Journal of Medical Systems (2022) 46: 34
https://doi.org/10.1007/s10916-022-01821-3

EDUCATION & TRAINING

A Comparative Utility Score for Digital Health Tools

Joshua C. Burton[1,2] · Samantha Regala[2] · Deonte Williams[1,2] · Aditi Desai[2] · Han He[2] · Oliver Aalami[3,4] · Edward R. Mariano[5,6] · Randall S. Stafford[7] · Seshadri C. Mudumbai[5,6]

Received: 13 January 2022 / Accepted: 19 April 2022 / Published online: 5 May 2022

Abstract

Digital health tools (DHT) are increasingly poised to change healthcare delivery given the Coronavirus Disease 2019 (COVID-19) pandemic and the drive to telehealth. Establishing the potential utility of a given DHT could aid in identifying how it could be best used and further opportunities for healthcare improvement. We propose a metric, a Utility Factor Score, which quantifies the benefits of a DHT by explicitly defining adherence and linking it directly to satisfaction and health goals met. To provide data for how the comparative utility score can or should work, we illustrate in detail the application of our metrics across four DHTs with two simulated users. The Utility Factor Score can potentially facilitate integration of DHTs into various healthcare settings and should be evaluated within a clinical study.

Keywords Digital health · Telehealth · Technology · Acceptance · Utility

Abbreviations

DHT	Digital health tools
COVID-19	Coronavirus Disease 2019
TAM	Technology Acceptance Model
TPB	Theory of Planned Behavior

This article is part of the Topical Collection on *Education & Training*

✉ Seshadri C. Mudumbai
mudumbai@stanford.edu

Joshua C. Burton
Joshua.burton@va.gov

Samantha Regala
sregala@stanford.edu

Deonte Williams
dwilliams@paloaltou.edu

Aditi Desai
aditihd@gmail.com

Han He
han.he@va.gov

Oliver Aalami
aalami@stanford.edu

Edward R. Mariano
emariano@stanford.edu

Randall S. Stafford
rstaff@stanford.edu

Introduction

The ongoing Coronavirus Disease 2019 (COVID-19) pandemic has focused intense attention on the role of telehealth in medicine and healthcare in general [1, 2]. Both primary care as well as specialist physicians and other healthcare professionals have had to adopt virtual visits, remote

1 Palo Alto University, 1791 Arastradero Road, Palo Alto, CA 94304, USA

2 Center for Innovation to Implementation, Veterans Affairs Palo Alto Health Care System, 795 Willow Road (152-MPD), Menlo Park, CA 94025, USA

3 Vascular Surgery Service, Veterans Affairs Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94402, USA

4 Department of Vascular & Endovascular Surgery, Stanford University School of Medicine, 780 Welch Rd, CJ350, Palo Alto, CA 94304, USA

5 Anesthesiology and Perioperative Care Service, Veterans Affairs Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94402, USA

6 Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, 291 Campus Drive, Stanford, CA 94305, USA

7 Department of Medicine, Stanford University School of Medicine, 291 Campus Drive, Stanford, CA 94305, USA

monitoring, and mobile technology to provide care while promoting social and physical distancing. Within telehealth, the term "digital health tool" (DHT) is broad in scope, referring to wearable accelerometers and activity trackers as well as mobile applications [3]. These DHTs have also been categorized within electronic health/*e*health or mobile health/mhealth [4, 5].

The role of DHTs extends to physical and mental health issues ranging from diabetes to depression, fall risk to mindfulness [4, 6, 7]. They can also be classified as active where the user inputs data, passive where the DHT collects data, or hybrid which is a mixture of passive and active [2]. The range of options within current DHTs provides opportunity to custom fit user needs and comfort levels. The presence of DHTs in daily life and the funding involved in their development therefore suggest that these tools will only continue to grow in number, specificity, and scope. The following categories may be useful for initial classification [8]:

- **Diagnosis**
- **Treatment**
- **Prevention and Wellness**
- **Prognosis**
- **Rehabilitation**
- **Behavioral Health**
- **Disease Management**
- **Public Health**

However, because DHTs vary considerably in their purpose, comparative evaluations may be necessarily difficult.

