (5/42, 12%), and longitudinal (4/42, 10%) approaches.

**Cancer Primary Sites and Predicted Symptoms**

Various primary cancer sites were studied, with head and neck cancers being the most prevalent (9/42, 21%). Breast cancer was the focus of 19% (8/42) of the studies, and lung cancer was studied in 17% (3/42) of the cases. The included studies included participants undergoing a range of treatments, including chemotherapy (9/42, 21%), radiotherapy (9/42, 21%), surgery (4/42, 10%), and investigations of posttreatment survivors (2/42, 5%). Of the 42 included studies, 10 unique symptoms were reported as outcome variables in the predictions. Those included were xerostomia (9/42, 14%) [27,30,34,49-51,55,58,60], depression (8/42, 13%) [22,33,37,41,43,45,52,59], pain (8/42, 13%) [20,25,26,35,37,40,42,56], fatigue (6/42, 10%) [23,24,29,38,42,44], anxiety (3/42, 5%) [33,46,52], sleep disturbance or insomnia (3/42, 5%) [26,28,52], nausea or vomiting (3/42, 5%) [17,26,29], weight loss (2/42, 3%) [47,53], cognitive impairment (2/42, 3%) [21,36], and diarrhea (2/42, 3%) [29,42].

One study reported multiple symptoms, including hypersensitivity [29], stomatitis [29], hand-foot syndrome [29], peripheral neuropathy [29], and constipation [29]. Another study delved into taste and general activity [40]. Individual studies were dedicated to each of the following symptoms: delirium [31], lung infection [32], lymphedema [39], well-being [37], odynophagia [54], social distress [26], spiritual pain [26], dyspnea [26], and hearing loss [57]. The distribution of these symptoms is depicted in Multimedia Appendix 5.

**Significant Candidate Predictors of Symptoms**

Numerous predictors were frequently used for predicting symptoms, which can be grouped into demographic features and clinical characteristics.

**Demographic Features**

The demographic features include age, sex, BMI, income, medical insurance, education, marital status, and zip code–level poverty.

**Clinical Characteristics**

The clinical characteristics include smoking and alcohol use, initial diagnosis, presence of cancer, stage of cancer, cancer course, tumor site, type and number of prior treatments, chemotherapy type, and radiotherapy dose and volume. Health conditions such as comorbidity, diabetes, hypertension, osteoarthritis, and coronary disease also play a significant role. In addition, psychological factors such as depression and anxiety, fatigue, sleep disturbance, and pain are considered. Other influential predictors encompass care fragmentation, polypharmacy, hormone levels, physical activity, diet, heart rate, and social support factors.

In our comprehensive analysis of 42 studies, all the detailed findings on common cancer symptoms are compiled in Figure 2. We provide a detailed analysis of the predictors for the 4 most frequently reported cancer symptoms—xerostomia, pain, depression, and fatigue—have been identified, each with its distinct set of influencing factors.
For xerostomia, age, gender, chemotherapy type, radiotherapy dose and volume, cancer stage, tumor site, and hypertension are crucial predictors. In the case of pain, factors such as age, BMI, smoking and alcohol habits, cancer site and stage, tumor site, diabetes, hypertension, osteoarthritis, coronary disease, physical activity, psychological factors, sleep disorders, and existing pain conditions emerge as significant. Significant predictors for depression include age; gender; education; cancer site and stage; economic factors such as insurance, income, and poverty level; marital status; initial diagnosis impact; comorbidities (diabetes, hypertension, osteoarthritis, and coronary disease); pain; social support; care fragmentation; polypharmacy; and various scale scores. Finally, for fatigue, the key predictors are existing fatigue and low energy, cancer site, sleep disturbances, age, income, education, chemotherapy type, tumor site, comorbidities, hypercholesterolemia, heart rate, hypoproteinemia, physical and psychological factors, pain, adverse drug reaction history, limited social support, Eastern Cooperative Oncology Group score, platelet distribution width, and erythropoiesis.

When examining the commonalities across these predictors for xerostomia, pain, depression, and fatigue, several factors stand out as particularly influential across multiple symptoms: age; gender; cancer site and stage; treatment-related factors such as the type of chemotherapy and radiotherapy; comorbidities such as diabetes, hypertension, and coronary disease; physical and psychological factors; and socioeconomic factors such as income and education level, demonstrating the impact of cancer treatments on symptom development. These common predictors underscore the complex, multifactorial nature of symptom manifestation in patients with cancer, necessitating a comprehensive approach to their management and care.

ML Algorithms and Validation Metrics

Of the 42 studies analyzed, 7 (17%) used a single ML algorithm, whereas 35 (83%) used multiple algorithms. The most effective models, in terms of performance, were logistic regression (LR; 9/42, 17%), random forest (7/42, 13%), artificial neural networks (5/42, 9%), decision trees (DTs; 5/42, 9%), and extreme gradient boosting (3/42, 6%). For validation methods, 10-fold cross-validation was the most used (14/42, 31%), followed by 5-fold cross-validation (5/42, 11%), 3-fold cross-validation (2/42, 4%), and 8-fold cross-validation (1/42, 2%). The primary evaluation metric across these studies was the area under the curve, which was adopted in 24% (26/42) of the studies. A visual representation of the leading ML models along with the validation and evaluation metrics used in the study presents in Multimedia Appendix 6.

Discussion

Principal Findings

In this review, we present the first systematic analysis of ML applications for predicting the development of cancer symptoms. We explore the most frequently studied cancer sites and delve into the intricacies of ML procedures. Breast, head or neck, and lung cancers are the most frequently studied sites in current research, with xerostomia, depression, pain, and fatigue being the most prominent symptoms. The application of various ML techniques is on the rise, with data acquisition and preprocessing being pivotal for successful ML models. While a range of
algorithms, from traditional methods such as LR and DT to advanced ones such as DL, are used, there is a growing emphasis on data quality, external validation, and a standardized approach to model evaluation. The future of ML in cancer symptom prediction looks promising, with a need for collaborative efforts among oncologists, data scientists, and patient groups, combined with more comprehensive research on lesser-studied cancer sites and standardized methodologies.

Regarding the cancer sites covered in the studies, breast, head or neck, and lung cancers emerged as the most frequently researched primary cancer sites. The range of symptoms and side effects that patients experienced varied from one study to another. Some symptoms depended on the specific cancer site and the treatments patients received. For example, xerostomia, which can either arise from the tumor itself or manifest as a treatment side effect, has a significant impact on patients’ dental health and compromises antimicrobial functions [61]. However, most symptoms were not directly attributed to a particular cancer site or treatment.

Our review revealed a notable emphasis on predicting xerostomia in 14% (9/42) of the studies, despite head and neck cancers being less prevalent. The notable emphasis on predicting xerostomia in ML research, despite the lower prevalence of head and neck cancers, is likely due to advancements in integrating ML with CT imaging. CT imaging is a pivotal tool in the diagnosis and treatment planning of head and neck cancers. The integration of ML with CT imaging has opened new possibilities for more accurately predicting side effects such as xerostomia. ML techniques, when applied to CT images, can potentially identify patterns and indicators that are not easily discernible by human observers. This capability can lead to earlier and more precise predictions of xerostomia, thereby enabling better preventive measures and treatment planning to mitigate this side effect. Therefore, the focus on xerostomia in ML research, in the context of head and neck cancers, is likely driven by the opportunities presented by combining ML with advanced imaging techniques.

Depression, a widespread emotional challenge for people with cancer [62,63], was the focus of prediction in many studies (8/24, 13%). Similarly, pain, a recurrent concern for palliative care patients [64] and survivors of cancer [65,66], was the subject of prediction in >13% (8/24) of the studies. Fatigue, prevalent across all age groups with cancer [67,68], was highlighted in 6 (10%) of the 42 studies reviewed.

In terms of the ML approaches used in the studies, a plethora of techniques were used to construct these predictive models, spanning all phases of the ML process, from data collection and preprocessing to feature and algorithm selection, model training, testing, and evaluation. The process of data acquisition is pivotal for the development of ML models, thereby emphasizing the importance of an adequate sample size. Upon reviewing 42 studies, we discerned that the most frequent sample sizes for ML applications ranged between 100 and 1000 samples. More advanced ML techniques necessitate larger data sets to bolster robustness and mitigate the risk of overfitting. Alarmingly, certain studies in our review used ML with comparably smaller data sets, introducing the risk of model overfitting and potential biases in the subsequent performance metrics [69]. Challenges tied to sample size might impede the creation of sturdy and trustworthy ML models [70]. Data preprocessing is indispensable to yield clean and interpretable data, which is a cornerstone for proficient ML models. Data cleaning approaches encompass addressing missing values, tackling data noise, and data normalization. Within health care data sets, noisy or absent data are frequently a by-product of inaccuracies in manual entries or instrument recordings made by medical personnel or ancillary staff [71]. However, most of the reviewed studies lacked comprehensive descriptions of their data cleaning methodologies or strategies for handling noisy data and normalization, constrained by word or page limits in publications.

Given the crucial importance of data quality in developing ML models, it is essential for researchers to focus equally on effective data preparation and choosing suitable algorithms. Future endeavors would benefit from exhaustive procedural documentation made available on public platforms such as GitHub. In a research context, GitHub can be used for sharing and collaborating on various aspects of a research project, including but not limited to code. It allows researchers to maintain version control of their scripts, data analysis procedures, and even documentation. This feature is particularly beneficial for replicating studies and verifying results, as it provides a transparent view of the methodologies and analyses used.

Overloading an ML model with excessive features can undermine its ability to differentiate between pertinent data and superfluous noise, leading to the challenge often referred to as the “curse of dimensionality.” The goal of feature engineering is to mitigate model complexity, expedite the training process, reduce the data’s dimensionality, and avert overfitting [72]. By streamlining the model with a curated set of predictors, it becomes more accessible and transparent, emphasizing the importance of feature selection during data preparation. Our review pinpointed the most frequently used significant predictors in cancer symptom prediction. The efficacy of prediction models is heavily influenced by the number and interplay of the relevant predictors. Factors such as age, gender, type and number of previous treatments, cancer location, cancer stage, chemotherapy type, dosage and volume of radiotherapy; chronic conditions such as diabetes and hypertension; concurrent diseases; and symptoms including depression, anxiety, fatigue, pain, and sleep disturbances have consistently featured as determinants in numerous predictive frameworks. Our review of cancer symptom prediction underscored age as a pivotal factor, associated with predominant symptoms such as depression, pain, xerostomia, and fatigue. While numerous elements, from gender to type of treatment and cancer stage, influence the predictive models, it is the prominence of age that consistently emerges as a cornerstone predictor. As we delve deeper into this field, even with the introduction of newer determinants and correlations, the centrality of age in these frameworks remains indisputable.

Regarding algorithm selection, traditional methods often struggle with handling high-dimensional data and processing extensive information. To tackle these challenges, researchers have...
increasingly shifted toward innovative ML algorithms that are renowned for their robust predictive power and strong generalization capacities. These sophisticated algorithms excel at delving deep into data and discerning intricate interrelationships among variables. To navigate the multifaceted landscape of modeling challenges, it is advantageous for researchers to leverage a diverse array of ML algorithms. Most studies used multiple predictive models, with techniques such as LR, RF, ANN, and DT consistently delivering stellar results. The introduction of advanced ML techniques, such as DL and ensemble classifiers, provides promising opportunities to elevate prediction accuracy in future research.

After their design, the ML models undergo training and testing on different data sets. However, these models can grapple with issues such as overfitting and underfitting. Overfitting occurs when a model becomes overly complex, which leads to increased variance and reduced clarity. In contrast, underfitting results from an oversimplified model, causing it to overlook key data patterns and diminish its predictive capacity. Therefore, the ideal learning model should strike a balance between the optimal variance and justifiable bias. To mitigate these issues, the common strategy is to divide the data set into training and testing subsets, followed by internal or external validation. While most studies in our review used internal validation, only 1 study reported external validation [58], which was demonstrated on a small cohort of 25 patients with head and neck cancer. Although its performance is typically lower than evaluations using the original data sets, external validation remains crucial for gauging ML models [72]. It is a crucial step in ensuring that the model’s performance is not just limited to the conditions and data it was originally trained on but also applicable and reliable in broader, real-world clinical settings. This approach serves to verify the model’s efficacy and generalizability across different patient populations and settings.

Understanding and interpreting ML models continue to pose challenges. Determining the variables that significantly impact symptom prediction can be elusive due to the intricate prediction processes. Many studies gauge the performance of ML models using metrics that examine their ability to distinguish between 2 classes. From our systematic review of 42 studies, the area under the curve emerged as the predominant metric for the prediction models. Other metrics included accuracy, sensitivity, specificity, positive predictive value, root mean square error, and negative predictive value. These metrics provide a holistic view of a model’s efficacy, facilitating its refinement and enabling more precise predictions. However, the diverse emphasis on distinct metrics in numerous studies underscores the need for a uniform approach to evaluating ML models in cancer symptom prediction.

As interest grows in using ML for predicting cancer symptoms, there are several areas that merit deeper investigation. A crucial area is broadening the range of studied cancer sites and more comprehensively correlating symptoms with various treatment methods. To fully understand symptom prediction, it is essential that future studies delve into lesser-explored or infrequently studied cancer sites. Furthermore, the methodologies used for data preprocessing and cleaning should be documented more thoroughly, focusing on best practices to ensure data integrity. As data are foundational to ML models, transparent and detailed preprocessing can improve the reliability and repeatability of these models. Although our analysis highlighted common predictors for symptom forecasting, examining potentially underrepresented or emerging indicators could refine these models further. On the algorithmic front, exploring hybrid ML methods that merge the strengths of multiple algorithms might be particularly beneficial for cancer symptom prediction. Standardizing evaluation metrics across studies would also provide clarity and facilitate a more accurate comparison of various ML techniques. To genuinely progress, collaborations among oncologists, data scientists, and patient advocacy groups are vital to ensure that the developed models are technically robust and clinically pertinent. With these insights, ML stands poised to transform cancer care, creating treatment plans based on patient-focused and accurate symptom prediction models.

**Limitations**

This review is not without its limitations. Although we established clear inclusion and exclusion criteria, potential biases in the studies we analyzed could inherently limit our review. We might have missed or excluded relevant studies due to inadequate information or the absence of keywords in their titles or abstracts. Many of the studies we reviewed did not specify the cancer site, potentially limiting the accuracy and applicability of our findings to specific cancer types. The broad range of predictors used across the studies also made it difficult to draw definitive conclusions about the most influential factors in predicting cancer symptoms using ML algorithms. As such, readers should interpret these results cautiously, given this variability.

**Conclusions**

ML offers an intriguing potential for predicting cancer symptoms, thereby preemptively mitigating the associated challenges. Predicting the symptoms that people with cancer might experience and determining their onset throughout their treatment journey is a pivotal clinical issue that can enhance patients’ quality of life. Notably, all studies in our review were published after 2017, highlighting the nascent nature of this research area. Our investigation primarily sought to outline the ML methodologies harnessed for symptom prediction in people with cancer. While ML techniques hold an edge over traditional statistical approaches by virtue of their prowess in analyzing vast data sets and gauging the efficacy of diverse prediction models, certain impediments such as a limited pool of symptoms; suboptimal data preparation; challenges in feature engineering; and complexities in ML algorithm design, validation, and evaluation can constrain the broad applicability of these predictive models. Future research should pivot toward amplifying the efficacy of ML strategies. This enhancement can be achieved by harnessing expansive, high-caliber data sets; tapping into innovative technologies for data refinement; and sculpting refined models. Harnessing ML can potentially free health care practitioners—including doctors, nurses, and clinic personnel—to accentuate the human touch in managing cancer symptoms.
Acknowledgments

The authors would like to express their profound gratitude to their esteemed colleagues and academic mentors. Their invaluable insights, unwavering support, and dedicated time significantly enriched the authors’ research journey. Special acknowledgment is reserved for the myriad of researchers and study participants whose dedication and data have underpinned this comprehensive review. Their collective efforts have made this work possible. The authors extend special thanks to Jennifer Deberg, a specialist at the Hardin Library for the Health Sciences, for her invaluable support in selecting search terms and databases.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The detailed search strategy for the databases and the Boolean expressions used.

[DOCX File, 30 KB - cancer_v10i1e52322_app1.docx ]

Multimedia Appendix 2

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.

[PDF File (Adobe PDF File), 942 KB - cancer_v10i1e52322_app2.pdf ]

Multimedia Appendix 3

The distribution and overlap of 42 studies across the databases.

[PNG File, 88 KB - cancer_v10i1e52322_app3.png ]

Multimedia Appendix 4

The data extracted from 42 studies.

[DOCX File, 31 KB - cancer_v10i1e52322_app4.docx ]

Multimedia Appendix 5

Number of studies per cancer symptoms.

[PNG File, 62 KB - cancer_v10i1e52322_app5.png ]

Multimedia Appendix 6

Visual overview of the machine learning models and metrics.

[PNG File, 74 KB - cancer_v10i1e52322_app6.png ]

References


36. Zhou SP, Fei SD, Han HH, Li JJ, Yang S, Zhao CY. A prediction model for cognitive impairment risk in colorectal cancer patients receiving chemotherapy: less may be more. Fatigue 2021;9(1):14-32 [FREE Full text] [Medline: 34076372]


Abbreviations

CT: computed tomography
DL: deep learning
DT: decision tree
LR: logistic regression
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

©Nahid Zeinali, Nayung Youn, Alaa Albashayreh, Weiguo Fan, Stéphanie Gilbertson White. Originally published in JMIR Cancer (https://cancer.jmir.org), 19.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cancer, is properly cited. The complete bibliographic information, a link to the original publication on https://cancer.jmir.org/, as well as this copyright and license information must be included.
Need for Culturally Competent and Responsive Cancer Education for African Immigrant Families and Youth Living in the United States

Olufunmilola Abraham, BPharm, MS, PhD; Adeola Agoke, MA, PhD; Kazeem Sanuth, MA, PhD; Abimbola Fapohunda, MPH, MS, DrPH; Motolani Ogunsanya, BPharm, PhD; Megan Piper, BA, MClinPsych, PhD; Amy Trentham-Dietz, MS, PhD

Social and Administrative Sciences Division, School of Pharmacy, University of Wisconsin-Madison, Madison, WI, United States
African Cultural Studies, University of Wisconsin-Madison, Madison, WI, United States
National African Language Resource Center, Indiana University Bloomington, Bloomington, IN, United States
Behavioral and Community Health Sciences, School of Public Health, University of Pittsburgh, Pittsburgh, PA, United States
College of Pharmacy, Health Sciences Center, University of Oklahoma, Oklahoma City, OK, United States
Department of Medicine and Center for Tobacco Research and Intervention, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States
Population Health Sciences and Carbone Cancer Center, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
Olufunmilola Abraham, BPharm, MS, PhD
Social and Administrative Sciences Division
School of Pharmacy
University of Wisconsin-Madison
Room 2515 Rennebohm Hall
777 Highland Avenue
Madison, WI, 53705
United States
Phone: 1 6082634498
Fax: 1 6082625262
Email: olufunmilola.abraham@wisc.edu

Abstract

Cancer prevalence data for Black Americans is monolithic and fails to consider the diverse cultures and backgrounds within that community. For instance, African immigrants constitute a meaningful proportion of the foreign-born Black immigrants in the United States (42%), but the prevalence of cancer in the African immigrant community itself is unknown. Therefore, without accurate cancer prevalence data, it is impossible to identify trends and other key factors that are needed to support the health of African immigrants and their children. Moreover, it is impossible to understand how the culture and language of subgroups influence their cancer-related health behavior. While research in this area is limited, the existing literature articulates the need for culturally responsive and culturally tailored cancer education for African immigrants and their adolescent children, which is what we advocate for in this viewpoint paper. Existing projects demonstrate the feasibility of culturally responsive programming for adults; however, few projects include or focus on adolescents or children born to African immigrants. To best meet the needs of this understudied community, researchers must use culturally competent interventions alongside familiar, usable media. For adolescents, technology is ubiquitous thus, the creation of a culturally tailored digital intervention has immense potential to improve cancer awareness and prevention for youth and their community. More research is needed to address many of the existing research gaps and develop a rich understanding of the unique experience of cancer among African immigrant families that can be used to inform intervention development. Through this viewpoint, we review the current state of cancer-related research among African immigrant families in the United States. In this paper, we acknowledge the current knowledge gaps and issues surrounding measurement and then discuss the factors relevant to designing an educational intervention targeted at African immigrants and the role of African immigrant youth.

(JMIR Cancer 2024;10:e53956) doi:10.2196/53956

https://cancer.jmir.org/2024/1/e53956

 JMIR Cancer 2024 | vol. 10 | e53956 | p.40
(page number not for citation purposes)
KEYWORDS
African immigrant; youth; adolescent; adolescents; teen; teens; teenager; teenagers; cancer awareness; health disparities; culturally competent education; African; Black; immigrant; cultural; culturally; cancer; oncology; patient education; awareness; prevention; disparity; disparities

Introduction
Cancer has a profound impact on the experience of health for many in the United States that only continues to grow. Research has demonstrated the escalating rates of early-onset cancer diagnosis among women and the alarming decreasing rates for men and Black people; most commonly in breast, thyroid, and colorectal cancer [1]. Early diagnosis and prompt treatment of cancer are critical to improved public health. The decreasing rate of early cancer detection and response is introducing a significant health inequity among Black people in the United States. African-born immigrants and their children comprise a meaningful portion of the US population. The paper aims to describe the existing research gaps and experiences of cancer among African immigrant families and highlight the need to design and tailor cancer education for African immigrant families.

There was a surge in the African immigrant population between 1970 and 2015 [2]. This migration pattern has continued, with the African immigrant population growing from 881,000 in 2000 to 2.0 million in 2019, comprising 42% of the US foreign-born Black population. African immigrants have tended to settle in 4 main cities in the United States: Washington DC, New York City, Minneapolis or St Paul, and Atlanta [3]. Prior research has established that most African immigrants come from Western (35.7%) and Eastern Africa (29.6%) [4]: from Nigeria (13.7%), Ethiopia (10.8%), Ghana (7.8%), and Kenya (5.5%) [5]. Therefore, Nigeria and Ethiopia constitute the top birthplaces of African immigrants in the United States.

In 2011, one of the first peer-reviewed papers on African immigrant health was published. It highlighted the growing population of African immigrants in the United States and the need to study their health care needs and practices since chronic diseases, including cancer, were poorly understood [3]. There is a growing research interest in African immigrant health, especially by researchers who are first- and second-generation African immigrants themselves, especially in light of the myriad of factors that impact African immigrants’ health, and that of their families, including the trauma of immigration, conflicting cultural contexts between African family dynamics and those common in the United States, diet and lifestyles, culture, religion, and spirituality. These constructs make up the richness of immigrants’ lives and continue to impact their health behaviors, health care experiences, and decision-making about their health practices after they move to the United States. Therefore, attention must be given to these factors. These factors also influence how African immigrants interact with and navigate the health care system, making it essential to understand how these factors can negatively impact health care system involvement.

The number of health-related areas influenced by immigration only grows as their length of stay in the United States increases [6]. Given the unique health experiences of African immigrants, the need to better understand the cancer-related health of African immigrants is imperative. The United States has begun to address disparities in immigrant health, such as affordances of health care following the implementation of the Affordable Care Act or state-level legislation allowing immigrants (especially the young, pregnant persons, and older people) to be eligible for state Medicare. States that have expanded care for immigrant children have seen reductions in no-insurance rates and rates of families forgoing medical care [7]. However, there is a current dearth of initiatives directly designed for African immigrants. With a deeper understanding of cancer in the African immigrant community, we can create novel, innovative, and culturally specific educational tools to support African immigrant families and improve current and future African immigrant community cancer health.

Discerning Cancer Prevalence Among African Immigrants

Overview
Uncovering cancer prevalence and awareness among African immigrants is challenging. Racial and ethnic minority groups are underrepresented in health research overall, contributing to persistent health disparities in the United States [8]. Cancer death rates among Black people continue to increase and so does the risk of developing cancers [9]. In the past few decades, there have been focused efforts to conduct research and draft policies to address health disparities within immigrant communities. However, there is a discernable lack of research on African immigrants’ (and their children’s) health related to cancer prevention and awareness in the United States. The challenge is due to limited resources allocated to minority issues and a lack of effort to distinguish the population as different and unique from other minority populations [10,11].

