

ARTICLE OPEN



Mobile app validation: a digital health scorecard approach

Ramy Sedhom¹, Michael J. McShea², Adam B. Cohen^{3,4}, Jonathan A. Webster⁵ and Simon C. Mathews⁶

While digital health solutions continue to grow in number and in complexity, the ability for stakeholders in healthcare to easily discern quality lags far behind. This challenge is in part due to the lack of a transparent and standardized approach to validation. Evaluation of mobile health applications (apps) is further burdened by low barriers to development and direct-to-user marketing, leading to a crowded and confusing landscape. In this context, we investigated the pragmatic application of a previously described framework for digital health validation, the Digital Health Scorecard, in a cohort of 22 popular mobile health oncology apps. The apps evaluated using this framework performed poorly, scoring 49.4% across all evaluation criteria as a group. Performance across component domains varied considerably with cost scoring highest at 100%, usability at 56.7%, technical at 37.3%, and clinical at 15.9%. Satisfaction of prospectively determined end-user requirements derived from patient, family, and clinician consensus scored 37.2%. While cost outperformed consistently and usability was adequate, the results also suggested that apps suffered from significant technical limitations, were of limited clinical value, and generally did not do what end users wanted. These large gaps further support the need for transparent and standardized evaluation to help all stakeholders in healthcare improve the quality of mobile health.

npj Digital Medicine (2021)4:111 | <https://doi.org/10.1038/s41746-021-00476-7>

INTRODUCTION

Digital health technologies, including mobile health applications (apps), hold great potential to improve American health and healthcare¹—there is nearly ubiquitous use of smartphones by Americans and an ever-growing and increasingly sophisticated suite of health apps. These apps are providing a wide range of medical functions that span the care continuum from prevention to diagnosis to care management^{2,3}. They are also increasingly demonstrating an impact on clinical outcomes with recent successes in cancer⁴, cardiac health⁵, and mental health⁶. The adoption of these digital solutions is further amplified by their accessibility, low cost, and personalized features. In addition, their ability to provide practical functions such as health education, tracking of symptoms and side effects, appointment management, and social support make them compelling healthcare tools.

While the range and number of healthcare apps directly accessible to patients continue to grow, it remains difficult to reliably discern quality, which is hindered by the lack of standardized evaluation and validation processes^{7,8}. Historically, only a fraction of health apps are appropriate for use⁹, which is concerning since patients are influenced by the health information they discover on their own⁸. In addition, common online metrics (e.g., star rating or the number of downloads) do not correlate with clinical utility or validity¹⁰. While identifying quality in mobile apps is an issue for all stakeholders in medicine, it is more acutely experienced by patients and physicians who directly seek these solutions. As the technology capabilities underpinning these apps increase, so do the potential risks. This trend makes the need for nimble evaluation even greater. Several recent studies have provided a broad overview of mobile apps in oncology^{11,12} and others have focused on provider^{13,14} and patient perspectives¹⁵ in general; however, very few have focused on the pragmatic implementation of app evaluation. Moreover, widely

cited generalized mobile app evaluation frameworks, like the Mobile App Rating Scale (MARS)¹⁶, are underdeveloped in the area of clinical appropriateness. The App Quality Assessment Tool or Health-Related Apps (AQUA)¹⁷ combines MARS with Enlight¹⁸, an assessment tool that does incorporate therapeutic concepts. However, this tool, as with its component frameworks, does not include key technical elements beyond privacy and security of data. The THESIS approach¹⁹ uses the domains of transparency, health content, technical content, security/privacy, usability, and subjective ratings. However, it does not incorporate clinical evidence. While these frameworks have advanced quality assessment, a pragmatic quality assessment tool that incorporates end-user requirements as well as technical, usability, and clinical dimensions is still needed.

This study's primary objective was to test the practical application of a previously reported validation framework⁷ for mobile apps with results presented within a digital health scorecard. Oncology apps were selected as the primary focus area since cancer patients represent a diverse population and are likely to seek out health information¹⁹. In addition, offerings in oncology are diverse and vast, with over one thousand mobile apps available^{20,21}.

RESULTS

Overall summary

The process of identification and inclusion of apps is outlined in Fig. 1. We examined a total of 18 Apple and 4 Android apps. Six were focused on education, 5 on prevention, 6 on social support, and 5 in our sample ($n = 20$, 91%) made no clinical claim, and 2 (9%) had misleading/erroneous information. The mean total score for all apps evaluated across all domains and end-user requirements was 49.4%. Individual domain scores varied with a mean

¹Memorial Sloan Kettering Cancer Center, Department of Medicine, New York, NY, USA. ²Next Generation Care Delivery, National Health Mission Area, The Johns Hopkins University Applied Physics Lab (APL), Laurel, MD, USA. ³Health Technologies, National Health Mission Area, The Johns Hopkins University Applied Physics Lab (APL), Laurel, MD, USA. ⁴Johns Hopkins Medicine, Department of Neurology, Baltimore, MD, USA. ⁵Johns Hopkins Medicine, Department of Oncology, Division of Hematology, Baltimore, MD, USA. ⁶Johns Hopkins Medicine, Department of Internal Medicine, Division of Gastroenterology, Baltimore, MD, USA. ✉email: smathe14@jh.edu

