

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Evaluating Evaluation Frameworks: A Review of Frameworks for Assessing Health Apps

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-047001
Article Type:	Original research
Date Submitted by the Author:	17-Nov-2020
Complete List of Authors:	Lagan, Sarah; Beth Israel Lahey Health Sandler, Lev; Beth Israel Lahey Health Torous, John; Beth Israel Lahey Health
Keywords:	PSYCHIATRY, Telemedicine < BIOTECHNOLOGY & BIOINFORMATICS, Information management < BIOTECHNOLOGY & BIOINFORMATICS

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Evaluating Evaluation Frameworks: A Review of Frameworks for Assessing Health Apps

Sarah Lagan BA¹, Lev Sandler BA¹, John Torous MD*¹

Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School

*

Division of Digital Psychiatry
Rabb 2 Building, Department of Psychiatry
Beth Israel Deaconess Medical Center
Boston, MA, 02115, USA
jtorous@bidmc.harvard.edu

Word Count: 3,149

Objectives

Despite an estimated 300,000 mobile health apps on the market, there remains no consensus around helping patients and clinicians select safe and effective apps. In 2018, our team drew upon existing evaluation frameworks to identify salient categories and create a new framework endorsed by the American Psychiatric Association (APA). We have since created a more expanded and operational framework (MIND) that aligns with the APA categories but includes objective and auditable questions (105). We sought survey the existing space, conducting a review of all mobile health app evaluation framework published since 2018, and demonstrate the comprehensiveness of this new model by comparing it to existing and emerging frameworks.

Methods

We conducted a systematic review of mobile health app evaluation frameworks and compiled a collection of more than 1,701 questions across 79 frameworks. We compared and grouped these questions using the MIND framework as a reference. We sought to identify the most common domains of evaluation while assessing the comprehensiveness and flexibility- as well as any potential gaps- of MIND.

Results

New app evaluation frameworks continue to emerge and expand. Since our 2019 review of the app evaluation framework space, more frameworks include questions around privacy (43) and clinical foundation (57), reflecting an increased focus on issues of app security and evidence base. The majority of mapped frameworks overlapped with at least half of the MIND categories. The results of this search have informed a database (apps.digitalpsych.org) that users can access today.

Conclusion

As the number of app evaluation frameworks continues to rise, it is becoming difficult for users to select both an appropriate evaluation tool and to find an appropriate health app. This review provides a comparison of what different app evaluation frameworks are offering, where the field is converging, and new priorities for improving clinical guidance.

Strengths and limitations of this study

- We conducted a systematic review of mobile health app evaluation frameworks.
- Searches of three databases, as well as the grey literature and government websites, were conducted.
- 79 health app evaluation frameworks met the inclusion criteria and questions from each framework were compiled to create a list of 1,701 questions.
- The 1,701 questions were compared and mapped to a reference framework (MIND) in order to identify commonalities and gaps in categories of consideration.

Introduction

The past five years have seen a proliferation of both mobile health apps and proposed tools to rate such apps. While these digital health tools hold great potential, concerns around privacy, efficacy, and credibility, coupled with a lack of strict oversight by governing bodies, have highlighted a need for frameworks that can help guide clinicians and consumers to make informed app choices. Although the United States' Food and Drug Administration (FDA) has recognized the issue and is piloting a pre-certification program that would prioritize app safety at the developer level,¹ this model is still in pilot stages and there has yet to be an international consensus around standards for health apps, resulting in a profusion of proposed frameworks across governments, academic institutions, and commercial interests.

In 2018, our team drew upon existing evaluation frameworks to identify salient categories from existing rating schemes and create a new framework.² The American Psychiatric Association's (APA) App Evaluation Model was developed by harmonizing questions from 45 evaluation frameworks and selecting 38 total questions that mapped to five categories: background information, privacy and security, clinical foundation, ease of use, and interoperability. This APA model subsequently has been utilized by many diverse stakeholders given its flexibility in guiding informed decision making.^{3,4,5,6,7} However, the flexibility of the model also created a demand for a more applied approach that offered users more concrete information instead of placing the onus entirely on a clinician or provider.

Patient and Public Involvement

Thus, since the framework's development, the initial 38 questions have been operationalized into 105 new objective questions that invite a binary (yes/no) or numeric response by a rater.⁸ These questions align with the categories proposed by the APA model but are more extensive and objective, with, for example, "app engagement" operationalized into eleven different engagement styles to select. Questions were created with input of clinicians, patients, family members, researchers, and policy makers. These 105 questions are sorted into six categories (App Origin and Functionality, Inputs and Outputs, Privacy and Security, Clinical Foundation, Features and Engagement, Interoperability and Data Sharing) are intended to be answerable for any trained rater – clinician, peer, end user –and inform the public facing Mhealth Index and Navigation Database (MIND), where users can view app attributes and compare ratings (see figure 1 below).

FIGURE 1

Like the APA model, MIND shifts the app evaluation process away from finding one "best" app, and instead guiding users towards an informed decision based on selecting and placing value on the clinically-relevant criteria that account for the needs and preferences of each patient and case. The goal is not for a patient or clinician to consider all 105 questions but rather be able to access a subset of questions that appear most appropriate for the current use case at hand. Thus, MIND

1
2
3 offers a useful tool to compare evaluation frameworks to given its development derived from
4 other frameworks and composition of discrete questions that aim to be objective and
5 reproducible. It also offers an actionable resource for any user anywhere in the world to engage
6 with app evaluation, providing tangible results in the often more theoretical world of app
7 evaluation.
8
9

10 Recent systematic reviews have illustrated the growing number of evaluation tools for digital
11 health devices, including mobile health apps.^{9,10,11} Given the rapidly evolving health app space
12 and the need to understand what aspects are considered in evaluation frameworks, we have
13 sought to survey the landscape of existing frameworks. Our goal was to compare the categories
14 and questions of from others frameworks to 1) identify common elements between them, 2)
15 identify if gaps in evaluation frameworks have improved since 2018, and 3) and assess how
16 reflective our team's MIND framework is in the current landscape. We thus aimed to map every
17 question from the 2018 review, as well as questions from new app evaluation frameworks that
18 have emerged since, using the questions of MIND as a reference. While informing our own
19 efforts around MIND, the results of this review offer broad relevance across all of digital health,
20 as understanding the current state of app evaluation helps inform how any new app may be
21 assessed, categorized, judged, and adopted.
22
23
24
25
26

27 **Methods**

28
29 We followed a three-step process in order to identify and compare frameworks to MIND. This
30 process included (1) assembling all existing frameworks for mobile medical applications, (2)
31 separating each framework into the discrete evaluation questions comprising it, and (3) mapping
32 all questions to the 105 MIND framework questions as a reference.
33
34

35 *Search Strategy and Selection Criteria*

36 We started with the 45 frameworks identified in the 2018 review by Moshi et al.⁹ and included
37 34 frameworks that have emerged since our initial analysis of the space that was conducted in
38 2018 and published in 2019.² To accomplish this, we conducted an adapted systematic review
39 based on the Moshi criteria to identify recent frameworks. Although MIND focuses on mental
40 health apps, its considerations and categories are transferable to health apps more broadly, and
41 thus there was no mental health specification in the search terms.
42
43

44
45 References were identified through searches of PubMed, EMBASE, and PsychINFO with the
46 search terms ((mobile application) OR (smartphone app)) AND ((framework) OR (criteria) OR
47 (rating)) and publication date between January 2018 through October 2020. We also identified
48 records beyond the database search by seeking frameworks mentioned in subsequent and recent
49 reviews^{5,12,13} and surveying the grey literature and government websites. Papers were selected
50 for inclusion if they met the predetermined eligibility criteria -- presenting an evaluation
51 framework for mobile health apps with patient, clinician, or end user facing questions. Two
52 reviewers (SL and JT) screened the literature separately and applied the inclusion criteria. The
53 data extracted from the papers included: Author and dates of publication, source affiliation,
54 country of origin, name of framework, study design, description of framework, intended
55
56
57
58
59
60

audience/user, and framework scoring system. Articles were screened if they described the evaluation of a single app, did not present a new framework (instead conducting a review of the space or relying upon a previous framework), the framework was focused on developer instead of clinicians or end users, was implementation and not evaluation focused, was not a framework for health apps, and was a satisfaction survey instead of an evaluation framework. The data selection process is outlined in figure 2.

FIG 2

The 34 frameworks identified in the search were combined with the 45 frameworks from the 2018 review for a total of 79 frameworks for consideration. To our knowledge, this list comprehensively reflects the state of the field at the time of assembly. However, we do not claim it to be exhaustive, as frameworks are constantly changing, emerging, and sunsetting, with no central repository. The final list of frameworks assembled can be found in Appendix A.

Mapping

Each resulting framework was reviewed and compiled into a complete list of the it's unique questions. The 79 frameworks yielded 1,701 questions in total. Several of the original 45 frameworks focused exclusively on in-depth privacy considerations (evaluating privacy and security practices rather than the app itself),¹⁴ and after eliminating these checklists that did not facilitate app evaluation by a clinician or end user, 70 total frameworks were mapped in entirety to the MIND framework.

In mapping questions, clinical judgment was necessary as not every question was an exact, word-for-word match. The authors thus used discretion when it came to matching questions to MIND, and discussed each decision to confirm mapping placement. Disputes were brought to a third reviewer for final consideration. "Is data portable and interoperable?"¹⁵, for example, would be mapped to the question "can you email or export your data?" "Connectivity"¹⁶ was mapped to "Does the app work offline?" and "Is the arrangement and size of buttons/content on the screen zoomable if needed"¹⁷ was mapped to "is there at least one accessibility feature?" Questions about suitability for the "target audience" were mapped to the "patient facing" question in MIND.

Results

Framework Type

The aim of this review was to identify and compare mobile health app rating frameworks, assessing overlap and exploring changes and gaps relative to both previous reviews and to the MIND framework. Of the 70 frameworks ultimately assessed and mapped, the majority 39 (55.7%) offered models for evaluating mobile health apps broadly. Seven (10%) considered mental health apps, while six (8.5%) focused on apps for diabetes management. Other evaluation

focuses included apps for asthma, autism, concussions, COVID-19, dermatology, eating disorders, heart failure, HIV, pain management, infertility, and sickle cell disease (table 1).

Focus of Evaluation	Number of Frameworks
Asthma	2 (2.86%)
Autism	1 (1.43%)
Concussion	1 (1.43%)
COVID-19	1 (1.43%)
Dermatology	1 (1.43%)
Diabetes	6 (8.57%)
Eating Disorders	1 (1.43%)
General Mobile Health	39 (55.71%)
Heart Failure/Cardiac Rehabilitation	3 (4.29%)
HIV	2(2.86%)
Mental Health	7 (10%)
Pain Management	2 (2.86%)
Pediatric or Adolescent Health	2 (2.86%)
Reproductive Endocrinology/Infertility	1 (1.43%)
Sickle Cell Disease	1 (1.43%)

Table 1: *Number of disease specific and general app evaluation frameworks, with general mobile health frameworks constituting more than half of identified frameworks.*

Mapping

In total, we reviewed 1,701 questions from 79 different app frameworks, comparing them to the MIND framework as a reference point from which to compare and assess overlap and gaps. We mapped questions from 70 app evaluation frameworks against the six categories and 105 questions of MIND (see appendix B). We examined the number of frameworks that addressed each specific MIND category and identified areas of evaluation that are not addressed by MIND. Through the mapping process, we were able to gauge the most common questions and categories across different app evaluation frameworks.

We sorted the questions into MIND's six different categories – App Origin & Functionality, Inputs & outputs, Privacy & Security, Evidence & Clinical Foundation, Features & Engagement Style, and Interoperability & Data Sharing – in order to assess the most common broad areas of consideration. Across frameworks, the most common considerations were around privacy/security and clinical foundation, with 43 frameworks posing at least one question around the app's privacy protections, and 57 of the frameworks containing at least one question to evaluate evidence base or clinical foundation, as denoted in table 2. Fifty-nine frameworks covered at least two of the MIND categories, with the majority of frameworks overlapping with at least four of MIND categories.

Category of Evaluation	Number of Frameworks Addressing Category
App Origin / App Functionality	42
Inputs and Outputs	17
Privacy / Security	43
Evidence / Clinical Foundation	57
Features / Engagement Style	29
Interoperability and Data Sharing	23

Table 2: *The questions from all frameworks were mapped to the reference framework (MIND) sorted into its six categories, with this table denoting how many frameworks had questions that could be sorted into each of the categories.*

We then took a more granular look at the questions from each of the 70 frameworks, matching questions one-by-one to questions of the MIND framework when possible. On an individual question level, specific questions about the presence of a privacy policy, security measures in place, supporting studies, and patient-facing (or target population) tools were the most prevalent, with representation from 20, 25, 27, and 28 frameworks respectively for each question. Each of the 70 frameworks had at least one question that mapped to MIND. The most common questions, sorted into their respective categories, are depicted in Figure 3 and Table 3, while the full list of mapped questions can be found in Appendix B.

FIG 3

Every question was examined but not every question in every framework could be matched to a corresponding question in MIND, and some questions fell outside one of the six categories. For example, 18 frameworks continue to present the subjective question of “is the app easy to use” which will vary depending on the person and use case. MIND also does not offer questions related to other objective questions to which answers are not readily available such as “How were target users involved in the initial design and usability evaluations of the app?”¹⁸ While questions such as this are of high importance, lack of easily accessible answers creates a dilemma in their present utility for app evaluation. Furthermore, some questions such as economic analysis were not covered by MIND but by other frameworks and represent a similar dilemma in that actual data to base evaluation on is often lacking. Aside from subjective questions, other pronounced absences MIND were questions about customizability (addressed by 7 other frameworks) and advertising (9 frameworks). Although MIND does ask about customizability in part by encouraging raters to consider accessibility features (and some frameworks ask about the ability to customize in conjunction with accessibility features¹⁹), MIND does not pose a question around the user’s ability to tailor or customize app content, nor does it ask questions about the presence of advertisements on an app. Other questions unaddressed by MIND were about the user’s ability to contact the producer or developer to seek guidance about app use. Variations of this question include “is there a way to feedback user comments to the app developer?” MIND also does not pose any questions regarding instructions in the app or the existence of a user guide.²⁰ Finally, it does ask about speed of app functionality. This variant of question asks, “is the app fast and easy to use in clinical settings?”¹⁵ Figure 3

above, and Table 3 below present additional details on categories and questions both inside and outside the MIND reference framework.