Issues With Measurement
Most research on immigrant health in the United States has focused on Latinos and Asian Americans [12-15]. Similarly, most cancer-related research in the United States has used race, and Hispanic or Latino ethnic affinity, regardless of the differing histories of migration, as the basis of categorizing research participants. Therefore, there is limited knowledge about African immigrant health in the United States, especially on cancer awareness, cancer care, and overall health outcomes. Accurate prevalence of cancer among African immigrants is unknown, available literature mostly focuses on databases that have combined data for African-born immigrants and US-born Black people. This makes it difficult to identify African immigrants and to provide accurate evidence of the extent and impact of cancer within their communities [16]. This practice facilitates a monolithic view of people with African heritage; therefore, it discourages granularity of analysis and limits health services’
researchers’ ability to address African immigrant-specific health challenges and examine related research questions.

Some researchers have started to address the overgeneralization of categorizing all African immigrants as “Black” by focusing on their country of origin or time since immigration or assimilation or acculturation [17,18]. Assimilation allows immigrants to integrate into the social, linguistic, and cultural fabric of the host society. However, acculturation experience differs across immigrant groups. The Hispanics, specifically Mexican Americans, constitute an immigrant group in the United States with a robust acculturation. Safran Williams in his classification of diasporas describes Mexican American as “not true diaspora” [19]. This is because of their immigration history with the United States [19]. Further, Spanish is the dominant language of the Mexicans and is also the most popular foreign language in the United States. As a result, acculturation for Mexican immigrants is steady and impacts the strength of research and health interventions for this immigrant group [20-22]. Contrarily, the cultural and linguistic significance of African immigrant identities do not share the same history and recognition both in the United States social milieu and in the US health care system especially those relating to cancer education and research. African immigrants have an existing cultural identification from their homeland and their languages do not have the same recognition as that of Hispanic Americans. Nonetheless, the effort to acculturate among African immigrants accounts for the experiences such as changes in diet, modified language practices, and using the health care services for access to information, treatment, and care. The acculturation process is also layered with the African immigrants’ spirituality and how it influences their reception of health care treatment. Careful attention to the cultural practices of African immigrants and their relevance to health intervention will largely impact the outcomes in cancer awareness and education.

The issue of having a monolithic “Black” category affects the extraction of research data on African immigrants [16]. Some progress is being made in this area. For instance, I study promoted awareness and accessibility to screening for chronic diseases among African immigrants living in Georgia [17]. Other research has discussed African immigrants’ health and allostatic load score as it relates to cardiovascular, metabolic, and immune systems [18]. Finally, a scoping review identified additional socio-ecological challenges faced including the lack of culturally competent health care, distrust of the health care system, challenges navigating the US health system, and the burdensome cost of care [16].

What We Do Know About African Immigrants and Cancer

Accurate prevalence rates of cancer in African immigrants in the United States are lacking. Evidence suggests high cancer prevalence in their countries of origin, especially breast and cervical cancer for women and prostate cancer for men [23-25]. More research is needed to understand the prevalence of cancer within immigrant families and how their immigration may influence cancer prevalence.

The experiences and needs of African immigrants are unique [17,26-32]. Sociocultural factors underlie the experience of cancer in the African immigrant community. The stigma of being diagnosed with cancer, lack of cancer awareness, limited or no screening (especially among African immigrant women), and limited familiarity with prevention strategies and treatment technologies available may be contributing to the high prevalence of cancer [24,30]. These factors lead to late-stage diagnoses because of a lack of access to health care, lower education levels, and cultural and religious beliefs regarding cancer [33,34]. Studies also found that African-born women have limited knowledge and exposure to breast cancer screening information before their arrival in the United States [30,34,35], which can impact their preventative and cancer screening behaviors. Existing research has also explored cancer mortality among adults across different Black ethnic groups—African, African American, and Caribbean—showing some mortality and prevalence differences between these groups [36].

Further, 1 study has found that income, among other factors, plays a significant role in the population’s understanding of colorectal cancer [37]. With a focus on breast and cervical cancer screening, other studies examined the knowledge and perspectives of African immigrants [38,39]. Their findings underscore significant factors impacting the decision to seek preventative screening measures among African immigrants, including fatalism, lack of cancer knowledge, stigma, length of stay in the United States, provider gender, and privacy concerns [40-43]. Another study examined prostate cancer risk experiences among West African men and shed light on the modifiable risk factors implicated in prostate cancer mortality and morbidity [44]. A study of cervical cancer awareness among African immigrant women in Iowa City highlighted factors such as fear, languages spoken, and education as barriers to preventative treatment [45]. Considering the available research and prevalent factors that limit cancer prevention knowledge and behavior it is imperative to develop culturally, and linguistically appropriate cancer education programs aimed at increasing awareness and screening of cancer. In summary, while research has begun to address differences in African immigrant health, the differences are many which will require further study and consensus.

Lack of Cancer Awareness Among Youth and African Immigrants

In 2008, it was estimated that the 82% of the US population increase between 2005 and 2050 would be attributed to immigrants and their descendants [26]. Despite an increase of African immigrants’ offspring in the United States, little is known about these second-generation individuals born and raised in the United States (with at least 1 foreign-born parent), regarding their health beliefs, perceptions, and practices. This is understandable as little is known about their parents regarding these factors. A study that explored beliefs and lifestyle behaviors relating to healthy living and diet among middle-aged adults in the immigrant population indicated that little is known about the beliefs, perceptions, and practices of diet and exercise among young African immigrants [46]. Young adults of African
immigrant descent are part of the future, and attention needs to be paid to their well-being.

It is unknown if children of African immigrants are being educated about cancer by their parents, their communities, their health care providers, or in schools. Cancer is often termed as a taboo subject in most African homes and communities. This is further compounded by other barriers such as access to care, quality of care, communication gaps, lack of education, lack of affordable health care, lack of transportation, socioeconomic status, shame and stigma, and cultural and religious beliefs [47]. Nonetheless, some children of African immigrants become aware of cancer when close family members or friends are diagnosed. With limited cancer awareness and the vulnerability of African immigrants regarding cancer, youth, and their parents must be educated using culturally competent, tailored, and responsive family-oriented cancer education initiatives that build on the strengths of these immigrant cultures as well as address the barriers to cancer prevention behaviors.

Although the limited research reviewed above examines cancer among Black immigrant men and women, there is no substantial body of research that addresses cancer education and awareness among first and second-generation African immigrant adolescents in the United States. A lack of knowledge about youth immigrants and second-generation African immigrants can put this population at a disadvantage as compared to their peers. Cancer awareness among African immigrants and youth studies, including older and younger Somali women, use age as a factor for examining standardized prevalence of cardiovascular disease risk factors among both African immigrants and African Americans [47,48]. Although age is an important factor to consider, this work does not focus on youth. Another study, rather than age, used the year of residence in the United States to examine self-reported health problems among African immigrant adults [49].

While several studies have begun to address cancer research among the African population broadly, the significant paucity of research that focuses on the youth of African immigrant families in the United States leaves a critical gap in cancer awareness and prevalence research. To our knowledge, no studies have sought to examine or address cancer awareness among the youth of African immigrant families, nor interventions for cancer awareness and education. The youth of African immigrant families in the United States constitute an important population that is instrumental in creating awareness about the prevalence of cancer within their community. To access the youth groups of African immigrant descent in the United States, it is expedient to identify cultural and age-relevant educational tools for creating awareness about the prevalence of cancer disease.

**Existing Studies on the Promotion of Cancer Awareness and Education Among African Immigrants**

**Overview**

There is evidence of studies that promote cancer health education among African immigrants and other minority groups using various culturally tailored approaches and technologies. The success of a community-academic partnership model at community faith-based centers is effective for immigrant women in learning about breast cancer [50]. Moreover, health education programs in community-based settings have indicated strong potential. Further, 2 studies involving interpreters and culturally targeted communication, showed increased breast cancer knowledge and an improvement in screening for breast cancer for immigrant and multicultural women [51,52].

Study findings have demonstrated the importance of culturally tailored educational tools and different approaches to reduce cancer-related disparities. These studies provide strong evidence supporting the use of culturally relevant educational materials, patient navigation programs, peer-to-peer education, education programs, videos, and cofacilitated health promotion forums in promoting preventative and cancer screening behaviors [33,53-62]. Together these projects shed light on some of the few, yet variable opportunities for successful community-engaged research with African immigrant families.

Furthermore, some studies have demonstrated the potential of technology in promoting cancer awareness and education among African immigrants. Mobile devices, tablets, and computers have been used to address common cultural and linguistic barriers to cancer screening. Mobile health initiatives, culturally tailored messaging, language support, mobile apps, short message services, and text messages have all proven effective in impacting cancer screening behaviors [18,63-65]. Some of these initiatives could be adapted into family-based programs where young African immigrants could learn in familiar spaces using ubiquitous and widely acceptable technologies such as serious games.

**Global Health Perspectives and Solutions for Culturally Competent Care Among African Immigrants**

Health care approaches for immigrant populations require adaptation and cultural competence to serve diverse communities effectively. Parallel analysis of the US health care models with those of other nations like Canada and Australia offers a framework to evaluate and refine strategies to address health disparities among African immigrants. Canada and Australia have made strides in fostering inclusive health strategies that can inform US health care practices, particularly in providing culturally competent care to African immigrants.

For example, in Canada, health care delivery to immigrant populations acknowledges the necessity of cultural competence. Canada’s universal health care system actively integrates culturally tailored interventions. The Canadian government has pushed for strategies that involve community engagement and representation in health decision-making, enhancing the cultural appropriateness of health care services [66]. Using community health workers who share the same cultural background as immigrants has been a breakthrough, acting as a bridge between health care providers and immigrant communities [67]. These community health workers facilitate communication, understanding, and trust—essential elements in promoting the health and well-being of immigrant populations [68].
Further, Australia’s approach to immigrant health pivots on inclusivity and health equity to deliver services that are respectful of and responsive to diverse patients’ health beliefs, practices, and needs [69]. A notable instance is the Victorian Immigrant and Refugee Women’s Coalition’s efforts, which engage women directly to educate about health issues, including cancer awareness [70]. Australian health policies aim to address the language barriers and the diverse cultural contexts that can influence health care usage and outcomes. In contrast, the United States continues to grapple with creating a standardized approach for culturally competent care throughout its health care system.

While there are pockets of exemplary practices, such as using patient navigators in cancer care to assist patients from minority backgrounds, there is not a universal health care mandate specifically aimed at immigrant health [67]. Instead, the United States relies on a patchwork of local initiatives and federal guidelines, such as those by the Office of Minority Health which established the National Standards for Culturally and Linguistically Appropriate Services in health and health care [71]. In conclusion, both the Canadian and Australian models underscore the importance of cultural competence and systemic support in improving immigrant health outcomes. They demonstrate that effective immigrant health strategies require the integration of culturally informed practices across all stages of health care—from preventive education to treatment. This implies adopting multifaceted approaches that can cater to the unique cultural, linguistic, and religious elements that define African immigrant communities.

**Youth: the Bridge for Culturally Tailored Cancer Education**

Given their positionality, first through second-generation African immigrant youth are at a unique nexus from which they can bridge health gaps related to cancer that arise from their heritage and sociocultural contexts. Cultural tailoring acknowledges the broad culture but identifies specific strategies for reaching specific individuals. These groups of individuals have insights into the linguistic and cultural practices of their families as well as those of the society they live in. Due to their positionality, the youth are motivated to embrace language awareness, which emphasizes the interrelatedness of language, culture, and social structures [72]. The interrelatedness of cultural meanings and linguistic signs allows for the tailoring of educational content that addresses distinctive groups. The adolescents of African immigrant families are a product of the diverse linguistic and cultural interactions that occur through transnational migration and globalization.

To engage with youth and form a robust bridge between coexisting sociocultural systems to improve African immigrant community health, research should focus on methods that are familiar and usable for adolescents. A ubiquitous facet of adolescent life is technology. There is increasing interest in serious games (ie, games that serve an educational or developmental purpose aside from pure entertainment) as a learning medium. Although innovative interventions including serious games are becoming popular, they are not traditionally designed and tailored to meet the cultural and health needs of minoritized populations such as African immigrant families. Systemic reviews of serious games indicate limitations that need to be addressed [73-76]. It will be beneficial for health services’ researchers to use a participatory design approach when designing cancer education and intervention tools for African immigrant families. Such a collaborative approach will allow African immigrant families to partner in the co-design of technologies such as serious games and facilitate the creation of a culturally competent and responsive learning medium. Youth from African immigrant families typically have a hybrid of identities which necessitate the use of education technologies such as serious games in ways that speak to their lived experiences and families’ cultural heritage and realities. Therefore, there is a need to tailor educational resources using technology platforms that would engage the linguistic and sociocultural realities of the African immigrant population. Interventions to improve cancer outcomes in African immigrants, especially among youth, are necessary.

Youth and community members from other minority populations in the Northwest Arctic region of Alaska participated in community-based participatory action research honoring indigenous ways, creating a Sharing Circle used to understand community priorities and develop culturally relevant cancer education that could be incorporated into school curriculum. It is an opportunity for youth involvement in culturally relevant health promotion efforts to address health disparities in cancer [77].

**Culturally Tailored Education for African Immigrant Youth**

**Overview**

Developing educational resources for African immigrant youth brings into focus the question of curricular content and pedagogical approaches that fit this group. The connection of educational content with cultural identities is espoused in the framework of culturally relevant pedagogy (CRP) [78]. CRP encompasses multiple concepts related to students’ academic achievements and social inequalities, but its central tenet is the interconnection of theories and cultures in manners that will “empower students intellectually, socially, emotionally, and politically by using cultural referents to impart knowledge, skills, and attitudes [79].” African immigrants and people of historically marginalized cultures are unique and deserving of an educational approach that is aligned with their needs. It offers liberatory education which inspires the learners to become social commentators, advocates, and critical consumers of knowledge while empowering control over one’s health. The use of such an approach will be beneficial in disseminating and promoting cancer education in the community.

The pedagogical approaches to achieving culturally tailored education may derive from CRP and adopt effective strategies that will merge critical consciousness and cultural connections in the learning content. CRP proposes three components that must be integrated to achieve learning: (1) a focus on youth learning and academic success, (2) developing youth’s cultural
Competence to assist them in developing positive ethnic and social identities, and (3) supporting youth’s critical consciousness or their ability to recognize and critique societal inequalities.

Researchers have described examples of targeted and tailored strategies, techniques, and procedures for successful intervention with a variety of populations [80]. These researchers identified linguistic, community-engaged, and sociocultural strategies as important to reaching a particular community. Building on this knowledge, we identify four approaches that a cancer education intervention that the youth of African immigrant heritage can draw on, namely: (1) linguistic and cultural markers, (2) belief system and religious affiliation, (3) hybrid nationality, and (4) age-related learning preferences. With a deeper understanding of how these factors, concerning cancer health, shape the identities, beliefs, and behaviors of African immigrant youth in the United States, we may be able to create culturally competent educational tools for cancer awareness and prevention.

**Linguistic and Cultural Markers**

African immigrants, having come from different countries with diverse colonial histories, have distinct languages. The native languages of African immigrants play an important part in their identity. The youth of African immigrants assimilate the language and cultures of the host society while leveraging their cultural and linguistic heritage for optimum survival, a process that yields linguistic and cultural hybridity.

The complexity that underlies the African immigrants’ linguistic and cultural identities in the United States should inform approaches to developing culturally competent education for youth and their families to improve overall health outcomes. It is expedient to target cancer-awareness information by incorporating aspects of the home languages of African immigrants—especially Western and Eastern Africa [5]. For example, the Swahili language would be accessible to immigrant families of East African origin, and Pidgin English for families with West African heritage. Appropriate learning mediums for cancer awareness for African immigrant youth should intersect with the linguistic and cultural practices of the African immigrant population.

**Belief System and Religious Affiliation**

In a 2021 report, the Pew Research Center stated that African immigrants in the United States are more religious than other Black Americans, even though Black Americans are more religious than Americans of other races [81]. Further broken down into specific practices, the report noted that African immigrants value attending religious services weekly, more than other Black Americans: “around half of the African immigrants living in the United States (54%) say they attend religious services at least weekly, compared with about 3-in-10 United States-born (32%) and Caribbean-born (30%) Black adults.”

Similar to language, culture, and national consciousness, the belief systems and religiosity of African immigrants will have a major imprint on their young children. Health information tailored specifically to religiosity will not only be responsive to African immigrants’ cultural perspectives, but it may also improve engagement with pedagogical materials. Moreover, studies are scarce on the intersection of African immigrants’ religious practices and responses to health care education about cancer, thereby illustrating another gap in research that may ultimately improve the approach to cancer education among distinctive ethnic and racial groups. Additionally, there is a shortage of research on the religious practices of African immigrants, highlighting another research gap that could ultimately enhance approaches to cancer education among distinct ethnic and racial groups.

**Hybrid Nationality and Afropolitanism**

African immigrants in the United States, have diverse origins from one of the 54 nations of Africa, many of which are multiethnic. These diverse ethnic identities house unique cultural and linguistic features within and outside the individual nation’s borders. While African immigrants actively engage with the dominant Western traditions of the society they reside in, they also maintain their cultural customs. As a result, youth from African immigrant families often exhibit hybrid language use, blending the host language with elements of African culture, including specific exclamations and colloquial forms rooted in African cultural beliefs. This linguistic and cultural hybridity is significant in addressing the existing gap in cancer awareness research among African immigrant families and fosters a sense of community within the African immigrant population in the United States.

The concept of Afropolitanism defines Africans as an integral part of the global community rather than separate from it. This concept refers to the empowerment associated with a blended, polyethnic, and cosmopolitan identity [80]. Afropolitanism iterates Africans’ awareness of their origins and the consciousness of the cultural ambiguities that occur because of their integration into the host society. This understanding impacts African immigrants’ response to cancer education and approaches to accessing health care for cancer treatment. Their cultural and spiritual beliefs are still very much prominent in their perspective on cancer disease. This consciousness could, however, be tapped into for possible changes and adaptations among this immigrant group. The summary of the African immigrants’ complex experience is iterated in the term, “Afropolitan.” Afropolitan describes an individual whose identities are deeply rooted in their diverse, transcultural experiences, reflecting youth linguistic and cultural practices within African immigrant families [68,69]. African immigrants’ hybrid language and cultural identities necessitate the development of health educational tools and technologies that integrate African cultural perspectives and engage these youth in learning and retaining health information in a culturally responsive manner.

**Age-Related Learning Preferences**

Consideration for age-appropriateness in technology is not unique to African immigrant youth; however, the connection of this factor to digital literacy, access, and equity makes it critical to examine further and worthy of discussion. A report by the Migration Policy Institute on immigrant learning with digital technology has identified uneven access to digital resources for youth aged between 15 and 17 years who are either...
immigrants themselves or have at least one immigrant parent [82]. Research suggests that factors like work, language, and familial influence affect how immigrant youth use technology for learning [83]. Given the versatility of the adolescent age group with technology, they have increased access to vital information on health issues and diseases that are prevalent within their community. More important is their access to their heritage culture as well as the culture of their residing society. As a result, youth play a vital role as intermediaries, connecting with their families to promote cancer awareness within their communities.

Textbox 1. Summary of key main areas for future research.

**Priorities for future cancer prevention and control research focused on African immigrant populations**

- Disaggregate study populations according to country or region of origin to improve cultural tailoring and knowledge.
- Develop family-oriented educational initiatives including programs for children.
- Use community-engaged approaches including partnerships with faith-based organizations.
- Leverage emerging technology for recruiting study participants and delivering educational messages while accounting for barriers to access.
- Align cancer awareness information with language and cultural markers specific to the population.
- Consider the global African community and hybrid African and American cultural practices.
- Incorporate relevant religious and spiritual beliefs and practices to enhance cancer education effectiveness.
- Consider youth and adolescents as intermediaries for increasing cancer awareness among family members.
- Explore the potential for interagency collaboration (Centers for Disease Control and Prevention, Centers for Medicare & Medicaid Services, Health Resources and Services Administration, and National Institutes of Health) to address cancer-related health challenges for African immigrant families.

To achieve the goal of increasing cancer awareness among African immigrant families, 1 strategy involves creating a culturally tailored serious game. Serious games offer opportunities to build upon the research base of effective approaches to reduce the cancer burden by focusing on youth and leveraging technology. Research is crucial that examines the language use of youth from African immigrants in the United States. Previous research has already categorized most African immigrants living in the United States into Western (35.71%) and Eastern Africa (29.61%) groups, which could serve as a basis for examining youth cancer awareness within each group [5]. Open-ended ethnographic interviews could be used to identify the nuanced cultural and linguistic practices of the youth of African immigrant families. The heterogeneity of Africa’s cultural identities could result in a new monolithic idea of Black subgroups in the United States, the importance of beginning this inquiry cannot be delayed. Detailed demographic questionnaires and open response forms can allow for flexibility in how studies aggregate and allow for new divisions and aggregations of African immigrants. However, it is noted that immigration by African countries is unequal with many African immigrants arriving from Western and Eastern African countries [5].

Additionally, recruitment strategies are particularly important in the success of this line of research and will need to be evaluated. As immigrant populations are “Hard-to-Reach,” using innovative ways to reach a target population is also important [84]. During the COVID-19 pandemic, online recruitment using Facebook (Meta), Instagram (Meta Platforms), and WhatsApp (WhatsApp LLC) was an effective recruitment strategy especially because it built on existing communication and information-sharing norms within the African immigrant community. Further research should use and evaluate multiple recruitment streams.

Findings from such research endeavors will have a meaningful impact on the strategies for developing culturally tailored educational content such as a serious game, to create awareness about cancer among African immigrant families in the United States. A culturally adapted serious game has immense potential to be instrumental in improving awareness and cancer prevention strategies in African immigrant families.

**Conclusion**

The importance of culturally tailored cancer education for African immigrants is underscored by uncertainty. Issues surrounding the measurement of cancer prevalence in African immigrant populations exacerbate the uncertainty of how cancer affects the African immigrant population in the United States. The existing, yet limited research on the topic suggests that African immigrants, especially adolescents, have unique experiences that lie at the nexus of their traditional culture and the complex novelty of the US health care system for immigrants. Research demonstrates the impact of cultural beliefs (such as fatalism and stigmatization of cancer diagnoses among African cultures) and lack of knowledge about cancer and cancer screening compounds to affect access to screening and care for African immigrants. Further research specifically targeting African immigrants and their youth can not only disentangle...
the unique struggles of African immigrants but also allow for the tailoring of education to provide maximal impact to vulnerable populations.

While recognizing our lack of knowledge and the uncertainty around the experience of cancer in the United States for African immigrants and advocating that increased research is the needed foundation for alleviating health disparities, more difficult work is ahead. It is integral for health scientists, health care providers, African culture scholars, and communities of African immigrants to come together for sustained research activity. These transdisciplinary associations will aid in the collection of data specific to African immigrants, but also the community engagement needed to co-design a culturally sensitive educational intervention. This will be no small task and require the dedicated work of many experts alongside and within the African immigrant community to forge long-term relationships that can facilitate recruitment, retention, and meaningful knowledge generation for the African immigrant community in the context of cancer experience.

Acknowledgments

Generative artificial intelligence technology was not used in the development or editing of this paper.

Authors’ Contributions

OA conceptualized the original idea for this paper, led the gathering of data and team expertise, and contributed to drafting and leading the revising of the final paper. All coauthors contributed to the literature review and drafting of this paper.

Conflicts of Interest

None declared.