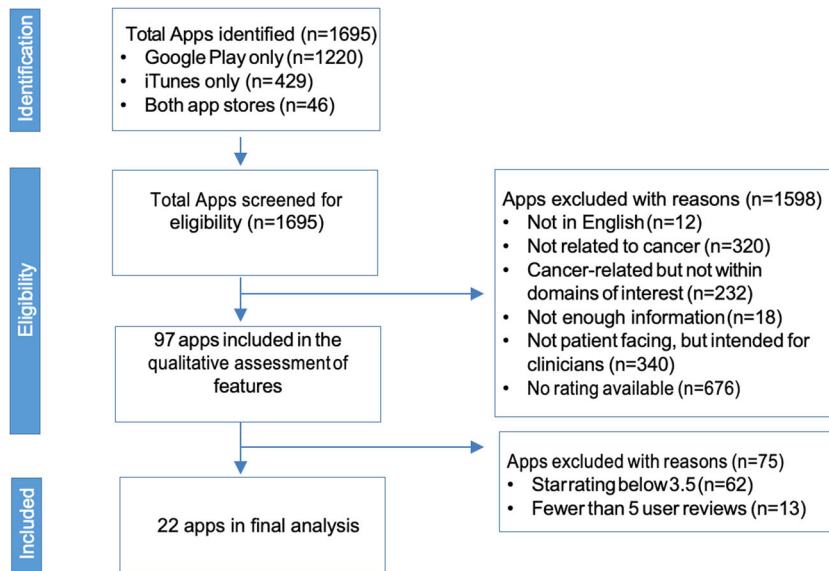


Fig. 1 Mobile app selection process. Out of 1695 apps identified, 22 apps were eligible based on exclusion criteria outlined in figure.

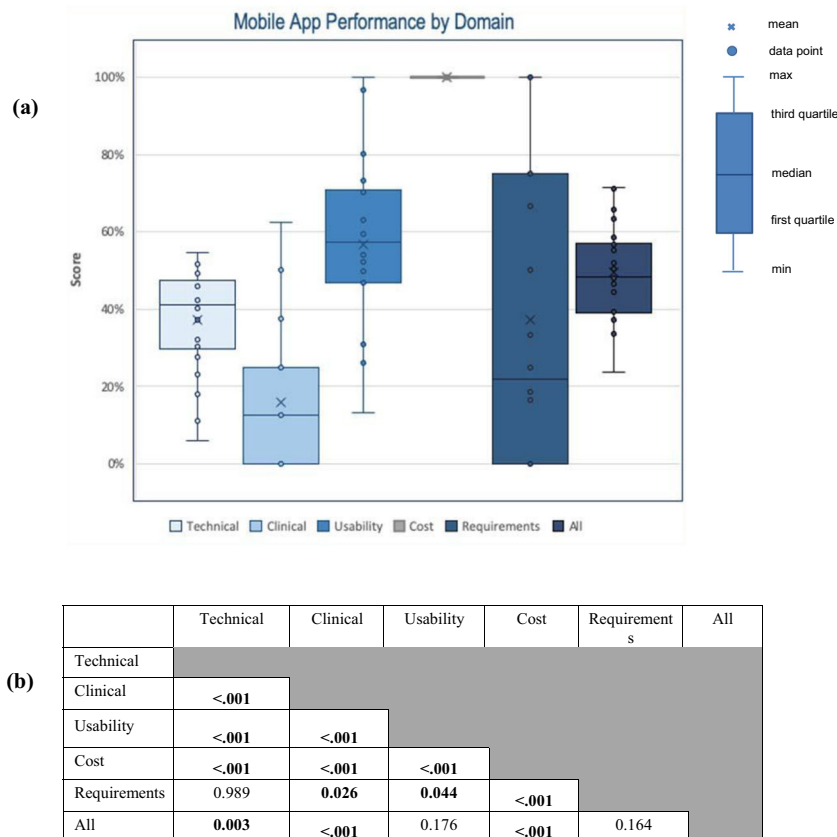


Fig. 2 Mobile app performance by domain. **a** Mobile app performance across domains of technical, clinical, usability, cost, and end-user requirements. **b** Statistical comparison of performance among domain areas and composite score with p -values shown.

score of 37.3% for technical, 15.9% clinical; 56.7% for usability, 100% for cost, and 37.2% for end-user requirements.

There was considerable inter-app variation across domain and end-user performance (Fig. 2a). The majority of inter-category score differences were statistically significant (Fig. 2b). When excluding the cost domain, the mean total score across all apps was 36.8%.

Focus areas

Performance by focus area also varied with mean scores of 51.4% for education, 36.9% for prevention, 53.1% for social support, and 55.2% for tracking with significant heterogeneity across domains and end-user requirement scores as well (Fig. 3a) with most statistical differences appreciated when comparing against the