Question	Number of Frameworks Addressing Question
Where does the app come from / who is the app developer?	26
How much does the app cost?	16
Does the app have at least one accessibility feature?	11
Is the app available on multiple platforms?	9
Has the app been updated recently / when was the last update?	9
How many downloads does the app have / how many reviews?	9
Does the app work offline?	7
Is there a privacy policy?	20
Does the app declare data use and purpose?	18
Does the app report security measures in place?	25
Does the app claim to meet HIPAA (or an analogous health data protection regulation)?	10
Is the app content well-written, correct, and relevant?	25
Does the app do what it claims?	18
Is the app patient facing?	28
Is there evidence of the app's efficacy?	27
Is the app collaborative with a provider?	8
Does the app offer gamification?	4
Can you email or export your data?	10
Is the app easy to use?	18
Is the app visually appealing?	10
Is the app age appropriate?	4
Does the app allow for customization and/or personalization?	7
Was the app developed in collaborated with target users?	4
Is there a manual or set of instructions for the app?	8
Is there any advertising within the app?	9
Risk/Benefit Analysis	2
Speed of Functionality	5
Is there a way to contact the developer with questions or concerns?	6

Table 3: Commonly addressed questions among those that could be mapped to the MIND reference framework (blue), and those that could not (green).

Discussion

1
2
3 As mobile health apps have proliferated, choosing the right one has become increasingly
4 challenging for patient and clinician alike. While app evaluation frameworks can help sort
5 through the myriad of mobile health apps, the growing number of frameworks further
6 complicates the process of evaluation. Our review examined the largest number of evaluation
7 frameworks to date with the goal of assessing their unique characteristics, gaps, as well as
8 overlap with the 105 questions in MIND. We identified frameworks for evaluating a wide range
9 of mobile health apps – some focused on general mobile health, some specific and addressing
10 specific disease domains like asthma, heart failure, mental health, or pain management.
11 Despite the different disease conditions they addressed, there was substantial overlap among the
12 frameworks, especially around clinical foundation and privacy and security. The most common
13 category addressed was clinical foundation, with 57 of the evaluation frameworks posing at least
14 one question regarding evidence base. More than half of the frameworks also addressed privacy
15 and/or security and app functionality or origin.
16
17
18

19 The widespread focus on clinical foundation and privacy represents a major change in the space
20 since 2018, when our team analyzed an initial review of 45 health app evaluation frameworks
21 and found that the most common category of consideration among the different frameworks was
22 usability, with short term usability highly overrepresented compared to privacy and evidence
23 with base. In this 2018 review, there were 93 unique questions corresponding to short-term
24 usability but only 10 to the presence of a privacy policy. Although many frameworks continue
25 to consider usability, our current review suggests the most common questions across frameworks
26 now concern evidence, clinical foundation, and privacy. This shift may reflect an increased
27 recognition of the privacy dangers some apps may pose.
28
29
30

31 This review illustrates the challenges in conceiving a comprehensive evaluation model. A
32 continued concern in mobile health apps is engagement,⁶ and it is unclear whether any
33 framework adequately predicts engagement. Another persistent challenge is striking a balance
34 between transparency/objectivity and subjectivity. Questions that prompt consideration of
35 subjective user experiences may limit the generalizability and standardization of a framework, as
36 the questions inherently reflect the experience of the rater. An app's ease of use, for example,
37 will differ significantly depending on an individual's level of comfort and experience with
38 technology. However, subjective questions around user friendliness, visual appeal, and interface
39 design may be of greatest concern to an app user, and most predictive of engagement with an
40 app.²¹ Finally, a thorough assessment of an app is only feasible if information about the app is
41 available. For example, some questions with clinical significance, such as the consideration of
42 how peers or target users may be involved in app development, are not easily answerable by a
43 health app consumer. Overall, there is a need for more data and transparency when it comes to
44 health apps. App evaluation frameworks, while thorough, rigorous, and tailored to clinical app
45 use, can only go so far without transparency on the part of app developers.²²
46
47
48

49 The analysis additionally highlighted the flexibility and comprehensiveness of the MIND
50 framework, which was used as a reference framework in this review, in diverse contexts. The
51 MIND categories are inclusive of a wide range of frameworks and questions. Even without
52 including any subjective questions in the mapping process, each of the 70 frameworks that were
53 ultimately mapped had some overlap with MIND, and many of the 1,700 questions ultimately
54 included were mapped exactly with a MIND question. Although MIND was initially
55
56
57
58
59
60

conceptualized as an evaluation tool specifically for mental health apps, the coherence between MIND and diverse types of app evaluation frameworks, such as those for concussion²³, heart disease²⁴, and sickle cell anemia²⁵, demonstrates how the MIND categories can encompass many health domains. Condition-specific questions, for example, are a good fit for the “Features & Engagement” category of MIND.

The results of our analysis suggest while numerous new app evaluation frameworks continue to emerge, there is a naturally appearing standard of common questions asked across all. While different use cases and medical subspecialties will require unique questions to evaluate apps, there are set of common questions around aspects like privacy and level of evidence that are more universal. MIND appears to cover a large subset of these questions and thus may offer a useful starting point for new efforts as well as means to consolidate exiting efforts. Advantages of the more objective approach offered by MIND is that it can be represented as a research database to facilitate discovery of apps while not conflicting with local needs, personal preferences, or cultural priorities.²⁶

Limitations

Our work is not the first to compare app evaluation frameworks. Recently, several reviews have compared how different mobile health app evaluation models address privacy,¹²¹⁴¹¹ and another database (<https://search.appcensus.io/>) focuses exclusively on compiling privacy assessments of Android apps. We chose to exclude app evaluation frameworks that focused exclusively on in-depth privacy considerations and were unusable by a clinician or layperson as our goal was more comprehensive app evaluation. This decision is not to reject considerations of privacy and security that are of critical importance, but rather to narrow the focus to frameworks that are usable in the hands of the public today and can be used to inform clinical decisions. In addition, MIND was initially tailored to mental health, and thus does not encompass thorough disease-specific criteria for other conditions, such as asthma, diabetes, and sickle cell anemia – though such questions may be easily integrated. Finally, subjective questions, especially those around ease of use and visual appeal, are difficult to standardize but may be among the most important features driving user engagement with mental health apps.²¹

Conclusion

Our work demonstrates expansion of app evaluation frameworks. By illustrating how the MIND overlaps with many of these existing and emerging frameworks – we suggest the practical need for consolidation. Although specific disease tailored mobile health apps require specialized app evaluation *questions*, concerns around accessibility, privacy, clinical foundation, and interoperability are nonspecific. If the full potential of digital health can be realized, there is a need for increased collaboration among industry, government, and academia in order to ensure that the highest quality digital health tools reach the public. We emphasize that this effort is just a first step, and highlight the need for interdisciplinary continued communication amongst diverse digital health stakeholders in order to best serve the public.

Funding Statement: This work was supported by a gift from the Argosy Foundation

Declaration of Interests: None declared. JT serves as chair of the APA's Health IT committee, which created the APA app evaluation framework but receives no compensation or support for this volunteer work.

Data Availability Statement: All data are pretended in appendix A and B.

Author Contributions: SL and JT designed the procedure. SL and JT screened articles for eligibility. SL and LS compiled and mapped questions from frameworks. SL and JT composed manuscript.

Figure Legends

Figure 1: A screenshot of MIND highlighting several of the app evaluation questions (green boxes) and ability to access more

Figure 2: Framework identification through database searches (PubMed, EMBASE, PsychINFO) and other sources (reviews since 2018, grey literature, government websites).

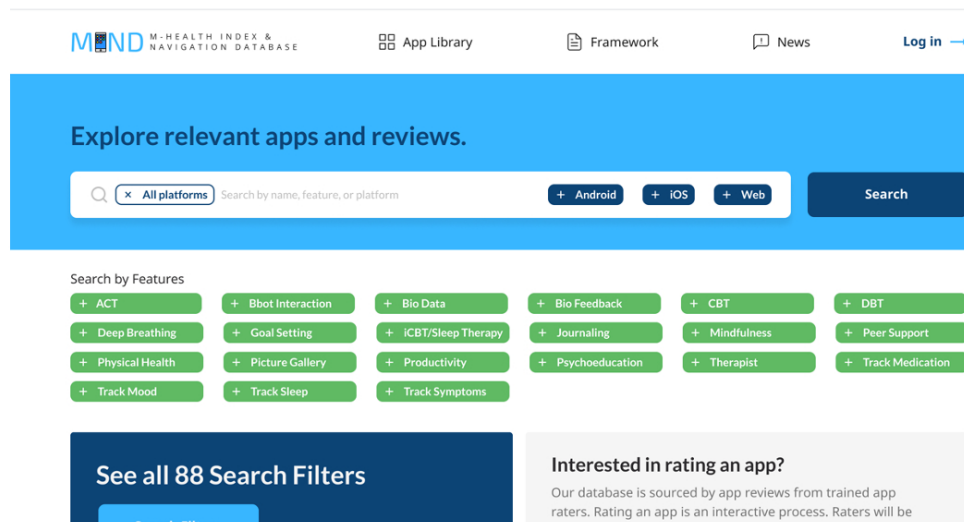
Figure 3: The most commonly addressed questions, grouped within the categories of MIND. The blue triangle constitutes MIND and its six main categories, while the green trapezoid represents questions pertaining to usability or ease of use, which are not covered by MIND.

References:

1. Health C for D and R. Digital Health Policies and Public Health Solutions for COVID-19. *FDA*. Published online August 3, 2020. Accessed September 14, 2020. <https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/digital-health-policies-and-public-health-solutions-covid-19>
2. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. *The Lancet Digital Health*. 2019;1(2):e52-e54. doi:10.1016/S2589-7500(19)30013-5
3. Ondersma SJ, Walters ST. Clinician's Guide to Evaluating and Developing eHealth Interventions for Mental Health. *Psychiatr res clin pract*. 2020;2(1):26-33. doi:10.1176/appi.prcp.2020.20190036
4. Bergin A, Davies EB. Technology Matters: Mental health apps – separating the wheat from the chaff. *Child and Adolescent Mental Health*. 2020;25(1):51-53. doi:10.1111/camh.12363
5. Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. *BMC Medicine*. 2019;17(1):231. doi:10.1186/s12916-019-1461-z
6. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. *npj Digit Med*. 2019;2(1):54. doi:10.1038/s41746-019-0129-6
7. Magee JC, Adut S, Brazill K, Warnick S. Mobile App Tools for Identifying and Managing Mental Health Disorders in Primary Care. *Curr Treat Options Psychiatry*. 2018;5(3):345-362. doi:10.1007/s40501-018-0154-0

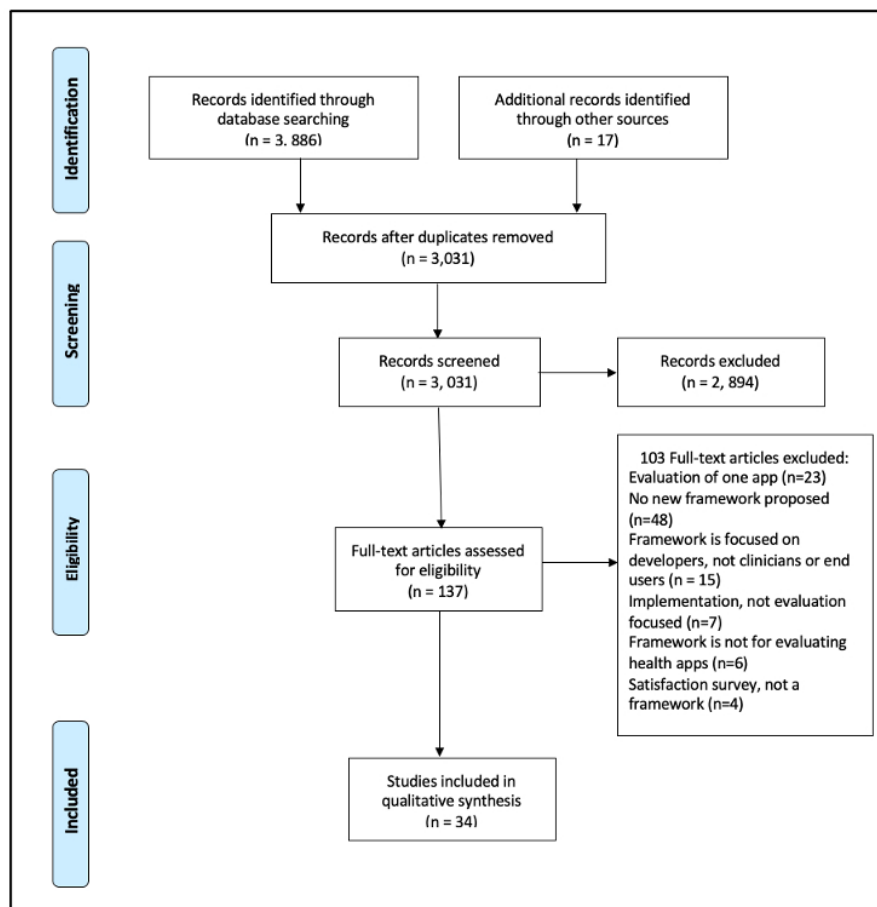
- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
8. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. *npj Digital Medicine*. 2020;3(1):1-8. doi:10.1038/s41746-020-00312-4
9. Moshi MR, Tooher R, Merlin T. SUITABILITY OF CURRENT EVALUATION FRAMEWORKS FOR USE IN THE HEALTH TECHNOLOGY ASSESSMENT OF MOBILE MEDICAL APPLICATIONS: A SYSTEMATIC REVIEW. *International Journal of Technology Assessment in Health Care*. 2018;34(5):464-475. doi:10.1017/S026646231800051X
10. Coravos A, Doerr M, Goldsack J, et al. Modernizing and designing evaluation frameworks for connected sensor technologies in medicine. *npj Digital Medicine*. 2020;3(1):1-10. doi:10.1038/s41746-020-0237-3
11. Nurgalieva L, O'Callaghan D, Doherty G. Security and Privacy of mHealth Applications: A Scoping Review. *IEEE Access*. 2020;8:104247-104268. doi:10.1109/ACCESS.2020.2999934
12. Benjumea J, Roper J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy Assessment in Mobile Health Apps: Scoping Review. *JMIR mHealth and uHealth*. 2020;8(7). doi:10.2196/18868
13. Levine DM, Co Z, Newmark LP, et al. Design and testing of a mobile health application rating tool. *npj Digital Medicine*. 2020;3(1):1-7. doi:10.1038/s41746-020-0268-9
14. Hussain M, Zaidan AA, Zidan BB, et al. Conceptual framework for the security of mobile health applications on Android platform. *Telematics and Informatics*. 2018;35(5):1335-1354. doi:10.1016/j.tele.2018.03.005
15. McMillan B, Hickey E, Patel MG, Mitchell C. Quality assessment of a sample of mobile app-based health behavior change interventions using a tool based on the National Institute of Health and Care Excellence behavior change guidance. *Patient Education and Counseling*. 2016;99(3):429-435. doi:10.1016/j.pec.2015.10.023
16. Shaia KL, Farag S, Chyjek K, Knopman J, Chen KT. An Evaluation of Mobile Applications for Reproductive Endocrinology and Infertility Providers. *Telemed J E Health*. 2017;23(3):254-258. doi:10.1089/tmj.2016.0079
17. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. *JMIR mHealth and uHealth*. 2015;3(1):e27. doi:10.2196/mhealth.3422
18. Wykes T, Schueller S. Why Reviewing Apps Is Not Enough: Transparency for Trust (T4T) Principles of Responsible Health App Marketplaces. *Journal of Medical Internet Research*. 2019;21(5):e12390. doi:10.2196/12390
19. Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support technology selection in digital health research. *Transl Behav Med*. Published online May 23, 2019. doi:10.1093/tbm/ibz074
20. Noee M, Baba akbari A, Olyaeemanesh A, Mobinizadeh M. Prioritizing the Potential Applications of Mobile-Health in the Iranian Health System. *Journal of Research in Health Sciences*. 2020;20. doi:10.34172/jrhs.2020.08
21. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. *Health Informatics J*. 2020;26(3):2042-2066. doi:10.1177/1460458219896492
22. Rodriguez-Villa E, Torous J. Regulating digital health technologies with transparency: the case for dynamic and multi-stakeholder evaluation. *BMC Med*. 2019;17. doi:10.1186/s12916-019-1447-x

