Why reviewing apps is not enough: Transparency for Trust (T4T): The Principles of Responsible Mental Health App Marketplaces

Increased scrutiny of health apps is necessary to protect consumers and promote effective products

Til Wykes, Stephen M Schueller

King's College London, De Crespigny Park, London SE5 8AF, Til Wykes Professor and Vice Dean Psychology and Systems Science, Center for Behavioral Intervention Technologies, Northwestern University, 750 N. Lake Shore Drive, 10th Floor, Chicago, IL 60611 United States of America, Stephen M Schueller Assistant Professor of Preventive Medicine

Correspondence to: T Wykes til.wykes@kcl.ac.uk

Digital therapeutics are being touted as having the potential to transform healthcare by improving people's experience, increasing effectiveness, and reducing costs. A few digital aids are recommended and integral to health services, but much of the e-health field depends on overselling[1]. This business plan seems to be working as the digital health field was estimated to be worth $25b globally in 2017[2] with one US survey finding that 58% of smartphone users have downloaded at least one health app[3]. The overselling of health apps, which may provide little benefit and some potential harm, needs the health community's immediate attention. With little formal regulation, a light touch approach to consumer protection is necessary to give customers a modicum of information as a basis for choosing from the vast array of “so-called” health apps.

We believe that simple, but informative, evidence should be available at the point of downloading and propose only four aspects that represent the critical information required for responsible health app marketplaces. We refer to these four principles – privacy and data security, development practices, feasibility, and health benefits – as the Transparency for Trust (T4T) principles. These T4T principles draw from several sources including patient and regulatory perspectives, recent systematic reviews and experimental studies (e.g.[3-6]). We focus on a fast growing segment of digital therapeutics – mental health apps. Stigma and discrimination experienced by these consumers reduces their help seeking and even when they do seek help the worldwide scarcity of specialty mental health services is likely to reduce their contacts. So these consumers are left to navigate the app stores alone. However, our proposed principles are applicable to the whole panoply of health apps.

Why do we need some simple principles?

**Exponential Growth and Poor Regulation:** The recent WPA-Lancet Psychiatry Commission[7] concluded that although digital psychiatry offers enormous potential to rethink how services are provided, there are large roadblocks in its way. A further Lancet Psychiatry Commission suggested that digital therapeutics could provide benefit now to complement current mental health treatments and aid self-management[8]. But they also may delay effective treatment potentially resulting in increased health care costs and lost productivity. There is also evidence that some apps are not only ineffective and unsafe but do not meet users privacy and security expectations[9-11].

Last year 78,000 new health apps were added bringing the numbers of health apps available in these unregulated market place to 325,000 [4]. Formal regulation is remarkably light and restricted to a narrow definition of a health app[5, 12, 13]. Even when an app is regulated, we cannot be sure that it will work. For example, the United States Food and Drug Administration (FDA) recently approved the first behavioral health app, reSET, for the treatment of substance use by using
evidence from a clinical trial of a web-based version of the treatment not the app itself[14]. In the UK, the CQC issued guidance last year for digital healthcare providers, but these concentrate on safety[15]. The Medicines and Healthcare Products Regulatory Agency (MHRA) provides CE marking for medical devices and Certificates of Free Sale but leaves review to the National Information Board. As it stands many mental health apps are marketed with few checks and even regulatory approval appears to offer little confidence on whether that specific product was ever directly evaluated.

More complex regulation has been proposed. The National Institute of Health and Clinical Excellence (NICE) is curating an NHS app store. This is a burdensome process and so few apps will be assessed in any year. Currently the library has 78 apps (with only 18 for mental health), which is a minute subset of the 325,000 available[16]. In the US, the FDA launched a pre-certification pilot program involving nine companies to speed the approval process but this will evaluate the developer and their practices rather than focusing on the product. Apple Inc has introduced additional requirements for medical apps for developers but these focus mainly on measurement accuracy. There is a middle way to fill this important, and now yawning gap, in consumer information. Health app marketplaces could take a lead by providing relatively simple guidance.