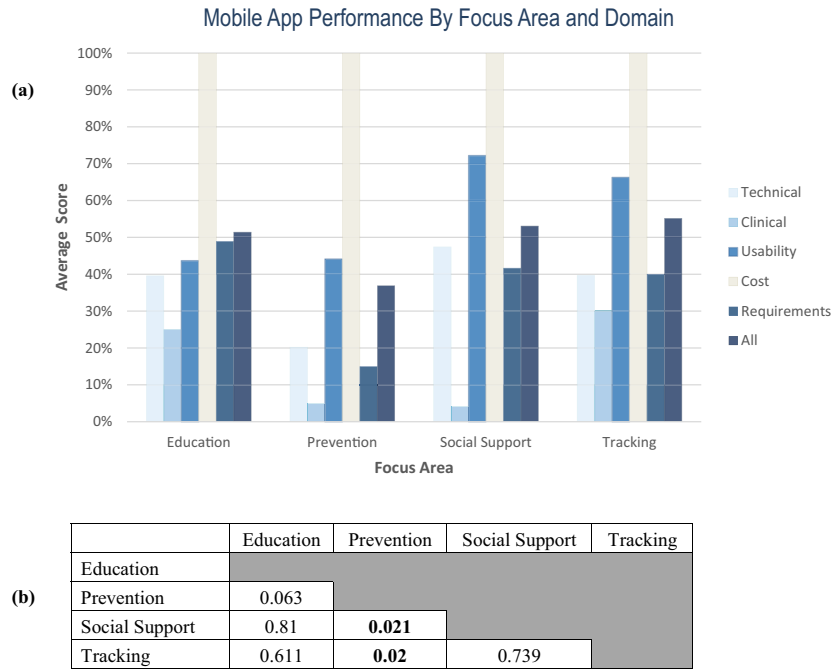


Fig. 3 Mobile app performance by focus area and domain. **a** Mobile app performance by focus area and domain areas. Mean scores across technical, clinical, usability, cost, and end-user requirement domains are represented in bar graphs. **b** Statistical comparison among focus areas.

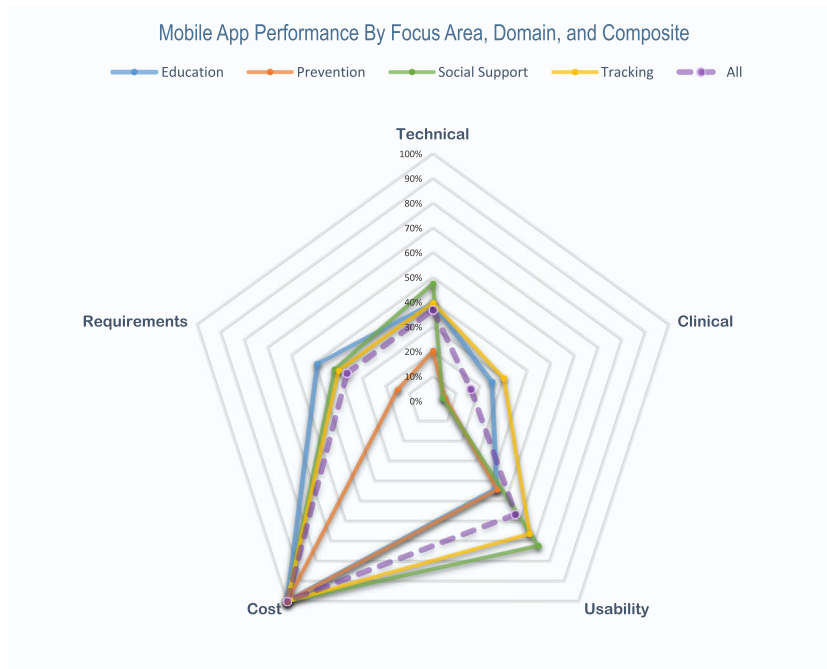


Fig. 4 Mobile app performance by focus area, domain, and composite. Distance from center to each point of the pentagon indicates performance in each respective domain domain. Colors represent app focus areas and dashed purple line represents composite performance of all apps.

prevention focus area (Fig. 3b). Compared to the composite average domain scores for all apps, the prevention category underperformed in all domains except cost, which was equal across all focus areas (Fig. 4). Tracking apps outperformed the all-app composite in all domains and end-user requirements, while education apps outperformed the average in all domains and end-user requirements, except usability. Social support apps

underperformed in the clinical domain but were otherwise above average.

Domain areas

Apps performed best in the domain of cost where all scored a maximum value of 100%. All apps tested were free. Usability was