References


65. Simich L. Health literacy and immigrant populations. Public Health Agency of Canada and Metropolis Canada. 2009. URL: https://citeserx.ist.psu.edu/document?repid=rep1&fype=pdf&doi=e2ce08b55090a7c701e73b91e51505a9e91f4899 [accessed 2024-02-14]


75. Abraham et al. JMIR CANCER 2024 | vol. 10 | e53956 | p.50 https://cancer.jmir.org/2024/1/e53956


Abbreviations

**CRP:** culturally relevant pedagogy
Iterative Patient Testing of a Stimuli-Responsive Swallowing Activity Sensor to Promote Extended User Engagement During the First Year After Radiation: Multiphase Remote and In-Person Observational Cohort Study

Eileen H Shinn¹, PhD; Adam S Garden², MD; Susan K Peterson¹, PhD; Dylan J Leupi³, BS; Minxing Chen³, MS; Rachel Blau⁴, PhD; Laura Becerra⁶, MS; Tarek Rafeedt⁵; Julian Ramirez⁵, PhD; Daniel Rodriquez⁵, PhD; Finley VanFossen⁶, MA; Sydney Zehner¹, BS; Patrick P Mercier⁶, PhD; Joseph Wang⁵, PhD; Kate Hutcheson²,³, PhD; Ehab Hanna⁷, MD; Darren J Lipomi⁵, PhD

¹Department of Behavioral Science, University of Texas, MD Anderson Cancer Center, Houston, TX, United States
²Department of Radiation Oncology, University of Texas, MD Anderson Cancer Center, Houston, TX, United States
³Department of Chemistry and Biochemistry, College of Science, University of Notre Dame, South Bend, IN, United States
⁴Department of Biostatistics, University of Texas, MD Anderson Cancer Center, Houston, TX, United States
⁵Department of Nano and Chemical Engineering, University of California, San Diego, CA, United States
⁶Department of Electrical and Computer Engineering, University of California, San Diego, CA, United States
⁷Department of Head and Neck Surgery, University of Texas, MD Anderson Cancer Center, Houston, TX, United States

Corresponding Author:
Eileen H Shinn, PhD
Department of Behavioral Science
University of Texas
MD Anderson Cancer Center
1155 Herman Pressler
Unit 1330, PO Box 301439
Houston, TX, 77230-1330
United States
Phone: 1 713 745 0870
Fax: 1 713 745 4286
Email: e.shinn@mdanderson.org

Abstract

Background: Frequent sensor-assisted monitoring of changes in swallowing function may help improve detection of radiation-associated dysphagia before it becomes permanent. While our group has prototyped an epidermal strain/surface electromyography sensor that can detect minute changes in swallowing muscle movement, it is unknown whether patients with head and neck cancer would be willing to wear such a device at home after radiation for several months.

Objective: We iteratively assessed patients’ design preferences and perceived barriers to long-term use of the prototype sensor.

Methods: In study 1 (questionnaire only), survivors of pharyngeal cancer who were 3-5 years post treatment and part of a larger prospective study were asked their design preferences for a hypothetical throat sensor and rated their willingness to use the sensor at home during the first year after radiation. In studies 2 and 3 (iterative user testing), patients with and survivors of head and neck cancer attending visits at MD Anderson’s Head and Neck Cancer Center were recruited for two rounds of on-throat testing with prototype sensors while completing a series of swallowing tasks. Afterward, participants were asked about their willingness to use the sensor during the first year post radiation. In study 2, patients also rated the sensor’s ease of use and comfort, whereas in study 3, preferences were elicited regarding haptic feedback.

Results: The majority of respondents in study 1 (116/138, 84%) were willing to wear the sensor 9 months after radiation, and participant willingness rates were similar in studies 2 (10/14, 71.4%) and 3 (12/14, 85.7%). The most prevalent reasons for participants’ unwillingness to wear the sensor were 9 months being excessive, unwanted increase in responsibility, and feeling self-conscious. Across all three studies, the sensor’s ability to detect developing dysphagia increased willingness the most compared
to its appearance and ability to increase adherence to preventive speech pathology exercises. Direct haptic signaling was also rated highly, especially to indicate correct sensor placement and swallowing exercise performance.

**Conclusions:** Patients and survivors were receptive to the idea of wearing a personalized risk sensor for an extended period during the first year after radiation, although this may have been limited to well-educated non-Hispanic participants. A significant minority of patients expressed concern with various aspects of the sensor’s burden and its appearance.

**Trial Registration:** ClinicalTrials.gov NCT03010150; https://clinicaltrials.gov/study/NCT03010150

**KEYWORDS**
user-centered design; patients with head and neck cancer; dysphagia throat sensor

**Introduction**

**Background**

In 2021, approximately 32,000 Americans developed laryngeal or pharyngeal cancer, which has a 5-year survival rate of 61% for all stages combined [1]. Management of these cancers often includes high-dose intensity-modulated radiation therapy (IMRT) designed to spare pharyngeal muscles and reduce the incidence of radiation-associated dysphagia (swallowing difficulty) [2]. Still, a range of studies have reported that roughly 60% of patients receiving IMRT developed long-term swallowing problems within 2 years after radiation had ended, ranging in intensity from inability to swallow solid food without compensatory strategies to being completely feeding tube dependent [3-10].

As with most chronic conditions, early detection and intensive swallowing therapies are key to preventing long-term dysphagia [11-26], especially if patients are adherent to swallowing therapy instructions [27]. However, noninvasive screening procedures for early detection of radiation-associated fibrosis do not yet exist in the United States. Instead, gold standard modified barium swallow (MBS) and fiberoptic endoscopic evaluation of swallowing tests are typically ordered after the patient begins to complain of difficulties with swallowing [12]. Furthermore, preventive swallowing therapies are not always prescribed prior to the development of radiation-associated dysphagia [28-30]. Unfortunately, once radiation-associated dysphagia is clinically detected, there is little hope of fully restoring normal function [11,31,32].

To detect radiation-associated dysphagia before it becomes permanent, it is necessary to monitor changes in swallowing function much more frequently than is currently possible in the clinical setting. Subclinical change in swallowing activity or risk for dysphagia could be assessed during standard cancer surveillance visits, but increasing the periodicity of these visits would increase patient burden by requiring more frequent travel to the medical center for swallowing imaging and tests. Frequent at-home monitoring with wearable sensors between scheduled surveillance visits could address this gap in monitoring, at-home monitoring with wearable sensors between scheduled visits could address this gap in monitoring, surveillance visits could address this gap in monitoring, especially if the sensors were designed to support decision-making regarding initiation of intensive speech language therapies [33]. To this end, researchers have developed myriad devices that can be worn on the skin and measure a range of mechanical, optical, biochemical, electrical, or acoustic signals with high fidelity [34-39]. However, sensor performance alone is not sufficient for improving health outcomes as patient engagement is also important [40]. Within the specific context of preventing dysphagia in survivors of head and neck cancer, repeated at-home monitoring over a period of months if not years is necessary to demonstrate a clinical advantage over current treatment paradigms. Unfortunately, most mobile technologies fail to engage patients over sustained durations, with most mobile health (mHealth) interventions for chronic disease reporting steep declines in use, some as high as 95% within the first few weeks, depending on the technology and context [41-43]. The most frequently cited reasons for discontinued use are decreased interest in the technology after its novelty abates, perceived lack of usefulness relative to burden, poor implementation of user experience, and frustration with technical issues [44-47].

To counter these barriers, it is widely agreed that user-centered testing be conducted in a sustained and iterative fashion during the design and development of new health technology. User-centered testing assesses the human technology interface by evaluating how well the technology incorporates into end users’ daily routines, habits, and capabilities, known loosely as user acceptability [40,48]. Beyond acceptability, technologies should be designed to maximize their potential to effect changes in patients’ attitudes and health behaviors. Oinas-Kukkonen and Harjumaa’s [49] persuasive system design model describes four categories of persuasive design principles that optimize the likelihood of health behavior change: task support (personalized design features that make it easier for users to achieve their goals), social support (leveraging interpersonal learning, eg, via online community forums), dialogue support (providing feedback to the user in a manner that helps the user move toward their goal, eg, with praise and rewards), and system credibility (the perceived clinical expertise embedded within the sensor output) [49]. Relatively few mHealth interventions conduct user-centered testing during technology development, which may be one reason for diminishing patient engagement and eventual abandonment [50-53]. In the US market, the user abandonment rate of fitness trackers is 50% within 6-12 months [44,54]. Patient abandonment rates are higher for those 70 years and older; one study found that 43% of their sample had abandoned their sleep and activity trackers within the first 2 weeks of use [55].

A recent review of 51 mHealth intervention studies targeting chronic diabetes, cardiovascular, or pulmonary diseases noted that diminished patient engagement was prevalent and posed a
significant threat to effective use of the technology. Accordingly, nonsignificant effects on clinical markers outweighed significant findings two-to-one [42]. Therefore, our study explicitly addressed the design of a wearable sensor with the future intended use of home-based assessment for 9 months, starting with the third month after radiation to the 12th month. All design preferences and opinions were solicited within the context of sustaining engagement with the sensor for 9 months during the first year since repeated measurements over time would be needed to detect patterns of developing dysphagia in posttreatment patients.

**Goal of This Study**

We assessed patient needs and preferred characteristics regarding the design of a wearable sensor to deliver personalized risk of dysphagia. Specifically, we assessed perceived barriers to wearing the sensor for 9 months, starting from the third month after the end of radiation treatment (to allow for healing from radiation skin burn) until the 12th month post treatment, and the impact of proposed design features on willingness to wear the sensor. In the first of three iterative user-centered tests, we surveyed a large cohort of survivors of head and neck cancer who were 4-5 years past radiation treatment to assess the perceived need for the sensor and desired design features for future prototypes. In study 2, we assessed user acceptability for a wired prototype sensor within a small sample of long-term survivors, oversampled for radiation-associated dysphagia. Finally, in the third user test, we tested a revised prototype on a sample of patients with head and neck cancer undergoing active treatment to get a better sense of competing priorities during a fraught time in their lives. The revised prototype included more elastic and comfortable materials for the strain sensor and custom-made dry electromyography (EMG) sensors, as opposed to commercial sensors. During the third test, we repeated our questions about user acceptability and willingness to wear the sensor for 9 months, as well as new questions about bidirectional feedback in the form of haptic (vibration) signaling.

**Methods**

**Study 1**

**Design and Eligibility**

Survivors of head and neck cancer who were still alive and who were already enrolled in a psychosocial parent study were asked to answer a questionnaire about a hypothetical throat sensor. Men and women were eligible for the parent study if they had received radiation with curative intent for oropharyngeal (stage II-IVb), laryngeal (II-IVb), hypopharyngeal (I-IVb), or nasopharyngeal cancer (I-IVb), or an unknown primary cancer with cervical metastases; were at least 2 years post treatment; were 18 years or older; and spoke English. Men and women were excluded if they had treatment for previous head and neck cancer; a history of previous head and neck surgery (previous biopsy, tonsillectomy, or tracheotomy were allowed); other cancer diagnoses, except nonmelanoma skin cancer; or a history of current oropharyngeal dysphagia unrelated to cancer diagnosis (eg, dysphagia due to underlying neurogenic disorder).

**Recruitment and Data Collection Procedures**

For the psychosocial parent study, all eligible patients were approached for recruitment at the radiation clinic’s radiation education class after being identified at the weekly multidisciplinary tumor board conference. The accrual rate for entry into the original parent study was 77%; demographic and disease information was collected at baseline. Those patients who were already enrolled in the psychosocial parent study and still alive (n=234) were contacted by phone to determine if they would answer optional questions about a hypothetical sensor to be worn on the throat. Patients who did not return calls after 5 attempts or did not have working phone numbers were not approached further for enrollment into study 1. After obtaining informed consent, participants completed the optional questionnaire administered either by REDCap, telephone, or mail at a single time point [56]. For mailed questionnaires, a research staff’s phone number was provided if the patient had any questions about the questionnaire.

**Measures**

Demographic information regarding age, race/ethnicity, employment, income, and marital status were obtained by questionnaire. Disease stage was abstracted from the medical record. Participants then completed a questionnaire. The first page of the questionnaire showed a photograph of the proposed sensor (Figure 1A) and a diagram of the sensor’s placement on the neck (Figure 1B), a brief description of the sensor’s purpose, and the proposed timeline of wearing the sensor every weekend from the third month post radiation to the 12th month post radiation for a total of 9 months.
Figure 1. Appearance of the hypothetical and actual sensor prototypes. (A) Study 1 respondents were shown a photograph of the proposed sensor and (B) its proposed location on the neck. (C) Study 2’s graphene strain sensor prototype, supported on 13-µm-thick polyimide tape (the contact surface is silicone) and placed on the submental region probing muscle contraction. (D) Study 3’s soft polymer strain sensor, now placed under the laryngeal prominence to capture movement during swallowing.

Main Outcome: Willingness to Wear the Sensor
For studies 1-3, the study questionnaire asked whether the patient would have been willing to wear the sensor for 9 months during the first year after radiation, starting in month 3 post treatment. This time point was asked about since it would give sufficient time for the skin on their neck to have healed from radiation skin burn. Participants were then asked whether they would have been willing to wear the sensor for the entire 9 months, every other week, or every weekend during the 9-month period, and then a series of branched logic true-false questions about reasons for willingness versus unwillingness to wear the sensor. Next, using a 3-point Likert scale response format, all participants rated whether changes in the sensor design (either unobtrusive appearance or the ability to receive feedback about risk for dysphagia) would change the individual’s willingness or unwillingness to wear the sensor every weekend for 9 months. Additional comments or suggestions were also solicited as free text.

Study 2
Design and Eligibility
A second sample of survivors of head and neck cancer who were 2 to 10 years post radiation and attending surveillance visits at MD Anderson gave informed consent and enrolled into the study during a 1-week period; testing was constrained to a 1-week period in which visiting graduate engineering students from the University of California San Diego traveled to MD Anderson for on-patient equipment testing. The eligibility criteria for study 2 were the same as for study 1; however, we oversampled for patients with a Dynamic Imaging Grade of Swallowing Toxicity score >0, indicating radiation-associated dysphagia that had been verified with MBS [57]. The oversampling was done to gauge the accuracy of the prototype sensor in distinguishing between dysphagic survivors and survivors without dysphagia [58]. For every dysphagic participant, we recruited a nondysphagic patient matched for age and sex. For patients who declined participation, deidentified disease information, demographics, and reason for refusal were noted in the study record.

Procedure and Assessment
A wired prototype graphene strain sensor coupled with a wired surface EMG sensor was placed on the patient to obtain muscle movement measurements during a series of swallowing tasks of various bolus textures, as described previously (Figure 1C) [58]. Immediately after the on-throat sensor test, patients were asked to answer six questions about the sensor’s discomfort, ease of use, and associated embarrassment using a 5-point Likert scale ranging from strongly disagree to strongly agree. Patients
were again asked whether they would be willing to wear the sensor for 9 months (but now for once a month on the weekends) with branching questions asking for reasons for willingness versus unwillingness. Patients were again asked to rate the impact of sensor unobtrusiveness and predictive dysphagic feedback on willingness to wear the sensor for extended periods. Finally, demographic information regarding age, race, and marital status were abstracted from the medical record. All testing sessions were conducted at the Head and Neck Cancer Center at MD Anderson.

**Study 3**

**Design and Eligibility**

Similar eligibility, consent, and testing procedures were used in study 3. However, eligible patients were more likely to be approached during active treatment for throat cancer, whereas studies 1 and 2 recruited long-term survivors. Study 3’s sensor (Figure 1D) was revised to have better skin conformation and comfort; standard surface EMG electrodes were now replaced with flexible custom dry electrodes, whereas the strain sensor was supported on a silicon substrate [59].

**Assessment Procedures**

After completion of the on-throat sensor test, patients were also asked the same questions used in study 1 regarding willingness to wear the sensor for 9 months and whether changes in the sensor’s appearance and feedback capability would change their minds about their willingness to wear the sensor. In addition, participants were interviewed regarding the helpfulness of future capability of the sensor itself to give immediate haptic feedback in three different scenarios: to indicate correct placement of the sensor, to indicate correct performance of a particular swallowing exercise, and to indicate quality of swallowing during at-home testing of various bolus textures. Their answers were transcribed, categorized, and coded into three categories (0: not helpful; 1: helpful under certain conditions; 2: helpful).

**Analysis**

Descriptive statistics (eg, proportions, means, ranges, and SDs) were computed for the process evaluation and participant satisfaction data, together with 95% CIs. To assess the external validity of the study, demographic and disease information was compared between respondents and nonrespondents in study 1 (Table 1) and between participants and refusers in studies 2 and 3 (data not shown; available data for participants in studies 2 and 3 is shown in Table 2). All questionnaire responses were analyzed with SPSS (version 26; IBM Corp).
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Potentially eligible survivors (from parent study)</th>
<th>Survivors who completed the questionnaire</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total (N=234) Nonrespondent (did not participate) Respondent</td>
<td>Total (n=138) Willing to wear for 9 mo Unwilling</td>
<td></td>
</tr>
<tr>
<td>What is your age? (years)</td>
<td></td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>Participants, n (%)</td>
<td>234 96 (41.0) 138 (58.9) 138 22 (15.9) 116 (84.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>57.4 (10.0) 56.6 (9.8) 58 (10.1) 58 (10.1) 55.2 (9.4) 58.5 (10.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (min-max)</td>
<td>58 (18-83) 56 (30-79) 59 (18-83) 59 (18-83) 56.5 (35-75) 59 (18-83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is your ethnic background? n (%)</td>
<td></td>
<td></td>
<td>.003</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>21 (9.1) 15 (16.0) 6 (4.4) 6 (4.4) 0 (0.0) 6 (5.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>210 (90.9) 79 (84.0) 131 (95.6) 131 (95.6) 22 (100.0) 109 (94.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td>.23</td>
</tr>
<tr>
<td>African American</td>
<td>10 (4.3) 6 (6.4) 4 (2.9) 4 (2.9) 1 (4.5) 3 (2.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>1 (0.4) 1 (1.1) 0 (0.0) 0 (0.0) 0 (0.0) 0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>6 (2.6) 1 (1.1) 5 (3.6) 5 (3.6) 1 (4.5) 4 (3.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>1 (0.4) 0 (0.0) 1 (0.7) 1 (0.7) 0 (0.0) 1 (0.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>213 (92.2) 86 (91.5) 127 (92.7) 127 (92.7) 20 (90.9) 107 (93.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>Some college and lower</td>
<td>112 (48.9) 54 (58.1) 58 (42.6) 58 (42.6) 7 (31.8) 51 (44.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>117 (51.1) 39 (41.9) 78 (57.4) 78 (57.4) 15 (68.2) 63 (55.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td>.60</td>
</tr>
<tr>
<td>Full-time/part-time</td>
<td>145 (63.3) 57 (61.3) 88 (64.7) 88 (64.7) 17 (77.3) 71 (62.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not employed</td>
<td>84 (36.7) 36 (38.7) 48 (35.3) 48 (35.3) 5 (22.7) 43 (37.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td>.50</td>
</tr>
<tr>
<td>Single living alone/married but living apart/separated/divorced/widow</td>
<td>46 (20.0) 21 (22.1) 25 (18.5) 25 (18.5) 4 (18.2) 21 (18.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single but living with significant other/married living with spouse</td>
<td>184 (80.0) 74 (77.9) 110 (81.5) 110 (81.5) 18 (81.8) 92 (81.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation, n (%)</td>
<td></td>
<td></td>
<td>.07</td>
</tr>
<tr>
<td>Professional/managerial</td>
<td>143 (71.9) 51 (63.0) 92 (78.0) 92 (78.0) 16 (88.9) 76 (76.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail/service/labor</td>
<td>44 (22.1) 24 (29.6) 20 (16.9) 20 (16.9) 2 (11.1) 18 (18.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student/unemployed</td>
<td>12 (6.0) 6 (7.4) 6 (5.1) 6 (5.1) 0 (0.0) 6 (6.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is your income before taxes? (US $), n (%)</td>
<td></td>
<td></td>
<td>.007</td>
</tr>
<tr>
<td>&lt;30,000</td>
<td>38 (18.9) 24 (30.4) 14 (11.5) 14 (11.5) 0 (0.0) 14 (13.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30,000-50,000</td>
<td>31 (15.4) 13 (16.5) 18 (14.8) 18 (14.8) 2 (10.5) 16 (15.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50,000-75,000</td>
<td>28 (13.9) 9 (11.4) 19 (15.6) 19 (15.6) 2 (10.5) 17 (16.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;75,000</td>
<td>104 (51.7) 33 (41.8) 71 (58.2) 71 (58.2) 15 (78.9) 56 (54.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage of disease, n (%)</td>
<td></td>
<td></td>
<td>.24</td>
</tr>
<tr>
<td>Stages I or II</td>
<td>76 (32.5) 27 (28.1) 49 (35.5) 49 (35.5) 7 (31.8) 42 (36.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stages III or IV</td>
<td>158 (67.5) 69 (71.9) 89 (64.5) 89 (64.5) 15 (68.2) 74 (63.8)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Demographic/disease comparisons between willing and unwilling participants (studies 2 and 3).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Study 2</th>
<th>Study 3</th>
<th>P value</th>
<th>Study 2</th>
<th>Study 3</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total sample (n=14)</td>
<td>Willing to wear for 9 mo (n=10)</td>
<td>Unwilling (n=4)</td>
<td>Total sample (n=14)</td>
<td>Willing to wear for 9 mo (n=12)</td>
<td>Unwilling (n=2)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>61.6 (11.5)</td>
<td>61.2 (12.3)</td>
<td>62.8 (11.0)</td>
<td>.83</td>
<td>62.4 (12.3)</td>
<td>61.0 (11.9)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>.55</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Asian</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
<td>2 (14)</td>
<td>2 (17)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>13 (92.9)</td>
<td>9 (90)</td>
<td>4 (100)</td>
<td></td>
<td>12 (86)</td>
<td>10 (83)</td>
</tr>
<tr>
<td>More than one race</td>
<td>1 (7.1)</td>
<td>1 (10)</td>
<td>0 (0)</td>
<td></td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>What is your ethnic background? n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>3 (21.4)</td>
<td>2 (20)</td>
<td>1 (25)</td>
<td></td>
<td>1 (7)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>11 (78.6)</td>
<td>8 (80)</td>
<td>3 (75)</td>
<td></td>
<td>13 (93)</td>
<td>11 (92)</td>
</tr>
<tr>
<td>Occupation, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managerial/professional</td>
<td>2 (14)</td>
<td>2 (20)</td>
<td>0 (0)</td>
<td></td>
<td>7 (50)</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Retail, service, operator</td>
<td>9 (64)</td>
<td>6 (60)</td>
<td>3 (75)</td>
<td></td>
<td>6 (43)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>Student or unemployed</td>
<td>3 (21)</td>
<td>2 (20)</td>
<td>1 (25)</td>
<td></td>
<td>1 (7)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/living with significant other</td>
<td>12 (86)</td>
<td>8 (80)</td>
<td>4 (100)</td>
<td></td>
<td>12 (86)</td>
<td>12 (100)</td>
</tr>
<tr>
<td>Single/divorced/widowed/separated</td>
<td>2 (14)</td>
<td>2 (20)</td>
<td>0 (0)</td>
<td></td>
<td>2 (14)</td>
<td>1 (50)</td>
</tr>
<tr>
<td>Dysphagic status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dysphagic (DIGEST(a)&gt;0)</td>
<td>7 (50)</td>
<td>5 (50)</td>
<td>2 (50)</td>
<td>&gt;.99</td>
<td>7 (50)</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Not dysphagic (DIGEST=0)</td>
<td>7 (50)</td>
<td>5 (50)</td>
<td>2 (50)</td>
<td></td>
<td>N/A(b)</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Disease stage, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-II</td>
<td>1 (8)</td>
<td>0 (0)</td>
<td>1 (25)</td>
<td></td>
<td>1 (7)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>III-IV</td>
<td>12 (92)</td>
<td>9 (90)</td>
<td>3 (75)</td>
<td></td>
<td>13 (93)</td>
<td>11 (92)</td>
</tr>
</tbody>
</table>

\(a\)DIGEST: Dynamic Imaging Grade of Swallowing Toxicity.
\(b\)N/A: not applicable.

Ethical Considerations
All study materials and procedures were approved by the institutional review board at MD Anderson Cancer Center’s institutional review board (protocol 2016-0597). All enrolled participants signed informed consent forms before testing began. All study data were deidentified, and no compensation was provided for participation.