**What is wrong with current systems for reviewing Health Apps?**

Most proposed evaluations (e.g. Mobile App Rating Scale (MARS)[17], Enlight assessment tools[18]) assess usability, aesthetics, content, user engagement, and available research evidence and others have been adding to this list[19]. These systems are useful because they facilitate multi-faceted and thorough evaluations of apps but they fall short of allowing clear recommendations. In fact more recent evidence from Canada involving service users demonstrated that a high MARS rating would not on its own provide enough information to allow service users to form a decision on whether to download an app[6]. Advisory bodies such as NICE in the UK make a determination of what is likely to be cost-effective (effective, cost relative to benefit and other comparison treatments) before they recommend its use in the UK NHS. Consumers, however, want to make choices based on simpler information. One for-profit company, ORCHA, provides reviews based on “current standards, regulation and good practice” but their overall score does not allow a consumer to decide on which components are important to them. PsyberGuide, a nonprofit organisation also provides reviews which include a service user focus but does not receive data directly from app developers. But apart from their lack of fulfilling all users’ expectations, no method provides clear information at the point of sale and a potential consumer would have to search in two places for the information they need to make a choice. We propose only four aspects of apps that represent the critical information required for responsible health app marketplaces. These four principles, deemed the Transparency for Trust (T4T) principles, are privacy and data security, development characteristics, feasibility data, and benefits.

**Transparency for Trust (T4T) principles**

*Privacy and Data Security* are a primary concern for patients[20, 21] and its importance has only become more salient with recent events such as the Facebook and Cambridge Analytica scandal. Although the European Union General Data Protection Regulation is strong, weaker regulations elsewhere results in varying protections internationally. One review prompted the closure of the NHS app store when it was discovered that accredited apps were not encrypting data adequately and did not explicitly describe the personal data leaving the app[20]. Happtique, an early app certification company met a similar fate when several of its certified apps were hacked demonstrating the inadequacy of its processes to evaluate privacy and data security. Many apps rely
on selling the data they collect for their business plan which jeopardizes personal privacy[20]. There is evidence of poor practice resulting in fines for selling sensitive information to Lottery companies and fraudsters[21]. Privacy concerns change with the evolving technology even though device operating systems are moving towards encryption on the device by default. Nevertheless, users’ need information about data leaving the app to make informed decisions about their willingness to provide sensitive health information[22, 23].

Although full formal audits are needed to ensure apps follow their stated procedures[24], even requiring developers to list their privacy and data security procedures would be a significant step forward on raising standards[25]. We propose three questions: 1) what data leaves the device? 2) how is that data stored? (e.g., de-identified, encrypted), 3) who will have access to that data? It should be clear what, if any, data is being sold, to whom, and what steps are taken to ensure that users cannot be identified by that data.

**Development Characteristics** describe how the app was developed and our recommendations conceptually overlap with those of the FDA’s precertification pilot and the MHRA in the UK. Good developmental practices would involve patients from the beginning and at all stages of development and testing. These processes should also include people from the intended target audience. This may seem obvious, but unfortunately development practices often rely on convenience samples, for example soliciting feedback from stressed college students rather than individuals with depression[26]. Depression apps should include people with depression which may seem obvious but independent usability evaluations have demonstrated that many popular commercial apps are frustrating and challenging for members of the intended audience raising questions about their prior involvement and the potential for the app to benefit this community[27].

Developers should outline their design and development process and clearly describe how patients were involved. Our three questions are: 1) how were target users involved in the initial design, 2) how were target users involved in usability evaluations, 3) has usability been independently evaluated?

**Feasibility** evaluations should address how people use the app (usability and user experience), how long they use it (engagement), and whether any serious adverse concerns are discovered (safety). These aspects provide information on how people use the app, including expectations on frequency and length of use. This information is also vital to assess benefits. It would not be possible to run a drug trial or market a drug without some concept of the dosing frequency and expected therapeutic dose and the same should be true with health apps.

Again we have three questions: 1) what proportion of users continue to use the app after two weeks? 2) what adverse events occurred and what was the rate of those events? 3) has feasibility been independently evaluated? We propose a two week test not because it represents a likely therapeutic dose, but because very few users persist in using a health app after the first week[28]. A standard metric, such as two weeks, could promote cross app comparisons in engagement. Similar to usability testing, independent evaluation of apps is the key to promote transparency and confidence in findings. Independent evaluations could be carried out by service user groups which could further strengthen service user involvement in the process of development and evaluation. Transparency could be further facilitated by making these datasets available to the research community.