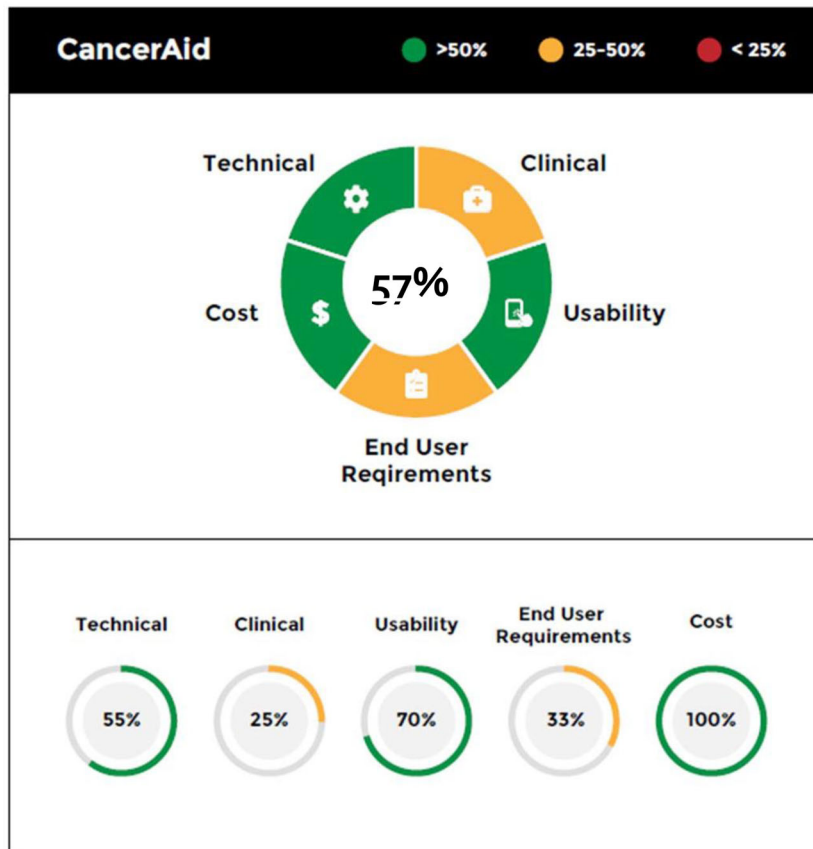


Fig. 5 Digital health scorecard, individual summary. Individual domains are depicted in a segmented wheel and color coded based on performance scores. The overall score for the app is depicted in the center of the segmented wheel.

Digital App Score Card Comparison

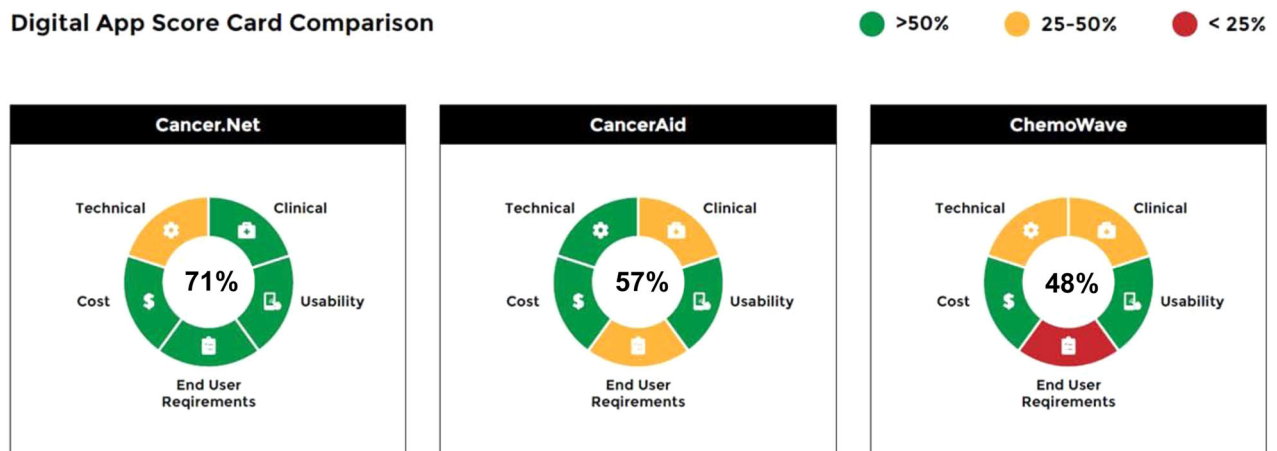


Fig. 6 Digital health scorecard, comparison view. A digital health scorecard comparison for three apps is depicted, with individual domains represented in segmented wheels, and color coded based on performance scores.

the second highest performing domain, with high subdomain scores in visual design and notifications, alerts, and alarms. The mean scores in these two categories were 68 and 73%, respectively. The technical domain's highest scoring subdomain was privacy with an average score of 48% across all apps. The remaining three subdomains scored in the low 30% range. The clinical domain had the lowest overall performance—its highest subdomain score was credibility which averaged 31%. The two apps that were scored "X" for misleading/false information were assigned a numeric value of 0 so as not to artificially inflate the

total cohort domain score by excluding them. Education apps scored the highest with respect to performance on end-user requirements with an average score of 58%, followed by social support at 42%, tracking at 40%, and prevention at 15%.

Figure descriptions

Figures 5–7 provide different visual representations of the digital health scorecard which represent composite assessments based on the scoring above. Figure 5 is a visual representation for the