Results
Overview
Prior to patient user testing, our study incorporated design input from multiple disciplines, including behavioral scientists, speech pathologists, radiation oncologists, and engineers. Initially, our primary concerns were to develop a wearable device that would not injure skin sensitized by radiation and have an uncomplicated application and removal procedure. Various invasive sensors, such as those worn inside the mouth, were dropped from consideration after it was realized that patients would possibly need to use the device during radiation and later at home during the first year post treatment. During study 1, we gathered patient reactions to a photograph of a sensor (Figure 1), whereas in studies 2 and 3, prototype versions were tested on survivors and patients in the clinic (Figure 1). The racial breakdown of the overall study sample (N=234) was non-Hispanic White (n=213, 92.2%), African American (n=10, 4.3%), Asian American (n=6, 2.6%), American Indian/Alaska
Native (n=1, 0.4%), and Native Hawaiian/Pacific Islander (n=1, 0.4%).

**Study 1**

Research staff contacted 234 eligible participants to complete study 1’s questionnaire, either via REDCap or by mail; 138 (59%) participants completed the questionnaire (Figure 2). Participants in study 1 were primarily non-Hispanic White and married, and their mean age was 57.4 (SD 10) years (Table 1). Median time since end of radiation treatment was 4 years and 26 days (Table 1). Analyses of responders versus nonresponders showed that responders were more likely to be non-Hispanic, have a bachelor’s degree, and have higher annual income; differences in race, age, and disease stage were not significantly different (Table 1).

**Figure 2.** Recruitment CONSORT (Consolidated Standards of Reporting Trials) for study 1 (n=138).

![Recruitment CONSORT](image)

**Survivor Preferences for Wearable Throat Sensor**

Of the 138 respondents, 115 (83.3%) agreed that they would have been willing to wear the sensor for 9 months during the first year after radiation. However, patients were not willing to wear the sensor during the workweek due to fear of coworkers or strangers asking about the sensor. Instead, they were willing to wear the sensor on weekends, but only for one weekend a month as opposed to every weekend. When presented with several potential reasons explaining their willingness to wear the sensor, nearly all participants cited altruism, whereas 88% (92/105) cited interest in the sensor technology and 77% (75/97) thought that the sensor would help them adhere to their preventive swallowing exercises (Table 3). For example, several patients commented that the personalized feedback from the sensor would provide additional motivation to adhere to their preventive swallowing exercises:

- *It would push me to do my exercises diligently...*
- *It would get me on the ball and do my exercises more often...*
- *It would give me the information I can use to fight back the scar tissue problem. And see the importance of my neck exercises.*
Table 3. Studies 1-3: Number of patients endorsing reasons for willingness/unwillingness to wear the sensor every weekend for 9 months.

<table>
<thead>
<tr>
<th>Reasons for willingness/unwillingness to wear the sensor for 9 months&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Would wear sensor, n (%)</th>
<th>Would not wear sensor, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>False</td>
</tr>
<tr>
<td><strong>Study 1 (n=138)&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The technology of the patch sounds interesting.</td>
<td>92 (87.6)</td>
<td>13 (12.4)</td>
</tr>
<tr>
<td>Wearing the patch would have reminded me to do my swallowing exercises.</td>
<td>75 (77.3)</td>
<td>22 (22.7)</td>
</tr>
<tr>
<td>I wanted to help with MD Anderson’s research.</td>
<td>108 (99.1)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>My skin was still sensitive during that time.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I wouldn’t want to put on and take off the patch every weekend.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I wouldn’t want to wear the patch for 9 months.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I would feel uncomfortable if people noticed the patch and ask me questions or wanted to talk about it.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I was being asked to participate in too many studies.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>It would have added to my daily responsibilities.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>It would have been a reminder of my cancer treatment.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I would not be able to see my data from the patch.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Study 2 (n=14)&lt;sup&gt;d&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The technology of the patch sounds interesting.</td>
<td>8 (80)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>Wearing the patch would have reminded me to do my swallowing exercises.</td>
<td>10 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>I wanted to help with MD Anderson’s research.</td>
<td>10 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>My skin was still sensitive during that time.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I wouldn’t want to put on and take off the patch every weekend.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I wouldn’t want to wear the patch for 9 months.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I would feel uncomfortable if people noticed the patch and ask me questions or wanted to talk about it.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I was being asked to participate in too many studies.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>It would have added to my daily responsibilities.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>It would have been a reminder of my cancer treatment.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I would not be able to see my data from the patch.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Study 3 (n=14)&lt;sup&gt;f&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The technology of the patch sounds interesting.</td>
<td>12 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Wearing the patch would have reminded me to do my swallowing exercises.</td>
<td>12 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>I wanted to help with MD Anderson’s research.</td>
<td>12 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>My skin was still sensitive during that time.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I wouldn’t want to put on and take off the patch every weekend.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I wouldn’t want to wear the patch for 9 months.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I would feel uncomfortable if people noticed the patch and ask me questions or wanted to talk about it.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I was being asked to participate in too many studies.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>It would have added to my daily responsibilities.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>It would have been a reminder of my cancer treatment.</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>I would not be able to see my data from the patch.</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<sup>a</sup>Participants were asked the following question “Which of the following reasons would motivate you to wear the sensor every weekend for 9 months after radiation?”

---

https://cancer.jmir.org/2024/1/e47359
In study 1, 155 (83.5%) participants indicated that they would wear the sensor, while 23 (16.5%) participants indicated that they would not wear it.

In study 2, 10 (71.4%) participants indicated that they would wear the sensor, while 4 (28.5%) participants indicated that they would not wear it.

In study 3, 12 (85.7%) participants indicated that they would wear the sensor, while 2 (14.3%) participants indicated that they would not wear it.

Others valued the additional information that the sensor would provide:

I would be curious to know what is going on with my body...
I would have liked to have known what was happening to my throat...
It’s my neck! Why wouldn’t I want to know?

Among the 22 participants who indicated that they would have been unwilling to wear the sensor, nearly 90% (24/28, 85.7%) of all unwilling participants cited the lengthy duration of having to wear the sensor and 57% (16/28) disliked the idea of having to wear the sensor every weekend. The photograph of the proposed sensor had large black letters embedded within the sensor (Figure 1) to contain its wiring; over half of the unwilling participants objected to the sensor being noticeable enough that others would want to ask questions about its purpose. Just under one-third of unwilling participants disliked the idea of being reminded of their cancer treatment during the first year after radiation (Table 3). Participants who were unwilling to wear the sensor for 9 months did not have any significant demographic or clinical differences compared to participants who expressed willingness to wear the sensor. When asked whether changing the sensor’s appearance to that of a Band-Aid would impact willingness, 29% (4/14) of all study 1 participants agreed that this would increase their willingness, whereas 71% (10/14) stated that unobtrusive appearance would not affect their willingness (mean 2.45, SD 0.87; Figure 3):

Cosmetics is the least of my worries when I am going through treatment and fighting for my life.

When asked about the sensor’s proposed function of delivering individual risk for dysphagia, the majority of the sample (21/28, 75%) agreed that this feature would increase their willingness (mean 1.5, SD 0.88; Figure 3). Notably, half of the free-text comments indicated that had they been able to measure muscle fibrosis earlier, they would have been more diligent about performing their prescribed swallowing exercises. Some simply wrote that they wanted the sensor to be available so that future patients would understand that the risk of dysphagia was high:

I would like to see this in ACTION NOW

Study 2

Within a 1-week period, a convenience sample of 20 potentially eligible survivors of oropharyngeal cancer who were nonmetastatic and able to speak English were approached at their surveillance visit for enrollment into the study. To test the sensor’s performance in distinguishing between normal and dysphagic swallowing patterns, survivors who had developed severe dysphagia as a result of their radiation were oversampled for study 2. Potentially eligible survivors were first identified in the electronic medical record; approached during a surveillance visit; and if consented, scheduled with the engineers for the sensor testing session in a clinic exam room. Three patients refused to participate, citing fatigue or disinterest; all were White, 2 were male, and 1 was female, and their age ranged from 63 to 74 years. Two of the patients were dysphagic and the third was nondysphagic. All three had been diagnosed with late-stage oropharynx cancer (data not shown). A total of 17 (85%) patients agreed, but 1 patient subsequently dropped out due to receiving news of cancer recurrence (Figure 4). Another 2 participants experienced scheduling conflicts; informed consent was obtained from the remaining 14 participants.
Consistent with this cancer type’s demographic profile, the average age of the sample was 61 years, with 12 male participants and 2 female participants. Three participants were Hispanic or Latino and 3 were of non-White race (Table 2). Specific cancer diagnoses included cancer of the oropharynx (9/14, 64%), larynx (3/14, 21%), and nasopharynx (1/14, 7%), and unknown primary cancers (1/14, 7%). The average time since completion of radiation treatment was 47.9 months, and half of the sample had received a diagnosis of radiation-associated dysphagia (Table 2).

Figure 4. Recruitment flowchart for study 2 (n=14).

After wearing the sensor, 10 of the 14 (71%) patients indicated that they would have been willing to wear the sensor for 9 months of the first year post radiation. The most prevalent reasons for willingness were wanting to help future patients detect developing dysphagia and wanting to help MD Anderson research (Table 3). Of the 4 (29%) patients who did not think they would have been willing to wear the sensor, the most popular reason for unwillingness was study burden, specifically, that 9 months was too long of a testing period and the increased responsibilities associated with the sensor. Using a 5-point Likert response scale, patient ratings of discomfort (mean 1.21, SD 0.42), embarrassment (mean 1.14, SD 0.36), and difficulty in application and removal (mean 1.5, SD 0.52) were minimal (Table 4). Therefore, these questions were not repeated in the next phase of user testing.

Table 4. Study 2’s mean patient ratings for sensor discomfort, embarrassment, and difficulty of application (n=14), and study 3’s mean patient ratings of helpfulness for haptic signaling (n=14).

<table>
<thead>
<tr>
<th>Study 2 (n=14)</th>
<th>Patient ratings, mean (SD)</th>
<th>Rangea</th>
</tr>
</thead>
<tbody>
<tr>
<td>The sensor was uncomfortable to wear.</td>
<td>1.21 (0.426)</td>
<td>1.0-5.0</td>
</tr>
<tr>
<td>The sensor would be difficult for me to use at home.</td>
<td>1.5 (0.519)</td>
<td>1.0-5.0</td>
</tr>
<tr>
<td>I thought the experiment was fun.</td>
<td>3.79 (0.893)</td>
<td>1.0-5.0</td>
</tr>
<tr>
<td>The testing session was embarrassing.</td>
<td>1.14 (0.363)</td>
<td>1.0-5.0</td>
</tr>
<tr>
<td>I am good about doing my swallowing exercises every day.</td>
<td>3.27 (1.51)</td>
<td>1.0-5.0</td>
</tr>
<tr>
<td>I believe it is important for me to do as many of my swallowing exercises as possible.</td>
<td>4.46 (1.13)</td>
<td>1.0-5.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study 3 (n=14)</th>
<th>Patient ratings, mean (SD)</th>
<th>Rangea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would it help for the sensor itself to vibrate when you put it in the right spot on your throat?</td>
<td>1.85 (0.376)</td>
<td>0-2.0</td>
</tr>
<tr>
<td>Do you think it would be helpful to have the sensor vibrate once you did your swallowing exercise correctly?</td>
<td>2.00 (0.000)</td>
<td>0-2.0</td>
</tr>
<tr>
<td>Do you think that having the sensor process your swallowing data and give you feedback about the quality of your swallowing would help?</td>
<td>1.46 (0.877)</td>
<td>0-2.0</td>
</tr>
</tbody>
</table>

aFor study 2, the scale ranged from 1 (strongly disagree) to 5 (strongly agree). For study 3, the scale ranged from 0 (no) to 2 (yes).
Study 3
As with study 2, a convenience sample of 14 participants were recruited within a 1-week period to assess user preferences to the updated sensor prototype. As in the previous two studies, the majority of patients were diagnosed with oropharyngeal cancer (11/14, 79%). Unlike the previous two studies, 11 of the 14 (78.6%) were undergoing radiation at the time of testing; the remaining 2 participants were 1-5 year survivors (data not shown). The long-term dysphagic status was not yet known for patients on active treatment. A total of 17 participants were eligible and approached to participate in the sensor study. Two patients refused, both being White and male: 1 patient was aged 76 years and had been diagnosed with late-stage oropharyngeal cancer 2 years prior and the other was aged 23 years and was in the third week of radiation for late-stage oropharynx cancer (data not shown). A total of 15 (83%) patients agreed to participate and gave informed consent. One participant developed an acute illness episode the following day and was, therefore, unable to complete the sensor test, leaving 14 participants who completed user testing (Figure 5). Study 3’s sample was primarily male (12/14, 86%) and non-Hispanic White (12/14, 86%) with an average age of 62 years (Table 2).

As with the previous studies, the majority of patients (12/14, 86%) indicated willingness to wear the sensor for 9 months during the first year post radiation. Wanting to help future patients detect developing dysphagia and wanting to help MD Anderson research were the most prevalent reasons for willingness to wear the sensor (Table 3). As in study 2, the most oft-cited reasons for unwillingness were that of study burden (lengthy testing period and increase in daily responsibilities; Table 3). Patients’ opinions regarding the helpfulness of haptic feedback were obtained for 13 of the 14 participants. All 13 participants thought it would be helpful for the sensor to vibrate when placed in the correct spot on the neck (mean 1.85, SD 0.38) as well as when swallowing exercises were performed correctly (mean 2.0, SD 0.00; Table 4). A total of 11 (85%) participants felt it would be helpful for the sensor to give haptic feedback of swallow quality during at-home testing (mean 1.5, SD 0.88; Table 4).

Discussion
Principal Findings
To our knowledge, this is the first study to assess evaluations from patients with head and neck cancer of a wearable throat sensor in clinical settings with separate cohorts at varying time points along their treatment trajectory. Across all studies, the overall willingness to wear the sensor for 9 months during the first year after radiation was high and the perceived need was rated highly. However, study 1’s results should be interpreted with caution since the participation rate was 59%, with non-Hispanic and higher income/education patients more likely to complete the questionnaire. While study 2 and 3 used convenience samples for user testing, accrual rates were high (88%), even for those undergoing active treatment at the time of approach.

Direct comparison of our results with other works is not possible since the vast majority of published data regarding wearable devices equipped with mechanical, optical, biochemical, electrical, or acoustic sensors are pilot studies conducted with graduate students in a laboratory under highly controlled conditions [60-64]. While it did not test actual user engagement over repeated time points, it did gather patients’ opinions about the likelihood that they would wear the sensor for a period of several months. This question was asked in study 1 for patients who were only exposed to a photo of the proposed sensor, whereas patients and survivors in study 2 were asked this question after wearing the actual sensor while swallowing boluses of varying textures in a controlled setting. When searching for comparable studies that address extended user engagement with health technologies, the extant literature is limited to nonsensor research with mobile websites or apps [65] and to real-world studies of fitness tracker abandonment rates in healthy adults; these studies tend to describe a steep decline in user engagement over time. It is possible that our high rates of expressed willingness to wear the sensor for 9 months is due to the perceived usefulness of this device for this highly specialized problem.

Since the majority of participants (137/166, 83%) expressed willingness to wear the sensor for 9 months, data from those participants who were unwilling provided valuable insight into the potential barriers to its long-term use. Across all three studies, nearly 86% (24/28) of the unwilling participants perceived the 9-month testing period as too long. The second-most prevalent reason, that the sensor’s appearance would provoke unwanted attention, was endorsed by 56% (15/27) of the unwilling participants. The third-most frequent
reason was an unwanted increase in daily responsibilities (16/26, 62%). This was also borne out by spontaneous comments in study 3, when nearly all 14 patients communicated a preference for a more streamlined one-step application process, rather than the separate applications for the strain sensor and surface EMG electrodes. On the other hand, several of the unwilling participants were much more willing to wear the sensor for 9 months if the sensor could provide individual dysphagic risk feedback and were made more unobtrusive in appearance (Figure 3). These findings are consistent with other mHealth reports citing multiple aspects of participant burden [48] and social implications of the technology’s appearance [66] as being relevant constructs to user engagement.

Bidirectional Communication

Our data confirmed two other persuasive design principles: the desire for bidirectional communication (dialog support) with a team of clinical experts (system credibility). In all three studies, a large proportion of patients endorsed the rationale for the sensor (study 1: 115/138, 83.5%; study 2: 10/14, 71.4%; study 3: 12/14, 85.7%); ie, that sensor data be processed and sent back with contextual explanations of their risk of dysphagia development). Furthermore, of the three proposed persuasive design features, feedback about dysphagia risk had the greatest impact in increasing willingness among all participants (Table 4). These findings point to the importance of fostering a sense of connectedness and reassurance between the user and the technology so that patients’ association between their own health behaviors and subsequent health outcomes can be continually reinforced [42]. Future plans for implementation include data visualization of near-time individualized risk for dysphagia in the form of an app that can be linked with the throat sensor. When asked about direct haptic communication with the sensor itself, patients in study 3 rated haptics as helpful, especially when unsure about correct placement on the throat and whether preventive exercises were being done correctly (Table 4). One patient commented that he was never really sure if he was performing the exercises correctly at home and was “just winging it.”

Sensor and Adherence to Exercises

The majority of participants (97/119, 82%) agreed that the sensor would serve as a reminder for them to do their speech pathology swallowing exercises. While the main goal of the sensor is to provide earlier detection of radiation-associated dysphagia, reminding patients to complete their swallowing exercises at home to counteract the development of dysphagia could be an additional benefit to this developing technology. Since personalized risk information is generally not sufficient in itself to increase exercise adherence per se [67], further user-centered testing would be needed to assess preferred modes of sensor feedback (eg, within an app or coupled with virtual coaching) [68].

Limitations

Our study was conducted solely with survivors and patients attending clinical visits at MD Anderson, which generally requires high-quality insurance for access. Generalizability of our results are further limited by examining the demographic patterns among respondents versus nonrespondents in study 1. A total of 38% (21/55) of the eligible survivors did not complete the questionnaire despite repeated contact by the study team; nonresponders were significantly more likely to be Hispanic (P<.003), without a bachelor’s degree (P=.02), and of lower annual household income compared to respondents (P=.007). This is consistent with Rising et al’s [69] recent analysis of National Cancer Institute’s 2018 Health Information National Trends (HINTS) population survey data showing that nonusers of personal mHealth technologies were more likely to be older than 65 years and have lower incomes. Given the challenge of sustaining patient engagement in mHealth technology, future research should target these patients who fit within the above demographic profiles. Finally, the sample sizes for study 2’s and 3’s on-patient testing were constrained by the need to complete all testing within 1-week periods, as the sensors were applied/tested by visiting engineers and not MD Anderson research staff. It is conceivable that larger sample sizes might have produced a wider variation in response to the sensor’s features and perceived usefulness.

Conclusion

Large proportions of non-Hispanic well-educated patients with high-quality insurance and above-average incomes were receptive to the idea of wearing a personalized risk sensor for an extended period during the first year after radiation. User ratings of discomfort and difficulty were minimal; however, a significant minority of patients expressed concern with various aspects of the sensor’s burden and its appearance.

Acknowledgments

We would like to acknowledge our patients who participated in the study and Evalyne W Kamunyo for her assistance with study 1.

The primary funding for this study was provided by the National Institute of Dental Craniofacial Research (NIDCR DE019141). DJ Lipomi acknowledges support from the National Science Foundation (grant CBET-2223566). Support was provided, in part, by the Assessment, Intervention and Measurement Shared Resource through a Cancer Center Support Grant (CA16672; principal investigator: P Pisters, MD Anderson Cancer Center), from the National Cancer Institute, from the National Institutes of Health, and through the Duncan Family Institute for Cancer Prevention and Risk Assessment.

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No 898571, awarded to RB.
Data Availability
The data that support the findings of this study are available from the corresponding author EHS upon reasonable request.

Conflicts of Interest
None declared.

References


52. Turesson C, Liedberg G, Björk M. Development of a digital support application with evidence-based content for sustainable return to work for persons with chronic pain and their employers: user-centered agile design approach. JMIR Hum Factors 2021 Jan 07;9(1):e23402 [FREE Full text] [doi: 10.2196/mhealth.9832] [Medline: 33410758]


Abbreviations

- **EMG**: electromyography
- **HINTS**: Health Information National Trends
- **IMRT**: intensity-modulated radiation therapy
- **MBS**: modified barium swallow
- **mHealth**: mobile health
Toxic Relationships Described by People With Breast Cancer on Reddit: Topic Modeling Study

Cara Anne Davidson¹, MSc; Richard Booth², PhD; Kimberley Teresa Jackson², PhD; Tara Mantler³, PhD

¹Department of Health and Rehabilitation Sciences, Faculty of Health Sciences, Western University, London, ON, Canada
²Arthur Labatt Family School of Nursing, Faculty of Health Sciences, Western University, London, ON, Canada
³School of Health Studies, Faculty of Health Sciences, Western University, London, ON, Canada

Corresponding Author:
Cara Anne Davidson, MSc
Department of Health and Rehabilitation Sciences
Faculty of Health Sciences
Western University
1151 Richmond St
London, ON, N6A 3K7
Canada
Phone: 1 519 661 2111 ext 85541
Email: cda_vid53@uwow.ca

Abstract

Background: Social support is essential to promoting optimal health outcomes for women with breast cancer. However, an estimated 12% of women with breast cancer simultaneously experience intimate partner violence (IPV; physical, psychological, or sexual abuse by an intimate partner). Women who experience IPV during breast cancer may lack traditional social support, and thus seek out alternative sources of support. Online community forums, such as Reddit, can provide accessible social connections within breast cancer–specific communities. However, it is largely unknown how women with breast cancer use Reddit to describe and seek support for experiences of IPV.

Objective: This study aims to explore how patients with breast cancer describe toxic relationships with their partners and immediate family members on Reddit.

Methods: This exploratory, cross-sectional, topic-modeling study analyzed textual data from 96 users in the r/breastcancer subreddit in February 2023. The meaning extraction method, inclusive of principal component analysis, was used to identify underlying components. Components were subjected to sentiment analysis and summative content analysis with emergent categorical development to articulate themes.

Results: Seven themes emerged related to toxic relationships: (1) contextualizing storytelling with lymph nodes, (2) toxic behavior and venting emotions, (3) abandonment and abuse following diagnosis, (4) toxic relationships and social-related fears, (5) inner strength and navigating breast cancer over time, (6) assessing social relationships and interactions, and (7) community advice and support. Toxic relationships were commonly characterized by isolation, abandonment, and emotional abuse, which had profound emotional consequences for patients. Reddit facilitated anonymous venting about toxic relationships that helped patients cope with intense feelings and stress. Exchanging advice and support about navigating toxic relationships during breast cancer were core functions of the r/breastcancer community.

Conclusions: Findings emphasized the value of Reddit as a source of social support for patients with breast cancer experiencing toxic relationships. Clinicians who understand that many patients with breast cancer experience toxic relationships and considerable psychological sequelae are better prepared to support their patients’ holistic well-being. Further investigation of Reddit as a possible resource for advice, information, and support has the potential to help inform clinical practice and subsequently, patient health outcomes.

(JMIR Cancer 2024;10:e48860) doi:10.2196/48860

KEYWORDS
breast cancer; intimate partner violence; meaning extraction method; Reddit; sentiment analysis; social media; social support; toxic relationships; topic modelling
Introduction

Breast cancer has a way of making existing cracks in relationships even wider. Just like water will fill a crack in the road, freeze, and create a larger gap, breast cancer tends to permeate all parts of our lives and distance us from people with whom we have troubled relationships. [Original poster #85]

Projected rates of breast cancer in Canada have remained consistent over the past 5 years, with estimates that approximately 1 in 8 women will develop breast cancer in their lifetime and breast cancer will account for 25% of all new cancer cases [1-3]. Women’s experiences of breast cancer are influenced by the social determinants of health, particularly their social environment [4,5]. Among patients with breast cancer, strong social relationships have been found to act as a buffer to stress [6] and help to improve treatment effectiveness, psychological functioning, coping, survival, and quality of life, as well as prevent cancer recurrence [7-9]. Conversely, weak or nonexistent social relationships have been broadly linked to long-term psychological distress [10] and an increased risk of breast cancer progression, recurrence, and mortality [11,12]. However, there is a need for research that explores connections between social relationships and breast cancer outcomes among diverse populations and social contexts.

