Actionable App Evaluation: Objective Standards to Guide Assessment and Implementation of Digital Health **Interventions**





THE DIVISION OF DIGITAL PSYCHIATRY AT BIDMC Beth Israel Deaconess Medical Center



HARVARD MEDICAL SCHOOL TEACHING HOSPITAL

Outline







The Need for More Rigorous App Evaluation

- Many good apps exist and can augment care.
- But there are also many dangerous apps, including those that expose or sell personal health data.
- Most are not tightly regulated (categorized as "Health & Fitness," not as medical devices) and many make false claims about effectiveness.



- Most apps do **NOT** claim to be HIPAA compliant
- In one study, only 50% of apps shared data securely.
- 80% shared health-related data to third parties, with the remaining 20% storing data on the phones.



HARVARD MEDICAL SCHOOL TEACHING HOSPITAL

Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and Privacy Analysis of Mobile Health Applications: The Alarming State of Practice. IEEE Access. 2018 Jan 29.





Huckvale K, Torous J, Larsen ME. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. JAMA Netw Open. Published online April 19, 20192(4):e192542. doi:10.1001/jamanetworkopen.2019.2542



CHIATRY







- "A security researcher said he was forced to take down a blog post describing an apparent bug in Talkspace's website"
- "within hours of Jackson publishing his findings...Talkspace sent Jackson a cease and desist letter, accusing the researcher of defaming Talkspace"





Molly Osberg and Dhruv Mehrotra 2/19/20 1:39PM • Filed to: BETTER HELP \/

- "...sensitive information does end up being shared—all with the ostensible goal of better tracking user behavior, and perhaps giving social media companies an easy way to see who's feeling depressed."
- "[BetterHelp]...brings up questions about how a person's intimate, supposedly private sessions might be exploited by advertisers, an industry that isn't exactly known for operating in good faith."



Exaggerated Claims of Effectiveness



Stimulate Vital Areas of the Brain and Heal Naturally.

Prepare Yourself For Coronavirus! Download Flu treatment From Our Vital Tones Physical app.

Vital Tones has discovered an evolutionary brainwave technology which can stimulate specific parts of the brain and the body.

- This app purports to provide treatment for Depression,
 Bipolar Disorder, and
 Schizophrenia
- It provides no content or links related to COVID-19, despite its app store claim.



Exaggerated Claims of Effectiveness

 Although 59 apps claimed to be effective at diagnosing a mental health condition or improving symptoms, only one app included a citation to published literature

Coding element	n (%) of apps
3. Positive claims	59 (81%)
3.a. Claims of effectiveness	47 (64%)
3.a.I. Detection or diagnosis	7 (10%)
3.a.ii. Improvement in symptoms or mood	22 (30%)
3.a.iii, Improvement in self-management	26 (36%)
3.b. Claims of acceptability	33 (45%)
4. Supporting statements	47 (64%)
4.a. Scientific language	32 (44%)
4.a.i. Specific technique described	24 (33%)
4.a.ii. Evidence from study using app	2 (2.7%)
4.a.iii, Citation to scientific literature	1 (1.4%)
4.b. Technical expertise	23 (32%)
4.b.i. Certification or accreditation	0
4.b.li. Prizes or awards	2 (2.7%)
4.b.iii. Credible developers	18 (25%)
4.b.iv. Credible endorsements	3 (4.1%)
4.c. Lived experience design	10 (14%)
4.c.i. Lived experience involvement	6 (8.2%)
4.c.il. Lived experience developer	5 (6.8%)
	A A LOOPART
4.d. "Wisdom of the crowd"	14 (19%)

Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, Reda B. Using science to sell apps: evaluation of mental health app store quality claims. NPJ digital medicine. 2019 Mar 22;2(1):1-6. AND /



The Perils of Misinformation



- "Nonexistent or inaccurate suicide crisis helpline phone numbers were provided by mental health apps downloaded more than 2 million times."
- Only 5 of 69 apps offered all 6 evidence-based strategies for suicide prevention.
- Few consequences for releasing health apps containing inaccurate or non-evidenced based information.

Martinengo, L., Van Galen, L., Lum, E., Kowalski, M., Subramaniam, M., & Car, J. (2019). Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. *BMC medicine*, *17*(1), 1-12.