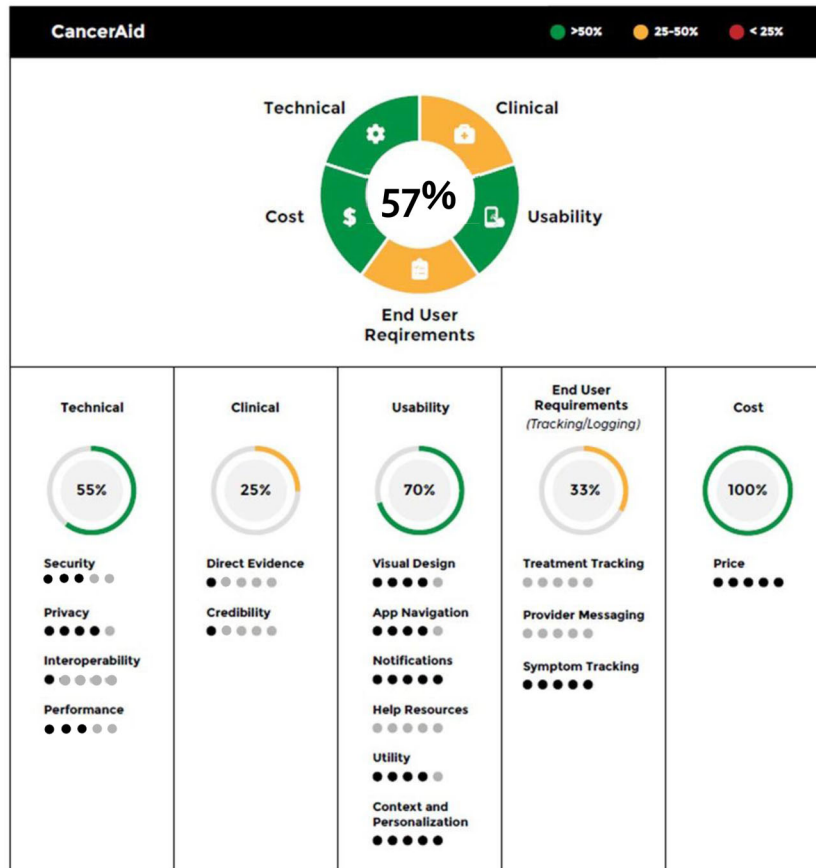


Fig. 7 Digital health scorecard, granular view. Number in wheel at top indicates overall performance with individual domain performance summarized in segmented portions of wheel. Performance by domain is detailed below the wheel with overall domain score indicated with subdomain performance included below.

individual app, CancerAid. The 4 domains and end-user requirement assessment are depicted in a segmented wheel, and color coded based on performance scores. In this example, CancerAid had the following performance scores: technical 55%, clinical 25%, usability 70%, end user requirements 33%, and cost 100%. The overall score for the app was 57%, and depicted in the center of the segmented wheel. A digital health scorecard comparison for three apps is illustrated in Fig. 6.

Figure 7 offers a more granular description of performance across each individual domain.

Subdomain scores were normalized and rounded to the nearest whole number and assigned zero to five bullets depending on performance in that subdomain. As a result of rounding, the fraction of completed bullets per domain does not always equal the exact percentage score for that domain. However, normalization was performed to facilitate easier comparisons among domains and subdomains. In the example provided, CancerAid (a tracking app) allows users to track symptoms, but does not meet the end-user requirements necessary for tracking treatment or messaging providers. Similarly, in the technical domain, CancerAid had strong security, privacy, and performance features, but interoperability was limited.

DISCUSSION

The quality of mobile health apps that physicians and patients encounter varies markedly. Our findings demonstrated that popular oncology apps generally did not meet high quality standards across multiple domains and end-user requirements. These broad deficiencies call into question whether widespread

use of these solutions is appropriate or beneficial. There was significant variability in performance across individual domains and focus areas. That the apps evaluated were popular suggests that individual domain attributes may have attracted users instead of complete solutions or that users may not be aware of app quality metrics.

Within the cohort evaluated in this study, mobile apps scored poorly in the technical and clinical domains and in end-user requirements—that is, they had major technical deficiencies, showed little clinical benefit, and did not satisfy what the users wanted them to do. With all apps being free, the cost domain did not serve as a useful discriminator of value or quality. However, the lack of cost may in part explain the popularity of these solutions. Usability scored the second highest, which may reflect the broader maturity of app design relative to attributes in other domains and may explain app utilization. For example, basic aspects of functionality, including navigation and notifications, have been rigorously iterated upon in other industries and have reached a competitive standard that is not seen in other domains. This may be a function of these attributes being directly consumer facing, whereas elements such as privacy and security, which are also broadly applicable to other industries but less visible to users, were low for most oncology apps. Apps also generally scored poorly with interoperability and performance measures. This highlights the ongoing need to more broadly implement minimum technical standards in these areas.

Clinical performance was low and driven by the low number of apps making any clinical claim. This finding likely reflects the present state of mobile apps in oncology and highlights a significant gap in the current solution landscape. In addition to

most apps not reaching a high clinical bar, some made misleading or false claims. This latter scenario presents a potentially dangerous situation particularly as apps become more sophisticated through advancements in technology such as image and movement sensors and machine learning based artificial intelligence. This presents an opportunity for medical societies, academic centers, and other trusted entities to partner with developers to produce trustworthy, high-quality, and evidence-based solutions.

The vast majority of apps did not adequately satisfy end-user requirements. The cohort average was 37.2% which highlights the significant discrepancy between what is being built for patients and physicians and what these stakeholders actually want apps to do. Incorporating these stakeholder perspectives in the product development lifecycle may help improve the overall functionality and effectiveness of the product. These results and the findings more generally demonstrated the wide disparity between what is popular and what is high quality. An initial step in bridging this gap is making the gap transparent to all stakeholders. Without systematic and transparent evaluation and a clear determination of end-user needs, patients, physicians, and other stakeholders have no reliable way to assess app quality. Our results also demonstrate that users cannot discriminate quality by using assessments available through the traditional digital gatekeepers—app stores.

There are key limitations of the framework and results described. The oncology setting and the resulting focus areas do not reflect the entire range of mobile health apps available. Thus, this study does not represent a comprehensive assessment of all oncologic use cases, settings, and needs. In addition, the end-user requirements reflected the priorities from a sample of oncology conditions and were collected from a small group of patients and providers—other stakeholders and larger, more diverse groups were not represented. In the setting where the end-user perspective from the appropriate clinical and solution specific context is characterized by real-world data, this information could serve as a more representative substitute for the design groups. Thus, assessing both more conditions (and apps) and a larger, more diverse patient population would allow more generalizable conclusions. However, the narrower approach in this study allowed an assessment of the unique needs and use cases for the apps available for evaluation. Another limitation was the inability to discriminate quality using the cost domain. This challenge may continue to persist in the general evaluation of mobile apps where price, set up, and maintenance costs are generally low or non-existent. However, other areas of health technology are unlikely to have this same degree of homogeneity, and direct integration of mobile apps into broader care delivery solutions may introduce more discriminatory capabilities in the cost domain.