- 1
2
3 23. Lee H, Sullivan SJ, Schneiders AG, et al. Smartphone and tablet apps for concussion road warriors (team
4 clinicians): a systematic review for practical users. *Br J Sports Med*. 2015;49(8):499-505.
5 doi:10.1136/bjsports-2013-092930
6
7 24. Farnia T, Jaulent M-C, Steichen O. Evaluation Criteria of Noninvasive Telemonitoring for Patients With Heart
8 Failure: Systematic Review. *Journal of Medical Internet Research*. 2018;20(1):e16. doi:10.2196/jmir.7873
9
10 25. Shah N, Jonassaint J, Castro LD. Patients Welcome the Sickle Cell Disease Mobile Application to Record
11 Symptoms via Technology (SMART). *Hemoglobin*. 2014;38(2):99-103. doi:10.3109/03630269.2014.880716
12
13 26. Hoffman L, Wisniewski H, Hays R, et al. Digital Opportunities for Outcomes in Recovery Services (DOORS): A
14 Pragmatic Hands-On Group Approach Toward Increasing Digital Health and Smartphone Competencies,
15 Autonomy, Relatedness, and Alliance for Those With Serious Mental Illness. *J Psychiatr Pract*. 2020;26(2):80-
16 88. doi:10.1097/PRA.0000000000000450
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



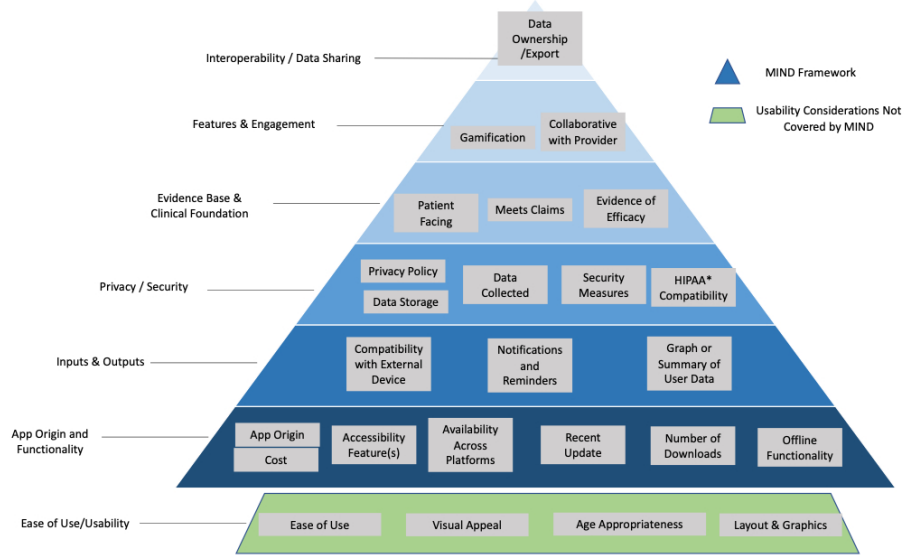
1: A screenshot of MIND highlighting several of the app evaluation questions (green boxes) and ability to access more

381x207mm (72 x 72 DPI)



2: PRISMA Diagram

291x287mm (72 x 72 DPI)



3: The most commonly addressed questions, grouped within the categories of MIND. The blue triangle constitutes MIND and its six main categories, while the green trapezoid represents questions pertaining to usability or ease of use, which are not covered by MIND.

390x266mm (72 x 72 DPI)

Appendix A: List of Reviewed Frameworks

PsyberGuide(1)(2), Mobile App Rating Scale (MARS)(3), AppScript Score(4), MindTech toolkit(5), NHS Digital Development Lab(6), Digital Assessment Questionnaire(7), PAS 277:2015(8), Cambridge Health Alliance(9), mHIMSS (10), NICE Behavior Change Guidance(11), AMA Principles for Safe Apps(12), Mindtools.io, ORCHA(13), Xcertia (14)(15), Guidelines for Mental Health App Evaluation Framework(16), mHealth Quality Label(17), Medical App Evaluation(18), Mobile Health Evaluation Framework(19), APPLICATIONS scoring system(20), Evaluating Sickle Cell Mobile App(21), mHealth App Evaluation for HIV(22), mHealth App Evaluation for HIV (23), mHealth App Evaluation for Pain Management(24), Evaluation of Mobile Clinical Applications(25), Quality of Experience (QoE) Survey(26), Concussion App Evaluation, Apps for Pain Management(27), Evaluation Tool for Healthcare Smartphone Applications(28), IMS: Patient Apps for Improved Healthcare(29), Mobile Apps for Asthma(30), Assessing Mobile Health App Quality(31), Certification Program for App Quality and Safety(32), Apps and Eating Disorders(33), Evidence-based Mobile Medical Applications in Diabetes(34), Diabetes Self-Management Applications(35), Framework for Evaluating Mobile Applications for Cardiac Rehabilitation(36), PIS: graphical classification tool for mHealth apps(37), Usability Evaluation of Mobile Applications for Diabetics(38), DiaDigital Apps(39), Applications of Mobile-Health in Iranian Health System(40), A Health Technology Assessment Module for Evaluating Mobile Medical Applications(41), Transparency4Trust(42), Graded Review of Dermatology Apps(43), What Makes a Good Health App?(44), Development a Guide for Mobile Health-Related Apps (45), Assessment Framework for COVID-19 Apps (46), Assessment of the Transparency and Reliability of Health Information Dissemination(47), Medical Mobile App Classification(48), Framework for Digital Support for the Autism Community (49), mHealth App Trustworthiness Checklist(50), THESIS(51), Framework for Evaluating Quality of mHealth Apps for Adolescent Users(52), ABACUS(53), Alberta Rating Index(54), Evaluation Framework for Digital Health Interventions(55), Mobile Applications Recommendations(56), Assessment Framework for Quality of Asthma Smartphone Applications (57), Global Digital Health Scorecard(58), NASS(59), Australian NSQDMH(60), Express Scripts Digital Health Formulary (61), Medical App Checker (62), Modernizing and Designing Evaluation Frameworks (63), Enlight (64), Improving the Quality of Apps for Patient Use(65), Assessment Framework for e-Mental Health in Canada(66), A Decision-Making Checklist to Support Technology Selection(67), Canadian Mental Health Commission (68), Trust4App(69), Evaluation Criteria (70), IDEAS(71)

Appendix B: Question-by-Question Analysis of MIND

Question	Frameworks with representative question
App Origin	26
Does it come from the government?	6
Does it come from a for-profit company or developer?	4

1	Does it come from a non-profit company?	3
2	Does it come from a trusted healthcare company?	6
3	Does it come from an academic institution?	3
4	App Functionality	42
5	Does it work on Apple(iOS)?	9
6	What is the Apple version?	0
7	What is the oldest iOS version supported?	0
8	What was the Apple release date?	1
9	When was the last Apple (IOS) update?	7
10	Has the apple version been updated in the last 180 days?	9
11	Number of reviews on Apple store?	9
12	Rating (number of stars) on Apple store?	9
13	App size on iOS?	2
14	Does it work on Android?	9
15	What is the Android version?	0
16	What is the oldest Android version supported?	0
17	What was the Google play store release date?	1
18	When was the last Android update?	7
19	Has the android version been updated in the last 180 days?	9
20	Number of reviews on google play store?	9
21	Rating (number of stars) on google play store?	9
22	App size on android?	2
23	Does the app work offline?	7
24	Does it have at least one accessibility feature (like adjust text size, text to voice, or colorblind color scheme adjuster)?	11
25	Does it have a web version?	5
26	Does it work with Spanish?	0
27	Does it work with a language other than English or Spanish?	5
28	Is the app totally free?	8
29	What is the cost up front?	16
30	Are there in-app purchases?	6
31	Is it a subscription (recurrent/monthly/annual)?	3
32	Inputs & Outputs	17
33	Input: surveys?	2

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1		
2		
3		
4	Input: Diary?	1
5	Input: Geolocation?	0
6		
7	Input: contact list?	0
8		
9	Input: Camera?	0
10		
11	Input: Microphone?	0
12		
13	Input: step count?	0
14	Input: external devices (e.g. a wearable sending direct data)?	4
15		
16	Input: social network?	1
17	Output: notifications?	6
18	Output: psychoeducational references/information?	0
19		
20	Output: social network?	1
21		
22	Output: reminders?	3
23		
24	Output: graphs of data?	3
25	Output: summary of data (in text or numbers)?	3
26		
27	Output: link to formal care/coaching?	2
28	Privacy & Security	43
29		
30	Is there a privacy policy?	20
31	Does the app declare data use and purpose?	18
32	Does the app report security measures in place?	25
33		
34	Is PHI shared?	9
35		
36	Is de-identified data shared?	5
37		
38	Is anonymized/aggregate data shared?	5
39		
40	Can you opt out of data collection?	4
41		
42	Can you delete your data?	3
43		
44	Is the user data stored only on the device?	5
45	Is the user data stored on a server?	9
46		
47	Does the app have a crisis management feature?	4
48	Does the app claim it meets HIPAA (or analogous national standard for patient/PHI privacy protection)	10
49		
50	Reading level of the privacy policy (what grade reading level)?	4
51		
52	Does the app use 3rd party vendors (i.e. google analytics, etc)?	4
53	Evidence & Clinical Foundation	57
54		
55	Is the app content well-written, correct, and relevant?	25
56	Does the app appear to do what it claims to do?	18
57		
58		
59		
60		

1		
2		
3		
4	Is the app patient facing?	28
5	How many feasibility/usability studies?	14
6		
7	What is the highest feasibility impact factor?	1
8		
9	How many evidence/efficacy studies?	27
10		
11	What is the highest efficacy impact factor?	4
12		
13	Can the app cause harm?	7
14	Does the app provide any warning for use?	11
15	Features & Engagement Style	29
16		
17	Features: mood tracking?	4
18	Features: medication tracking?	2
19		
20	Features: sleep tracking?	1
21		
22	Features: physical exercise tracking?	1
23		
24	Features: psychoeducation?	3
25	Features: journaling?	0
26	Features: picture gallery/hope board?	0
27		
28	Features: mindfulness?	1
29		
30	Features: deep breathing?	0
31		
32	Features: iCBT or sleep therapy?	1
33	Features: CBT?	0
34		
35	Features: ACT?	0
36	Features: DBT?	0
37		
38	Features: peer support?	1
39		
40	Features: connection to coach/therapist?	1
41	Features: biodata?	2
42		
43	Features: goal setting/habits?	3
44		
45	Features: physical health exercises?	2
46	Features: Chatbot interaction (like with virtual character)?	0
47		
48	Features: Biofeedback with sense data (eeg, HRV, skin conductance, etc)?	2
49		
50	Engagement style: user generated data?	4
51	Engagement style: chat/message based?	1
52		
53	Engagement style: is it a screener/assessment?	3
54	Engagement style: real time response?	3
55		
56	Engagement style: Asynchronous response?	0
57		
58		
59		
60		

Engagement style: gamification (points, badges)?	7
Engagement style: videos?	2
Engagement style: audio/music/scripts?	2
Engagement style: AI support?	0
Engagement style: peer support?	6
Engagement style: network support?	5
Engagement style: Collaborative with provider/other?	8
App Use	8
Is it a self-help/self-management tool?	6
Is it a reference app?	1
Is it intended for hybrid use with a clinician in conjunction with treatment plan?	3
Interoperability & Data Sharing	23
Do you own your data?	8
Can you email or export your data?	10
Can you send your data to a medical record?	5

Full Framework Reference List

1. Neary M, Schueller SM. State of the Field of Mental Health Apps. *Cognitive and Behavioral Practice*. 2018 Nov 1;25(4):531–7.
2. About One Mind PsyberGuide [Internet]. One Mind PsyberGuide. [cited 2020 Oct 19]. Available from: <https://live-one-mind-psyberguide.pantheonsite.io/about-psyberguide/>
3. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. *JMIR Mhealth Uhealth* [Internet]. 2015 Mar 11 [cited 2020 Oct 19];3(1). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4376132/>
4. AppScript | Discover, Deliver & Track Digital Health [Internet]. [cited 2020 Oct 19]. Available from: <https://www.appscript.net/score-details>
5. Toolkit for Appraising Digital Mental Health Products - Technology theme - MindTech [Internet]. 2017 [cited 2020 Oct 19]. Available from: <https://web.archive.org/web/20170701091502/http://mindtech.org.uk/projects/71-digital-mental-health-tools-evaluation-criteria-technology-theme.html>
6. Betton V, Craven M, Davies B, Martin J, Nelissen N, Simons L. Framework for the effectiveness evaluation of mobile (mental) health tools. :42.
7. Digital-Assessment-Questions-V2.1-Beta-PDF.pdf [Internet]. [cited 2020 Oct 19]. Available from: <https://developer.nhs.uk/wp-content/uploads/2018/09/Digital-Assessment-Questions-V2.1-Beta-PDF.pdf>