Intimate partners (eg, spouses and significant others) and immediate family members (eg, parents and siblings) are perceived as the most important social supports for patients with breast cancer [13,14], as they provide essential social-emotional, tangible, affection, and positive social interaction support [15]. For example, partners commonly serve as the primary caregivers of patients with cancer [16]. However, not all social relationships are supportive [17]. Patients who experience intimate partner violence (IPV) may face a lack of support due to the abusive behaviors of their partner [18]. IPV, understood as physical, psychological, or sexual abuse within the context of coercive control by an intimate partner [19], concurrently affects an estimated 12.5% of patients with breast cancer [20]—and this is likely to be an underestimation given underreporting of IPV [21]. Similarly, patients may be negatively affected by an unsupportive (but not necessarily abusive) partner [22], as well as abusive or unsupportive family members [23,24]. Aside from the patients themselves, immediate female family members are often most affected by a breast cancer diagnosis; unsupportive reactions often include being in denial about the diagnosis and abandoning the patient [25].

Toxic relationships are characterized by conflict, competition, undermining, disrespect, and a lack of cohesiveness [26]. Toxic relationships encompass unsupportive and abusive dynamics in both romantic (eg, a partner) and platonic (eg, a family member) contexts and are associated with emotional distress [26], which imparts numerous downstream mental and physical health consequences [27]. To compensate for unmet support needs, patients with breast cancer may expand their social networks via the internet, including social media and online forums. A popular means of accessing information and support related to breast cancer awareness, literacy, and treatment [29-33]. The use of online breast cancer forums grew exponentially between 2006 and 2010, growing from an estimated 282,000 new posts per year to over 1,270,000 new posts per year [34] and continues to increase over a decade later [35,36]. Despite data availability and the potential for knowledge advancement [33], research on patient social media use, particularly in the context of toxic relationships, is underexplored.

Reddit, the world’s third most popular social media platform, is an online forum dedicated to community-building, news dissemination, and discussion facilitation [37]. The Reddit platform consists of topic-specific subreddits (ie, forums), where all content is user-generated. Users subscribe to subreddits that interest them to see more related content. Users can post content, as well as comment and vote on others’ content. To join Reddit, users create a username and password—no identifiable information is required. Reddit’s capacity for anonymous participation and long-form, conversational content makes the platform a rich source of self-reported textual data [38]. The Reddit platform includes breast cancer–specific spaces that offer access to psychosocial support (eg, r/breastcancer), presenting a unique and valuable opportunity to explore how patients with breast cancer navigate toxic relationships after diagnosis. Previous research has provided preliminary insights into how patients with breast cancer use Reddit [39], but there is a notable gap in the literature regarding how patients with breast cancer describe toxic relationships and their psychosocial impacts on Reddit. Studying social media data has the potential to generate significant advances in knowledge [33], which can inform improvements to psychosocial support for patients with breast cancer experiencing toxic relationships and enhance care providers’ ability to promote patient well-being. Accordingly, this study sought to explore how people with breast cancer describe toxic relationships with their partners and immediate family members on Reddit.

Methods

Design

This exploratory, cross-sectional, topic-modeling study was conducted from December 2022 to February 2023 and aimed to explore how patients with breast cancer describe toxic relationships with their partners and immediate family members on Reddit. As of February 2023, the public r/breastcancer subreddit, established in 2011, included 13,900 subscribers and self-identified as a support and information group for people who have been diagnosed with breast cancer and their caregivers and loved ones. While Reddit generally attracts young White men of high socioeconomic status [38], demographics vary by subreddit and r/breastcancer is hypothesized to be largely composed of women [40].

Ethical Considerations

This study was deemed exempt from oversight by the author’s institutional ethics review board because all data were gathered from the public domain (per Article 2.2 of the Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans). The subreddit at the center of this study was public at the time of data collection and writing, meaning that any person could...
access its content at any time. It was therefore determined that r/breastcancer users had no expectation of privacy, negating the need for oversight by an ethics review board.

**Data Collection**

This subreddit was scraped for textual data from posts and comments using the Python Reddit application programming interface wrapper. No date limits were imposed. An iterative approach to keyword-based searching extracted posts (n=187) related to toxic relationships with partners and immediate family members. Two keyword strings were combined to scrape data: String 1 included words associated with a toxic relationship (eg, narcissist, boundaries, abuse, violence, assault,셰터, affair, divorce, toxic, abandon, and manipulate) and string 2 included words that identified people of interest in the immediate family of the user (eg, abuser, spouse, partner, marriage, significant other, parent, and sibling). To be scraped, posts were required to include a minimum of 1 keyword from both string 1 and string 2 (see Textbox 1).

**Textbox 1. Keywords included in the final iteration of the search strategy.

<table>
<thead>
<tr>
<th>String 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>narcissist; boundaries; abuse; abusive; abusing; abused; violent; violence; assault; assaulting; assaulted; harass; harassed; harassing; lie; lied; neglect; unsupportive; not supportive; not supporting; no support; cheating; cheated; cheating; affair; divorce; divorcing; break up; breaking up; broke up; toxic; abandoned; manipulate; manipulated; emotionally unavailable; disown; alone; selfish; strained</td>
</tr>
<tr>
<td>String 2</td>
</tr>
<tr>
<td>abuser; husband; wife; partner; hubby; marriage; girlfriend; boyfriend; gf; bf; SO; significant other; spouse; mom; mother; mum; dad; father; parent; parents; sibling; sis; sister; brother</td>
</tr>
</tbody>
</table>

The scraped posts were then screened for eligibility by one of the authors (CAD), such that posts were ineligible if they addressed anyone other than a partner or immediate family member, were of an administrative nature posted by a moderator, were posted by a user who did not have breast cancer, aimed to exclusively seek or share medical information, or described toxic relationships outside of the context of breast cancer. After screening, 36 posts were eligible for inclusion. Eligible posts were scraped for comments (n=601), of which 98 were relevant (as determined by CAD using the eligibility criteria described above used to filter posts). Textual data were compiled into packets, where 1 packet represented the total relevant contributions (ie, posts and comments) from a single user, with an average of 260 words per packet. The final data corpus included 96 unique users with 36 posts and 98 comments (see Figure 1).

**Figure 1.** Final data corpus diagram.

Reddit users have no expectation of privacy in public subreddits and have agreed to the platform’s end user license agreement; all Reddit user content is subject to use by third parties at any time [41]. However, recommended ethical practices aim to protect participant privacy by censoring usernames and avoiding direct quotes through exclusion or paraphrasing to prevent reverse-searching [38,42]. Accordingly, within this study, users were assigned an original poster (OP) number, and reported quotes were reworded to convey their original meaning and style but protect the OP’s identity. For example (fictitious), “My
mom has never bothered checking on me” could become “My mother doesn’t ever ask how I’m doing.”

**Analysis**

**Multistaged Approach**

A 2021 systematic analysis by Proferes et al [39] identified that computational-driven textual analysis (which includes topic modeling) was the primary means of knowledge generation using Reddit data. However, the authors also identified that such analyses are enhanced by the addition of qualitative and mixed methods analyses that account for contextual details [39]. Accordingly, the data corpus was subjected to a 3-stage, mixed methods analysis that used (1) the meaning extraction method (MEM), (2) qualitative sentiment analysis, and (3) summative content analysis (Figure 2).

**Figure 2.** The 3 stages of analysis. PC: principal component.

---

**Stage 1: Meaning Extraction Method**

The MEM is a form of topic modeling useful for social media data exploration [43] and for generating large sample sizes of participants that are traditionally difficult to recruit [38]. Within other breast cancer–related research using Reddit data, the MEM has been described as a cost-effective means of identifying common themes described by patients [39]. The results of this method have been found to be similar in content and utility to those of traditional research methods in this domain (focus groups) [39].

The MEM identifies word clusters that co-occur in a data corpus, providing an efficient means of extracting meaningful patterns in language within high volumes of natural language data [38,43]. The Meaning Extraction Helper developed by Boyd [44] was used to analyze the textual packet data corpus, inclusive of the removal of common closed and open class words (<7.5%) and content word retention (≥5%), producing a binary output of each retained content word per OP (eg, 0=absent and 1=present). Boyd [45] also developed an open-access script for the R open-access statistical software (R Foundation for Statistical Computing), which was adapted for a principal component analysis (using a varimax rotation [43]). This produced a 9-component model that was considered acceptable ($\chi^2=3357.40, df=304, P<2.2e–16$, and KMO=0.538 [43]). Using a scree plot analysis, components 1 to 7 were retained. The 7 retained components explained 84.16% of the variance—a high proportion for a natural language application [43]. A high loading threshold of 0.50 was imposed on the content words within each component to promote thematic clarity and reduce cross-loading. The final components with refined content words were considered sufficiently strong (≥3 content words per component [46]).

**Stage 2: Qualitative Sentiment Analysis**

Qualitative sentiment analysis aims to assess the affective valence of components and their content words [47]. Modern qualitative sentiment analysis (ie, internet-based) is an increasingly popular and effective method of interpreting user-generated social media content [48]. Using the *syuzhet* R package [49] and Afinn sentiment lexicon of –5 (negative sentiment) to 5 (positive sentiment [50]), a total model and 7 component-specific sentiment scores were computed based on content words. The content word *tchp* was changed to *combination drug cancer therapy* for the algorithm because it cannot assess acronyms.

**Stage 3: Summative Content Analysis**

Summative content analysis with emergent categorical development was used to articulate patterns and themes within the textual data packets for each of the 7 refined components [51]. Component categories (referred to as themes) were
inductively developed to describe their overall message, inclusive of the use of sentiment scores to contextualize positioning. The 6-step approach to trustworthy thematic analysis by Nowell et al [52], rooted in the trustworthiness theory of Lincoln and Guba [53,54] was adopted.

**Summary of Analysis**

Quantitative topic modeling (with MEM) was combined with qualitative sentiment and content analysis to produce a comprehensive analytical framework capable of providing an overall interpretive assessment of the data corpus. The r/breastcancer subreddit includes thousands of textual data sources, requiring the combination of complex methods to efficiently target and isolate meaningful, manageable patterns from the large volume of natural language data [38,43]. The MEM is a computational method specifically developed to facilitate efficient filtering of large textual data sets, however, a second stage of qualitative or mixed methods–based analysis is recommended to facilitate deeper exploration and interpretation in context [39]. Accordingly, sentiment analysis was applied within MEM-generated principal components to facilitate the assessment and incorporation of considerations of user’s emotions and situational contexts. Following MEM and sentiment analysis, content analysis was used to deeply explore principal components through the lens of their socioemotional contexts to enrich interpretation and understanding. In sum, this combined mixed methods framework aimed to produce holistic, contextualized insights from MEM-generated categories, which is well-suited to complex, dynamic social media data.

**Results**

**Overview of Themes**

Seven distinct but related themes emerged from descriptions of toxic relationships by patients with breast cancer on Reddit, presented in order of explained variance proportion (highest to lowest) as follows: (1) contextualizing storytelling with lymph nodes, (2) toxic behavior and venting emotions, (3) abandonment and abuse following diagnosis, (4) toxic relationships and social-related fears, (5) inner strength and navigating breast cancer over time, (6) assessing social relationships and interactions, and (7) community advice and support. The overall corpus sentiment score was –4, indicative of very negative sentiment. Theme-specific sentiment scores \([\bar{s}]\) reflect the average valence of retained content words within each component.

**Theme 1: Contextualizing Storytelling With Lymph Nodes**

*I’ll have to get my lymph nodes removed next, among other things. Treatment is lonely and miserable.* [OP 2]

The first theme was classified as neutral \([\bar{s}]=0.00\) and included lymph, node, and pick as key content words. Lymph nodes functioned as context indicators in users’ stories about toxic relationships to highlight their temporality within cancer treatment. For example, one user was undergoing chemotherapy while navigating a toxic relationship with their mother. This OP prefaced their post by sharing, “After a lot of treatment, my cancer went from grade 3 to grade 1. My lymph nodes shrunk as well” (OP 66).

They then went on to disclose unsupportive behavior from their mother, stating, “My mom doesn’t think I’m capable of making my own decisions—but I am. I’ve picked excellent physicians and made it to all of my appointments” (OP 66).

**Theme 2: Toxic Behavior and Venting Emotions**

*I’m going to vent because I think it’s better to write than to cry...* [OP 65]

The second theme was classified as neutral \([\bar{s}]=0.00\) and described toxic relationships that the user experienced a strong emotional reaction to, which prompted them to vent their emotions on Reddit. Key content words included boundary, effort, vent, upset, and stress. Users reported a variety of toxic behaviors, such as boundary violations and disrespectful or abusive actions. Venting was commonly used to cope with powerful negative emotions associated with toxic relationships.

Users felt unsupported when their partners or families reacted to their diagnosis by becoming detached or distressed to the extent of relying on the patient for support. To illustrate, one user expressed disappointment in their father’s silence after diagnosis, stating, “My dad isn’t there for me. I guess I shouldn’t be surprised, he’s always been like this” (OP 45).

Other users were frustrated with bearing the emotional burden for others regarding their cancer. For example, one OP resented their husband for expecting them to manage his emotions, sharing, “I did my best to explain that I needed him to be my rock. He got upset... he wanted us to be mutually supportive. But he doesn’t have cancer... I do!” (OP 62).

Some OPs described being disrespected and emotionally abused following their diagnosis. For instance, one OP shared that their partner told them, “Lately, you aren’t sexually desirable to me without your natural breasts. I miss them and how they felt... probably even more than you do” (OP 65).

Similarly, another OP disclosed experiencing emotional and verbal abuse from their partner both before and after their breast cancer diagnosis. This OP shared feeling extremely upset that just 2 weeks after their diagnosis, their partner asked them, “How long are you going to pull the breast cancer card?” (OP 85).

Toxic relationships described within this theme were strongly associated with venting, that is, posting negative, emotionally charged content. For example, an OP trying to cope with being isolated by their family prefaced their story by writing, “Heads up that this is a massive, sad vent post. Sorry but I feel like I need to shout into the void” (OP 34).

**Theme 3: Abandonment and Abuse Following Diagnosis**

*Anyone else dealing with an emotionally abusive spouse before and during cancer? I’m trying to get away and he’s being awful.* [OP 40]
Theme 3 was classified as slightly negative (⋯ = −1.00) and captured how patients in toxic relationships were abandoned or emotionally abused by their partners following their diagnosis. Key content words included devastate, experience, and abuse. Patients who navigated abandonment or abuse concurrently with a breast cancer diagnosis reported feeling emotionally devastated.

Abandonment was especially common after disclosing a breast cancer diagnosis. For example, one OP shared that their husband abandoned them on the way home from their diagnosis appointment, stating, “He said he won’t look after the kids and plans on leaving” (OP 9). Other users were abandoned as treatment began. Many users who shared stories of abandonment described emotional whiplash, characterized by a sudden, unexpected transition from feeling secure in their relationship to feeling betrayed following abandonment. As illustrated by one user, “He made me feel cared for, loved, and safe... until I said I was considering a mastectomy. Then he shut me out” (OP 12).

The emotional impacts of betrayal were devastating. An OP whose long-term partner unexpectedly broke their promise to stick by them during treatment shared, “I am completely devastated. I am infuriated. He and my body betrayed me. I am so furious” (OP 86).

Of partners who stayed following a diagnosis, many subjected the patient to emotional abuse. One OP was told that they deserved their cancer, recounting, “He used my cancer against me by saying I got it because I’m weak and that’s just natural selection at work. He told me not to bother with treatment and to just let nature run its course” (OP 76).

Other experiences involved infidelity, threats of child abandonment, accusations of faking symptoms, and coercion in treatment choices. Emotional abuse was repeatedly described as devastating. For example, an OP whose spouse had been emotionally abusive for years posted, “What can I do to stop feeling devastated that my husband feels I should be punished all the time?” (OP 40).

Theme 4: Toxic Relationships and Social-Related Fears

Do any of you also feel like the emotional consequences of breast cancer are almost worse to deal with than the physical? [OP 66]

The fourth theme was classified as slightly negative (⋯ = −0.75) and focused on social-related fears associated with breast cancer. Key content words included biopsy, tchp, and scare. Patients’ fear stemmed from anticipating or experiencing a negative reaction to their breast cancer by a toxic family member or partner. For example, an OP who disclosed a toxic family shared dreading their reaction to their cancer, expressing, “The fear of how my family will react to my breast cancer diagnosis is nearly as overwhelming as the actual diagnosis” (OP 81).

Other users felt scared because they had already experienced an unsupportive reaction by a toxic family member or partner to their cancer. For example, one OP felt scared and hopeless after being gaslit by their partner about their diagnosis, sharing, “He was trying to tell me that my breast cancer was all in my head, despite having seen my biopsy results and meeting with multiple members of my medical team” (OP 3). Similarly, an OP whose family neglected to support them after learning of their diagnosis expressed, “My family doesn’t care about me or my breast cancer. It makes me feel scared and alone” (OP 34).

Theme 5: Inner Strength and Navigating Breast Cancer Over Time

I thought to myself that if my cancer ever came back, I’d rather deal with it alone than with a person like that. [OP 32]

Theme 5 was classified as slightly positive (⋯ = 1.00) and highlighted how breast cancer was disruptive to the lives of patients. Key content words included future, matter, and strength. Users described how health and social adversity influenced their inner strength. Toxic relationships that emerged after diagnosis were especially trying for patients. For instance, one OP expected their partner’s support as they began cancer treatment (as their partner had promised). However, the OP’s partner abruptly took back their commitment, leaving the OP to navigate cancer alone: “They sent me a message the next day and said they don’t want anything to do with me” (OP 12).

Inner strength emerged as a dynamic construct that was both challenged by experiencing a breast cancer diagnosis and toxic relationships and enhanced by surviving these adverse experiences. Many users believed that surviving breast cancer concurrently with exposure to toxic relationships was a testament to their inner strength. For example, one OP attributed their inner strength to recovering from breast cancer while navigating a lack of empathy and support from their spouse. This OP stated, “I feel 100% confident that I am a strong, intelligent woman who can face almost anything” (OP 64), while sharing that they had received a new cancer diagnosis. Inner strength also enabled users to regain a sense of control over how they were going to navigate living with a breast cancer diagnosis. For example, an OP who was abandoned by their partner after being diagnosed stated, “I finally felt strong enough to delete his contact information because I couldn’t stop myself from calling him—it was the best choice I could have made” (OP 47).

Theme 6: Assessing Social Relationships and Interactions

I am immensely grateful for you all for helping me navigate a chaotic and frustrating moment. [OP 85]

The sixth theme was classified as marginally positive (⋯ = 0.20) and described how OPs assessed their social relationships and interactions. Key content words included conversation, response, listen, regret, and grateful. Users assessed the quality of social support from family based on whether they felt judged, subjected to toxic positivity, or made to listen to unsolicited advice. For example, an OP with an emotionally unsupportive family shared, “I think a lot of family think it’s helpful when they shove positivity down our throats. What we really need is support and someone to listen without trying to solve all our problems” (OP 82).
For some users, responses to breast cancer unveiled toxic relationships that they regretted having to face. For example, an OP with unsupportive parents shared, “I regret that my breast cancer forced me to confront that my parents never have and still don’t support me how I need them to” (OP 4). However, OPs who discovered both toxic and supportive relationships during cancer expressed gratitude for the sources of support they did have. As one OP stated, “Sometimes I get jealous of people whose parents love and support them, but then I remember the rest of my friends and family who showed up for me when I needed them, and I’m grateful” (OP 32). The subreddit community was repeatedly praised by users because it was such a valuable source of support. For instance, one OP shared, “I am endlessly grateful for the knowledge and resilience of this community” (OP 53).

**Theme 7: Community Advice and Support**

I know what it feels like to be abandoned. I could tell you all the red flags in a man’s behavior... but just trust me—it’s better to be alone. You dodged a MASSIVE bullet. A person who lacks compassion about your breast cancer is NOT a good life partner. Please message me if you need someone to vent to. I really do understand...and you’ve got this. [OP 89]

The seventh theme was classified as marginally negative (r = -0.17) and characterized a core function of r/breastcancer: providing advice and support. Key content words included money, quit, and follow. The subreddit facilitated advice regarding various topics, especially related to navigating financial matters and treatment options in the context of a toxic relationship.

Numerous users offered money-related advice to OPs facing difficult financial situations because of toxic relationships. Situations included financial coercion, exploitation, and manipulation following cancer disclosure and managing finances during separation from a toxic partner. For instance, one OP was abandoned by their partner during a joint real estate purchase. A community member with self-professed real estate expertise strongly advised the OP against continuing with the investment, writing, “I’m begging you... please DO NOT sign anything else! Lose your money... that’s not important... please do not continue with this purchase” (OP 89).

Members also counseled OPs about postmastectomy reconstruction by offering advice on how to reduce social pressure and prioritize personal preferences. For example, one OP shared how they resisted their partner’s pressure to follow reconstruction, stating, “I made him look at photos of reconstruction to show him that it’s not a free boob job and can be ugly. He changed his tune real quick” (OP 39).

Members who were ultimately pressured into reconstruction strongly encouraged OPs to follow their instincts. For example, one member who was coerced into reconstruction by their husband advised, “I constantly wish I went flat instead. If I had to do it again I would listen to my gut and go flat” (OP 8).

Similarly, it was common to share advice about treatment adherence. Many OPs struggling with a lack of support expressed wanting to quit treatment. While members empathized with users and understood their feelings, they ultimately encouraged continuing. For example, one OP shared, “I’m just sick of this. I’m pretty sure I’m done with it all” (OP 34).

The community offered empathy, such as, “When I was in the middle of your treatment, I was frustrated too and tried to quit every week” (OP 7), as well as advice, for example, “Don’t stop treatment without a good reason. It’s a gift in spite of tough side effects because it keeps us alive” (OP 48).

**Discussion**

**Principal Results**

This study explored the use of the r/breastcancer subreddit by patients to describe toxic relationships with their partners and immediate family members. Themes highlighted patients’ lived experiences of toxic relationships, emotional impacts, and support from the subreddit community. A key finding was that many people with breast cancer sought out the r/breastcancer subreddit to share their experiences of toxic relationships, often including descriptions of abandonment, isolation, and emotional abuse within this context. Further, this study presented compelling evidence that toxic relationships impart profound emotional consequences for patients and that some patients cope with these strong emotions through online venting. This work also emphasized the value of online communities like Reddit as alternative, complementary sources of support for patients experiencing toxic relationships.

**Comparison With Prior Work**

**Abandonment and Betrayal as Common Experiences**

These findings suggest that abandonment is a common experience for patients with breast cancer following diagnosis. Prior research has lacked consensus regarding the risk of abandonment among patients with breast cancer after diagnosis [55,56]. Generally, however, women are more likely to be abandoned by a partner after being diagnosed with a serious medical illness [57]. Further, distancing is the most prevalent unsupportive response experienced by a patient following their breast cancer diagnosis [23]. Fears and feelings of abandonment following diagnosis are also well-documented within breast cancer research [58-60]. Given this understanding, and considering that Reddit data can be regarded as an authentic representation of user experiences [61], it is reasonable to conclude that these findings are suggestive of an increased risk of abandonment for patients with breast cancer.

A novel finding was the occurrence of emotional whiplash, where a patient was initially promised support by their partner but was later abandoned unexpectedly. The emotional transition from security to betrayal was repeatedly reported as devastating. There is limited research describing betrayal in the context of abandonment and breast cancer, but it is known that feelings of betrayal in this context can reduce the desire for future relationships [62]. Broadly, the loss, disruption, and deterioration of social ties are some of the most stressful experiences a patient with cancer can face [6], which makes abandonment a serious risk factor for reduced mental health [63]. Comprehensive cancer care entails stress-reducing psychosocial interventions [63], but...
a limited understanding of the psychological effects of betrayal hinders clinicians’ ability to optimally manage abandonment-related stress.