There is considerable variability in how the health technology evaluation approach herein could be executed and adapted to other technologies and applications. For example, we chose criteria based on standard, best clinical practice and made subjective decisions regarding the thoroughness of evaluation (i.e., number and type of evaluation criteria). The approach we chose represented a balance between efficiency and comprehensiveness. Financial, human, and time resources will also impact the criteria used. As a result, in different stakeholder contexts, the range and depth of evaluation could vary. We recommend establishing consensus-based tiers of evaluation for each domain to maintain consistency and transparency. A minimum set of criteria may also be useful to ensure the validity of the assessment.

Regardless, making these criteria transparent and widely available will be key to achieving consensus and driving higher standards for mobile health app development.

The digital health scorecard and its underlying evaluation framework offer a pragmatic, requirements-driven, and impact-

focused approach to evaluating the digital health landscape. Our study provides a practical demonstration for implementing standards-based assessment that produces accessible and meaningful results for clinicians and patients. While future research is still needed to validate the digital health scorecard approach, this study is a formative step in advancing a theoretical concept to a real-world application. The road to validating digital health will take resources, collaboration, and time. We believe an important next step is to establish an evaluation process that reflects the real-world needs of stakeholders in healthcare. The digital health scorecard represents one possible path.

METHODS APP SELECTION PROCESS

A summary of the digital health scorecard approach below is outlined in Fig. 8. We first used a structured search query using 6 search terms to identify cancer related apps on US Android Google Play and Apple iTunes store. The search terms used were “cancer”, “cancer support”, “cancer care”, “breast cancer”, “prostate cancer”, and “best cancer apps”. Apps were eligible to be included in this review if they met all of the following criteria: (1) apps that aimed to support patients, (2) apps related to cancer, (3) apps with a star rating, (4) apps that were in English, and (5) apps that were targeted into one of our four domains of interest. These domains of interest were identified as most important through stakeholder engagement with patients and clinicians in an iterative process, and include prevention, management, social support, and education. This resulted in a preliminary review of 97 apps by a multidisciplinary team. We aimed to include apps that were used by a large number of patients or individuals, and further excluded apps that had a star rating below 3.5 and/or fewer than 5 user reviews. We then identified 22 oncology apps for substantive evaluation based on the composite ranking of the highest ratings and number of reviews as of July 15, 2019. Information about this apps retrieved from the search was entered into an electronic spreadsheet. The information entered included name of the app, name of the developer, cost, app store or stores in which the app was available, and the search term or terms that retrieved the app. *App Evaluation Process* App validation was then performed by applying a previously described framework⁷, which encompassed evaluation across the following domains: technical, clinical, usability, and cost. In addition, per the framework, end-user requirements were identified and assessed. The app focus areas and end-user requirements were determined and validated through design thinking sessions with key stakeholders detailed below. Each app was downloaded and tested by two independent app developers who completed the technical, usability, and end-user requirement evaluations for each app. A group of three physicians performed the clinical domain evaluation—including two hematology–oncology specialists. Purchase price was recorded at the time of download for all apps. The domains of evaluation are summarized briefly below. A detailed summary of the individual elements of evaluation are provided in Supplement 1.

Technical validation encompassed evaluation of security, privacy, interoperability, and performance⁷. The scoring for the technical domain included 25 individual criterion, which were scored 0 (none) to 2 (full) for a maximum domain score of 50.

Clinical validation referred to an assessment of evidence supporting any clinical claims using an approach based on the GRADE system²². Scoring for the clinical domain included the two subdomains of direct evidence and credibility. For the former, if the technology did not make a clinical claim to impact a health outcome, a score of 0 was given. If the technology provided any false or misleading clinical information, a score of “X” was given. If a clinical claim was made, a maximum of 4 points were given if the technology was supported by a randomized clinical trial or meta-analysis, and a minimum score of 1 if the clinical claim was made