Current approaches-predicting use and adherence

Attempts to answer the questions surrounding which DHTs work and why one should choose one over the other have been scant. Current approaches draw from theoretical models of technology and/or psychology and have primarily focused on predicting adoption or use. The Technology Acceptance Model (TAM) assumes that the decision to use a specific object is based on its perceived ease of use as well as the perceived usefulness of the object [9–11]. Since its inception, TAM has undergone several iterations, most recently the Unified Theory of Acceptance and Use of Technology which improved predictive ability by including social variables [11]. However, several recent studies modified the original TAM to explore acceptance of DHTs and related this to intention to use rather than actual use [12, 13]. Another defined use minimally (at least once) although the focus of this study was not solely on use [11]. Critiques of TAM include the interchanging of terms like acceptance, adoption, and use, and direct use (particularly consistent use) is not typically measured [12]. Moreover, because TAM was developed for measurement of acceptance in environments where technology use is mandatory, it is unclear how this may impact this model's utility.

Another predictive model of use is the Theory of Planned Behavior (TPB) [13]. TPB includes attitude toward a behavior (e.g., using a DHT), the subjective norm of a behavior, and one's perceived ability to attain the desired outcomes. For example, in the context of utilizing a mindfulness application, one study was able to successfully predict use through TPB [12, 14]. Furthermore, an application developed to manage non-specific backpain based on TPB saw greater reductions in back pain than control groups [15]. Another study showed that application usage possessed strong correlations with attitudes and perceived behavioral control, though not as much influence through social means [16]. TPB can add predictive value regarding use, but it is currently unclear for how this approach can be practically deployed.

While iterations of TAM and TPB are informative, the most glaring gap in our understanding of the effectiveness of DHTs is the lack of a metric that combines multiple, key aspects of the DHT. For example, use of a DHT possesses an inextricable link to the attainment of a health goal. Some reports have suggested wearable technology use declines or ceases after six months, with similar estimates for smartphone apps [17, 18]. The current literature evaluating wearable devices indicates little benefit of the devices on chronic disease health outcomes [19]. Any attempt to quantify the utility of a DHT must consider how much and what type of use is necessary to achieve a goal. Another gap is related to satisfaction. Most phones come preprogrammed with health-oriented applications (e.g., Apple Health, Samsung health). If someone is satisfied with a product, they may be more likely to recommend it to others which may indicate a higher likelihood of becoming socially acceptable (a component of both TPB and TAM).

Recommendations

We propose a formula that includes objective and subjective data to inform the effectiveness of a DHT (Fig. 1). Combining objective (adherence) and subjective (satisfaction) data could provide the foundation of a metric to appraise consumers and researchers on the relative usefulness of a particular tool. Healthcare practitioners and consumers could then access these scores in selecting the best DHT to meet their unique health needs.

We define this metric as a DHT's Utility Factor score which can then be applied to multiple DHTs. In general,

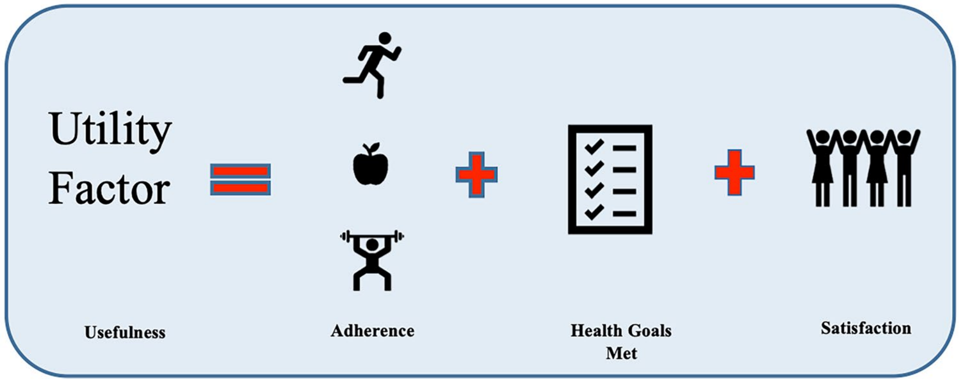

Fig. 1 Utility Factor Score. Adherence is defined as the amount of use per digital health tool (DHT). Health goals defined as type of health goal met. Satisfaction defined as whether a user would recommend a specific DHT

the utility score should be a function of the goal or outcomes achieved, the user's satisfaction with the DHT, and the adherence to the DHT. We suggest the following equation as a candidate approach to calculate this score:

$$U = [(x|G|)(z(S+1))(\frac{1}{yA})] * 100$$

Here, $\boldsymbol{U}$ represents the utility factor score between 0 and 100 of a particular DHT or its ability to provide a benefit to a patient; specifically, the goal in mind (e.g., weight loss). $\boldsymbol{U}$ is a function of the absolute value of a health goal that is met (amount of weight lost), $\boldsymbol{G}$; satisfaction with the DHT is indicated by recommending the product to another person, $\boldsymbol{S}$**;** and adherence, $\boldsymbol{A}$. Here, $\boldsymbol{A}$ can be viewed as a measure that incorporates important aspects of how individuals *use* a particular DHT. Specifically, adherence is defined as use density $\boldsymbol{U_d}$ (number of uses in a day which we define, $\mathbf{D_i}$**,** and longest number of consecutive days used, $\mathbf{D_c}$) **plus** use duration, $\boldsymbol{U_m}$, or the number of months where the device was used at least once a week. $\boldsymbol{A}$ is represented below:

$$\boldsymbol{A} = \boldsymbol{U}_\mathrm{d} + \boldsymbol{U}_\mathrm{m}$$

with $\mathrm{U}_d = \mathrm{D_i} + \mathrm{D_c}$,

Finally, we classify $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$, as variables representing the relative weight each factor contributes to the overall utility factor score. While the weights can be calibrated to different situations and to healthcare practitioner or consumer preferences, we suggest the following general approach to weighting the weighting of inputs of goals, adherence, and satisfaction:

$$\boldsymbol{x} > \boldsymbol{y} > \boldsymbol{z}$$

The attainment of health goals should be weighted heaviest. If a health goal is not met, then its utility factor score (i.e., the benefit of a DHT) should be zero because it does not help. Adherence should be weighted greater than satisfaction because use puts a burden on the consumer and often requires input, a major issue given that adherence rates can be low. Here, it should be noted that what maintains adherence is beyond the scope of this paper. We are more focused on identifying a metric to inform consumers of the potential for a DHT to provide some benefit, i.e., its utility. Finally, satisfaction is given the distinction of $\boldsymbol{z(S+1)}$ because if using a DHT allows the user to meet a health goal, dissatisfaction should not completely eliminate its utility.

To help compare similar DHTs which might occur when accessing a DHT in an App Store for example, we suggest the following approach:

1. First, normalize U scores to 0–100 using min–max normalization:

$$\textbf{\textit{New value}} = (\textbf{\textit{value}} - \textbf{\textit{min}})/(\textbf{\textit{max}} - \textbf{\textit{min}}) * 100$$

After U values are collected, both a minimum and maximum are identified. Each value based on the above formula would then normalize to a range from 0 to 100.

2. Averages for a given category could then be provided by the App Store.

Overall, these metrics can help healthcare practitioners and consumers select a DHT to achieve individual goals across a spectrum of specific tools and identify which specific individuals will define the Goals met and Adherence metrics. We also suggest that the goal or "outcome" is what primarily should drive the weighting approach and secondarily healthcare practitioner or consumer preferences. For example, a goal of a DHT (like Headspace) to help reduce anxiety may have more frequent adherence needs than another DHT to help reduce weight that might be daily. This would therefore require a change in the adherence weighting. Because we consider the weight of each variable's contribution, different scores could be considered a "good" score for DHTs themselves and the specific goals attained. For example, one score could be considered "good" for a phone application aimed at reducing anxiety whereas a "good" score for a wearable designed to aid in weight loss might be different. These scores could be obtained from users during regular visits to their healthcare practitioners and then averaged across different tools and goals. The flexibility of these metrics allows for clarity over the large number of DHTs available for consumers.

Application of metrics to simulated users and potential limitations

To provide data for how the comparative utility score can or should work, we illustrate in detail the application of our metrics across four DHTs with two simulated users. Our rationale in selecting these DHTs was to identify a group that were commonly available; widely used; and potentially initiated in either consumer or clinical settings for a variety of purposes:

- Fitbit, for weight management and exercise. [20]
- Headspace, for assisting with better sleep hygiene and stress management [21].
- mySugr, for aiding diabetes and blood sugar management [22].
- Mindshift, for help with anxiety management using cognitive behavioral therapy [23].

For the purpose of demonstration and based on the clinical experience of the authors, our research.

(A) developed two simulated users and their histories. To assist illustration and comparison, the goals of both persons were set up to be identical. The goals were to lose weight, sleep better, lower blood sugar, and anxiety management.

- User 1: 68 year-old white male. Presenting concerns: Type II diabetes, high blood pressure, sleep issues.
- User 2: 38-year-old Hispanic female. Presenting Concerns: Type I diabetes, anxiety, low physical activity.