**Anonymous Venting Enables Disclosure of Toxic Relationships**

The central role of venting within the r/breastcancer community highlighted the unique socioemotional needs of patients with breast cancer in the context of toxic relationships. Toxic relationships impart emotional consequences that can be difficult to navigate and cope with [26]. Venting is a disinhibitory, emotion-focused strategy for coping with stress [64,65]. Venting can be considered a form of expressive writing, that is, writing that describes a deeply personal experience [66], which is well-evidenced to facilitate coping with psychological distress [67]. Online venting was consistently described as cathartic among patients in this study, aligning with prior evidence of patients with breast cancer seeking support in online communities during periods of stress [68] and perceiving reduced stress after they vent online [69]. Further, patients with breast cancer who self-manage their emotions by narrating their experiences are known to experience strong psychological benefits [70].

It might be expected that the stigma attached to breast cancer and toxic relationships would hinder disclosure [58,71], however, seeking out group-oriented support is reportedly most common for diseases considered stigmatizing [72]. The latter position is consistent with this study, as venting posts often included stigmatized thoughts and feelings (eg, wanting to ‘give in’ to cancer or discussing abuse without wanting to leave the relationship). Further, it appeared that Reddit’s capacity for anonymity created a sense of safety that made patients comfortable disclosing information considered stigmatizing, which is consistent with existing evidence [73]. Overall, patients appeared to perceive anonymous venting via Reddit as an effective, safe strategy for coping with stress from toxic relationships. Interventions that aim to promote coping among this patient population would likely benefit from integrating anonymity to encourage uninhibited self-expression.

**Advice About Navigating Toxic Relationships**

Validating the feelings of other users, as well as soliciting and providing advice regarding toxic relationships, were core activities within r/breastcancer. It was previously known that participation in online forums contributes to the practical, informative, and emotional empowerment of patients with breast cancer [74]. However, this study uniquely identified that community members on Reddit often urged OPs to leave or go against the wishes of their abusive partner. While well-intentioned, this advice may not always be safe or practical. Leaving an abusive partner can be the most dangerous time in the relationship due to an increased risk of retaliation [75]. Similarly, acting in a manner that might antagonize an abuser can initiate or escalate relationship discord and consequently increase the risk of violence [76]. Furthermore, patients who depend on an abusive partner (eg, for caregiving, access to health insurance, and transportation to appointments [77]) may be unable to leave or risk the relationship by acting defiantly [78]. Resultantly, relationship advice received on Reddit by patients with abusive partners may have been incompatible with their reality or suboptimal in promoting their safety.

This indicates a knowledge gap concerning safety planning within r/breastcancer; safety planning can be understood as the development of strategies to reduce the risk of abuse and enhance support [79]. Safety planning is a proven, widely endorsed health promotion intervention that is effective both within an abusive relationship and after leaving [80,81]. Considering the prevalence of abuse among patients with breast cancer [20] and that many seek support in online forums such as Reddit [39], it could be useful to raise awareness of safety planning within r/breastcancer as a health promotion strategy. Further, considering the importance attributed to inner strength by patients in this study, building awareness of strengths-based approaches to safety planning [82] could be particularly useful. For example, community moderators could pin relationship-oriented resources (eg, hotlines and informative websites) as the top comment under posts about challenging, potentially toxic relationships. However, a needs assessment would be best suited to developing an IPV-related intervention considered acceptable and effective within r/breastcancer.

**Clinical Implications**

Psycho-oncology care teams play a critical role in optimizing health outcomes for patients with breast cancer, yet the emotional well-being of patients with cancer is often underreported and underexplored [83]. Patient-reported social media data offers real-time insights into patient experiences and needs which can be beneficial for informing clinical practice [33,83].

Clinicians who understand that many of their patients with breast cancer are negatively affected by toxic relationships are better prepared to support their emotional well-being. Acquiring knowledge about practices and resources that foster coping and inner strength, including venting and safety planning, can contribute to improved patient outcomes.

Some clinicians may be unfamiliar with the advantages of online forums for patients, but recognizing the potential benefits could enhance care [84]. Recommending Reddit as a possible source of advice, information, and support could be a valuable addition to clinical practice for patients navigating breast cancer and toxic relationships. However, because digital literacy is often overlooked in breast cancer care [85], clinicians who concurrently promote digital literacy can empower their patients to access online communities and ultimately, improve their health outcomes.

**Limitations**

There are limitations to this work. First, the analysis was conducted by a single researcher, which may have introduced bias in data interpretation. The analysis also relied heavily on automated methods that may have been inadequate in fully capturing nuance or interpreting context cues in textual data. Second, these data are self-reported, which may have resulted in biased perspectives. While users in this sample self-identified as patients with breast cancer, it was not possible to validate this. These data may have inadvertently included content from online robots or people without breast cancer, and thus may not

https://cancer.jmir.org/2024/1/e48860

JRIM Cancer 2024 | vol. 10 | e48860 | p.76

(page number not for citation purposes)
accurately reflect the experiences of the target population. Additionally, these data were scraped from a single social media platform and may not be representative of the experiences of patients who use other social media platforms, do not use Reddit to discuss their personal lives, or lack access to an internet-enabled device. No demographic information was available to further contextualize findings. It is important to note that these results only relate to experiences of emotional abuse, as physical and sexual abuse were not represented in the data. Furthermore, all participants could write in English, were digitally literate, and had access to the internet, meaning that the findings may not represent the experiences of patients who are nonanglophone or lack technological access or literacy. Caution should be used when applying these findings to other patients with breast cancer.

Conclusions
This study identified that toxic relationships described by patients with breast cancer on Reddit were common and characterized by abandonment, abuse, and unsupportive behaviors. Patients often experienced profound emotional reactions to this form of social stress and anonymous venting on Reddit was described as an effective coping mechanism. Some patients described breast cancer and toxic relationships as adverse experiences that ultimately enhanced their inner strength. Overall, the r/breastcancer community appeared to be a means of exchanging advice, information, and support for patients experiencing toxic relationships. Clinicians who understand that their patients may be negatively affected by toxic relationships are better prepared to support their holistic well-being. Further investigation of Reddit as a possible source of advice, information, and support has the potential to help inform clinical practice and subsequently, improve patient health outcomes.

Data Availability
The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
CAD was involved in the conceptualization, data curation, formal analysis, investigation, methodology, and both original draft preparation and review and editing of the manuscript. RB, KTJ, and TM contributed to the conceptualization, supervision, methodology, and review and editing of the manuscript.

Conflicts of Interest
None declared.

References


44. Boyd RL. MEH: Meaning Extraction Helper (Version 2.3.0). Ryanboid.io. URL: https://www.ryanboid.io/software/meh [accessed 2023-02-10]


Abbreviations

IPV: intimate partner violence  
MEM: meaning extraction method  
OP: original poster
Improving Concordance Between Clinicians With Australian Guidelines for Bowel Cancer Prevention Using a Digital Application: Randomized Controlled Crossover Study

Tsai-Wing Ow, MBBS; Olga Sukocheva, BSc, PhD; Peter Bampton, MBBS; Guruparan Iyngkaran, MBBS, PhD; Christopher K Rayner, MBBS, PhD; Edmund Tse, MBBS, PhD

1Department of Gastroenterology and Hepatology, Royal Adelaide Hospital, Adelaide, Australia
2Faculty of Health and Medical Sciences, University of Adelaide, Adelaide, Australia
3Department of Gastroenterology and Hepatology, Royal Melbourne Hospital, Melbourne, Australia

Corresponding Author:
Tsai-Wing Ow, MBBS
Department of Gastroenterology and Hepatology
Royal Adelaide Hospital
Port Road
Adelaide, 5000
Australia
Phone: 61 70740000
Fax: 61 70746247
Email: tsai-wing.ow@sa.gov.au

Abstract

Background: Australia’s bowel cancer prevention guidelines, following a recent revision, are among the most complex in the world. Detailed decision tables outline screening or surveillance recommendations for 230 case scenarios alongside cessation recommendations for older patients. While these guidelines can help better allocate limited colonoscopy resources, their increasing complexity may limit their adoption and potential benefits. Therefore, tools to support clinicians in navigating these guidelines could be essential for national bowel cancer prevention efforts. Digital applications (DAs) represent a potentially inexpensive and scalable solution but are yet to be tested for this purpose.

Objective: This study aims to assess whether a DA could increase clinician adherence to Australia’s new colorectal cancer screening and surveillance guidelines and determine whether improved usability correlates with greater conformance to guidelines.

Methods: As part of a randomized controlled crossover study, we created a clinical vignette quiz to evaluate the efficacy of a DA in comparison with the standard resource (SR) for making screening and surveillance decisions. Briefings were provided to study participants, which were tailored to their level of familiarity with the guidelines. We measured the adherence of clinicians according to their number of guideline-concordant responses to the scenarios in the quiz using either the DA or the SR. The maximum score was 18, with higher scores indicating improved adherence. We also tested the DA’s usability using the System Usability Scale.

Results: Of 117 participants, 80 were included in the final analysis. Using the SR, the adherence of participants was rated a median (IQR) score of 10 (7.75-13) out of 18. The participants’ adherence improved by 40% (relative risk 1.4, \( P < .001 \)) when using the DA, reaching a median (IQR) score of 14 (12-17) out of 18. The DA was rated highly for usability with a median (IQR) score of 90 (72.5-95) and ranked in the 96th percentile of systems. There was a moderate correlation between the usability of the DA and better adherence (\( r_s = 0.4; P < .001 \)). No differences between the adherence of specialists and nonspecialists were found, either with the SR (10 vs 9; \( P = .47 \)) or with the DA (13 vs 15; \( P = .24 \)). There was no significant association between participants...
who were less adherent with the DA (n=17) and their age (P=.06), experience with decision support tools (P=.51), or academic involvement with a university (P=.39).

Conclusions: DAs can significantly improve the adoption of complex Australian bowel cancer prevention guidelines. As screening and surveillance guidelines become increasingly complex and personalized, these tools will be crucial to help clinicians accurately determine the most appropriate recommendations for their patients. Additional research to understand why some practitioners perform worse with DAs is required. Further improvements in application usability may optimize guideline concordance further.

(JMIR Cancer 2024;10:e46625) doi:10.2196/46625

KEYWORDS
colorectal cancer; guidelines; colorectal cancer screening; digital application; questionnaire; application; cancer prevention; prevention; cancer; bowel cancer; surveillance; clinical vignette quiz; usability; Australia

Introduction

Australia’s National Health and Medical Research Council (NHMRC) screening and surveillance guidelines for colorectal cancer have become substantially more complex with their latest revision [1,2]. This is due to a shift toward personalized recommendations through detailed risk stratification based on an individual’s history of polyps or a family history of cancer. As a result, the guidelines now describe up to 230 different screening or surveillance scenarios, requiring clinicians to navigate through multiple tables to determine an appropriate recommendation. While implementing these changes can considerably improve resource use, this complexity may be a barrier to adherence, limiting the benefits of the guidelines [3,4]. Consequently, there is a need for tools to support clinicians using these guidelines. However, few of these tools have been adequately evaluated.

Several approaches have previously been considered to assist clinicians in determining appropriate bowel cancer prevention guideline recommendations. In the United States, where the complexity of polyp surveillance guidelines is the most similar to those of Australia, researchers have primarily focused on developing methods to assist clinicians in determining the appropriate advice, with a particular emphasis on automating the extraction of clinical information from electronic records to determine guideline-concordant recommendations [5-7]. In clinical practice, this resulted in a small but significant improvement in the rate of guideline-concordant recommendations (84.6% vs 77.4%) [7]. In Australia, print-based educational interventions for screening and surveillance, targeted at patients and clinicians, respectively, have had a minimal impact on improving guideline adherence [8,9]. By contrast, a nurse-led decision-making model has been the most successful intervention, increasing the rate of guideline-concordant recommendations from 83% to 97% [10]. Although successful, these options are associated with substantial costs for setup and maintenance and are not easily scalable beyond individual health services. Furthermore, how they perform when applied to the recently revised Australian guidelines is unclear.

Smartphone- or web-based digital applications (DAs) can be developed cheaply and are readily scalable. However, there are limited studies evaluating their effectiveness in supporting clinician adherence to complex bowel cancer prevention guidelines. Khan et al [11] showed that a DA was able to improve medical students’ knowledge of US colorectal cancer screening guidelines. However, their study was not randomized and did not control for the improvement in scores merely due to repeated exposure to the same clinical questions. In another study, a DA was evaluated by 6 endoscopists assessing a total of 58 colonoscopies [12]. As this was a small pilot study primarily focused on assessing the attitudes of potential users to guide the development of a new DA, it is difficult to draw meaningful conclusions about the potential benefit of the tool in improving guideline concordance.

In Australia, some of the DAs developed in response to the complexity of the latest surveillance guidelines include polyp guide, polyp app, and CRCwebapp [13-15]. These 3 tools provide greater ease of use by not requiring users to work through the risk tables manually. To the best of our knowledge, only CRCwebapp has been validated against all 230 possible case scenarios due to its use as a research tool in a previously published study [3]. However, none of these have been evaluated for their ability to improve the rate of guideline concordance among clinicians.

We hypothesized that a DA could improve clinician adherence to Australian screening and surveillance guidelines. To test this, we conducted a randomized controlled crossover study to compare the proportion of guideline-concordant decisions made by clinicians using either the CRCwebapp DA or the standard resource (SR).

Methods

Study Design and Setting

We enrolled practicing Australian clinicians to our online randomized controlled crossover clinical vignette questionnaire between July 1, 2020, and August 1, 2021. Participants were asked to provide guideline-concordant recommendations for 2 sets of clinical vignettes using either the SR or the DA. All participants were provided with an orientation that was tailored according to their experience with the guidelines. The clinical vignettes and order in which the tools were used were randomized. A study portal was used to present the vignettes, and this provided participants with access to both the SR and DA. After completing questions related to the clinical vignettes with both the SR and DA, the System Usability Scale (SUS) questionnaire was administered.

https://cancer.jmir.org/2024/1/e46625
Inclusion Criteria
We included medical, surgical, or specialist nurse practitioners who were actively practicing in Australia during the study period.

Exclusion Criteria
Participants who were not actively involved in making screening or surveillance decisions for colorectal cancer in their clinical work were excluded.

Participant Orientation
We classified participants into 2 groups according to their familiarity with the guidelines. The nonspecialist group comprised primary care practitioners who had limited experience with the terminology and structure of the published guidelines. The specialist group comprised gastroenterologists, colorectal surgeons, and specialist nurse practitioners who were routinely using the current screening and surveillance guidelines in clinical practice. The orientation program was tailored according to the experience of each group, in order to reduce the impact of experience on participant scores and to reduce barriers to participation.

For nonspecialists, the necessary terminology pertaining to screening and surveillance was defined during a web seminar. This included degree of relationship in family history for screening protocols and the individual risk characteristics and classification of lesions for surveillance protocols. The seminar also included a breakdown of every decision table in the SR and the most efficient methods to navigate to each of these. Participants were also introduced to the 4 main pages of the DA and shown how to input data and where the results were presented. In contrast, the specialist orientation did not define the terminology, and the introductions to the SR and the DA were presented as optional videos available before the questionnaire.

Primary Outcome
The primary outcome was the proportion of correct screening and surveillance recommendations issued by participants in response to the clinical vignettes. Each vignette could receive a maximum score of 6, resulting in each participant being graded with a score out of 18 for each of 2 sets of 3 clinical vignettes.

Secondary Outcome
The secondary outcome was the usability of the DA. This was assessed using each participant’s response to the SUS. A score was determined for each participant and normalized in accordance with previously published methods [16].

Clinical Vignette Design
Three pairs of clinical vignettes were developed for the study (alpha and beta, gamma and theta, and delta and omega). Each vignette described the family history, medical comorbidities, and the number and characteristics of conventional adenomas or sessile serrated lesions identified over the preceding 2 colonoscopies. We avoided scenarios commonly highlighted in previous guidelines to reduce the likelihood that participants could answer according to their recollection of these [17]. Each pair of clinical vignettes focused participants on navigating identical sets of tables to balance for difficulty.

For each clinical vignette, participants were asked to determine the age and appropriate screening modality (stool testing or colonoscopy) based on the family history presented, the first and subsequent recommended surveillance intervals, and whether surveillance should be continued when considering the comorbidities of the patient if the age of the patient was >75 years at the time of the intended procedure. Each vignette received a score out of 6. Thus, each participant could receive a maximum score of 18 for each section.

Usability
We adapted the standard SUS questionnaire by changing the term “system” to “application” in order to focus participants on assessing the usability of the DA (Textbox 1). This comprised 10 standardized statements for which users were asked to indicate their level of agreement. Numerical scores provided by participants on a slider scale were translated into Likert scores: 0-20=strongly disagree (1); 21-40=disagree (2); 41-60=neither agree nor disagree (3); 61-80=agree (4); and 81-100=strongly agree (5). A total SUS score was calculated for each participant [18]. The scores were normalized to provide a percentile ranking of the usability of the DA, as described by Sauro and Lewis [16].

Textbox 1. System Usability Scale questionnaire adapted for the use of the digital application.

1. I think that I would like to use this application frequently.
2. I found the application unnecessarily complex.
3. I thought the application was easy to use.
4. I think that I would need the support of a technical person to be able to use this application.
5. I found that the various functions in this application were well integrated.
6. I thought there was too much inconsistency in this application.
7. I would imagine that people would learn to use this application very quickly.
8. I found the application very cumbersome to use.
9. I felt very confident using the application.
10. I needed to learn a lot of things before I could get going with this application.
DA Design

Each NHMRC screening and surveillance recommendation was coded into an Excel (Microsoft Corp) spreadsheet. We eliminated redundant user data entry by determining the minimum number of inputs necessary to calculate each recommendation. For screening decisions, this included 4 fields relating to the number and age of relatives with colorectal cancer and their relation (first or second degree) to the patient. For surveillance intervals, this included the number, type, and characteristics of the lesions found during the initial procedure. Subsequent surveillance intervals required 2 additional inputs: the initial surveillance interval and the type of lesion previously identified. An additional section, incorporating a list of potential patient comorbidities, was used to determine stopping rules.

A graphical user interface was applied using an open-source platform (Open as App), which would allow for the distribution of the DA as either a web page or smartphone app. Each type of calculation (screening, first surveillance, second surveillance, or stopping rules) was identified by a tab on the bottom of the screen. Sliders were used to input data on the number of lesions, and drop-down menus were used to provide details regarding the accompanying risk characteristics. The recommendations for screening, surveillance interval, or cessation of surveillance were provided at the bottom of each respective page. The answers provided by the digital calculator were validated by individually calculating all possible scenarios covered by the updated guidelines before recruitment.

SR for Screening and Surveillance

The SR was the official web publication of the latest guidelines for screening and surveillance for bowel cancer prevention in Australia by the NHMRC [1,2]. In addition to a written summary, it provides details regarding the development of and evidence for each recommendation. Also included are a series of colored risk stratification tables to guide users through screening, initial and follow-up surveillance, and stopping rules. For screening, 90 possible scenarios are defined according to the number of relatives with colorectal cancer as well as how closely they are related to the patient.

For initial surveillance colonoscopy, 37 separate scenarios are described across 3 tables according to the various combinations of “conventional adenomas” or “clinically significant serrated polyps” identified. A total of 140 scenarios are similarly characterized across an additional 9 tables to account for the possible combinations of “conventional adenomas” and “clinically significant serrated polyps” between 2 consecutive procedures. Determining the correct surveillance interval can thus require users to successfully navigate 2 consecutive tables.

Lastly, the rules for cessation of surveillance colonoscopy are detailed in a text table that uses a modified Charlson score. Scores are allocated according to age and the presence of comorbidities. Depending on the combination of age and severity of comorbid conditions, the benefit of continuing surveillance for patients may be deemed too low to justify the potential risks of colonoscopy.

Recruitment

Advertising flyers were created and distributed to the 3 local Primary Health Networks, social media (Facebook: Adelaide GP Referral Network, Medical Mums, and Mums To Be), general practitioner education providers (GPEX and GP Synergy), and directly to practice managers located within metropolitan Adelaide. Additional flyers for specialists were distributed to members of the Departments of Gastroenterology and Hepatology and Colorectal Surgery Departments at 4 major teaching hospitals in Adelaide, as well as to private specialist practices. Snowball sampling was used to aid in the recruitment of additional participants. Continuing professional development points and a certificate of completion were awarded as an incentive to improve recruitment.

Data Collection

The questionnaire was programmed using REDCap (Research Electronic Data Capture; Vanderbilt University) tools hosted at the University of Technology, Sydney, and accessed through the Australian Access Federation [19,20]. We collected data including each participant’s age, professional background (general practice, medical specialist or trainee, and surgical specialist or trainee), active affiliation with a university, and experience with tools supporting screening and surveillance guidelines. We scored the answers for each clinical vignette in the order in which they were completed and collected each participant’s responses to the SUS questionnaire regarding their experience with the DA on a digital spreadsheet for analysis according to a previously described methodology [15].

Randomization

Two randomly permuted schedules (primary care and specialist groups) were created for a crossover study with 2 interventions (DA vs SR) with equal allocation over 8 strata (combinations 1-8; Table 1). A total of 14 allocations were generated per stratum with a total of 112 allocations. Participants were randomized to use either the DA or SR as the first aid in a 1:1 ratio. The 2 allocation schedules were programmed into the REDCap software using branching logic tools. The randomization schema was generated using Microsoft Excel (version 16.66.1; Microsoft Corp).
### Statistics

Previously reported rates of adherence to Australian surveillance guidelines have ranged from 50.8% to 83% [4,10]. The impact of a nurse-led intervention improved the rate of guideline concordance by a factor of 1.17 relative to the non–nurse-led group [10]. On the basis of these results, we predicted a mean accuracy score of 60% with the SR and anticipated a 1.17 improvement in the rate of guideline concordance to 70% with the intervention (DA). Using an expected SD of 20%, an α of .05, and a statistical power of 0.8, the minimum necessary sample size required was calculated at 64 participants.

Descriptive statistics were used to characterize the data. A Kolmogorov-Smirnov test was applied to assess for normality of the data before the statistical analysis. A related-samples Wilcoxon signed rank test was used to compare the performances of participants with either the SR or the DA. An independent-samples Mann-Whitney U test was used to compare outcomes between specialists and nonspecialists. Spearman ρ was used to assess the relationship between usability and scores from the DA. χ² tests of independence were used to compare the allocation of participants between tools and clinical vignettes. The SPSS statistical software (version 22; IBM Corp) was used for all analyses.

### Ethical Considerations

The study protocol was reviewed and approved by the Central Adelaide Local Health Network Human Research Ethics Committee (CALHN Research Office reference 13438). The background, procedures, and aims of the study were provided to prospective participants via a digital participant information sheet before the commencement of the survey. Participants were informed that their consent to participate would be implied via completion and submission of the online questionnaire. All data collected were deidentified. No participants received financial compensation.

### Results

#### Participant Characteristics

In total, 117 participants initiated the questionnaire. The records of 37 participants were excluded from the primary analysis due to survey noncompletion. Of these, no components of the questionnaire were attempted in 8 cases. 25 participants completed the background survey but did not attempt the clinical vignette section, and 4 participants aborted the clinical vignette section before completion (Figure 1). These included 7 primary care doctors, 20 gastroenterologists, 1 surgeon, 1 nurse endoscopist, and 8 participants of unknown vocation. One additional participant aborted the study after completing the vignettes and was included in the primary analysis but not in the evaluation of the usability scores.
The remaining 80 participants, consisting of 43 primary care doctors and 37 specialist doctors (35 gastroenterologists and 2 surgeons), were included in the primary analysis. They had a median age of 38 (IQR 27-71) years. Fewer than half (35/80, 44%) held an affiliation with a university (27/37, 73% of specialists and 8/43, 19% of primary care doctors), and almost two-thirds (51/80, 64%) had previously used tools for screening and surveillance decisions in colorectal cancer (32/37, 87% of specialists and 19/43, 44% of primary care doctors; Table 2). The study flowchart shows how participants were randomized to 1 of 8 sequences of vignettes (Figure 1). Of the 80 included participants, 38 (48%) were assigned to use the DA as the first aid (Figure 2). Alpha, gamma, and delta were the first vignettes in their respective pairs in 48% (38/80), 56% (45/80), and 51% (41/80) of cases (Figure 2).
Figure 2. Allocation of tools and vignettes for the first set of clinical vignettes after randomization. Pearson $\chi^2$ tests of independence were used to assess the distribution order of tools (standard resource or digital application) and vignettes (alpha or beta, gamma or theta, and delta or omega) after excluding participants who did not complete the study. The analysis confirmed that the differences in the final allocation of participants at each stage after exclusions were not significant.
Table 2. Tools to aid decisions in colorectal cancer screening and surveillance (N=80).