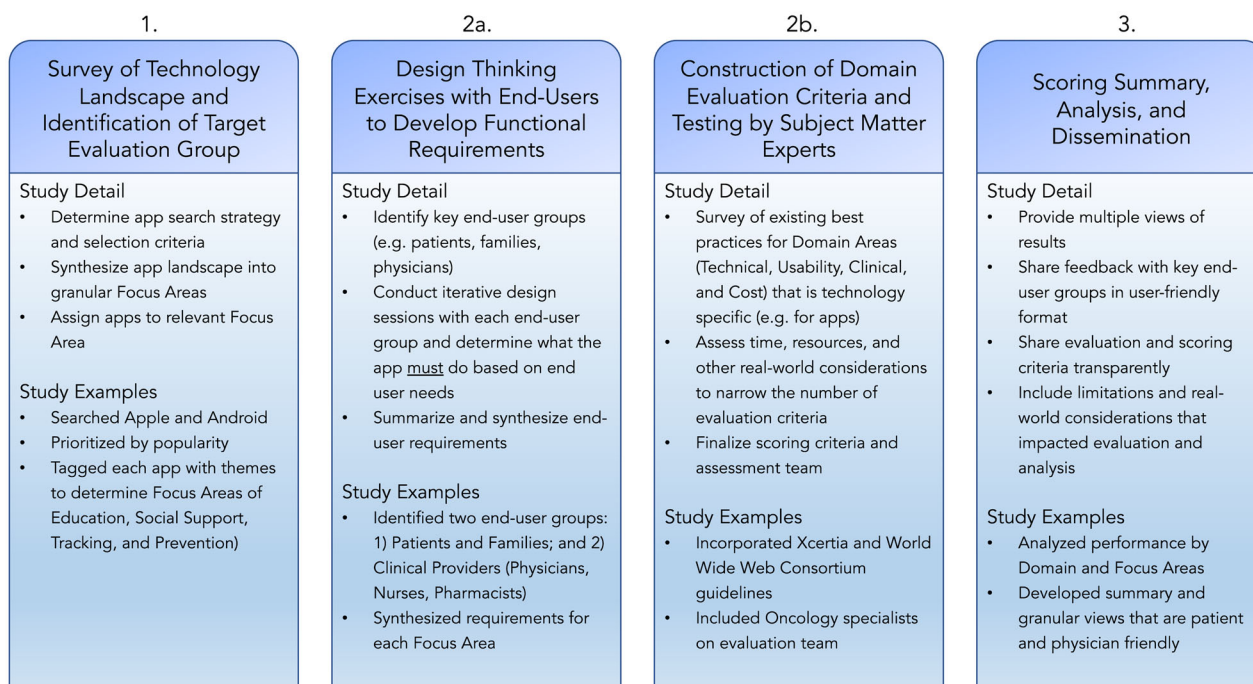


Fig. 8 Summary of digital health scorecard approach. Phases of evaluation are depicted, beginning with technology target assessment, followed by determination of requirements, development of assessment criteria, and finally, scoring and analysis.

based on expert opinion alone. Credibility was approximated by appraising any recognition by known affiliations, associations, or endorsements. Credibility scores range from a maximum of 4 (produced by recognized medical society or organization) to a minimum of 0 (not produced or recognized by a medical society, organization, institution, team, or individual). A maximum score for this domain was 8.

Usability assessment focused on whether the mobile technology was patient-oriented and the degree to which the app design followed good principles for positive user experience. The app was evaluated to determine ease of use for its intended purpose, the effort required to complete tasks, data entry burden, and whether users were allowed to control preferences such as notifications when appropriate. This perspective aligned with the World Wide Web Consortium (W3C) guidelines, which summarized a user-centered design process (UCD)²³. The key objectives state apps should be useful in helping users achieve their goals, effective (i.e., producing results with minimal user error), learnable (i.e., easy and intuitive to use), and likeable (i.e., enjoyable to use). Usability frameworks have been in place since the early days of software applications, most notably the System Usability Scale (SUS)²⁴, which has been used for 35 years. However, usability assessed with SUS does not include consideration of mobile app design specific concepts, or address user likeability. For our usability assessment, two digital health-specific evaluation frameworks were analyzed against the overarching objectives W3C, the Xcertia Guidelines²⁵, and the Node.Health User Experience (UX) Framework²⁶, both published in 2019. A pragmatic approach was taken based on incorporating the strongest elements of each framework, augmented with essential elements of consumer-based frameworks²⁷. The resulting scoring for the usability domain was divided into six subdomains (visual design and readability; app navigation; notifications, alerts, alarms; help resources; utility; and context and personalization). A total of 22 individual items were scored, with a range from 0 to 5 (maximum score = 110).

Cost in the validation framework incorporated a number of variables including face-value price, time for set up and use, additional training, and maintenance expenses. However, this

domain was simplified to price alone since the remaining variables did not apply or did not add any additional discriminatory value.

End-user requirements were determined through design thinking exercises conducted with end users, including patients with cancer, their caregivers, physicians, nurses, and pharmacists²⁸. These sessions determined the attributes most important to the end user for each focus area (education, social support, tracking, prevention). Education apps were scored on whether content was personalized and whether it was presented in easy to understand terms. Social support apps were scored on their ability to connect patients with community resources and an online community of similar patients. Tracking apps were scored on their ability to track individual treatment plans, their ability to connect patients with their healthcare team, and their ability to track symptoms over time. Prevention apps were expected to provide personalized risk assessments and risk mitigation techniques. Only the requirements in the focus area applicable to each app was scored.

Scoring and statistical approach

A full schematic of the scoring sheet is provided in Supplement 1. The total score for each app, which is shown as a percentage of total possible points, provides equal weight to each of the evaluation domains (technical, clinical, usability, cost, and end-user requirements). Likewise, subdomains within each of the evaluation domain were normalized and treated with equal weights. This allowed scoring to be independent of how many evaluation criteria were used for each subcategory. For example, the composite technical score was based on the weighted sum of normalized underlying subdomain scores: security (20), privacy (14), interoperability (10), and performance (6). Each subdomain score was calculated and weighted equally when determining the total average for the technical domain. Statistical comparisons of mean scores between domains categories and between focus areas were conducted using the two-tailed, unpaired *t*-test.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Upon reasonable request, our team will provide aggregated data used in the findings of this study.