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
8. British Standards Institution. Health and wellness apps: quality criteria across the life cycle : code of practice. 2015.
9. Hoffman L. Mobile App Technology Meets Collaborative Care [Internet]. Available from: http://app.ihp.org/FacultyDocuments/Events/Event-2930/Presentation-15621/Document-13034/Presentation_A1_B1_Mobile_App_Technology_benedetto.pdf
10. Selecting a Mobile App: Evaluating the Usability of Medical Applications. :32.
11. McMillan B, Hickey E, Patel MG, Mitchell C. Quality assessment of a sample of mobile app-based health behavior change interventions using a tool based on the National Institute of Health and Care Excellence behavior change guidance. *Patient Education and Counseling*. 2016 Mar;99(3):429–35.
12. AMA adopts principles to promote safe, effective mHealth applications [Internet]. American Medical Association. [cited 2020 Oct 19]. Available from: <https://www.ama-assn.org/press-center/press-releases/ama-adopts-principles-promote-safe-effective-mhealth-applications>
13. Ltd FS. ORCHA Review Process | Reviewing Health and Care Apps | ORCHA [Internet]. [cited 2020 Oct 19]. Available from: <https://www.orchha.co.uk/our-solution/the-orchha-review/#0>
14. mHealthIntelligence. Amid a Flood of New mHealth Apps, Xcertia Looks to Set Standards [Internet]. mHealthIntelligence. 2018 [cited 2020 Oct 19]. Available from: <https://mhealthintelligence.com/news/amid-a-flood-of-new-mhealth-apps-xcertia-looks-to-set-standards>
15. Interoperability & Health Information Exchange Committee | HIMSS [Internet]. 2016 [cited 2020 Oct 19]. Available from: <https://www.himss.org/resources/interoperability-health-information-exchange-committee>
16. Chan S, Torous J, Hinton L, Yellowlees P. Towards a Framework for Evaluating Mobile Mental Health Apps. *Telemed J E Health*. 2015 Dec;21(12):1038–41.
17. Yasini M, Marchand G. Mobile Health Applications, in the Absence of an Authentic Regulation, Does the Usability Score Correlate with a Better Medical Reliability? *Stud Health Technol Inform*. 2015;216:127–31.
18. Walsworth DT. Medical Apps: Making Your Mobile Device a Medical Device. *FPM*. 2012 Jun;19(3):10–3.
19. Developing a Framework for Evaluating the Patient Engagement, Quality, and Safety of Mobile Health Applications | Commonwealth Fund [Internet]. [cited 2020 Oct 19]. Available from: <https://www.commonwealthfund.org/publications/issue-briefs/2016/feb/developing-framework-evaluating-patient-engagement-quality-and>
20. Shaia KL, Farag S, Chyjek K, Knopman J, Chen KT. An Evaluation of Mobile Applications for Reproductive Endocrinology and Infertility Providers. *Telemed J E Health*. 2017;23(3):254–8.
21. Shah N, Jonassaint J, De Castro L. Patients welcome the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). *Hemoglobin*. 2014;38(2):99–103.
22. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown Iii W. Comparison of a User-Centered Design, Self-Management App to Existing mHealth Apps for Persons Living With HIV. *JMIR Mhealth Uhealth*. 2015 Sep 18;3(3):e91.

23. Robustillo Cortés M de las A, Cantudo Cuenca MR, Morillo Verdugo R, Calvo Cidoncha E. High quantity but limited quality in healthcare applications intended for HIV-infected patients. *Telemed J E Health*. 2014 Aug;20(8):729–35.
24. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs SJ, et al. Assessing the quality and usability of smartphone apps for pain self-management. *Pain Med*. 2014 Jun;15(6):898–909.
25. Murfin M. Know your apps: an evidence-based approach to evaluation of mobile clinical applications. *J Physician Assist Educ*. 2013;24(3):38–40.
26. Martínez-Pérez B, de la Torre-Díez I, Candelas-Plasencia S, López-Coronado M. Development and evaluation of tools for measuring the quality of experience (QoE) in mHealth applications. *J Med Syst*. 2013 Oct;37(5):9976.
27. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. “There’s a Pain App for That”: Review of Patient-targeted Smartphone Applications for Pain Management. *Clin J Pain*. 2015 Jun;31(6):557–63.
28. Jin M, Kim J. Development and Evaluation of an Evaluation Tool for Healthcare Smartphone Applications. *Telemed J E Health*. 2015 Oct;21(10):831–7.
29. patient-adoption-of-mhealth.pdf [Internet]. [cited 2020 Oct 19]. Available from: <https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf>
30. Huckvale K, Morrison C, Ouyang J, Ghaghda A, Car J. The evolution of mobile apps for asthma: an updated systematic assessment of content and tools. *BMC Med*. 2015 Dec;13(1):58.
31. Grundy QH, Wang Z, Bero LA. Challenges in Assessing Mobile Health App Quality: A Systematic Review of Prevalent and Innovative Methods. *Am J Prev Med*. 2016;51(6):1051–9.
32. Álvarez-Rementería JF. QUALITY AND SAFETY STRATEGY FOR MOBILE HEALTH APPLICATIONS. :18.
33. Fairburn CG, Rothwell ER. Apps and eating disorders: A systematic clinical appraisal. *Int J Eat Disord*. 2015 Nov;48(7):1038–46.
34. Drincic A, Prahalad P, Greenwood D, Klonoff DC. Evidence-based Mobile Medical Applications in Diabetes. *Endocrinol Metab Clin North Am*. 2016 Dec;45(4):943–65.
35. Demidowich AP, Lu K, Tamler R, Bloomgarden Z. An evaluation of diabetes self-management applications for Android smartphones. *J Telemed Telecare*. 2012 Jun;18(4):235–8.
36. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. *J Am Heart Assoc*. 2013 Nov 1;2(6):e000568.
37. Basilico A, Marceglia S, Bonacina S, Pincioli F. Advising patients on selecting trustful apps for diabetes self-care. *Computers in Biology and Medicine*. 2016 Apr 1;71:86–96.
38. Arnhold M, Quade M, Kirch W. Mobile Applications for Diabetics: A Systematic Review and Expert-Based Usability Evaluation Considering the Special Requirements of Diabetes Patients Age 50 Years or Older. *Journal of Medical Internet Research*. 2014;16(4):e104.
39. Kaltheuner M, Drossel D, Heinemann L. DiaDigital Apps: Evaluation of Smartphone Apps Using a Quality Rating Methodology for Use by Patients and Diabetologists in Germany. *J Diabetes Sci Technol*. 2018 Sep 28;13(4):756–62.

- 1
2
3 40. Noee M, Baba akbari A, Olyaeemanesh A, Mobinizadeh M. Prioritizing the Potential Applications of
4 Mobile-Health in the Iranian Health System. *Journal of Research in Health Sciences*. 2020 Mar 7;20.
5
6 41. Moshi MR, Tooher R, Merlin T. Development of a health technology assessment module for
7 evaluating mobile medical applications. *Int J Technol Assess Health Care*. 2020 Jun;36(3):252–61.
8
9 42. Wykes T, Schueller S. Why Reviewing Apps Is Not Enough: Transparency for Trust (T4T)
10 Principles of Responsible Health App Marketplaces. *Journal of Medical Internet Research*.
11 2019;21(5):e12390.
12
13 43. Masud A, Shafi S, Rao BK. Mobile medical apps for patient education: a graded review of available
14 dermatology apps. *Cutis*. 2018 Feb;101(2):141–4.
15
16 44. Dawson RM, Felder TM, Donevant SB, McDonnell KK, Card EB, King CC, et al. What makes a
17 good health “app”? Identifying the strengths and limitations of existing mobile application evaluation
18 tools. *Nurs Inq*. 2020;27(2):e12333.
19
20 45. Llorens-Vernet P, Miró J. Standards for Mobile Health–Related Apps: Systematic Review and
21 Development of a Guide. *JMIR mHealth and uHealth*. 2020;8(3):e13057.
22
23 46. Vokinger KN, Nittas V, Witt CM, Fabrikant SI, von Wyl V. Digital health and the COVID-19
24 epidemic: an assessment framework for apps from an epidemiological and legal perspective. *Swiss
25 Medical Weekly [Internet]*. 2020 May 17 [cited 2020 Oct 19];150(1920). Available from:
26 <https://smw.ch/article/doi/smw.2020.20282>
27
28 47. Huang Z, Lum E, Car J. Medication Management Apps for Diabetes: Systematic Assessment of the
29 Transparency and Reliability of Health Information Dissemination. *JMIR mHealth and uHealth*.
30 2020;8(2):e15364.
31
32 48. Nwe K, Larsen ME, Nelissen N, Wong DC-W. Medical Mobile App Classification Using the
33 National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health
34 Technologies: Interrater Reliability Study. *Journal of Medical Internet Research*. 2020;22(6):e17457.
35
36 49. Zervogianni V, Fletcher-Watson S, Herrera G, Goodwin M, Pérez-Fuster P, Brosnan M, et al. A
37 framework of evidence-based practice for digital support, co-developed with and for the autism
38 community: Autism [Internet]. 2020 Feb 6 [cited 2020 Oct 19]; Available from:
39 <https://journals.sagepub.com/eprint/4KEDZJ7YYCZAUHJ7B2NR/full>
40
41 50. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth App Trustworthiness
42 checklist. *DIGITAL HEALTH*. 2019 Jan 1;5:2055207619886463.
43
44 51. Levine DM, Co Z, Newmark LP, Groisser AR, Holmgren AJ, Haas JS, et al. Design and testing of a
45 mobile health application rating tool. *npj Digital Medicine*. 2020 May 21;3(1):1–7.
46
47 52. Jeminiwa RN, Hohmann NS, Fox BI. Developing a Theoretical Framework for Evaluating the
48 Quality of mHealth Apps for Adolescent Users: A Systematic Review. *J Pediatr Pharmacol Ther*.
49 2019;24(4):254–69.
50
51 53. McKay FH, Slykerman S, Dunn M. The App Behavior Change Scale: Creation of a Scale to Assess
52 the Potential of Apps to Promote Behavior Change. *JMIR mHealth and uHealth*. 2019;7(1):e11130.
53
54 54. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality
55 rating scales: a systematic review. *Disability and Rehabilitation: Assistive Technology*. 2020 Jan
56 8;1–10.
57
58
59
60

- 1
2
3 55. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for
4 digital health interventions. *it - Information Technology*. 2019 Oct 25;61(5–6):253–63.
5
6 56. Jisha RC, Krishnan R, Vikraman V. Mobile Applications Recommendation Based on User Ratings
7 and Permissions. In: 2018 International Conference on Advances in Computing, Communications
8 and Informatics (ICACCI). 2018. p. 1000–5.
9
10 57. Guan Z, Sun L, Xiao Q, Wang Y. Constructing an assessment framework for the quality of asthma
11 smartphone applications. *BMC Medical Informatics and Decision Making*. 2019 Oct 15;19(1):192.
12
13 58. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to
14 validation. *npj Digital Medicine*. 2019 May 13;2(1):1–9.
15
16 59. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A’Court C, et al. Beyond Adoption: A
17 New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the
18 Scale-Up, Spread, and Sustainability of Health and Care Technologies. *Journal of Medical Internet
19 Research*. 2017;19(11):e367.
20
21 60. National Safety and Quality Digital Mental Health Standards | Australian Commission on Safety and
22 Quality in Health Care [Internet]. [cited 2020 Oct 23]. Available from:
23 [https://www.safetyandquality.gov.au/standards/national-safety-and-quality-digital-mental-health-
24 standards](https://www.safetyandquality.gov.au/standards/national-safety-and-quality-digital-mental-health-standards)
25
26 61. Digital Health Formulary | Express Scripts [Internet]. [cited 2020 Oct 23]. Available from:
27 <https://www.express-scripts.com/corporate/tag/digital-health-formulary>
28
29 62. Medical App Checker: a Guide to assessing Mobile Medical Apps [Internet]. [cited 2020 Oct 23].
30 Available from: [https://www.knmg.nl/actualiteit-opinie/nieuws/nieuwsbericht/medical-app-checker-
31 a-guide-to-assessing-mobile-medical-apps.htm](https://www.knmg.nl/actualiteit-opinie/nieuws/nieuwsbericht/medical-app-checker-a-guide-to-assessing-mobile-medical-apps.htm)
32
33 63. Coravos A, Doerr M, Goldsack J, Manta C, Shurvey M, Woods B, et al. Modernizing and designing
34 evaluation frameworks for connected sensor technologies in medicine. *npj Digital Medicine*. 2020
35 Mar 13;3(1):1–10.
36
37 64. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: A Comprehensive Quality and
38 Therapeutic Potential Evaluation Tool for Mobile and Web-Based eHealth Interventions. *Journal of
39 Medical Internet Research*. 2017;19(3):e82.
40
41 65. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps
42 for patient use? *BMC Medicine*. 2018 Dec 3;16(1):225.
43
44 66. Zelmer J, Hoof K van, Notarianni M, Mierlo T van, Schellenberg M, Tannenbaum C. An Assessment
45 Framework for e-Mental Health Apps in Canada: Results of a Modified Delphi Process. *JMIR
46 mHealth and uHealth*. 2018;6(7):e10016.
47
48 67. Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support
49 technology selection in digital health research. *Transl Behav Med*. 2020 Oct 8;10(4):1004–15.
50
51 68. eMH_app_eng.pdf [Internet]. [cited 2020 Oct 23]. Available from:
52 https://www.mentalhealthcommission.ca/sites/default/files/2018-01/eMH_app_eng.pdf
53
54 69. Habib SM, Alexopoulos N, Islam MM, Heider J, Marsh S, Muehlhaeuser M. Trust4App: Automating
55 Trustworthiness Assessment of Mobile Applications. In: 2018 17th IEEE International Conference
56 On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International
57 Conference On Big Data Science And Engineering (TrustCom/BigDataSE). 2018. p. 124–35.
58
59
60

- 1
- 2
- 3
- 4 70. Farnia T, Jaulent M-C, Steichen O. Evaluation Criteria of Noninvasive Telemonitoring for Patients
- 5 With Heart Failure: Systematic Review. *Journal of Medical Internet Research*. 2018;20(1):e16.
- 6
- 7 71. Fedele DA, McConville A, Moon J, Thomas JG. Topical Review: Design Considerations When
- 8 Creating Pediatric Mobile Health Interventions: Applying the IDEAS Framework. *J Pediatr Psychol*.
- 9 2019 Apr 1;44(3):343–8.
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

For peer review only

BMJ Open

Evaluating Evaluation Frameworks: A Scoping Review of Frameworks for Assessing Health Apps

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-047001.R1
Article Type:	Original research
Date Submitted by the Author:	06-Mar-2021
Complete List of Authors:	Lagan, Sarah; Beth Israel Lahey Health Sandler, Lev; Beth Israel Lahey Health Torous, John; Beth Israel Lahey Health
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Research methods
Keywords:	PSYCHIATRY, Telemedicine < BIOTECHNOLOGY & BIOINFORMATICS, Information management < BIOTECHNOLOGY & BIOINFORMATICS

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Evaluating Evaluation Frameworks: A Scoping Review of Frameworks for Assessing Health Apps

Sarah Lagan BA¹, Lev Sandler BA¹, John Torous MD*¹

Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School

*corresponding author

Division of Digital Psychiatry
Rabb 2 Building, Department of Psychiatry
Beth Israel Deaconess Medical Center
Boston, MA, 02115, USA
jtorous@bidmc.harvard.edu

Word Count: 3,149

Objectives

Despite an estimated 300,000 mobile health apps on the market, there remains no consensus around helping patients and clinicians select safe and effective apps. In 2018, our team drew upon existing evaluation frameworks to identify salient categories and create a new framework endorsed by the American Psychiatric Association (APA). We have since created a more expanded and operational framework (MIND) that aligns with the APA categories but includes objective and auditable questions (105). We sought to survey the existing space, conducting a review of all mobile health app evaluation frameworks published since 2018, and demonstrate the comprehensiveness of this new model by comparing it to existing and emerging frameworks.