<table>
<thead>
<tr>
<th>Tool</th>
<th>Specialist (n=37), n (%)</th>
<th>Primary care (n=43), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiki.cancer Guideline (NHMRC&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>20 (54)</td>
<td>11 (26)</td>
</tr>
<tr>
<td>Polyp.guide</td>
<td>8 (22)</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Digital calculator</td>
<td>6 (16)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Media in endoscopy suite</td>
<td>19 (51)</td>
<td>_&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Polyp nurse support</td>
<td>3 (8)</td>
<td>_</td>
</tr>
<tr>
<td>Funding codes (Medicare)</td>
<td>2 (5)</td>
<td>_</td>
</tr>
<tr>
<td>The Royal Australian College of General Practitioners’ Redbook</td>
<td>—</td>
<td>6 (14)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (3)</td>
<td>2 (5%)</td>
</tr>
</tbody>
</table>

<sup>a</sup>NHMRC: National Health and Medical Research Council.

<sup>b</sup>Not available.

A Kolmogorov-Smirnov test of normality indicated that the scores of participants using the SR were normally distributed: D(80)=0.075; P=.20, while those of the DA were not: D(80)=0.152; P<.001. With the SR, the median (IQR) number of guideline concordant answers was 10 (7.75-13) out of 18. The use of the DA improved the number of correct recommendations to a median (IQR) of 14 (12-17) out of 18 (relative risk 1.4, P<.001; Figures 3 and 4). Lower performance with the DA compared with SR (n=17) was not associated with previous experience with screening and surveillance decision tools (P=.51), affiliation with a university (P=.39), or age (P=.06).

Figure 3. Comparison of spread of clinical vignette scores with either the standard resource (SR) or the digital application (DA). The participant scores when using the SR showed a normal distribution. A rightward shift in the distribution of the scores was observed with the use of the DA.
The median (IQR) SUS score for the DA was 90 (72.5-95), which equated to a top 4 percentile ranking among tested applications (Table 3). A moderate correlation between usability grade and DA results was observed using Spearman $\rho$ correlation coefficient ($r_s=0.4; P<.001; n=79$).

**Table 3.** System Usability Scale (SUS) grades and percentiles for participants using the digital application (n=79).

<table>
<thead>
<tr>
<th>Grade</th>
<th>SUS</th>
<th>Participants, n (%)</th>
<th>Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>&gt;78.8</td>
<td>51 (65)</td>
<td>85-100</td>
</tr>
<tr>
<td>B</td>
<td>72.6-78.8</td>
<td>7 (9)</td>
<td>65-84</td>
</tr>
<tr>
<td>C</td>
<td>62.7-72.5</td>
<td>11 (14)</td>
<td>35-64</td>
</tr>
<tr>
<td>D</td>
<td>51.7-62.6</td>
<td>6 (8)</td>
<td>15-34</td>
</tr>
<tr>
<td>F</td>
<td>0-51.7</td>
<td>4 (5)</td>
<td>0-15</td>
</tr>
</tbody>
</table>

**Sensitivity Analysis**

After excluding those who did not complete the study, differences in the randomization of participants regarding order of use of the tools (SR vs DA) and clinical vignettes (alpha vs beta, gamma vs theta, and delta vs gamma) were not significant (Figure 2). Additionally, there was no difference ($P=.55$) between the median number of guideline concordant recommendations according to whether the clinical vignettes were posed to participants: first (12, IQR 8.75-15) or second (13, IQR 9-16; Figure 5). Similarly, no difference was observed between the performance of specialists and primary care doctors, either with the SR (10 vs 9; $P=.47$) or with the DA (13 vs 15, $P=.24$; Figures 6 and 7).
Figure 5. Box and whisker plot of clinical vignette scores according to the order they were answered (first or second). A related-samples Wilcoxon signed rank test was used to compare the results achieved in the first and second set of questions indicating no significant difference ($P=.55$). Thus, increasing familiarity with the format of the questionnaire did not improve the scores achieved by participants.

Figure 6. Box and whisker plot of clinical vignette scores according to the vocational training of the participants (specialist or primary care) using the standard resource (SR). An independent-samples Mann-Whitney U test was used to compare the results of specialists with primary care doctors using the SR. There was no significant difference ($P=.47$) in the performance of participants based on their previous training in either specialist or primary care.
Figure 7. Box and whisker plot of clinical vignette scores according to the vocational training of the participants (specialist or primary care) using the digital application (DA). An independent-samples Mann-Whitney U test was used to compare the results of specialists with primary care doctors using the DA. There was no significant difference ($P=.24$) in the performance of participants based on their previous training in either specialist or primary care.

**Discussion**

**Principal Findings**

The findings of this study showed that the adherence of clinicians with Australia’s current screening and surveillance guidelines in their current form is limited. This was significantly improved when clinicians used a DA to assist their navigation of these complex guidelines. These findings were independent of the clinicians’ level of specialization, age, university affiliation, or experience with the use of other decision support tools. However, greater adherence was associated with better DA usability ratings, highlighting the importance of this attribute as a potential target to further bolster clinician guideline adherence.

Australia’s screening and surveillance guidelines are among the most complex worldwide. With the increasing trend toward personalized health care and our growing knowledge of colorectal cancer risk factors, guidelines are likely to continue increasing in complexity. For clinicians, navigating these guidelines in busy practices can be challenging. Even under the controlled conditions of our testing environment, participants could only provide appropriate recommendations in slightly over half of the questions when evaluating the scenarios only with the SR. These findings are consistent with another recent report that assessed the concordance of surveillance recommendations with current guidelines [4]. Because the adherence to previous relatively more straightforward guidelines was already known to be suboptimal, it could be anticipated that rates of adherence may be even lower as their complexity increases. This could undermine their potential benefits in the care of patients and the allocation of limited colonoscopy resources in Australia.

DAs can play an important role in supporting the implementation of Australia’s complex bowel cancer prevention guidelines. Not only do they improve the ability of clinicians to provide guideline-concordant recommendations, as demonstrated by our study, but they can be developed at a relatively low cost and are scalable to a national level. Furthermore, they can be updated with future revisions of the guidelines, ensuring that clinicians can continue to make decisions that are in keeping with the latest evidence.

Despite their clear advantages, the role of DAs in supporting complex guideline adoption has received little attention in the literature. To date, only 2 studies have evaluated DAs in assisting medical personnel with the application of bowel cancer screening and surveillance guidelines. However, these were assessed in relation to US guidelines and are limited by their small size and lack of a randomized controlled methodology. To our knowledge, our study is the first to evaluate a DA using a rigorous randomized controlled crossover design.

Participants provided discordant recommendations in 22% of clinical decisions despite assistance from the DA. However, as the DA used in this study had been validated across all the possible scenarios provided by the guidelines, we considered other factors that may have contributed to this. Our results showed that poor performance with the DA relative to the SR was not associated with participant age, academic experience, or prior experience with similar tools. One area that may have
contributed was DA usability. Although the DA scored very well in the SUS, ranking at or above the 96th percentile of tested systems, there was still a relatively large spread of scores (median 90, IQR 72.5-95) and a moderate correlation between SUS scores and participant performance. This suggests that improvements directed at improving usability for those who scored the DA less well could bolster the adherence rate of clinicians with guidelines; however, the magnitude of overall improvement may be small. Therefore, additional research to gather the opinions of participants who found the interface difficult to use and quantify the degree of progress achieved by addressing these is required.

Human error is another potential factor contributing to the rate of discordant answers. Despite simplifying the process of determining guideline-concordant recommendations, the DA still requires individuals to extract relevant and appropriate data from sometimes complex patient histories. Although human error remains an inevitable component of any interface requiring human input, natural language processing software, which has been used in prior US-based studies, could provide a valuable adjunct to a mobile app [5,6]. This would retain the scalability and portability of the DA but would require additional research, development, and testing before it could be implemented. Such a tool could provide a better balance of the advantages of the tools tested thus far.

Strengths
Our study design accounted for the possibility that participants could improve their performance in the clinical vignettes simply due to increasing experience with the questionnaire, by randomizing the order of use of the 2 aids (SR or DA). Furthermore, although the clinical vignettes were designed in pairs that were balanced for difficulty, the order in which each pair was presented to the participant was randomized to limit the risk of bias. The vignettes also focused on clinical scenarios with updated and distinct recommendations within the guidelines, requiring participants to determine the correct answer solely through navigation of the SR or DA.

Another strength of our study was the ability to cater for participants with varying levels of familiarity with guidelines. Our participants included specialists, who are accustomed to using colorectal cancer screening and surveillance guidelines in their everyday practice, and nonspecialists, whose breadth of clinical practice typically limits their experience with specialty guidelines. As these differences may have impacted participant vignette questionnaire scores, particularly those encountered without the DA, we tailored the introductory briefings to provide nonspecialists with additional information in the structured seminars. When the 2 groups were compared, no significant differences between them were observed, either with the SR or with the DA. Although this indicated that the potential effect of experience had been controlled for during our study, it is not possible to say whether this resulted from our differential approach to participant briefing, as this was not an outcome that was measured during our study.

Additionally, we were able to control successfully for potential confounders by randomizing both the order of the questions and the tools used by participants during the vignettes. This was used to address the possibility that participant scores may have improved over time and that the clinical vignettes may not have been completely balanced in their difficulty. Our sensitivity analysis showed that the distribution of questions and tools remained balanced, even after exclusion of participants, and that increasing participant experience with the questionnaire did not result in higher scores.

Limitations
While the vignettes intentionally challenged participants to navigate the breadth of the decision tables, only a limited number of scenarios are typically encountered in clinical practice. More than 95% of patients will be classified in the lowest risk category for screening based on family history, while most colonoscopies in Australia will detect few or no significant lesions [1]. Thus, participant performance in this study may not be indicative of real-world application. However, adherence rates to current surveillance guidelines, which have been reported at 50.8%, closely resemble the scores obtained using the SR in our study [4]. Clarification of the real-world efficacy of the DA will require further studies, for example, through prospective randomized nested case-controlled studies involving both primary and specialist group practices.

Our study was also prone to sampling bias. Despite efforts to circulate advertising material for the study via social media, education providers, and hospitals, only 117 participants visited the questionnaire website, and the recruitment rate was slow. The diversity of our sample group was also affected, with surgeons outnumbered by gastroenterologists in the specialist group (2-35). Due to our specific subject matter, it is possible that our participants held favorable views toward technology that may not be representative of the greater community of medical professionals. Although these challenges are not uncommon among studies recruiting clinical personnel as participants, the generalizability of our findings may be limited [21].

Finally, as both resources were readily available for public access at the time of the study, it was not possible to restrict participants to using the tools in the prerandomized order specified. Our intention-to-treat analysis may therefore have underestimated the potential differences in the scores obtained by users in the trial.

Conclusions
Australia’s bowel cancer screening and surveillance guidelines have become increasingly complex, posing a challenge for clinicians trying to make appropriate recommendations. Currently, the available options to assist them are costly and need more scalability. DAs represent an inexpensive and scalable solution that enhances guideline concordance among clinicians. Further development and assessment of these tools could improve screening and surveillance outcomes and optimize resource use in an era of increasingly complex and personalized care.
Acknowledgments

We would like to acknowledge the participants who kindly donated their time for this study, without which it would not have been possible.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

TO was responsible for study conceptualization, data curation, formal analysis, investigation, methodology, project administration, programming of the online questionnaire and digital application, and writing and review of the manuscript. OS was involved in the formal analysis and review and editing of the manuscript. PB and GI were involved in the conceptualization of the study. CR and ET were involved in the conceptualization, supervision, and review and editing of the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT checklist.

[PDF File (Adobe PDF File), 70 KB - cancer_v10i1e46625_app1.pdf ]

References


Abbreviations

DA: digital application
NHMRC: National Health and Medical Research Council
REDCap: Research Electronic Data Capture
SR: standard resource
SUS: System Usability Scale

Edited by T de Azevedo Cardoso; submitted 08.03.23; peer-reviewed by N Ribeiro, J Makin; comments to author 19.09.23; revised version received 28.11.23; accepted 18.01.24; published 22.02.24.

Please cite as:
Ow TW, Sukocheva O, Bampton P, Iyngkaran G, Rayner CK, Tse E
Improving Concordance Between Clinicians With Australian Guidelines for Bowel Cancer Prevention Using a Digital Application: Randomized Controlled Crossover Study
JMIR Cancer 2024;10:e46625
URL: https://cancer.jmir.org/2024/1/e46625
doi:10.2196/46625
PMID:38238256

©Tsai-Wing Ow, Olga Sukocheva, Peter Bampton, Guruparan Iyngkaran, Christopher K Rayner, Edmund Tse. Originally published in JMIR Cancer (https://cancer.jmir.org), 22.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cancer, is properly cited. The complete bibliographic information, a link to the original publication on https://cancer.jmir.org/, as well as this copyright and license information must be included.
Assessing the Quality, Privacy, and Security of Breast Cancer Apps for Arabic Speakers: Systematic Search and Review of Smartphone Apps

Dari Alhuwail1,2, PhD; Aisha Alhouti1, MSc; Latifah Alsarhan1, BS

1Information Science Department, College of Life Sciences, Kuwait University, Sabah AlSalem University City, Kuwait
2Health Informatics Unit, Dasman Diabetes Institute, Dasman, Kuwait

Corresponding Author:
Dari Alhuwail, PhD
Information Science Department
College of Life Sciences
Kuwait University
P.O. Box 5969
Sabah AlSalem University City, Safat 13060
Kuwait
Phone: 965 24633214
Email: dari.alhuwail@ku.edu.kw

Abstract

Background: Breast cancer is a widespread disease, and its incidence is rapidly increasing in the Middle East and North Africa region. With the increasing availability of smartphone apps for various health purposes, breast cancer apps have emerged as tools for raising awareness, providing support, and empowering women affected by this disease. These apps offer many features, including information on breast cancer risk factors, self-examination guides, appointment reminders, and community support groups or hotlines. Using apps raises the risk of privacy and security issues, and we hope that examining these features of the apps will contribute to the understanding of how technology can be used to improve these apps and provide insights for future development and improvement of breast cancer apps.

Objective: This study aims to critically review the quality, privacy, and security of breast cancer apps available to Arabic speakers.

Methods: Similar to several recent studies, we used a systematic search for apps available in Google Play and Apple App stores using both the web interface and the built-in native stores installed on smartphones. The search was conducted in mid-December 2022 in Arabic using the following keywords: breast cancer, breast cancer treatment, breast cancer disease, breast cancer symptoms, breast cancer screening, and breast test. These preidentified search terms are based on earlier work concerning the top searched breast cancer topics by Arabic speakers through Google’s search engine. We excluded apps that did not have an Arabic interface, were developed for non-Arabic speakers, were paid, needed a subscription, or were directed toward health care workers. The Mobile App Rating Scale was used to evaluate the quality of the apps concerning their engagement, functionality, aesthetics, and information. A risk score was calculated for the apps to determine their security risk factors.

Results: Only 9 apps were included, with most (6/9, 67%) being supported by advertisements and categorized as informational. Overall, the apps had low numbers of downloads (>10 to >1000). The majority of the included apps (8/9, 89%) requested dangerous access permissions, including access to storage, media files, and the camera. The average security score of the included apps was 3.22, while only 2 apps provided information about data security and privacy. The included apps achieved an overall average quality score of 3.27, with individual dimension scores of 4.75 for functionality, 3.04 for information, 3.00 for aesthetics, and 2.32 for engagement.

Conclusions: The limited availability of breast cancer apps available to Arabic speakers should be a call to action and prompt health care organizations and developers to join forces and collaboratively develop information-rich, usable, functional, engaging, and secure apps.

(JMIR Cancer 2024;10:e48428) doi:10.2196/48428
Introduction

Female breast cancer is among the most commonly diagnosed cancers worldwide, with a rate of approximately 2.3 million new cases [1]. Over the past decade, breast cancer incidence has been on the rise in the Middle East and North Africa (MENA) region [2]. Breast cancer is the most frequently diagnosed cancer (17.7%-19% of all types of cancer) in the region [3], and it accounts for 30% of all cancer cases [4]. The lack of cancer education and barriers to cancer screening are seen as major health problems [5]. Education is one of the most effective tools in the fight against female breast cancer; it can have positive effects on women’s practices, attitudes, and knowledge of the disease [6]. However, recent studies suggest a lack of resources and poor awareness of breast cancer in women in the MENA region [7]. While the digital delivery of health education content has been on the rise through different digital media (eg, websites and social media), the quality of Arabic content for female breast cancer remains poor [8-10].

Smartphone proliferation, ubiquity, and affordability, as well as the increasing availability of mobile apps, may be the long-awaited for “digiceuticals” or digital therapeutics [11-13]. Today, the number of health apps in smartphone app stores exceeds 325,000 [14] and will continue to rise, with estimates of more than 200 apps being added daily to app stores [15], covering a wide spectrum of health purposes, such as well-being, education, and disease management, including chronic conditions [16]. Several apps are available that can help individuals with breast cancer manage their condition. These apps have been used for the purposes of education [17-19], care management [20,21], prevention [22-24], and well-being [25,26]. These apps can be a valuable resource for individuals with breast cancer, helping them stay informed and connected to their care team and manage the symptoms and challenges associated with the disease.

The use of mobile health (mHealth) apps contributes to improving health literacy and facilitating communication between patients and their care providers [27]. Moreover, it improves patient well-being and helps caregivers make informed clinical decisions [28]. In fact, the use of such apps not only benefits patients while receiving treatment but also provides tools such as follow-up care and self-management for breast cancer survivors [29]. Patients’ need for self-management techniques is crucial, as it helps them to make their lives better by complying with the treatment needed and, as a result, accepting the disease [30].

However, despite the high number of downloads and star ratings of health apps, including those specifically targeting breast cancer, several challenges remain with respect to their quality and security. Evidence from the literature reports on the existence of poor-quality health apps that fall short with respect to (1) following evidence-based health guidelines and best practices, (2) involving experts and consumers in their development, and (3) demonstrating effectiveness based on empirical evidence, all of which ultimately can be potentially harmful to their users [14,31-33].

Additionally, health apps have been facing critical challenges related to their privacy, confidentiality, and security [14,34,35], especially given their nature of handling sensitive, personal, and health-related data [36]. These challenges have been magnified with the rise of cyberattacks through apps and mobile devices [37] and further highlighted by recent regulations such as the General Data Protection Regulation for member states of the European Union [38]. Such a regulation assesses the privacy score of mobile apps and identifies or measures apps’ privacy based on 14 components [39].

As the uptake of these apps increases, it becomes imperative for users to evaluate their quality and safety [40]. Despite the high prevalence of breast cancer among the population of the MENA region, evidence regarding the quality, privacy, and security of breast cancer apps available to Arabic speakers remains poor. This study aimed to conduct a systematic assessment of mobile breast cancer apps available for Arabic speakers to evaluate their functionality, quality, security, and data safety. To the best of our knowledge, no previous study has addressed this gap.

Methods

Overview

Using a similar approach to several recent studies [41-45] and to ensure scientific rigor, this study conducted a systematic search and content analysis of mobile breast cancer apps available for Arabic speakers. We searched both Google Play and Apple App stores between December 18 and 24, 2022.

Search Strategy

Initially, we used the Arabic search terms highlighted in Table 1 to search Google Play and Apple App stores. These search terms were selected based on earlier work that was published concerning the top searched breast cancer topics by Arabic speakers using Google’s search engine [8]. To ensure rigor, the researchers searched the app stores both through (1) the web interface and (2) natively on devices running the relevant operating system, thus mimicking how end users will discover such apps.
Table 1. The terms used to search for breast cancer apps available to Arabic speakers and their English translations.

<table>
<thead>
<tr>
<th>Arabic term</th>
<th>Translated term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>Breast cancer</td>
</tr>
<tr>
<td>Breast cancer screening</td>
<td>Breast cancer screening</td>
</tr>
<tr>
<td>Breast cancer treatment or therapy</td>
<td>Breast cancer treatment or therapy</td>
</tr>
<tr>
<td>Breast cancer detection or screening</td>
<td>Breast cancer detection or screening</td>
</tr>
<tr>
<td>Breast cancer disease</td>
<td>Breast cancer disease</td>
</tr>
<tr>
<td>Symptoms of breast cancer</td>
<td>Symptoms of breast cancer</td>
</tr>
<tr>
<td>Breast screening</td>
<td>Breast screening</td>
</tr>
</tbody>
</table>

Eligibility

App eligibility was determined by 2 independent researchers blinded to each other’s decisions, and the apps were initially screened based on the app’s name, the provided screenshots, and the app’s description. Discrepancies between researchers were resolved through consensus. Apps were included if they were free of charge, provided content and support for Arabic speakers, and were designed for use by consumers or patients; all apps were considered regardless of release or last update dates (Textbox 1). Apps were excluded if they were paid or were subscription-based, did not support Arabic speakers, or were designed for use by clinicians or health care workers.

Textbox 1. Inclusion and exclusion criteria for the apps.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free of charge</td>
</tr>
<tr>
<td>Available on Google Play or Apple App store</td>
</tr>
<tr>
<td>Designed for use by consumers or patients</td>
</tr>
<tr>
<td>Support Arabic speakers</td>
</tr>
<tr>
<td>Considered regardless of release or last update dates</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paid or subscription based</td>
</tr>
<tr>
<td>Designed for use by clinicians or health care workers</td>
</tr>
<tr>
<td>Does not support Arabic speakers</td>
</tr>
</tbody>
</table>

Data Extraction and Evaluation

Initially, all information provided by the app developers in the app stores was extracted to evaluate the descriptive features and the general characteristics of the included apps, which included the platform, developer name, update date, ratings, number of reviews, number of downloads, app category, and app permissions, as reported by the app developers. Afterward, 2 independent researchers downloaded the apps on their smartphones to assess the quality and privacy risks of the included apps.

We evaluated the quality of the included apps using a standardized form, the Mobile App Rating Scale (MARS), focusing on the following 4 dimensions: engagement, functionality, aesthetics, and information quality [46,47]. All scores were compared among 2 researchers, and the average score for each dimension was reported. To evaluate the apps’ privacy risks, we assigned scores to the permissions requested by the apps as reported by the app developers. The scores were informed by previous research, where the score risk is 0 for nonthreatening, 0.5 for potentially threatening, and 1 for threatening permissions [48]. Such permissions include access to restricted data, such as system state and user contact information, and restricted actions, such as connecting to a paired device and recording audio [49]. The 2 researchers independently carried out this evaluation and were unaware of each other’s scores; any discrepancies were resolved through consensus.

Results

Overview

The researchers followed the systemic steps, highlighted in Figure 1, resulting in the inclusion of 9 apps, all of which are Android apps found on the Google Play store.

Overall, the included apps were indicated to be appropriate for all ages and were either in the medical, education, lifestyle, personalization, or health and fitness categories as per Google Play store categorization (Table 2). Our investigation suggests that the apps were all informational in nature, mainly providing information about breast cancer. None of the included apps had a language option to make it available in more than 1 language.
At the time of data collection, the results show that the included apps had low overall downloads (>10 to >1000) and more than half (6/9, 67%) were supported by advertisements. Only 5 apps had reviews, with an average of 10.60 reviews and an average star rating of 4.78. Only 1 app was last updated in 2019, while the remaining apps were updated in the past 2 years.