(B) and identified a range of inputs for goals (G), satisfaction (S), and adherence (A; use density U_d (number of uses in a day D_i, and longest number of consecutive days used, D_c; plus use duration, U_m, or the number of months where the device was used at least once a week). Goals were set up to be quantitative (e.g., weight loss in pounds) and identified initially, followed by inputs for satisfaction, and adherence. To assist in interpretation of our examples, weights (e.g. x,y, and z) for goals, satisfaction, and adherence were set to equal 1.

(C) entered the inputs into an EXCEL (EXCEL 16.5, Microsoft, Redmond, WA) spreadsheet for scenario calculation.

The data inputs and calculations for the two simulated users are provided in Tables 1 and 2; utility scores between the two users for each of the DHTs are graphed in Fig. 2. Across all the DHTs, the relative role of the various factors and in particular the impact of goals in changing the overall utility score is evident. For example, though User 2 had a markedly lower utility score (0.5) for the mySgr DHT vsUser 11 (23.6). User 2 experienced a goal of 1 h on average per day in the number of hours with blood sugar in the normal range (70 to 99 mg/dL) vs. User 1's 10 h. Similar comparisons may be made for Headspace and the role of the goal of numbers of hours slept on average per day (User 2's utility score of 48.6 vs User 1's utility score of 16.7). In establishing these scores a patient/user should have sufficient experience with a given DHT and suggest that 2–4 weeks of routine would allow the time necessary to become comfortable with the device and troubleshooting a routine. We also suggest several approaches to store user metrics for a particular DHT including 1) an application itself where one would peruse a list of DHTs which are identified by the particular goal in mind; 2) the UF as a score that is advertised with an application (i.e., Apple App Store or Google Playstore); and 3) as a tool associated with facility or a health care system or (i.e. Veterans Healthcare Administration). De-identified user profile data or population statistics should also be available to the end-users as they interpret UF scores (such as in an App Store); for example, of the patients who provided feedback about Headspace, how many of them self-identified as having anxiety.

A potential limitation of our simulated datasets is the proposed equation for calculating utility scores. Alternative formulas can also be proposed but we nevertheless suggest that goals, satisfaction, and adherence be included. Another potential limitation is in how we defined goals (e.g., weight loss in pounds). Other definitions could be equally valuable, but our aim was to highlight the importance of defining goals quantitatively. Given potential concerns for honesty in how a user may enter data, one method to address this is to include language asking users at the start of the feedback process to commit to providing complete and accurate information [24].

Table 1 Utility Factor Scores for User 1 with Sample Inputs for Goals, Satisfaction, and Adherence

Digital Health Tool	U (Utility factor score)	G (Goals met)	S (Satisfaction)	A (Adherence)	Number of uses in a day (Di)	Longest number of consecutive days used (Dc)	Use Density (Ud)	Use duration (months) (Um)
Fitbit (weight management/ exercise)	18.3	4 (The numbers of pounds lost in 6 months)	4	109	1	100	101	8
Headspace (sleep/ stress management)	16.7	2 (The numbers of hours slept on average per day)	0	12	1	10	11	1
mySugr (diabetes and blood sugar management)	23.6	10 (the average number of hours with blood sugar in the normal range (70 to 99 mg/dL) per day	6	296	10	280	290	6
Mindshift (anxiety management)	13.3	2 (The number of days per week with mild anxiety (GAD-7 score < 5)	1	30	12	13	25	5

Utility Factor score defined as $U = [(x|G|)(z(S+1))(\frac{1}{yA})] * 100$. Weights (x,y,z) were each set to equal 1. Adherence (A) = Use density (Ud) + Use duration (Um); with Use density (Ud) = Number of uses in a day (Di) + Longest number of consecutive days used(Dc)

Table 2 Utility Factor Scores for User 2 with Sample Inputs for Goals, Satisfaction, and Adherence

Digital Health Tool	U (Utility factor score)	G (Goals met)	S (Satisfaction)	A (Adherence)	Number of uses in a day (Di)	Longest number of consecutive days used (Dc)	Use Density (Ud)	Use duration (months) (Um)
Fitbit (weight management/ exercise)	21.3	10 (The numbers of pounds lost in 6 months)	3	188	2	175	177	11
Headspace (sleep/ stress management)	48.6	6 (The numbers of hours slept on average per day)	2	37	2	30	32	5
mySugr (diabetes and blood sugar management)	0.5	1 (the average number of hours with blood sugar in the normal range (70 to 99 mg/dL) per day	0	216	6	200	206	10
Mindshift (anxiety management)	22.9	4 (The number of days per week with mild anxiety (GAD-7 score < 5)	1	35	5	26	31	4