Design

We conducted a scoping review of mobile health app evaluation frameworks.

Data Sources

References were identified through searches of PubMed, EMBASE, and PsychINFO with publication date between January 2018 through October 2020.

Eligibility Criteria

Papers were selected for inclusion if they met the predetermined eligibility criteria -- presenting an evaluation framework for mobile health apps with patient, clinician, or end user facing questions.

Data Extraction and Synthesis

Two reviewers screened the literature separately and applied the inclusion criteria. The data extracted from the papers included: Author and dates of publication, source affiliation, country of origin, name of framework, study design, description of framework, intended audience/user, and framework scoring system. We then compiled a collection of more than 1,701 questions across 79 frameworks. We compared and grouped these questions using the MIND framework as a reference. We sought to identify the most common domains of evaluation while assessing the comprehensiveness and flexibility- as well as any potential gaps- of MIND.

Results

New app evaluation frameworks continue to emerge and expand. Since our 2019 review of the app evaluation framework space, more frameworks include questions around privacy (43) and clinical foundation (57), reflecting an increased focus on issues of app security and evidence base. The majority of mapped frameworks overlapped with at least half of the MIND categories. The results of this search have informed a database (apps.digitalpsych.org) that users can access today.

Conclusion

As the number of app evaluation frameworks continues to rise, it is becoming difficult for users to select both an appropriate evaluation tool and to find an appropriate health app. This review provides a comparison of what different app evaluation frameworks are offering, where the field is converging, and new priorities for improving clinical guidance.

Strengths and limitations of this study

- This scoping review is the largest and most up to date review and comparison of mobile health app evaluation frameworks.
- The analysis highlighted the flexibility and comprehensiveness of the MIND framework, which was used as a reference framework in this review, in diverse contexts.
- MIND was initially tailored to mental health, and thus does not encompass thorough disease-specific criteria for other conditions, such as asthma, diabetes, and sickle cell anemia – though such questions may be easily integrated.
- Subjective questions, especially those around ease of use and visual appeal, are difficult to standardize but may be among the most important features driving user engagement with mental health apps.

Introduction

The past five years have seen a proliferation of both mobile health apps and proposed tools to rate such apps. While these digital health tools hold great potential, concerns around privacy, efficacy, and credibility, coupled with a lack of strict oversight by governing bodies, have highlighted a need for frameworks that can help guide clinicians and consumers to make informed app choices. Although the United States' Food and Drug Administration (FDA) has recognized the issue and is piloting a pre-certification program that would prioritize app safety at the developer level,¹ this model is still in pilot stages and there has yet to be an international consensus around standards for health apps, resulting in a profusion of proposed frameworks across governments, academic institutions, and commercial interests.

In 2018, our team drew upon existing evaluation frameworks to identify salient categories from existing rating schemes and create a new framework.² The American Psychiatric Association's (APA) App Evaluation Model was developed by harmonizing questions from 45 evaluation frameworks and selecting 38 total questions that mapped to five categories: background information, privacy and security, clinical foundation, ease of use, and interoperability. This APA model subsequently has been utilized by many diverse stakeholders given its flexibility in guiding informed decision making.^{3,4,5,6,7} However, the flexibility of the model also created a demand for a more applied approach that offered users more concrete information instead of placing the onus entirely on a clinician or provider.

Thus, since the framework's development, the initial 38 questions have been operationalized into 105 new objective questions that invite a binary (yes/no) or numeric response by a rater.⁸ These questions align with the categories proposed by the APA model but are more extensive and objective, with, for example, "app engagement" operationalized into eleven different engagement styles to select. These 105 questions are sorted into six categories (App Origin and Functionality, Inputs and Outputs, Privacy and Security, Clinical Foundation, Features and Engagement, Interoperability and Data Sharing) are intended to be answerable for any trained rater – clinician, peer, end user –and inform the public facing Mhealth Index and Navigation Database (MIND), where users can view app attributes and compare ratings (see figure 1 below). MIND thus constitutes a new framework based on the APA model, with an accompanying public facing database.

FIGURE 1

Recent systematic reviews have illustrated the growing number of evaluation tools for digital health devices, including mobile health apps.^{9,10,11} Given the rapidly evolving health app space and the need to understand what aspects are considered in evaluation frameworks, we have sought to survey the landscape of existing frameworks. Our goal was to compare the categories and questions composing other frameworks to 1) identify common elements between them, 2) identify if gaps in evaluation frameworks have improved since 2018, and 3) and assess how reflective our team's MIND framework is in the current landscape. We thus aimed to map every question from the 2018 review, as well as questions from new app evaluation frameworks that have emerged since, using the questions of MIND as a reference. While informing our own

1
2
3 efforts around MIND, the results of this review offer broad relevance across all of digital health,
4 as understanding the current state of app evaluation helps inform how any new app may be
5 assessed, categorized, judged, and adopted.
6
7
8
9

10 **Methods**

11 *Patient and Public Involvement*

12
13
14 Like the APA model, MIND shifts the app evaluation process away from finding one “best” app,
15 and instead guiding users towards an informed decision based on selecting and placing value on
16 the clinically-relevant criteria that account for the needs and preferences of each patient and case.
17 Questions were created with input of clinicians, patients, family members, researchers, and
18 policy makers. The goal is not for a patient or clinician to consider all 105 questions but rather be
19 able to access a subset of questions that appear most appropriate for the current use case at hand.
20 Thus, thanks to its composition of discrete questions that aim to be objective and reproducible,
21 MIND offers a useful tool to compare evaluation frameworks. It also offers an actionable
22 resource for any user anywhere in the world to engage with app evaluation, providing tangible
23 results in the often more theoretical world of app evaluation.
24
25
26

27 *Design*

28
29 We followed a three-step process in order to identify and compare frameworks to MIND. This
30 process included (1) assembling all existing frameworks for mobile medical applications, (2)
31 separating each framework into the discrete evaluation questions comprising it, and (3) mapping
32 all questions to the 105 MIND framework questions as a reference.
33
34

35 *Search Strategy and Selection Criteria*

36 We started with the 45 frameworks identified in the 2018 review by Moshi et al.⁹ and included
37 34 frameworks that have emerged since our initial analysis of the space that was conducted in
38 2018 and published in 2019.² To accomplish this, we conducted an adapted scoping review based
39 on the Moshi criteria to identify recent frameworks. Although MIND focuses on mental health
40 apps, its considerations and categories are transferable to health apps more broadly, and thus
41 there was no mental health specification in the search terms.
42
43
44

45 References were identified through searches of PubMed, EMBASE, and PsychINFO with the
46 search terms ((mobile application) OR (smartphone app)) AND ((framework) OR (criteria) OR
47 (rating)) and publication date between January 2018 through October 2020. We also identified
48 records beyond the database search by seeking frameworks mentioned in subsequent and recent
49 reviews^{5,12,13} and surveying the grey literature and government websites. Papers were selected
50 for inclusion if they met the predetermined eligibility criteria -- presenting an evaluation
51 framework for mobile health apps with patient, clinician, or end user facing questions. Two
52 reviewers (SL and JT) screened the literature separately and applied the inclusion criteria. The
53 data extracted from the papers included: Author and dates of publication, source affiliation,
54 country of origin, name of framework, study design, description of framework, intended
55
56
57
58
59
60

audience/user, and framework scoring system. Articles were screened if they described the evaluation of a single app, did not present a new framework (instead conducting a review of the space or relying upon a previous framework), the framework was focused on developer instead of clinicians or end users, was implementation and not evaluation focused, was not a framework for health apps, and was a satisfaction survey instead of an evaluation framework. The data selection process is outlined in figure 2.

FIG 2

The 34 frameworks identified in the search were combined with the 45 frameworks from the 2018 review for a total of 79 frameworks for consideration. To our knowledge, this list comprehensively reflects the state of the field at the time of assembly. However, we do not claim it to be exhaustive, as frameworks are constantly changing, emerging, and sunsetting, with no central repository. The final list of frameworks assembled can be found in Appendix A.

Mapping

Each resulting framework was reviewed and compiled into a complete list of its unique questions. The 79 frameworks yielded 1,701 questions in total. Several of the original 45 frameworks focused exclusively on in-depth privacy considerations (evaluating privacy and security practices rather than the app itself),¹⁴ and after eliminating these checklists that did not facilitate app evaluation by a clinician or end user, 70 total frameworks were mapped in entirety to the MIND framework.

In mapping questions, discussion was sometimes necessary as not every question was an exact, word-for-word match. The authors thus used discretion when it came to matching questions to MIND, and discussed each decision to confirm mapping placement. Two raters (SL, LS) agreed on mapping placement, and disputes were brought to a third reviewer (JT) for final consideration. “Is data portable and interoperable?”¹⁵, for example, would be mapped to the question “can you email or export your data?” “Connectivity”¹⁶ was mapped to “Does the app work offline?” and “Is the arrangement and size of buttons/content on the screen zoomable if needed”¹⁷ was mapped to “is there at least one accessibility feature?” Questions about suitability for the “target audience” were mapped to the “patient facing” question in MIND.

Results

Framework Type

The aim of this review was to identify and compare mobile health app rating frameworks, assessing overlap and exploring changes and gaps relative to both previous reviews and to the MIND framework. Of the 70 frameworks ultimately assessed and mapped, the majority 39 (55.7%) offered models for evaluating mobile health apps broadly. Seven (10%) considered mental health apps, while six (8.5%) focused on apps for diabetes management. Other evaluation

focuses included apps for asthma, autism, concussions, COVID-19, dermatology, eating disorders, heart failure, HIV, pain management, infertility, and sickle cell disease (table 1).

Focus of Evaluation	Number of Frameworks
Asthma	2 (2.86%)
Autism	1 (1.43%)
Concussion	1 (1.43%)
COVID-19	1 (1.43%)
Dermatology	1 (1.43%)
Diabetes	6 (8.57%)
Eating Disorders	1 (1.43%)
General Mobile Health	39 (55.71%)
Heart Failure/Cardiac Rehabilitation	3 (4.29%)
HIV	2(2.86%)
Mental Health	7 (10%)
Pain Management	2 (2.86%)
Pediatric or Adolescent Health	2 (2.86%)
Reproductive Endocrinology/Infertility	1 (1.43%)
Sickle Cell Disease	1 (1.43%)

Table 1: Number of disease specific and general app evaluation frameworks, with general mobile health frameworks constituting more than half of identified frameworks.

Mapping

We mapped questions from 70 app evaluation frameworks against the six categories and 105 questions of MIND (see appendix B). We examined the number of frameworks that addressed each specific MIND category and identified areas of evaluation that are not addressed by MIND. Through the mapping process, we were able to gauge the most common questions and categories across different app evaluation frameworks.

We sorted the questions into MIND's six different categories – App Origin & Functionality, Inputs & outputs, Privacy & Security, Evidence & Clinical Foundation, Features & Engagement Style, and Interoperability & Data Sharing – in order to assess the most common broad areas of consideration. Across frameworks, the most common considerations were around privacy/security and clinical foundation, with 43 frameworks posing at least one question around the app's privacy protections, and 57 of the frameworks containing at least one question to evaluate evidence base or clinical foundation, as denoted in table 2. Fifty-nine frameworks covered at least two of the MIND categories, with the majority of frameworks overlapping with at least four of MIND categories.

Category of Evaluation	Number of Frameworks Addressing Category

App Origin / App Functionality	42
Inputs and Outputs	17
Privacy / Security	43
Evidence / Clinical Foundation	57
Features / Engagement Style	29
Interoperability and Data Sharing	23

Table 2: *The questions from all frameworks were mapped to the reference framework (MIND) sorted into its six categories, with this table denoting how many frameworks had questions that could be sorted into each of the categories.*

We then took a more granular look at the questions from each of the 70 frameworks, matching questions one-by-one to questions of the MIND framework when possible. On an individual question level, specific questions about the presence of a privacy policy, security measures in place, supporting studies, and patient-facing (or target population) tools were the most prevalent, with representation from 20, 25, 27, and 28 frameworks respectively for each question. Each of the 70 frameworks had at least one question that mapped to MIND. The most common questions, sorted into their respective categories, are depicted in Figure 3 and Table 3, while the full list of mapped questions can be found in Appendix B.

FIG 3

Every question was examined but not every question in every framework could be matched to a corresponding question in MIND, and some questions fell outside one of the six categories. For example, 18 frameworks continue to present the subjective question of “is the app easy to use” which will vary depending on the person and use case. MIND also does not offer questions related to other objective questions to which answers are not readily available such as “How were target users involved in the initial design and usability evaluations of the app?”¹⁸ While questions such as this are of high importance, lack of easily accessible answers creates a dilemma in their present utility for app evaluation. Furthermore, some questions such as economic analysis were not covered by MIND but by other frameworks and represent a similar dilemma in that actual data to base evaluation on is often lacking. Aside from subjective questions, other pronounced absences MIND were questions about customizability (addressed by 7 other frameworks) and advertising (9 frameworks). Although MIND does ask about customizability in part by encouraging raters to consider accessibility features (and some frameworks ask about the ability to customize in conjunction with accessibility features¹⁹), MIND does not pose a question around the user’s ability to tailor or customize app content, nor does it ask questions about the presence of advertisements on an app. Other questions unaddressed by MIND were about the user’s ability to contact the producer or developer to seek guidance about app use. Variations of this question include “is there a way to feedback user comments to the app developer?” MIND also does not pose any questions regarding instructions in the app or the existence of a user guide.²⁰ Finally, it does ask about speed of app functionality. This variant of question asks, “is the app fast and easy to use in clinical settings?”¹⁵ Figure 3 above, and Table 3 below present additional details on categories and questions both inside and outside the MIND reference framework.