Stars and Download Metrics are Misleading

EXHIBIT 3

Correlations of mHealth app store ratings with apps' clinical utility and usability, 2014-15



SOURCE Authors' analysis of study data. **NOTES** The points were slightly randomly shifted horizontally and vertically to minimize overlap. Store ratings (on a scale of 0–5) are explained in the Notes to Exhibit 1. Clinical utility (on a scale of 0–10) refers to whether clinician reviewers (identified in the text) would recommend the app to a friend or colleague, with a rating of 0 meaning they would never recommend it. Usability (on a scale of 0 to 100, with a rating of 0 meaning the poorest usability) refers to nonclinician reviewers' (identified in the text) evaluations of how easy it would be for a consumer to use the app according to the System Usability Scale (see Note 20 in text).

K Singh et al. Many Mobile Health Apps Target High-Need, High-Cost Populations, But Gaps Remain. Health Affairs. 2016



Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. Journal of medical Internet research. 2019;21(9):e14567.

DIVISION

DIGITAL

SYCHIATRY

BIDMC





App Store Rank is Not Enough



Of the 27 apps that appear in a search on the app store for "schizophrenia"...

- 3 have been updated in the last 180 days.
- The second result on the list describes itself as a "game that will make you lose your mind"
- 26 have fewer than 100 ratings, and 24 have 10 or fewer.



App Store Ratings are Not Enough



- •"Break the silence on your hopelessness and depression by using a background that reflects your inner misery and despair"
- •Users can download phone backgrounds like:
 - I just don't care anymore
 - Crying without knowing why
 - You can't expect me to be fine
 - I disappoint myself
 - I don't like what I'm becoming



Numerous App Evaluation Schemes

NHS	Mind∧pps	Author, doze	Intended audience	Type of mobile application	Source affiliation	Scoring system in framework	
	MINGADDS	Albrecht, Van Jun & Pramann (20), 2013	Patients	MMA	Institute	No	Norway
	Apps til psykiske vanskeligheder	Anciety and Depression Association of America (ADAA) (38), 2016	Quality assurance for user protection	MMA	Private organization	Yes	USA
	Apps til psykiske valiskelighedel	Amhold, Quade & Kinch (57), 2014	Quality assurance for research setting	MMA	University	Yes	Germany
Diaital		Aungst et al. (32), 2014	Health professionals	MMA	University	No	USA
DIGITAL		Basilico et al. (41), 2016	Quality assurance for research setting	MMM	University	Yes	Holy
Digital		Beatty, Fukuoka & Whoaley (15), 2013	Patients, health professionals	MMA	University	No	USA
		BinDhim et al. (58), 2015	Regulators	mHealth	University	Yes	Australia
		Brooks et al. (39), 2015	Potients	MMA	University	Na	USA
		Chan et al. (18), 2015	Patients, health professionals	MMA	University	No	USA
		Chomutare et al. (59), 2011	Quality assurance for research setting	MMA	University	No	Norway
		Demidowich et al. (60), 2012	Quality assurance for research setting	AMMA	Medical school	Yes	USA
		Drincic et al. (16), 2016	Potients	MMA	University	No	USA
1773		Foitburn & Rothwell (33), 2015	Quality assurance for research setting	MMA	University	No	UK
Matical Antical Frederic		Ferrero-Alvarez-Remontenia (35), 2013	All stakeholders in mHealth	mHealth	Governmental	No	Spain
					organization		4-
		Gautham, Iyengar, & Johnson, C. W. (30), 2015	Health professionals	MMA	University	No	UK
		Gibbs et pl. (25), 2016	Quality assurance for research setting	MMA	University	No	UK
Behaviour change	in the second	Grundy et al. (26), 2016	Patients, health professionals, and app	mHealth	University	Yes	Australia
Benaviour change		and a stady stor	developers	the second	entities)	1.00	· · · · · · · · · · · · · · · · · · ·
Implementing NICE guidance		Hacking Medicine Institute (HMi) (34), 2016	Quality assurance for user protection	AMMA	Institute	Yes	USA
	ashp	Hoppe, Code & Carter, (61) (2016)	Patients	MMA	University	Yes	UK
		Hudkvale et al. (27), 2015	Quality assurance for research setting	MMA	University	Yes	UK
		UNS Institute for Healthcare Informatics (62), 2013	Quality assurance for user protection	mHealth	Institute	Yes	USA
October 2007	eReports	Jin & Kim (63), 2015	Health professionals	mHealth	University	Yes	Republic of
October 2007					and a second		Korea
		Kernishner at al (21) 2015	Busility arrunnes for recovery catting	mblash.	Iniuneity	No	IW

And many more here and more since 2018

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):464-75.