Utility Factor score defined as $U = [(x|G|)(z(S+1))(\frac{1}{yA})] * 100$. Weights (x,y,z) were each set to equal 1. Adherence (A) = Use density (Ud) + Use duration (Um); with Use density (Ud) = Number of uses in a day (Di) + Longest number of consecutive days used(Dc)

GAD General Anxiety Disorder 7 Scale

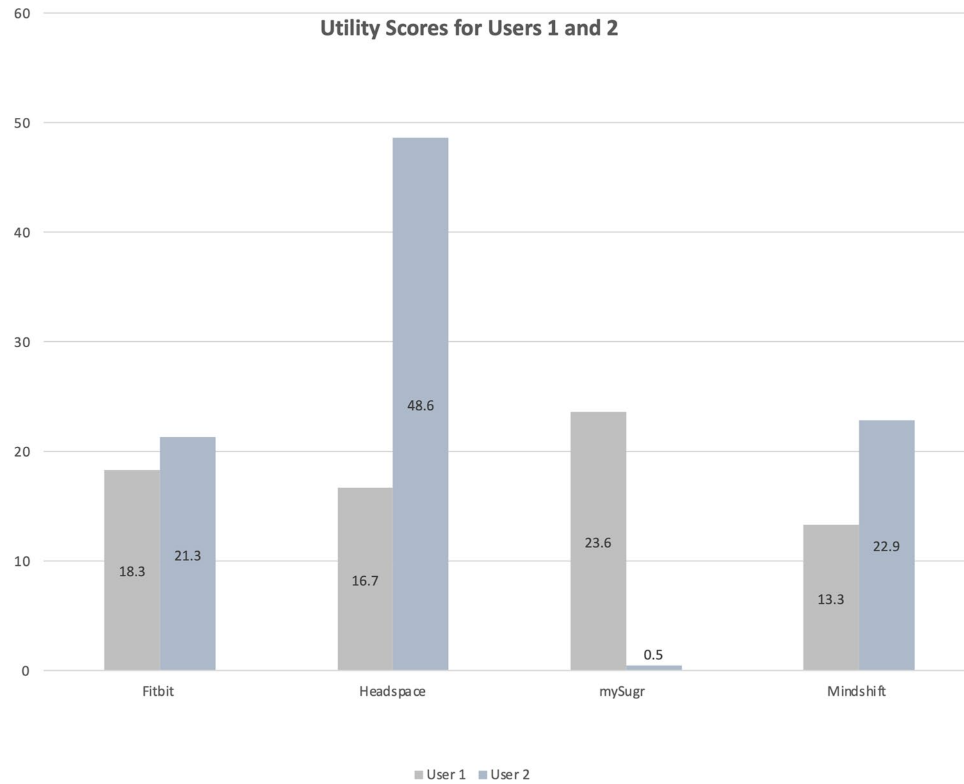

Fig. 2 Utility Factor Scores for Simulated Users 1 and 2. Utility factor scores are on the y-axis and digital health tool (DHT) type is on x-axis

Conclusions

DHT use is on the rise, but research is still in the nascent stages. Few studies distinguish between those who use DHTs once or twice a month from those who use daily, but this difference may be critically important.[21] Moreover, there are no studies that compare the types of health goals (e.g., reducing anxiety symptoms or H_bA1c levels) attained when using a DHT. And, despite the popularity of DHTs, few studies assess how satisfaction relates to health goals. We propose defining use, and linking it directly the types of health goals met and to adherence and satisfaction elements through the creation of a novel Utility Factor score that should quantify a DHT's ability to provide benefit to and inform consumers. Our simulations suggest that sustained attention should be made to the goals or outcomes achieved rather than simply satisfaction and adherence when evaluating DHTs. One advantage of our approach is its applicability to different types of DHTs with different health goals. The generalizability of defining utility as a function of health goals met, satisfaction, and adherence allows consumers and healthcare practitioners to select and deploy from a wide array of DHTs to address their needs. In summary, the Utility Factor Score can potentially facilitate integration of DHTs into various healthcare settings and should be evaluated within a clinical study. A clinical study would help elucidate many of the "nuts and bolts" details associated with use of the scoring algorithm, namely how to define G, how heavily to weight x,y,z, and who keeps the results data.

Authors contribution All authors have made substantial contributions to the conception and design of the work; drafted the work or substantively revised it; approved the submitted version; and agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Declarations

Conflicts of interest None of the authors have any conflicts of interest to declare. Funding for this study was derived from institutional funding.