Question	Number of Frameworks Addressing Question
App Origin / App Functionality: Where does the app come from / who is the app developer?	26
App Origin / App Functionality: How much does the app cost?	16
App Origin / App Functionality: Does the app have at least one accessibility feature?	11
App Origin / App Functionality: Is the app available on multiple platforms?	9
App Origin / App Functionality: Has the app been updated recently / when was the last update?	9
App Origin / App Functionality: How many downloads does the app have / how many reviews?	9
App Origin/App Functionality: Does the app work offline?	7
Privacy / Security: Is there a privacy policy?	20
Privacy / Security: Does the app declare data use and purpose?	18
Privacy / Security: Does the app report security measures in place?	25
Privacy / Security: Does the app claim to meet HIPAA (or an analogous health data protection regulation)?	10
Evidence / Clinical Foundation: Is the app content well-written, correct, and relevant?	25
Evidence / Clinical Foundation: Does the app do what it claims?	18
Evidence / Clinical Foundation: Is the app patient facing?	28
Evidence / Clinical Foundation: Is there evidence of the app's efficacy?	27
Features / Engagement Style: Is the app collaborative with a provider?	8
Features / Engagement Style: Does the app offer gamification?	4
Interoperability / Data Sharing: Can you email or export your data?	10
Is the app easy to use?	18
Is the app visually appealing?	10
Is the app age appropriate?	4
Does the app allow for customization and/or personalization?	7
Was the app developed in collaborated with target users?	4
Is there a manual or set of instructions for the app?	8
Is there any advertising within the app?	9
Risk/Benefit Analysis	2
Speed of Functionality	5
Is there a way to contact the developer with questions or concerns?	6

Table 3: Commonly addressed questions among those that could be mapped to the MIND reference framework (blue), and those that could not (green).

Discussion

As mobile health apps have proliferated, choosing the right one has become increasingly challenging for patient and clinician alike. While app evaluation frameworks can help sort through the myriad of mobile health apps, the growing number of frameworks further complicates the process of evaluation. Our review examined the largest number of evaluation frameworks to date with the goal of assessing their unique characteristics, gaps, as well as overlap with the 105 questions in MIND. We identified frameworks for evaluating a wide range of mobile health apps – some focused on general mobile health, some specific and addressing specific disease domains like asthma, heart failure, mental health, or pain management. Despite the different disease conditions they addressed, there was substantial overlap among the frameworks, especially around clinical foundation and privacy and security. The most common category addressed was clinical foundation, with 57 of the evaluation frameworks posing at least one question regarding evidence base. More than half of the frameworks also addressed privacy and/or security and app functionality or origin.

The widespread focus on clinical foundation and privacy represents a major change in the space since 2018, when our team analyzed an initial review of 45 health app evaluation frameworks and found that the most common category of consideration among the different frameworks was usability, with short term usability highly overrepresented compared to privacy and evidence with base. In this 2018 review, there were 93 unique questions corresponding to short-term usability but only 10 to the presence of a privacy policy. Although many frameworks continue to consider usability, our current review suggests the most common questions across frameworks now concern evidence, clinical foundation, and privacy. This shift may reflect an increased recognition of the privacy dangers some apps may pose.

This review illustrates the challenges in conceiving a comprehensive evaluation model. A continued concern in mobile health apps is engagement,⁶ and it is unclear whether any framework adequately predicts engagement. Another persistent challenge is striking a balance between transparency/objectivity and subjectivity. Questions that prompt consideration of subjective user experiences may limit the generalizability and standardization of a framework, as the questions inherently reflect the experience of the rater. An app's ease of use, for example, will differ significantly depending on an individual's level of comfort and experience with technology. However, subjective questions around user friendliness, visual appeal, and interface design may be of greatest concern to an app user, and most predictive of engagement with an app.²¹ Finally, a thorough assessment of an app is only feasible if information about the app is available. For example, some questions with clinical significance, such as the consideration of how peers or target users may be involved in app development, are not easily answerable by a health app consumer. Overall, there is a need for more data and transparency when it comes to health apps. App evaluation frameworks, while thorough, rigorous, and tailored to clinical app use, can only go so far without transparency on the part of app developers.²²

The analysis additionally highlighted the flexibility and comprehensiveness of the MIND framework, which was used as a reference framework in this review, in diverse contexts. The MIND categories are inclusive of a wide range of frameworks and questions. Even without including any subjective questions in the mapping process, each of the 70 frameworks that were

1
2
3 ultimately mapped had some overlap with MIND, and many of the 1,700 questions ultimately
4 included were mapped exactly with a MIND question. Although MIND was initially
5 conceptualized as an evaluation tool specifically for mental health apps, the coherence between
6 MIND and diverse types of app evaluation frameworks, such as those for concussion²³, heart
7 disease²⁴, and sickle cell anemia²⁵, demonstrates how the MIND categories can encompass many
8 health domains. Condition-specific questions, for example, are a good fit for the “Features &
9 Engagement” category of MIND.
10
11

12 The results of our analysis suggest while numerous new app evaluation frameworks continue to
13 emerge, there is a naturally appearing standard of common questions asked across all. While
14 different use cases and medical subspecialties will require unique questions to evaluate apps,
15 there are set of common questions around aspects like privacy and level of evidence that are
16 more universal. MIND appears to cover a large subset of these questions and thus may offer a
17 useful starting point for new efforts as well as means to consolidate exiting efforts. Advantages
18 of the more objective approach offered by MIND is that it can be represented as a research
19 database to facilitate discovery of apps while not conflicting with local needs, personal
20 preferences, or cultural priorities.²⁶
21
22
23
24

25 **Limitations**

26 Our work is not the first to compare app evaluation frameworks. Recently, several reviews have
27 compared how different mobile health app evaluation models address privacy,^{11,12,14} and another
28 database (<https://search.appcensus.io/>) focuses exclusively on compiling privacy assessments of
29 Android apps. We chose to exclude app evaluation frameworks that focused exclusively on in-
30 depth privacy considerations and were unusable by a clinician or layperson as our goal was more
31 comprehensive app evaluation. This decision is not to reject considerations of privacy and
32 security that are of critical importance, but rather to narrow the focus to frameworks that are
33 usable in the hands of the public today and can be used to inform clinical decisions. In addition,
34 MIND was initially tailored to mental health, and thus does not encompass thorough disease-
35 specific criteria for other conditions, such as asthma, diabetes, and sickle cell anemia – though
36 such questions may be easily integrated. Finally, subjective questions, especially those around
37 ease of use and visual appeal, are difficult to standardize but may be among the most important
38 features driving user engagement with mental health apps.²¹
39
40
41

42 **Conclusion**

43
44 Our work demonstrates expansion of app evaluation frameworks. By illustrating how the MIND
45 overlaps with many of these existing and emerging frameworks – we suggest the practical need
46 for consolidation. Although specific disease tailored mobile health apps require specialized app
47 evaluation *questions*, concerns around accessibility, privacy, clinical foundation, and
48 interoperability are nonspecific. If the full potential of digital health can be realized, there is a
49 need for increased collaboration among industry, government, and academia in order to ensure
50 that the highest quality digital health tools reach the public. We emphasize that this effort is just
51 a first step, and highlight the need for interdisciplinary continued communication amongst
52 diverse digital health stakeholders in order to best serve the public.
53
54
55
56
57
58
59
60

Funding Statement: This work was supported by a gift from the Argosy Foundation

Declaration of Interests: None declared. JT serves as chair of the APA's Health IT committee, which created the APA app evaluation framework but receives no compensation or support for this volunteer work.

Data Availability Statement: All data are pretended in appendix A and B.

Author Contributions: SL and JT designed the procedure. SL and JT screened articles for eligibility. SL and LS compiled and mapped questions from frameworks. SL and JT composed manuscript.

Figure Legends

Figure 1: *A screenshot of MIND highlighting several of the app evaluation questions (green boxes) and ability to access more*

Figure 2: *Framework identification through database searches (PubMed, EMBASE, PsychINFO) and other sources (reviews since 2018, grey literature, government websites).*

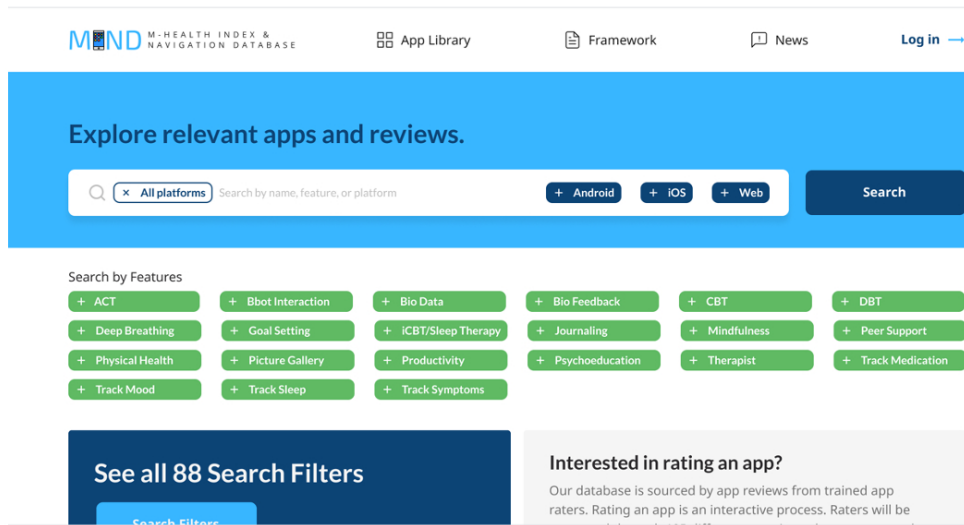
Figure 3: *The most commonly addressed questions, grouped within the categories of MIND. The blue triangle constitutes MIND and its six main categories, while the green trapezoid represents questions pertaining to usability or ease of use, which are not covered by MIND.*

References:

1. Health C for D and R. Digital Health Policies and Public Health Solutions for COVID-19. *FDA*. Published online August 3, 2020. Accessed September 14, 2020. <https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/digital-health-policies-and-public-health-solutions-covid-19>
2. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. *The Lancet Digital Health*. 2019;1(2):e52-e54. doi:10.1016/S2589-7500(19)30013-5
3. Ondersma SJ, Walters ST. Clinician's Guide to Evaluating and Developing eHealth Interventions for Mental Health. *Psychiatr res clin pract*. 2020;2(1):26-33. doi:10.1176/appi.prcp.2020.20190036
4. Bergin A, Davies EB. Technology Matters: Mental health apps – separating the wheat from the chaff. *Child and Adolescent Mental Health*. 2020;25(1):51-53. doi:10.1111/camh.12363
5. Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. *BMC Medicine*. 2019;17(1):231. doi:10.1186/s12916-019-1461-z
6. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. *npj Digit Med*. 2019;2(1):54. doi:10.1038/s41746-019-0129-6

- 1
- 2
- 3 7. Magee JC, Adut S, Brazill K, Warnick S. Mobile App Tools for Identifying and Managing Mental Health
- 4 Disorders in Primary Care. *Curr Treat Options Psychiatry*. 2018;5(3):345-362. doi:10.1007/s40501-018-0154-
- 5 0
- 6
- 7 8. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating
- 8 expert frameworks into objective metrics. *npj Digital Medicine*. 2020;3(1):1-8. doi:10.1038/s41746-020-
- 9 00312-4
- 10
- 11 9. Moshi MR, Tooher R, Merlin T. SUITABILITY OF CURRENT EVALUATION FRAMEWORKS FOR USE IN THE HEALTH
- 12 TECHNOLOGY ASSESSMENT OF MOBILE MEDICAL APPLICATIONS: A SYSTEMATIC REVIEW. *International*
- 13 *Journal of Technology Assessment in Health Care*. 2018;34(5):464-475. doi:10.1017/S026646231800051X
- 14
- 15 10. Coravos A, Doerr M, Goldsack J, et al. Modernizing and designing evaluation frameworks for connected sensor
- 16 technologies in medicine. *npj Digital Medicine*. 2020;3(1):1-10. doi:10.1038/s41746-020-0237-3
- 17
- 18 11. Nurgalieva L, O'Callaghan D, Doherty G. Security and Privacy of mHealth Applications: A Scoping Review. *IEEE*
- 19 *Access*. 2020;8:104247-104268. doi:10.1109/ACCESS.2020.2999934
- 20
- 21 12. Benjumea J, Ropero J, Rivera-Romero O, Dorrnoro-Zubiete E, Carrasco A. Privacy Assessment in Mobile
- 22 Health Apps: Scoping Review. *JMIR mHealth and uHealth*. 2020;8(7). doi:10.2196/18868
- 23
- 24 13. Levine DM, Co Z, Newmark LP, et al. Design and testing of a mobile health application rating tool. *npj Digital*
- 25 *Medicine*. 2020;3(1):1-7. doi:10.1038/s41746-020-0268-9
- 26
- 27 14. Hussain M, Zaidan AA, Zidan BB, et al. Conceptual framework for the security of mobile health applications on
- 28 Android platform. *Telematics and Informatics*. 2018;35(5):1335-1354. doi:10.1016/j.tele.2018.03.005
- 29
- 30 15. McMillan B, Hickey E, Patel MG, Mitchell C. Quality assessment of a sample of mobile app-based health
- 31 behavior change interventions using a tool based on the National Institute of Health and Care Excellence
- 32 behavior change guidance. *Patient Education and Counseling*. 2016;99(3):429-435.
- 33 doi:10.1016/j.pec.2015.10.023
- 34
- 35 16. Shaia KL, Farag S, Chyjek K, Knopman J, Chen KT. An Evaluation of Mobile Applications for Reproductive
- 36 Endocrinology and Infertility Providers. *Telemed J E Health*. 2017;23(3):254-258. doi:10.1089/tmj.2016.0079
- 37
- 38 17. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: A New
- 39 Tool for Assessing the Quality of Health Mobile Apps. *JMIR mHealth and uHealth*. 2015;3(1):e27.
- 40 doi:10.2196/mhealth.3422
- 41
- 42 18. Wykes T, Schueller S. Why Reviewing Apps Is Not Enough: Transparency for Trust (T4T) Principles of
- 43 Responsible Health App Marketplaces. *Journal of Medical Internet Research*. 2019;21(5):e12390.
- 44 doi:10.2196/12390
- 45
- 46 19. Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support technology
- 47 selection in digital health research. *Transl Behav Med*. Published online May 23, 2019.
- 48 doi:10.1093/tbm/ibz074
- 49
- 50 20. Noee M, Baba akbari A, Olyaeemanesh A, Mobinizadeh M. Prioritizing the Potential Applications of Mobile-
- 51 Health in the Iranian Health System. *Journal of Research in Health Sciences*. 2020;20.
- 52 doi:10.34172/jrhs.2020.08
- 53
- 54 21. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. *Health Informatics J*.
- 55 2020;26(3):2042-2066. doi:10.1177/1460458219896492
- 56
- 57
- 58
- 59
- 60