Table 1. Description of Frameworks





Potential for Harm with Lists and Static Ratings





DIVISION

DIGITAL

Lack of Concordance Among Existing Frameworks







Mindtools.io	Psyberguide	ORCHA
14/25 (56)	19/25 (76)	22/25 (88)
22,780	68,041	40,527
714	598	109



DIVISION

DIGITAL

604 Unique Questions from 45 Frameworks



Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. The Lancet Digital Health. 2019 Jun 1;1(2):e52-4.



Beth Israel Deaconess Medical Center HARVARD MEDICAL SCHOOL

From Ethics and 604 Questions to One Framework



Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. The Lancet Digital Health. 2019 Jun 1;1(2):e52-4.



Beth Israel Deaconess Medical Center

HARVARD MEDICAL SCHOOL

A Framework To Be Customized and Adapted

"By functioning as a filtering framework for evaluating apps, the APA App Evaluation Model is a useful tool for yielding a collection of apps for use in primary care."

"[Results]...suggest that frameworks designed to facilitate shared app decision-making, such as the American Psychiatric Association's (APA) app evaluation model and others, may be valuable tools for consumers, clinicians, and patients without providing ongoing expert reviews.

"App assessment tools, such as the newly developed APA framework, place the onus of assessing app quality and efficacy on the app users or their healthcare providers...these are important steps toward improved app quality..." Magee JC, Adut S, Brazill K, Warnick S. Mobile app tools for identifying and managing mental health disorders in primary care. Current treatment options in psychiatry. **2018** Sep 1;5(3):345-62.

Carlo AD, Ghomi RH, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ digital medicine. **2019** Jun 17;2(1):1-8.

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** Dec;17(1):1-2.



APA Framework in Action

Assessment Criteria

We use a set of digital assessment questions to make sure evidence-based, safe, and secure apps are published on the NYC Well app library. These questions emerge from the frameworks presented by AMA and APA to ensure industry best practice.

Base eligibility: The product must be publicly available, offer user support, and offer some portion of its services at no-cost to users. All crisis and app support must be U.S.-based.

Assessment

- · Security: data encryption, data storage practices, security measures
- Privacy: HIPAA compliance, privacy policy, de-identification of data, data access, data ownership effectiveness: claimed vs. actual benefits, published evidence, user feedback, value of content, content maintenance and accuracy
- Effectiveness: relevant outcomes research, appropriate measurement tracking, resource access, evidence-backed design and content, content maintenance
- Usability: ease of use, sustainability, customization, Internet requirements, platform requirements, accessibility, cultural sensitivity
- Data sharing: ability to share data with approved external parties (e.g. family, providers) on users' demand

If our internal panel assesses the product to have suitable security and privacy protocols, in addition to acceptable effectiveness data, a user testing group will examine the app for usability.





phipping

tan.

ciety

DIVISION AT BIDMC

Beth Israel Deaconess Medical Center

HARVARD MEDICAL SCHOOL TEACHING HOSPITAL

From Framework to Database





HARVARD MEDICAL SCHOOL

Respond to People – Not Artificial Scores





Actionable and Objective App Database

- Individual preferences and value guide app choice, making subjective metrics less standardized.
- What is a set of **objective** and **replicable** questions about an app?
- <u>What does "Ease of Use" even mean?</u> Whose value judgements are being imposed. What cultural assumptions are being made?



Rodriguez-Villa E, Torous J. Regulating digital health technologies with transparency: the case for dynamic and multi-stakeholder evaluation. BMC medicine. 2019



Beth Isra MIATRY DMC Medical

leaconess | 🐯 🛱

HARVARD MEDICAL SCHOOL TEACHING HOSPITAL 100+ Objective Questions

The questions are aligned with the levels of the APA pyramid but are designed for you to pick which matter when and where





User Pick What Matters to Them Today





YCHIATRY

BIDMO



HARVARD MEDICAL SCHOOL TEACHING HOSPITAL

App Origin & Functionality

Question
App Origin
Does it come from the government?
Does it come from a for-profit company or developer?
Does it come from a non-profit company?
Does it come from a trusted healthcare compnay?
Does it come from an academic institution?



https://docs.google.com/spreadsheets/d/1FO0Xv3W9jR0_unnfUzoL3QIzyX3NuNgRmC bAbw5zbMM/edit#gid=1245265579





App Origin & Functionality

Does the app work offline?

Does it have at least one accessibility feature (like adjust text size, text to voice, or colorblind color scheme adjuster)?

Does it work with Spanish?

Does it work with a language other than English or Spanish?

Is the app totally free?

What is the cost up front?

Are there in-app purchases?

Is it a subscription (recurrent/monthly/annual)?







App Inputs & Outputs

Input: surveys?

Input: Diary?

Input: Geolocation?

Input: contact list?