References

1. Torous J, Jan Myrick K, Rauseo-Ricupero N, Firth J (2020) Digital mental health and covid-19: Using technology today to accelerate the curve on access and quality tomorrow. J Med Internet Res Ment Health 7(3):e18848. https://doi.org/10.2196/18848
2. Fisk M, Livingstone A, Pit SW (2020) Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 22(6):e19264. https://doi.org/10.2196/19264

3. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al (2016) Evaluating digital health interventions: Key questions and approaches. Am J Preventive Med 51(5):843-851. https://doi.org/10.1016/j.amepre.2016.06.008
4. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL (2019) The history and future of digital health in the field of behavioral medicine. J Behavioral Med 42:67-83. https://doi.org/10.1007/s10865-018-9966-z
5. Safavi K, Mathews SC, Bates DW, Dorsey ER, Cohen AB (2019) Top-funded digital health companies and their impact on high-burden, high-cost conditions. Health Aff (Millwood) 38(1):115-23. https://doi.org/10.1377/hlthaff.2018.05081
6. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al (2016) Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 35(12):2310-2318. https://doi.org/10.1377/hlthaff.2016.0578
7. Richardson JE, Reid MC (2013) The promises and pitfalls of leveraging mobile health technology for pain care. Pain Med 14(11):1621-1626. https://doi.org/10.1111/pme.12206
8. Ronquillo Y, Meyers A, Korvek SJ (2022) Digital Health. StatPearls, Florida
9. Schnall R, Bakken S (2011) Testing the Technology Acceptance Model: HIV case managers' intention to use a continuity of care record with context-specific links. Inform Health Soc Care 36(3):161-172. https://doi.org/10.3109/17538157.2011.584998
10. Calvin KL, Severtson DJ, Karsh BT, Brennan PF, Casper GR, Sebern M, et al (2006) Development of an instrument to measure technology acceptance among homecare patients with heart disease In AMIA Annu Symp Proc 1053
11. Holden RJ, Karsh BT (2010) The technology acceptance model: Its past and its future in health care. J Biomedical Informatics 43:159-172. https://doi.org/10.1016/j.jbi.2009.07.002
12. Ajzen I (2011) The theory of Planned Behaviour: Reactions and reflections. Psychol Health 26(9):1113-1127. https://doi.org/10.1080/08870446.2011.613995
13. Crandall A, Cheung A, Young A, Hooper AP (2019) Theory-based predictors of mindfulness meditation mobile app usage: A survey and cohort study. J Med Internet Res mHealth uHealth 7(3):e10794. https://doi.org/10.2196/10794
14. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen T, Morgan R, et al (2015) Mobile-web app to self-manage low back pain: Randomized controlled trial. J Med Internet Res 17(1):e1. https://doi.org/10.2196/jmir.3130
15. Herrmann LK, Kim J (2017) The fitness of apps: a theory-based examination of mobile fitness app usage over 5 months. Mhealth 3(2). https://doi.org/10.21037/mhealth.2017.01.03
16. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M (2017) Determinants for sustained use of an activity tracker: Observational study. J Med Internet Res mHealth uHealth 5(10):e7311. https://doi.org/10.2196/mhealth.7311
17. Wang Q, Egelandsdal B, Amdam GV, Almli VL, Oostindjer M (2016) Diet and physical activity apps: Perceived effectiveness by app users. J Med Internet Res mHealth uHealth 4(2): e5114. https://doi.org/10.2196/mhealth.5114
18. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H (2018) Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements. Evid Based Ment Health 21(3):116-119. https://doi.org/10.1136/eb-2018-102891.
19. Jo A, Coronel BD, Coakes CE, Mainous AG, 3rd (2019 Is there a benefit to patients using wearable devices such as Fitbit or health apps on mobiles? A systematic review. Am J Med 132(12):1394–400. https://doi.org/10.1016/j.amjmed.2019.06.018
20. https://www.fitbit.com/global/us/home. Accessed 8 Sept 2021
21. https://www.headspace.com/. Accessed 10 Sept 2021
22. https://www.mysugr.com/en-us/. Accessed 12 Sep 2021
23. https://www.anxietycanada.com/resources/mindshift-cbt/. Accessed 10 Sept 2021
24. Vésteinsdóttir V, Joinson A, Reips UD, Danielsdottir HB, Thorarinsdottir EA, Thorsdottir F (2019) Questions on honest responding. Behav Res Methods 51(2):811-25. https://doi.org/10.3758/s13428-018-1121-9

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