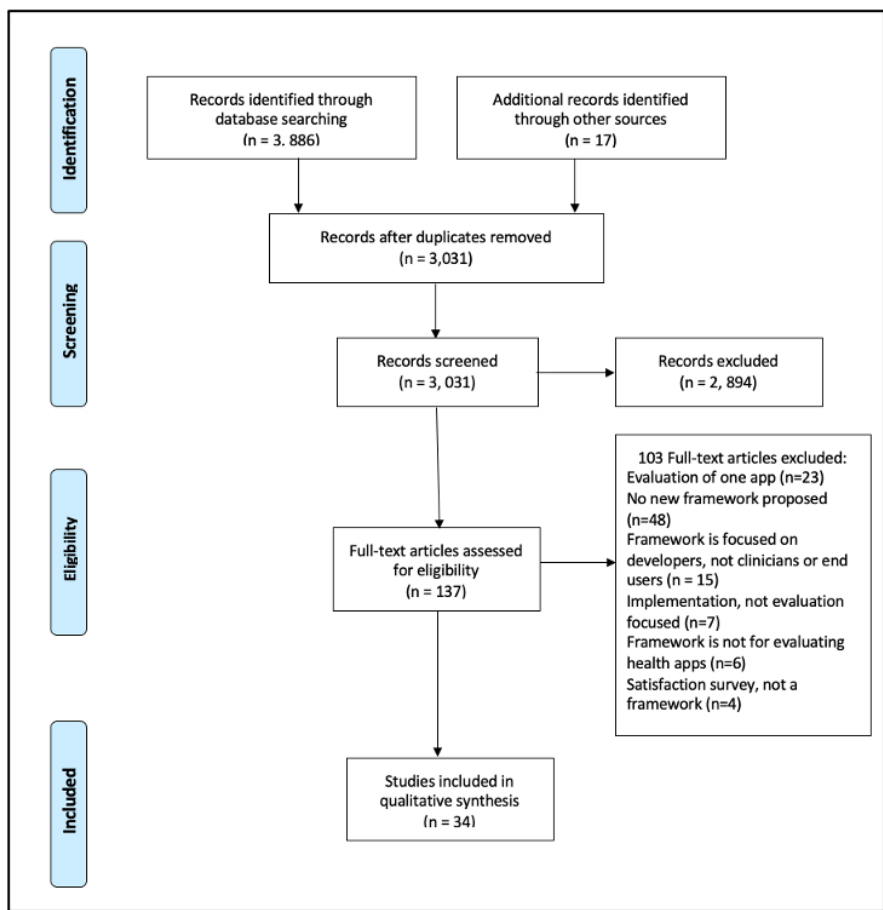
- 1
2
3 22. Rodriguez-Villa E, Torous J. Regulating digital health technologies with transparency: the case for dynamic and
4 multi-stakeholder evaluation. *BMC Med.* 2019;17. doi:10.1186/s12916-019-1447-x
5
6 23. Lee H, Sullivan SJ, Schneiders AG, et al. Smartphone and tablet apps for concussion road warriors (team
7 clinicians): a systematic review for practical users. *Br J Sports Med.* 2015;49(8):499-505.
8 doi:10.1136/bjsports-2013-092930
9
10 24. Farnia T, Jaulent M-C, Steichen O. Evaluation Criteria of Noninvasive Telemonitoring for Patients With Heart
11 Failure: Systematic Review. *Journal of Medical Internet Research.* 2018;20(1):e16. doi:10.2196/jmir.7873
12
13 25. Shah N, Jonassaint J, Castro LD. Patients Welcome the Sickle Cell Disease Mobile Application to Record
14 Symptoms via Technology (SMART). *Hemoglobin.* 2014;38(2):99-103. doi:10.3109/03630269.2014.880716
15
16 26. Hoffman L, Wisniewski H, Hays R, et al. Digital Opportunities for Outcomes in Recovery Services (DOORS): A
17 Pragmatic Hands-On Group Approach Toward Increasing Digital Health and Smartphone Competencies,
18 Autonomy, Relatedness, and Alliance for Those With Serious Mental Illness. *J Psychiatr Pract.* 2020;26(2):80-
19 88. doi:10.1097/PRA.0000000000000450
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



1: A screenshot of MIND highlighting several of the app evaluation questions (green boxes) and ability to access more

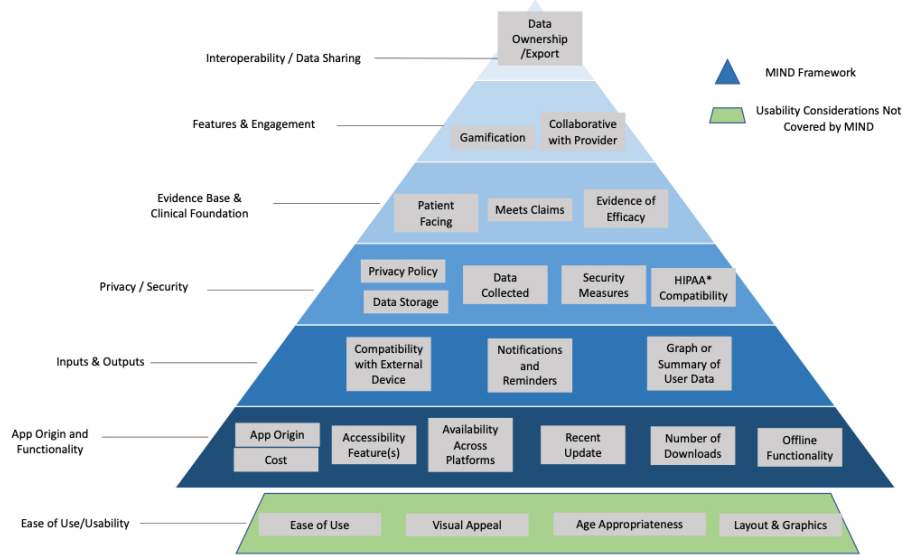
381x207mm (72 x 72 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



2: PRISMA Diagram

291x287mm (72 x 72 DPI)



3: The most commonly addressed questions, grouped within the categories of MIND. The blue triangle constitutes MIND and its six main categories, while the green trapezoid represents questions pertaining to usability or ease of use, which are not covered by MIND.

390x266mm (72 x 72 DPI)

Appendix A: List of Reviewed Frameworks

PsyberGuide(1)(2), Mobile App Rating Scale (MARS)(3), AppScript Score(4), MindTech toolkit(5), NHS Digital Development Lab(6), Digital Assessment Questionnaire(7), PAS 277:2015(8), Cambridge Health Alliance(9), mHIMSS (10), NICE Behavior Change Guidance(11), AMA Principles for Safe Apps(12), Mindtools.io, ORCHA(13), Xcertia (14)(15), Guidelines for Mental Health App Evaluation Framework(16), mHealth Quality Label(17), Medical App Evaluation(18), Mobile Health Evaluation Framework(19), APPLICATIONS scoring system(20), Evaluating Sickle Cell Mobile App(21), mHealth App Evaluation for HIV(22), mHealth App Evaluation for HIV (23), mHealth App Evaluation for Pain Management(24), Evaluation of Mobile Clinical Applications(25), Quality of Experience (QoE) Survey(26), Concussion App Evaluation, Apps for Pain Management(27), Evaluation Tool for Healthcare Smartphone Applications(28), IMS: Patient Apps for Improved Healthcare(29), Mobile Apps for Asthma(30), Assessing Mobile Health App Quality(31), Certification Program for App Quality and Safety(32), Apps and Eating Disorders(33), Evidence-based Mobile Medical Applications in Diabetes(34), Diabetes Self-Management Applications(35), Framework for Evaluating Mobile Applications for Cardiac Rehabilitation(36), PIS: graphical classification tool for mHealth apps(37), Usability Evaluation of Mobile Applications for Diabetics(38), DiaDigital Apps(39), Applications of Mobile-Health in Iranian Health System(40), A Health Technology Assessment Module for Evaluating Mobile Medical Applications(41), Transparency4Trust(42), Graded Review of Dermatology Apps(43), What Makes a Good Health App?(44), Development a Guide for Mobile Health-Related Apps (45), Assessment Framework for COVID-19 Apps (46), Assessment of the Transparency and Reliability of Health Information Dissemination(47), Medical Mobile App Classification(48), Framework for Digital Support for the Autism Community (49), mHealth App Trustworthiness Checklist(50), THESIS(51), Framework for Evaluating Quality of mHealth Apps for Adolescent Users(52), ABACUS(53), Alberta Rating Index(54), Evaluation Framework for Digital Health Interventions(55), Mobile Applications Recommendations(56), Assessment Framework for Quality of Asthma Smartphone Applications (57), Global Digital Health Scorecard(58), NASS(59), Australian NSQDMH(60), Express Scripts Digital Health Formulary (61), Medical App Checker (62), Modernizing and Designing Evaluation Frameworks (63), Enlight (64), Improving the Quality of Apps for Patient Use(65), Assessment Framework for e-Mental Health in Canada(66), A Decision-Making Checklist to Support Technology Selection(67), Canadian Mental Health Commission (68), Trust4App(69), Evaluation Criteria (70), IDEAS(71)

Appendix B: Question-by-Question Analysis of MIND

Question	Frameworks with representative question
App Origin	26
Does it come from the government?	6
Does it come from a for-profit company or developer?	4

Does it come from a non-profit company?	3
Does it come from a trusted healthcare company?	6
Does it come from an academic institution?	3
App Functionality	42
Does it work on Apple(iOS)?	9
What is the Apple version?	0
What is the oldest iOS version supported?	0
What was the Apple release date?	1
When was the last Apple (IOS) update?	7
Has the apple version been updated in the last 180 days?	9
Number of reviews on Apple store?	9
Rating (number of stars) on Apple store?	9
App size on iOS?	2
Does it work on Android?	9
What is the Android version?	0
What is the oldest Android version supported?	0
What was the Google play store release date?	1
When was the last Android update?	7
Has the android version been updated in the last 180 days?	9
Number of reviews on google play store?	9
Rating (number of stars) on google play store?	9
App size on android?	2
Does the app work offline?	7
Does it have at least one accessibility feature (like adjust text size, text to voice, or colorblind color scheme adjuster)?	11
Does it have a web version?	5
Does it work with Spanish?	0
Does it work with a language other than English or Spanish?	5
Is the app totally free?	8
What is the cost up front?	16
Are there in-app purchases?	6
Is it a subscription (recurrent/monthly/annual)?	3
Inputs & Outputs	17
Input: surveys?	2

1		
2		
3		
4	Input: Diary?	1
5	Input: Geolocation?	0
6		
7	Input: contact list?	0
8		
9	Input: Camera?	0
10		
11	Input: Microphone?	0
12		
13	Input: step count?	0
14	Input: external devices (e.g. a wearable sending direct data)?	4
15	Input: social network?	1
16		
17	Output: notifications?	6
18	Output: psychoeducational references/information?	0
19		
20	Output: social network?	1
21		
22	Output: reminders?	3
23	Output: graphs of data?	3
24		
25	Output: summary of data (in text or numbers)?	3
26		
27	Output: link to formal care/coaching?	2
28	Privacy & Security	43
29		
30	Is there a privacy policy?	20
31	Does the app declare data use and purpose?	18
32	Does the app report security measures in place?	25
33		
34	Is PHI shared?	9
35		
36	Is de-identified data shared?	5
37		
38	Is anonymized/aggregate data shared?	5
39		
40	Can you opt out of data collection?	4
41		
42	Can you delete your data?	3
43		
44	Is the user data stored only on the device?	5
45	Is the user data stored on a server?	9
46	Does the app have a crisis management feature?	4
47	Does the app claim it meets HIPAA (or analogous national standard for patient/PHI privacy protection)	10
48		
49	Reading level of the privacy policy (what grade reading level)?	4
50		
51	Does the app use 3rd party vendors (i.e. google analytics, etc)?	4
52		
53	Evidence & Clinical Foundation	57
54		
55	Is the app content well-written, correct, and relevant?	25
56	Does the app appear to do what it claims to do?	18
57		
58		
59		
60		

1		
2		
3		
4	Is the app patient facing?	28
5	How many feasibility/usability studies?	14
6		
7	What is the highest feasibility impact factor?	1
8		
9	How many evidence/efficacy studies?	27
10		
11	What is the highest efficacy impact factor?	4
12		
13	Can the app cause harm?	7
14		
15	Does the app provide any warning for use?	11
16	Features & Engagement Style	29
17	Features: mood tracking?	4
18	Features: medication tracking?	2
19		
20	Features: sleep tracking?	1
21		
22	Features: physical exercise tracking?	1
23		
24	Features: psychoeducation?	3
25		
26	Features: journaling?	0
27		
28	Features: picture gallery/hope board?	0
29		
30	Features: mindfulness?	1
31		
32	Features: deep breathing?	0
33		
34	Features: iCBT or sleep therapy?	1
35		
36	Features: CBT?	0
37		
38	Features: ACT?	0
39		
40	Features: DBT?	0
41		
42	Features: peer support?	1
43		
44	Features: connection to coach/therapist?	1
45		
46	Features: biodata?	2
47		
48	Features: goal setting/habits?	3
49		
50	Features: physical health exercises?	2
51		
52	Features: Chatbot interaction (like with virtual character)?	0
53		
54	Features: Biofeedback with sense data (eeg, HRV, skin conductance, etc)?	2
55		
56	Engagement style: user generated data?	4
57		
58	Engagement style: chat/message based?	1
59		
60	Engagement style: is it a screener/assessment?	3
	Engagement style: real time response?	3
	Engagement style: Asynchronous response?	0

Engagement style: gamification (points, badges)?	7
Engagement style: videos?	2
Engagement style: audio/music/scripts?	2
Engagement style: AI support?	0
Engagement style: peer support?	6
Engagement style: network support?	5
Engagement style: Collaborative with provider/other?	8
App Use	8
Is it a self-help/self-management tool?	6
Is it a reference app?	1
Is it intended for hybrid use with a clinician in conjunction with treatment plan?	3
Interoperability & Data Sharing	23
Do you own your data?	8
Can you email or export your data?	10
Can you send your data to a medical record?	5

Full Framework Reference List

1. Neary M, Schueller SM. State of the Field of Mental Health Apps. *Cognitive and Behavioral Practice*. 2018 Nov 1;25(4):531–7.
2. About One Mind PsyberGuide [Internet]. One Mind PsyberGuide. [cited 2020 Oct 19]. Available from: <https://live-one-mind-psyberguide.pantheonsite.io/about-psyberguide/>
3. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. *JMIR Mhealth Uhealth* [Internet]. 2015 Mar 11 [cited 2020 Oct 19];3(1). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4376132/>
4. AppScript | Discover, Deliver & Track Digital Health [Internet]. [cited 2020 Oct 19]. Available from: <https://www.appscript.net/score-details>
5. Toolkit for Appraising Digital Mental Health Products - Technology theme - MindTech [Internet]. 2017 [cited 2020 Oct 19]. Available from: <https://web.archive.org/web/20170701091502/http://mindtech.org.uk/projects/71-digital-mental-health-tools-evaluation-criteria-technology-theme.html>
6. Betton V, Craven M, Davies B, Martin J, Nelissen N, Simons L. Framework for the effectiveness evaluation of mobile (mental) health tools. :42.
7. Digital-Assessment-Questions-V2.1-Beta-PDF.pdf [Internet]. [cited 2020 Oct 19]. Available from: <https://developer.nhs.uk/wp-content/uploads/2018/09/Digital-Assessment-Questions-V2.1-Beta-PDF.pdf>