Input: Camera?

Input: Microphone?

Input: step count?

Input: external devices (e.g. a wearable sending direct data)?

Input: social network?



Output: notifications?

Output: psychoeducational references/information?

Output: social network?

Output: reminders?

Output: graphs of data?

Output: summary of data (in text or numbers)?

Output: link to formal care/coaching?





Privacy & Security



Is there a privacy policy? Does the app declare data use and purpose? Does the app report security measures in place? Is PHI shared? Is de-identified data shared? Is anonymized/aggregate data shared? Can you opt out of data collection? Can you delete your data? Is the user data stored only on the device? Is the user data stored on a server? Does the app have a crisis management feature? Does the app claim it meets HIPAA? Reading level of the privacy policy (what grade reading level)? Does the app use 3rd party vendors (i.e. google analytics, etc)?



Clinical Foundation

Evidence & Clinical Foundation

Does the app appear to do what it claims to do?

Is the app patient facing?

How many feasibility/usability studies?

What is the highest feasibility impact factor?

How many evidence/efficacy studies?

What is the highest efficacy impact factor?

Can the app cause harm?

bAbw5zbMM/edit#gid=1245265579

Does the app provide any warning for use?







App Features

Features: mood tracking?	
Features: medication tracking?	
Features: sleep tracking?	
Featuers: physical exercise tracking?	
Features: psychoeducation?	
Features: journaling?	
Features: picture gallery/hope board?	
Features: mindfulness?	
Features: deep breathing?	
Features: iCBT or sleep therapy?	
Features: CBT?	
Features: ACT?	
Features: DBT?	
Features: peer support?	
Features: connection to coach/therapist?	
Features: biodata?	
Features: goal setting/habits?	
Features: physical health exercises?	
Features: Bbot interaction (like with virtual character)?	
Features: Biofeedback with sense data (eeg, HRV, skin conductance, etc)?	
	-







Engagement Style



Engagement style: chat/message based?

Engagement style: is it a screener/assessment?

Engagement style: real time response?

Engagement style: Asynchronous response?

Engagement style: gamification (points, badges)?

Engagement style: videos?

Engagement style: audio/music/scripts?

Engagement style: AI support?

Engagement style: peer support?

Engagement style: network support?

Engagement style: Collaborative with provider/other?





Interoperability & Data Sharing

App Use

Is it a self-help/self-management tool?

Is it a reference app?

Is it intended for hybrid use with a clinician in conjunction with treatment plan?

Interoperability & Data Sharing

Do you own your data?

Can you email or export your data?

Can you send your data to a medical record?





Database Homepage

	APPS	RATE NEW APP
This data	base is based on the APA	Select filters and click search:
App E	valuation framework.	- Platforms -
Langle .	Shared Decision	Cost
Land & Sugar	Shared Declaron Making	- Developer Types
		- Féatures
	anna (stan) inning ann sinteann stan	RESET - SHOW ADVANCED FILTERS - SEARCH
individuals w digital health t	Evaluation Model is designed to equip ith the information they need to assess ools on their own and in the context of a	
include (1) acce	relationship. The five levels of analysis essibility; (2) privacy & security; (3) clinical engagement style; (5) therapeutic goal.	
For further	eading about how the framework was please refer to the following papers:	

On the homepage of the database, the user can filter mental health apps according to what characteristics matter to them: Cost? Privacy settings? Available features? YOU decide.

Division of Digital Psychiatry

©2020 Beth Israel Deaconess Medical Center



DIGITAL



A New Way of QUICKLY Finding a Clinically Relevant App

The database then returns the list of apps that meet the user's criteria.





Crowd-Sourced, Up-to-Date App Ratings



By answering the 100 questions, a trained rater can submit app metrics, allowing the app's entry in the database to be regularly updated and responsive to changes in the app overtime.

(

DIGITAL



Database in Action: Lack of Frequent Updates

- Among the top 104 returned apps on an iOS apple store search for "bipolar," only 41 (39%) had been updated in the last 180 days.
- Among the 24 apps that were returned on an app store search for schizophrenia, only 4 (16.6%) had been updated in the last 180 days.



Database in Action: Privacy Concerns in Schizophrenia Apps

- 42% of the top-returned apps is a search for "schizophrenia" lacked a privacy policy altogether.
- 25% of the apps specified in their privacy policy that personal information is shared with third parties.





Database in Action: Privacy Concerns in Bipolar Apps

- 31% of the top-returned apps is a search for "bipolar" lacked a privacy policy altogether.
- 33% of the apps specified in their privacy policy that personal information is shared with third parties.