Received: 8 October 2020; Accepted: 8 June 2021;

Published online: 15 July 2021

REFERENCES

- Pew Research Center. Mobile fact sheet. Available at: <https://www.pewresearch.org/internet/fact-sheet/mobile/>. Accessed September 22, 2020.
- Krebs, P. & Duncan, D. T. Health app use among US mobile phone owners: a national survey. *Jmir Mhealth Uhealth* **3**, 107–119 (2015).
- Cohen, A. B., Dorsey, E. R., Mathews, S. C., Bates, D. W. & Safavi, K. *npj Digit. Med.* **3**, Article number: 68 (2020).
- Larson, J. L., Rosen, A. B. & Wilson, F. A. The effect of telehealth interventions on quality of life of cancer patients: a systematic review and meta-analysis. *Telemed. E-Health* **24**, 397–405 (2018).
- Kotecha, D. et al. European Society of Cardiology smartphone and tablet applications for patients with atrial fibrillation and their health care providers. *Europace* **20**, 225–233 (2018).
- Barnett, I. et al. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology* **43**, 1660–1666 (2018).
- Mathews, S. C. et al. Digital health: a path to validation. *Npj Digit. Med.* **2**, Article number: 38 (2019).
- Mathews S. C. et al. Reply: Some considerations on digital health validation. *npj Digit. Med.* **2**, Article number: 103 (2019).
- Byambasuren, O., Sanders, S., Beller, E. & Glasziou, P. Prescribable mHealth apps identified from an overview of systematic reviews. *npj Digit. Med.* **1**, Article number: 12 (2018).
- Singh, K. et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. *Health Aff.* **35**, 2310–2318 (2016).
- Ana, F. A. et al. Mobile applications in oncology: a systematic review of health science databases. *Int. J. Med. Inform.* **133**, 104001 (2020).
- Jupp, J. C., Sultani, H., Cooper, C. A., Peterson, K. A. & Truong, T. H. Evaluation of mobile phone applications to support medication adherence and symptom management in oncology patients. *Pediatr. Blood Cancer* **65**, e27278 (2018).
- Kessel, K. A., Vogel, M. M., Schmidt-Graf, F. & Combs, S. E. Mobile apps in oncology: a survey on health care professionals' attitude toward telemedicine, mHealth, and oncological apps. *J. Med. Internet Res.* **18**, e312 (2016).
- Berkowitz, C. M., Zullig, L. L., Koontz, B. F. & Smith, S. K. Prescribing an app? Oncology providers' views on mobile health apps for cancer care. *JCO Clin. Cancer Inform.* **1**, 1–7 (2017).
- Kessel, K. A. et al. Mobile health in oncology: a patient survey about app-assisted cancer care. *JMIR mHealth uHealth* **5**, e81 (2017).
- Stoyanov, S. R. et al. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth uHealth* **3**, e27 (2015).
- O'Rourke, T., Pryss, R., Schlee, W. & Probst, T. Development of a multidimensional App-Quality Assessment Tool for Health-Related Apps (AQUA). *Digit. Psychol.* **1**, 13–23 (2020).
- Baumel, A., Faber, K., Mathur, N., Kane, J. M. & Muench, F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. *J. Med. Internet Res.* **19**, e82 (2017).
- Levine, D. M. et al. Design and testing of a mobile health application rating tool. *npj Digit. Med.* **3**, Article number: 74 (2020).
- Clauser, S. B., Wagner, E. H., Bowles, E. J. A., Tuzzio, L. & Greene, S. M. Improving modern cancer care through information technology. *Am. J. Prev. Med.* **40**, 198–207 (2011).
- IQVIA Institute. Global Oncology Trends 2018. Available at: <https://www.iqvia.com/institute/reports/global-oncology-trends>. Accessed September 22, 2020.
- Atkins, D. et al. Grading quality of evidence and strength of recommendations. *BMJ* **328**, Article 1490 (2004).
- World Wide Web Consortium (W3C) Web Accessibility Initiative (WAI). Web Content Accessibility Guidelines. Available at: <https://www.w3.org/WAI/standards-guidelines/wcag/>. Accessed September 22, 2020.
- Usability.gov Improving the User Experience. Available at: <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>. Accessed February 10, 2021.
- HIMSS. 2019 Board Approved Xcertia Guidelines Issued on - HIMSS. Available at: <https://www.himss.org/sites/hde/files/media/file/2020/04/17/xcertia-guidelines-2019-final.pdf>. Accessed September 22, 2020.
- Node.Health User Experience (UX) Measures. Available at: <https://nodehealth.org/ux-measures>. Accessed September 22, 2020.
- Norton DWPD, Digital Context 2.0, Gifted Press, LLC, 2015, Chapter 1, and 3.
- Garrubba, M. J. C., Yap, G. & Melder, A. Design thinking methodology to improve the patient experience: an evidence snapshot. Centre for Clinical Effectiveness, Monash Innovation and Quality, Monash Health, Melbourne, Australia (2016).

ACKNOWLEDGEMENTS

The authors would like to thank Jasmine McNeil, Patrick Ostendarp, and the entire Johns Hopkins Technology Innovation Center team for their invaluable help executing this study. The authors also thank the patients, family members, and hospital staff that participated in the user requirements sessions. This research received funding from Genentech grant.

AUTHOR CONTRIBUTIONS

All authors meet the criteria detailed in Author Instructions. S.C.M. was responsible for concept and design; analysis and interpretation of data; creation of figures and tables; drafting and revision of manuscript; final approval; and accountability for the work. R.S., M.J.M., A.B.C. and J.A.W. were responsible for elements of design and analysis and interpretation of data; drafting and revision of manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-021-00476-7>.

Correspondence and requests for materials should be addressed to S.C.M.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021