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
8. British Standards Institution. Health and wellness apps: quality criteria across the life cycle : code of practice. 2015.
9. Hoffman L. Mobile App Technology Meets Collaborative Care [Internet]. Available from: http://app.ihl.org/FacultyDocuments/Events/Event-2930/Presentation-15621/Document-13034/Presentation_A1_B1_Mobile_App_Technology_benedetto.pdf
10. Selecting a Mobile App: Evaluating the Usability of Medical Applications. :32.
11. McMillan B, Hickey E, Patel MG, Mitchell C. Quality assessment of a sample of mobile app-based health behavior change interventions using a tool based on the National Institute of Health and Care Excellence behavior change guidance. *Patient Education and Counseling*. 2016 Mar;99(3):429–35.
12. AMA adopts principles to promote safe, effective mHealth applications [Internet]. American Medical Association. [cited 2020 Oct 19]. Available from: <https://www.ama-assn.org/press-center/press-releases/ama-adopts-principles-promote-safe-effective-mhealth-applications>
13. Ltd FS. ORCHA Review Process | Reviewing Health and Care Apps | ORCHA [Internet]. [cited 2020 Oct 19]. Available from: <https://www.orchac.co.uk/our-solution/the-orchac-review/#0>
14. mHealthIntelligence. Amid a Flood of New mHealth Apps, Xcertia Looks to Set Standards [Internet]. mHealthIntelligence. 2018 [cited 2020 Oct 19]. Available from: <https://mhealthintelligence.com/news/amid-a-flood-of-new-mhealth-apps-xcertia-looks-to-set-standards>
15. Interoperability & Health Information Exchange Committee | HIMSS [Internet]. 2016 [cited 2020 Oct 19]. Available from: <https://www.himss.org/resources/interoperability-health-information-exchange-committee>
16. Chan S, Torous J, Hinton L, Yellowlees P. Towards a Framework for Evaluating Mobile Mental Health Apps. *Telemed J E Health*. 2015 Dec;21(12):1038–41.
17. Yasini M, Marchand G. Mobile Health Applications, in the Absence of an Authentic Regulation, Does the Usability Score Correlate with a Better Medical Reliability? *Stud Health Technol Inform*. 2015;216:127–31.
18. Walsworth DT. Medical Apps: Making Your Mobile Device a Medical Device. *FPM*. 2012 Jun;19(3):10–3.
19. Developing a Framework for Evaluating the Patient Engagement, Quality, and Safety of Mobile Health Applications | Commonwealth Fund [Internet]. [cited 2020 Oct 19]. Available from: <https://www.commonwealthfund.org/publications/issue-briefs/2016/feb/developing-framework-evaluating-patient-engagement-quality-and>
20. Shaia KL, Farag S, Chyjek K, Knopman J, Chen KT. An Evaluation of Mobile Applications for Reproductive Endocrinology and Infertility Providers. *Telemed J E Health*. 2017;23(3):254–8.
21. Shah N, Jonassaint J, De Castro L. Patients welcome the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). *Hemoglobin*. 2014;38(2):99–103.
22. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown Iii W. Comparison of a User-Centered Design, Self-Management App to Existing mHealth Apps for Persons Living With HIV. *JMIR Mhealth Uhealth*. 2015 Sep 18;3(3):e91.

23. Robustillo Cortés M de las A, Cantudo Cuenca MR, Morillo Verdugo R, Calvo Cidoncha E. High quantity but limited quality in healthcare applications intended for HIV-infected patients. *Telemed J E Health*. 2014 Aug;20(8):729–35.
24. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs SJ, et al. Assessing the quality and usability of smartphone apps for pain self-management. *Pain Med*. 2014 Jun;15(6):898–909.
25. Murfin M. Know your apps: an evidence-based approach to evaluation of mobile clinical applications. *J Physician Assist Educ*. 2013;24(3):38–40.
26. Martínez-Pérez B, de la Torre-Díez I, Candelas-Plasencia S, López-Coronado M. Development and evaluation of tools for measuring the quality of experience (QoE) in mHealth applications. *J Med Syst*. 2013 Oct;37(5):9976.
27. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. “There’s a Pain App for That”: Review of Patient-targeted Smartphone Applications for Pain Management. *Clin J Pain*. 2015 Jun;31(6):557–63.
28. Jin M, Kim J. Development and Evaluation of an Evaluation Tool for Healthcare Smartphone Applications. *Telemed J E Health*. 2015 Oct;21(10):831–7.
29. patient-adoption-of-mhealth.pdf [Internet]. [cited 2020 Oct 19]. Available from: <https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf>
30. Huckvale K, Morrison C, Ouyang J, Ghaghda A, Car J. The evolution of mobile apps for asthma: an updated systematic assessment of content and tools. *BMC Med*. 2015 Dec;13(1):58.
31. Grundy QH, Wang Z, Bero LA. Challenges in Assessing Mobile Health App Quality: A Systematic Review of Prevalent and Innovative Methods. *Am J Prev Med*. 2016;51(6):1051–9.
32. Álvarez-Rementería JF. QUALITY AND SAFETY STRATEGY FOR MOBILE HEALTH APPLICATIONS. :18.
33. Fairburn CG, Rothwell ER. Apps and eating disorders: A systematic clinical appraisal. *Int J Eat Disord*. 2015 Nov;48(7):1038–46.
34. Drincic A, Prahalad P, Greenwood D, Klonoff DC. Evidence-based Mobile Medical Applications in Diabetes. *Endocrinol Metab Clin North Am*. 2016 Dec;45(4):943–65.
35. Demidowich AP, Lu K, Tamler R, Bloomgarden Z. An evaluation of diabetes self-management applications for Android smartphones. *J Telemed Telecare*. 2012 Jun;18(4):235–8.
36. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. *J Am Heart Assoc*. 2013 Nov 1;2(6):e000568.
37. Basilico A, Marceglia S, Bonacina S, Pincioli F. Advising patients on selecting trustful apps for diabetes self-care. *Computers in Biology and Medicine*. 2016 Apr 1;71:86–96.
38. Arnhold M, Quade M, Kirch W. Mobile Applications for Diabetics: A Systematic Review and Expert-Based Usability Evaluation Considering the Special Requirements of Diabetes Patients Age 50 Years or Older. *Journal of Medical Internet Research*. 2014;16(4):e104.
39. Kaltheuner M, Drossel D, Heinemann L. DiaDigital Apps: Evaluation of Smartphone Apps Using a Quality Rating Methodology for Use by Patients and Diabetologists in Germany. *J Diabetes Sci Technol*. 2018 Sep 28;13(4):756–62.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
40. Noee M, Baba akbari A, Olyaeemanesh A, Mobinizadeh M. Prioritizing the Potential Applications of Mobile-Health in the Iranian Health System. *Journal of Research in Health Sciences*. 2020 Mar 7;20.
 41. Moshi MR, Tooher R, Merlin T. Development of a health technology assessment module for evaluating mobile medical applications. *Int J Technol Assess Health Care*. 2020 Jun;36(3):252–61.
 42. Wykes T, Schueller S. Why Reviewing Apps Is Not Enough: Transparency for Trust (T4T) Principles of Responsible Health App Marketplaces. *Journal of Medical Internet Research*. 2019;21(5):e12390.
 43. Masud A, Shafi S, Rao BK. Mobile medical apps for patient education: a graded review of available dermatology apps. *Cutis*. 2018 Feb;101(2):141–4.
 44. Dawson RM, Felder TM, Donevant SB, McDonnell KK, Card EB, King CC, et al. What makes a good health “app”? Identifying the strengths and limitations of existing mobile application evaluation tools. *Nurs Inq*. 2020;27(2):e12333.
 45. Llorens-Vernet P, Miró J. Standards for Mobile Health–Related Apps: Systematic Review and Development of a Guide. *JMIR mHealth and uHealth*. 2020;8(3):e13057.
 46. Vokinger KN, Nittas V, Witt CM, Fabrikant SI, von Wyl V. Digital health and the COVID-19 epidemic: an assessment framework for apps from an epidemiological and legal perspective. *Swiss Medical Weekly [Internet]*. 2020 May 17 [cited 2020 Oct 19];150(1920). Available from: <https://smw.ch/article/doi/smw.2020.20282>
 47. Huang Z, Lum E, Car J. Medication Management Apps for Diabetes: Systematic Assessment of the Transparency and Reliability of Health Information Dissemination. *JMIR mHealth and uHealth*. 2020;8(2):e15364.
 48. Nwe K, Larsen ME, Nelissen N, Wong DC-W. Medical Mobile App Classification Using the National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health Technologies: Interrater Reliability Study. *Journal of Medical Internet Research*. 2020;22(6):e17457.
 49. Zervogianni V, Fletcher-Watson S, Herrera G, Goodwin M, Pérez-Fuster P, Brosnan M, et al. A framework of evidence-based practice for digital support, co-developed with and for the autism community: Autism [Internet]. 2020 Feb 6 [cited 2020 Oct 19]; Available from: <https://journals.sagepub.com/eprint/4KEDZJ7YYCZAUHJ7B2NR/full>
 50. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth App Trustworthiness checklist. *DIGITAL HEALTH*. 2019 Jan 1;5:2055207619886463.
 51. Levine DM, Co Z, Newmark LP, Groisser AR, Holmgren AJ, Haas JS, et al. Design and testing of a mobile health application rating tool. *npj Digital Medicine*. 2020 May 21;3(1):1–7.
 52. Jeminiwa RN, Hohmann NS, Fox BI. Developing a Theoretical Framework for Evaluating the Quality of mHealth Apps for Adolescent Users: A Systematic Review. *J Pediatr Pharmacol Ther*. 2019;24(4):254–69.
 53. McKay FH, Slykerman S, Dunn M. The App Behavior Change Scale: Creation of a Scale to Assess the Potential of Apps to Promote Behavior Change. *JMIR mHealth and uHealth*. 2019;7(1):e11130.
 54. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. *Disability and Rehabilitation: Assistive Technology*. 2020 Jan 8;1–10.

- 1
2
3 55. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for
4 digital health interventions. *it - Information Technology*. 2019 Oct 25;61(5–6):253–63.
5
6 56. Jisha RC, Krishnan R, Vikraman V. Mobile Applications Recommendation Based on User Ratings
7 and Permissions. In: 2018 International Conference on Advances in Computing, Communications
8 and Informatics (ICACCI). 2018. p. 1000–5.
9
10 57. Guan Z, Sun L, Xiao Q, Wang Y. Constructing an assessment framework for the quality of asthma
11 smartphone applications. *BMC Medical Informatics and Decision Making*. 2019 Oct 15;19(1):192.
12
13 58. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to
14 validation. *npj Digital Medicine*. 2019 May 13;2(1):1–9.
15
16 59. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A’Court C, et al. Beyond Adoption: A
17 New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the
18 Scale-Up, Spread, and Sustainability of Health and Care Technologies. *Journal of Medical Internet
19 Research*. 2017;19(11):e367.
20
21 60. National Safety and Quality Digital Mental Health Standards | Australian Commission on Safety and
22 Quality in Health Care [Internet]. [cited 2020 Oct 23]. Available from:
23 [https://www.safetyandquality.gov.au/standards/national-safety-and-quality-digital-mental-health-
24 standards](https://www.safetyandquality.gov.au/standards/national-safety-and-quality-digital-mental-health-standards)
25
26 61. Digital Health Formulary | Express Scripts [Internet]. [cited 2020 Oct 23]. Available from:
27 <https://www.express-scripts.com/corporate/tag/digital-health-formulary>
28
29 62. Medical App Checker: a Guide to assessing Mobile Medical Apps [Internet]. [cited 2020 Oct 23].
30 Available from: [https://www.knmg.nl/actualiteit-opinie/nieuws/nieuwsbericht/medical-app-checker-
31 a-guide-to-assessing-mobile-medical-apps.htm](https://www.knmg.nl/actualiteit-opinie/nieuws/nieuwsbericht/medical-app-checker-a-guide-to-assessing-mobile-medical-apps.htm)
32
33 63. Coravos A, Doerr M, Goldsack J, Manta C, Shervey M, Woods B, et al. Modernizing and designing
34 evaluation frameworks for connected sensor technologies in medicine. *npj Digital Medicine*. 2020
35 Mar 13;3(1):1–10.
36
37 64. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: A Comprehensive Quality and
38 Therapeutic Potential Evaluation Tool for Mobile and Web-Based eHealth Interventions. *Journal of
39 Medical Internet Research*. 2017;19(3):e82.
40
41 65. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps
42 for patient use? *BMC Medicine*. 2018 Dec 3;16(1):225.
43
44 66. Zelmer J, Hoof K van, Notarianni M, Mierlo T van, Schellenberg M, Tannenbaum C. An Assessment
45 Framework for e-Mental Health Apps in Canada: Results of a Modified Delphi Process. *JMIR
46 mHealth and uHealth*. 2018;6(7):e10016.
47
48 67. Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support
49 technology selection in digital health research. *Transl Behav Med*. 2020 Oct 8;10(4):1004–15.
50
51 68. eMH_app_eng.pdf [Internet]. [cited 2020 Oct 23]. Available from:
52 https://www.mentalhealthcommission.ca/sites/default/files/2018-01/eMH_app_eng.pdf
53
54 69. Habib SM, Alexopoulos N, Islam MM, Heider J, Marsh S, Muehlhaeuser M. Trust4App: Automating
55 Trustworthiness Assessment of Mobile Applications. In: 2018 17th IEEE International Conference
56 On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International
57 Conference On Big Data Science And Engineering (TrustCom/BigDataSE). 2018. p. 124–35.
58
59
60

- 1
2
3 70. Farnia T, Jaulent M-C, Steichen O. Evaluation Criteria of Noninvasive Telemonitoring for Patients
4 With Heart Failure: Systematic Review. *Journal of Medical Internet Research*. 2018;20(1):e16.
5
6 71. Fedele DA, McConville A, Moon J, Thomas JG. Topical Review: Design Considerations When
7 Creating Pediatric Mobile Health Interventions: Applying the IDEAS Framework. *J Pediatr Psychol*.
8 2019 Apr 1;44(3):343–8.
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60