**Figure 1.** The flow diagram of the systematic search process to identify the relevant apps.

**Table 2.** Characteristics of the included breast cancer apps in the Google Play store.

<table>
<thead>
<tr>
<th>App number&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Stars</th>
<th>Reviews, n</th>
<th>Advertisement supported</th>
<th>Downloads, n</th>
<th>Rating&lt;sup&gt;b,c&lt;/sup&gt;</th>
<th>Updated on&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Category&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>No</td>
<td>&gt;100</td>
<td>3+</td>
<td>May 11, 2021</td>
<td>Medical</td>
</tr>
<tr>
<td>2</td>
<td>4.9</td>
<td>23</td>
<td>Yes</td>
<td>&gt;1000</td>
<td>3+</td>
<td>July 26, 2019</td>
<td>Medical</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>&gt;100</td>
<td>3+</td>
<td>April 18, 2020</td>
<td>Health and fitness</td>
</tr>
<tr>
<td>4</td>
<td>4.2</td>
<td>6</td>
<td>Yes</td>
<td>&gt;1000</td>
<td>3+</td>
<td>November 3, 2020</td>
<td>Education</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>8</td>
<td>No</td>
<td>&gt;500</td>
<td>3+</td>
<td>May 11, 2021</td>
<td>Medical</td>
</tr>
<tr>
<td>6</td>
<td>4.8</td>
<td>6</td>
<td>No</td>
<td>50</td>
<td>3+</td>
<td>July 11, 2022</td>
<td>Medical</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>500</td>
<td>3+</td>
<td>August 9, 2021</td>
<td>Lifestyle</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>10</td>
<td>3+</td>
<td>October 8, 2022</td>
<td>Personalization</td>
</tr>
<tr>
<td>9</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>10</td>
<td>3+</td>
<td>October 8, 2022</td>
<td>Personalization</td>
</tr>
</tbody>
</table>

<sup>a</sup>Arbitrary number to mask app name.

<sup>b</sup>Based on Google Play store.

<sup>c</sup>Content of apps considered suitable for age group indicated per Google Play store rating.

<sup>d</sup>At data collection.

<sup>e</sup>Not available.
App Permissions and Controls

As described in the Google Play store, the majority of the included apps (8/9, 89%) were requesting dangerous access permissions, including access to storage, media files, and camera permissions (Table 3). Additional permissions were exposed after the researchers downloaded the apps on the testing devices. These permissions included receiving data from the internet; viewing network connections; having full network access; running at start-up; controlling flashlight and vibration; preventing the device from sleeping; reading badge notifications; running foreground services; advertising permissions; reading location from media; playing and installing referrer application programming interface; and lastly, pairing with Bluetooth devices (Table S1 in Multimedia Appendix 1). The researchers considered all permissions and added them up to calculate the final risk score (Table 4).

On average, the security score for the included apps is 3.22 (total points possible: 9.5). The security scores ranged between 0 and 7.5; only 2 apps had a total security score of 0. The apps requested potentially dangerous permissions, namely full network access, advertising ID permission, read location from media collection, precise location (GPS and network-based), take pictures and videos, read the contents of the USB storage, modify or delete the contents of the USB storage, and view Wi-Fi connections.

Table 3. Included apps’ permissions as described in the Google Play store.

<table>
<thead>
<tr>
<th>App number</th>
<th>Location</th>
<th>Camera</th>
<th>Photos and media</th>
<th>Storage</th>
<th>Wi-Fi connection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Approximate location (network based)</td>
<td>Precise location (GPS and network based)</td>
<td>Take pictures and videos</td>
<td>Read the contents of your USB storage</td>
<td>Modify or delete the contents of your USB storage</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. The security score per app based on its permissions requested or declared.

<table>
<thead>
<tr>
<th>Permission</th>
<th>Weight&lt;sup&gt;a&lt;/sup&gt;</th>
<th>App 1</th>
<th>App 2</th>
<th>App 3</th>
<th>App 4</th>
<th>App 5</th>
<th>App 6</th>
<th>App 7</th>
<th>App 8</th>
<th>App 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receive data from the internet</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>View network connections</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Full network access</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Run at start-up</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Control flashlight</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Control vibration</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Prevent device from sleeping</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Read badge notification</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Run foreground service</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Advertising ID permission</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Read location from media collection</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Play install referrer API&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pair with Bluetooth devices</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Approximate location (network-based)</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Precise location (GPS and network-based)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Take pictures and videos</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Modify or delete the contents of your USB storage</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Read the contents of your USB storage</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Modify or delete the contents of your USB storage</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>View Wi-Fi connections</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total security score per app</td>
<td>9.5</td>
<td>0</td>
<td>2.5</td>
<td>7.5</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

<sup>a</sup>The threat weight was calculated following the guidance provided by Olmstead and Atkinson [48].

<sup>b</sup>API: application programming interface.

Data Safety

Only 3 apps provided information about data safety and how the data are handled (Table S2 in Multimedia Appendix 1). Specifically, only 2 apps reported not sharing user data with other companies or organizations; the same apps reported that their apps do not collect user data. On the other hand, only 1 reported sharing information such as location, health and fitness messages, as well as photos and videos. The 3 apps that declared information about data safety reported that the data are encrypted in transit. Only 1 app reported that the users can request to delete the data.

Apps’ Quality Rating

The researchers used MARS to evaluate the included apps’ quality. MARS uses 4 dimensions to assess the apps: engagement, functionality, aesthetics, and information. The average overall score rating of the included apps was 3.27 (Table 5). Notably, the evaluation showed that all the apps had a high rate in the functionality dimension, where they all scored 4.75; as for the engagement dimension, all apps had a meager score with an average of 2.32. Considering the other 2 dimensions, aesthetics and information, we can see that the scores vary between apps, where some have a high score and others have a low score, with an average of 3.00 and 3.04, respectively.
Table 5. Mobile App Rating Scale evaluation for Arabic breast cancer apps.

<table>
<thead>
<tr>
<th>App number</th>
<th>Engagement</th>
<th>Functionality</th>
<th>Aesthetics</th>
<th>Information</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.30</td>
<td>4.75</td>
<td>3.67</td>
<td>3.87</td>
<td>3.60</td>
</tr>
<tr>
<td>2</td>
<td>2.40</td>
<td>4.75</td>
<td>3.00</td>
<td>3.00</td>
<td>3.30</td>
</tr>
<tr>
<td>3</td>
<td>2.40</td>
<td>4.75</td>
<td>3.00</td>
<td>2.25</td>
<td>3.10</td>
</tr>
<tr>
<td>4</td>
<td>2.40</td>
<td>4.75</td>
<td>3.33</td>
<td>2.50</td>
<td>3.20</td>
</tr>
<tr>
<td>5</td>
<td>2.40</td>
<td>4.75</td>
<td>3.33</td>
<td>3.50</td>
<td>3.50</td>
</tr>
<tr>
<td>6</td>
<td>2.40</td>
<td>4.75</td>
<td>3.67</td>
<td>4.00</td>
<td>3.70</td>
</tr>
<tr>
<td>7</td>
<td>2.20</td>
<td>4.75</td>
<td>3.00</td>
<td>2.75</td>
<td>3.30</td>
</tr>
<tr>
<td>8</td>
<td>2.20</td>
<td>4.75</td>
<td>2.00</td>
<td>2.75</td>
<td>2.90</td>
</tr>
<tr>
<td>9</td>
<td>2.20</td>
<td>4.75</td>
<td>2.00</td>
<td>2.75</td>
<td>2.90</td>
</tr>
<tr>
<td>Total mean score</td>
<td>2.32</td>
<td>4.75</td>
<td>3.00</td>
<td>3.04</td>
<td>3.27</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

To the best of our knowledge, this work summarizes the most extensive collection of the currently available free-of-charge breast cancer apps for Arabic speakers on the Google Play and Apple App stores. Interestingly, this systematic investigation reveals that, at the time of data collection, no breast cancer apps were identified as available to Arabic speakers in the Apple App store. Additionally, none of the apps at the time of data collection provided language options, which can hinder the possibility of translating to multiple languages. Overall, the results of this research showed a lack of breast cancer apps available to Arabic speakers compared to apps available in Turkish [50] and Korean [51] for example.

The analyzed apps in this study are available for free, with the majority of these apps (6/9, 67%) being supported by advertisements. The apps varied in what they are able to access and control on the devices, as shown in Table 3. A total of 44% (4/9) apps use Wi-Fi, 33% (3/9) can access the device’s storage, 22% (2/9) have access to the device’s camera, and 11% (1/9) are able to access the media as well as the location. Only 33% (3/9) of the apps stated how the collected data would be used (Table S2 in Multimedia Appendix 1), while the rest of the apps did not specify any information.

The results of this study demonstrate that the quality of the considered apps is highly “functional” but less “engaging” (Table 5). The average overall score rating of the included apps (3.27) is slightly above average, suggesting that the majority of the apps may not be considered exceptional by consumers. While there seem to be positive aspects to the included apps, there is still room for improvement. Considering Arabic content related to breast cancer, recent evidence suggests the low quality of informational videos available on YouTube despite the high number of views [9]. Our findings provide evidence suggesting that nearly average-quality information content about breast cancer is available to Arabic speakers.

A recent Spanish study tested 6 apps (2 on iOS, 5 on Android, and 1 on both) using the MARS framework. The results of the study showed relatively diverse scores, with an objective quality mean score of 3.06 and a subjective quality mean score of 1.96 [52]. Another study evaluated the quality of mHealth apps for educational purposes in Iran using the MARS framework. The study reported a mean score of 4.01 for quality and 3.08 for subjective quality [53]. Turkish apps were also evaluated using the MARS framework, and the study found an average score of 3.31 [50], which is similar to the Spanish study’s results.

Although many research papers have addressed the importance of using mHealth apps to improve patients’ health, provide educational materials, enhance communication between patients and caregivers, and achieve a successful recovery, these studies have stated that such apps are exposed to several challenges and threats as well. For instance, middle- and high-income households had more access to the internet on their mobile devices compared to those with low incomes [28]. Therefore, patients and caregivers who have no or limited access to the internet may not be able to use the app’s features effectively, or they might not prefer to use a mobile phone for health-related purposes in general.

Another concern is that the process of storing and transferring personal health data through a mobile app could be insecure and might cause serious security and privacy issues [28,54]. Thus, mobile apps should be designed to accommodate a wide range of possible users while considering health knowledge, different levels of cultural needs, and linguistic requirements. Additionally, recent research suggests that assessing the apps’ quality, safety, and usability by involving patients and health care professionals will result in an ideal app that meets patients’ requirements and enhances the app’s overall safety as well [20].

The literature on breast cancer smartphone apps suggests that these apps can be a valuable resource of information for individuals with the disease [55]. These apps provide information on treatment options, support groups, and local resources. The apps also allow individuals to track their symptoms, set reminders for appointments, and record their progress throughout treatment [56]. Such features can significantly contribute to enhancing patients’ well-being [57]. Additionally, studies have found that these apps can improve communication with health care providers and improve self-efficacy and quality of life [30,58]. However, it is important to note that the quality and accuracy of the information provided...
by these apps can vary, and it is recommended to consult with a health care professional before using any app for managing a medical condition.

A recent study on breast health and breast cancer apps notes that although apps appear to be competitive and useful for patients, some major features have to be considered while developing these apps [59]. The features include notifications, reminders, symptoms tracking, and recording. The study also suggests designing the apps to be user-friendly, even for low-literacy patients, by adding an audio feature (text-to-speech). Developing features with audio support will not only help patients with low literacy but can also support multiple languages.

**Recommendations and Implications for Practice**

**Privacy and Security**

We would recommend that the developers of the apps be more transparent and state how the data will be used and that they should not have access to unnecessary data. We recommend that future breast cancer apps be available to Arabic speakers to justify the need for the permissions requested while also transparently disclosing the data safety handling measures to the app users. Security and privacy of apps are considered major requirements as they are accountable for sensitive patient data such as prescriptions, treatments, etc. Thus, to come up with robust apps that could ensure privacy and security appropriately, more evaluation techniques, as well as security mechanisms, should be analyzed and implemented on Arabic apps, in particular, to assess, measure, and control the apps’ security and privacy [60].

**Quality and Engagement**

Involving patients and health professionals in the app design phase is crucial. Several studies have addressed the idea that health applications should be developed and designed based on the combined efforts of health professionals, related academics, and patients [61]. To raise the quality of breast cancer apps, the inclusion of utility features such as appointment booking for mammograms and web-based consultations becomes necessary. In addition, it is recommended to improve health apps’ engagement by focusing on specific components such as personalized content, data visualization, reminders and notifications, educational material, self-management functions, and goal-setting features [62]. Providing users and patients with proper communication features and a well-designed interface leads to an ideal user experience as a result [30].

**Study Strengths and Limitations**

Similar to other studies, a rigorous multistep methodology mimicking systematic reviews is used in this study to assess the breast cancer apps that are available to Arabic speakers. Apps were thoroughly searched through both the web interface as well as the app stores natively on the devices, mimicking how end users will discover such apps. While the results provide an indication of the quality of the evaluated apps, additional investigations are required to consider patients’ perspectives about their views about the quality as well as the utility of such apps. Future studies can also involve rigorous assessments with respect to the security measures applied by breast cancer apps available to Arabic speakers.

This study only considered the publicly available apps and may have missed apps that are “prescribed” to patients or consumers or those that are developed locally by health care organizations. Another limitation, which is inherent to the search strategy used in this work as well as similar other work [63], is the fact that the search algorithms used by the app stores are nontransparent and can change without the public’s knowledge, potentially undermining the reproducibility of the outcomes. Lastly, the current state of the results as revealed by this work is likely to change quite rapidly since apps are regularly released, updated, and retired.

**Conclusions**

The battle against breast cancer is not over yet, and breast cancer apps can serve as valuable resources in this ongoing fight. The results of this systematic and thorough examination of breast cancer apps available for Arabic speakers reveal their limited existence at the time of study. The investigations evaluated these apps through the lenses of quality, privacy, and security, revealing that the included apps are rated as highly “functional” but at the same time are less “engaging.” The investigations also reveal that some apps were accessing unnecessary data and collecting information that was not relevant to the purpose of the app.

Developers of breast cancer apps that cater to Arabic speakers must focus on consumers’ preferences, demographics, usability, and the interface of their apps, as well as enhance measures and mechanisms of privacy and security for their apps. The low number of breast cancer apps available to Arabic speakers, as revealed in this study, should be a call to action for many health care organizations and developers to collaboratively develop information-rich, usable, functional, engaging, and secure apps.

**Data Availability**

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

**Authors’ Contributions**

DA conceptualized the study design and supervised the study. DA and AA participated in data assembly, analysis, and interpretation, and in the writing of the paper. LA contributed to data interpretation and in the writing of the paper. All authors read and approved the final paper.
Conflicts of Interest

None declared.

Multimedia Appendix 1

Included apps' declared permission and safety declarations.

References


54. Wright A. Evaluation of two mobile health apps for patients with breast cancer using the mobile application rating scale. Mhealth 2021;7:60 [Full text] [doi: 10.21203/rs.3.rs-992412/v1]
56. Nurgalieva L, O'Callaghan D, Doherty G. Security and privacy of mHealth applications: a scoping review. JMIR Mhealth 2021;7:60 [Full text] [doi: 10.21037/mhealth.20-161] [Medline: 34805391]
Web-Based Scaffolds: The Feasibility of a Constructivist Approach to Oncology Fellow Learning

Sam Brondfield¹, MAEd, MD; Matthew Schwede², MD; Tyler P Johnson², MD; Shagun Arora¹, MD

¹Department of Medicine, University of California, San Francisco, San Francisco, CA, United States
²Department of Medicine, Stanford University, Palo Alto, CA, United States

Corresponding Author:
Sam Brondfield, MAEd, MD
Department of Medicine
University of California, San Francisco
505 Parnassus Ave
Room M1286, Box 1270
San Francisco, CA, 94143
United States
Phone: 1 4155144783
Email: sam.brondfield@ucsf.edu

Abstract

In this 2-institution feasibility pilot, oncology fellows used and updated freely available web-based learning tools (scaffolds) in a constructivist fashion.

(JMIR Cancer 2024;10:e52501) doi:10.2196/52501

KEYWORDS

constructivist learning; scaffolded learning; graduate medical education; fellowship training; oncology; feasibility; medical education; pilot study; study; online learning; online tool; online tools; remote learning; e-learning; training; cancer

Introduction

Succinct and updated oncology fellow learning materials are lacking. Additionally, fellow didactic learning often takes the form of passive lectures, which is undesirable [1,2]. Constructivist learning, wherein learners construct their own knowledge, is rare for fellows.

We piloted “scaffolds”—succinct slide sets shared across oncology trainees—and evaluated feasibility [3,4]. Throughout training, fellows can update the shared scaffolds in a constructivist fashion, thereby providing updated resources for themselves and colleagues.

Methods

Study Design

Two institutions participated—University of California, San Francisco (UCSF), and Stanford University. From 2018 to 2019, SB—a UCSF oncologist—designed 12 scaffolds, using Google Slides covering the solid tumor chapters from the American Society of Clinical Oncology’s Self-Evaluation Program (ASCO-SEP) textbook [5]. Hematology, gynecologic oncology, and neuro-oncology were omitted for this pilot. Scaffolds included text and images synthesized from ASCO-SEP and National Comprehensive Cancer Center guidelines. For brevity, the slides instructed fellows to adhere to length limits when making edits.

We emailed scaffold links to all first- to third-year UCSF (n=21) and Stanford University (n=27) oncology fellows in July 2019 and July 2020. Use was optional, and fellows could access and update the scaffolds anonymously at any time. Updates were audited by SB.

In December 2021, to evaluate feasibility outcomes (fidelity: degree to which the innovation was implemented as intended; appropriateness: perceived fit of the innovation; self-efficacy: belief in the ability to execute the innovation’s goals) [6], we reviewed updates tracked in Google Slides and conducted 2 voluntary feedback focus groups (UCSF: facilitated by SB; Stanford University: facilitated by MS—a Stanford University oncology fellow) with 4 fellows each. Focus group size was determined by responses to recruitment emails. Consent and demographic information were obtained. Participants did not need to use the scaffolds, as we were also exploring barriers to use. Focus groups were recorded and professionally transcribed. SB and MS independently reviewed the transcripts and generated themes through iterative discussion [7].
The scaffolds were updated in 2023 by SB (available on Google Drive) [8].

**Ethical Considerations**

UCSF and Stanford University institutional review boards granted exemption (#20-31645) and approval (#57766), respectively. Participants received an information sheet and verbally consented before each focus group. Transcripts omitted personal identifiers, and interviewers never revealed participant identities to the rest of the study team. Participants received a US $10 electronic gift card.

**Results**

**Fidelity**

From July 2019 to December 2021, fellows made 60 updates (Table 1), ranging from new trials to changes in management; none were erroneous. SB made 9 edits for brevity.

Table 1. Number of updates to solid oncology scaffolds during the pilot period (July 2019 to December 2021).

<table>
<thead>
<tr>
<th>Scaffolds</th>
<th>Updates by fellows (N=60), n</th>
<th>Updates by auditor (N=9), n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder/kidney/adrenal</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Breast</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Gastrointestinal (lower)</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Gastrointestinal (upper)</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Germ cell</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Head/neck</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lung (nonsmall cell)</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Lung (small cell/other thoracic)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Melanoma</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Prostate</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Salivary/thyroid</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Sarcoma</td>
<td>12</td>
<td>1</td>
</tr>
</tbody>
</table>

**Appropriateness**

Focus group participants (N=8) were women and included Asian (n=3, 37.5%), White (n=3, 37.5%), Black (n=1, 12.5%), mixed-race (n=2, 25%), first-year (n=5, 62.5%), second-year (n=2, 25%), and third-year (n=1, 12.5%) fellows. Most (n=7, 87.5%) used the scaffolds. Qualitative analysis (Table 2) revealed that fellows felt the scaffolds were accessible and succinct learning tools, addressed the dearth of similar resources, served as effective preparation materials for clinical work and examinations, provided structured information for rapid reviews, and made interactions with complex resources easier.
Table 2. Qualitative analysis of transcripts from 2 oncology fellow focus groups (1 at the University of California, San Francisco, and 1 at Stanford University) that evaluated a pilot of solid oncology scaffolds (July 2019 to December 2021).

<table>
<thead>
<tr>
<th>Theme</th>
<th>Supportive quotation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td></td>
</tr>
<tr>
<td>Accessible, succinct resource</td>
<td>“[The scaffolds were] online and quickly accessible, for example on the shuttle on the way to work.”</td>
</tr>
<tr>
<td>Addressed the dearth of similar resources</td>
<td>“There are few resources currently available for oncology fellows. [The scaffolds] filled a niche not currently filled by other resources.”</td>
</tr>
<tr>
<td>Effective preparation materials for clinical work and examinations</td>
<td>“[The scaffolds] were a security blanket…helpful for clinic prep and inpatient consults.”</td>
</tr>
<tr>
<td>Structured information for rapid reviews</td>
<td>“[The scaffolds] were helpful in that they provided frameworks…and approaches.”</td>
</tr>
<tr>
<td>Easier subsequent use of more complex resources</td>
<td>“The guidelines felt less ‘foreign’ after reviewing the scaffolds…[the scaffolds] helped with knowledge retention from more complex resources.”</td>
</tr>
<tr>
<td><strong>Challenges</strong></td>
<td></td>
</tr>
<tr>
<td>Lack of fellow confidence in updating the scaffolds</td>
<td>“I wasn’t sure whether my learning points were important enough to add to the scaffold.”</td>
</tr>
<tr>
<td>Lack of fellow ownership over the scaffolds</td>
<td>“I think fellows are probably less likely to update the scaffolds if they don’t feel responsible for them.”</td>
</tr>
<tr>
<td>Too simple and broad to help with nuanced patient care</td>
<td>“Clinical care is so nuanced…the scaffolds may be too broad to help with some clinical situations.”</td>
</tr>
<tr>
<td><strong>Suggestions</strong></td>
<td></td>
</tr>
<tr>
<td>Improve visual appeal</td>
<td>“Maybe make them more visually appealing by including more figures or tables.”</td>
</tr>
<tr>
<td>Clarify purpose and the fact that scaffolds can be updated</td>
<td>“I would make it clear that the slides are editable and that fellows should update them.”</td>
</tr>
<tr>
<td>Facilitate opportunities for fellows to update scaffolds</td>
<td>“Asking fellows to update these might be good for their learning.”</td>
</tr>
</tbody>
</table>

**Self-Efficacy**

Qualitative analysis revealed barriers to updating the scaffolds—fellows’ lack of ownership over the scaffolds and low confidence regarding appropriate updates.

**Discussion**

**Principal Results**

This pilot explored the feasibility of implementing constructivist scaffolds for oncology fellows. We found evidence of fidelity and appropriateness and delineated next steps to optimize self-efficacy. The scaffolds [8] can be downloaded and modified to avoid generating institution-specific scaffolds from scratch. To promote ownership and confidence, we recommend assigning fellows to update the scaffolds under faculty mentorship.

Despite demonstrating superior outcomes when compared to passive lectures, constructivist learning is rarely studied at the fellowship level [9-11]. We recommend evaluating constructivist learning modalities, such as scaffolds, in graduate medical education to enhance learning outcomes.

**Limitations**

Though the focus groups suggested that multiple fellows used the scaffolds, Google Slides did not track how many fellows accessed or updated them. We did not incorporate multimedia components beyond images and tables (some needed to be removed before publication to respect copyright), nor did we include assessments in this pilot. We recommend that institutions consider incorporating multimedia content and assessments into the scaffolds. The number of focus group participants was small and not gender-diverse. Future studies should quantitatively evaluate usage patterns and user satisfaction to examine what factors drive utilization.

**Conclusion**

We piloted a novel constructivist approach to fellow learning and found evidence of feasibility. Oncology educators may use and modify the scaffolds [8] to jump-start constructivist education for fellows at their institutions. Educators in other fields may wish to apply this model to their specialties.
Acknowledgments
This study was funded by a University of California, San Francisco (UCSF), Academy of Medical Educators Education Innovations grant. The funder did not have a role in the study’s design, data collection, data analysis, data interpretation, manuscript writing, or decision to submit the manuscript for publication.

Data Availability
The data sets analyzed during this study are not publicly available due to institutional review board restrictions but are available from the corresponding author on reasonable request.

Authors’ Contributions
SB designed the scaffolds, conceived the study, conducted the quantitative analysis, and wrote the manuscript. SB and MS each conducted 1 focus group. SB and MS conducted the qualitative analysis. All authors contributed manuscript edits and approved the final manuscript for submission.

Conflicts of Interest
None declared.

References

Abbreviations
ASCO-SEP: American Society of Clinical Oncology’s Self-Evaluation Program
UCSF: University of California, San Francisco