**Mental Health Benefits** are apparent from rigorous evaluations using standardized and accepted outcomes for the target condition. Although many researchers have noted the mismatch between
the development cycle for mobile apps and traditional randomized controlled trials[29, 30], it is still the case that health apps presented as digital therapeutics require rigorous evaluation to back up their claims. The speed of development should not preclude such evaluations as suggested by some academics and designers[31]. We should be presented with direct evidence on an app’s safety and effectiveness because they are not merely mobile versions of websites even if they have similar content. People use apps differently including more frequently and in shorter bursts[32] and these differences could affect their impact. We mention this because recently the FDA approved the first behavioral health app, reSET, for the treatment of substance use by using evidence from a clinical trial of a web-based version of the treatment. Although triangulation of different sorts of data has been suggested (e.g. MindTech[33]) we believe that Mental Health apps should undergo a trial to determine their superiority to other treatment options, especially as many unsubstantiated claims have been made [34]. Advertising standards require evidence to support any claims made so these data fulfil both commercial and patient needs. Evaluations should also consider opportunity cost as using a health app may delay treatments that could be more beneficial, or a delay could worsen the health condition making it harder to treat. All these benefits and costs need to be weighed in the balance. Our three questions are 1) what is the impact on the health condition, 2) what percentage of users received either no benefit or deteriorated, and 3) are there specific benefits that outweigh any costs?

What would this look like for a mental health app?
We have inserted the information from two mental health apps with similar target audiences, one of which was named the app of the year in the iTunes store last year. We have extracted, where possible, the information on each of our principles. The differences are very clear especially in privacy and health benefits. But what these simple principles also provide is the ability for a consumer to trade-off the attributes. Some may want to know that their data is totally secure while others might want to allow some encrypted anonymized data to be transmitted if the effectiveness of the app is proven. Indeed, in a recent survey of participants recruited from a mood and anxiety disorder clinic many respondents were willing to allow an app to collect data directly from one’s phone including GPS, motion sensors, and screen state[35].

Table 1. Evaluating apps with the T4T principles

<table>
<thead>
<tr>
<th>Privacy and Security</th>
<th>Apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) What data leaves the device?</td>
<td>BlueIce*</td>
</tr>
<tr>
<td>2) How is that data stored?</td>
<td>All data are stored on the app and owned by the user.</td>
</tr>
<tr>
<td>3) Who will have access to that data?</td>
<td>Only user of the device where BlueIce is installed has access.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Development Characteristics</th>
<th>Apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) How were target users involved in the initial design of the app?</td>
<td>Co-produced by Oxford Health NHS Foundation Trust and young people with lived experienced.</td>
</tr>
<tr>
<td>2) How were target users</td>
<td>Not provided</td>
</tr>
</tbody>
</table>
involved in usability evaluations?

| 3) Has usability been independently evaluated? | No independent usability evaluations were conducted. | Independently evaluated by PsyberGuide.org |

**Feasibility**

| 1) What proportion of users continue to use the app after two weeks? | 93% of users kept using it. | No information provided |
| 2) What adverse events occurred in the test population and what was the rate of those events? | None found (clinicians did not withdraw user and users did not feel app use increased self-harm). | No information provided |
| 3) Has feasibility been independently evaluated? | No independent evaluations conducted. | No independent evaluations conducted. |

**Benefits**

| 1) What was the impact on clinical outcomes? | Significant reductions in depression and anxiety and 73% reduced self-harm after 12-weeks | No clinical outcomes research reported. |
| 2) What percentage of users received no benefit or deteriorated? | 27% reported no reductions in self-harm | No description provided of non-responders or users who deteriorated. |
| 3) Are the specific benefits worth the cost? | No information provided about the expected ratio of benefits to risks | No information provided about the expected ratio of benefits to risks |

* “BlueIce is a prescribed evidence-based app to help young people manage their emotions and to reduce urges to self-harm” ([https://www.oxfordhealth.nhs.uk/blueice/](https://www.oxfordhealth.nhs.uk/blueice/))

**“Calm is App of the Year on the Apple store in 2017” is for “Meditation and sleep” ([https://www.calm.com/](https://www.calm.com/))

---

**Are these principles different from those suggested by others?**

As we have said T4T were based on those suggested by others in recent years. However, they differ in terms of their form. So we have used information important to regulators, developers and health services but have overlaid the patient viewpoint taken from a number of different studies[35, 36]. Patients are after all the consumer group of interest. Their views do not necessarily coincide with expert groups as shown in the Delphi exercise by Zelmer and colleagues[6]. Privacy and security features in every assessment system and in regulations and is high on the list for patients, especially those with a mental health problem who may be more sensitive about information about them being shared. So it is included here but in the simplest terms and not buried in an incomprehensible privacy statement. Our principle for a “fit-for-purpose” app includes development with patients. This principle is often suggested[7] but rarely incorporated into app assessment. As we know that some commercial apps are complex and hard to use by the patient group they were intended for so we have valued this section highly. Effectiveness is often mentioned in many assessment systems but the promised effects are also dependent on the dose of the app and how intensively it is used. Patients need to consider this time constraint when deciding to make a purchase. Patients also want to know not how effective it is, but whether anyone doesn’t receive a benefit. This is also important to clinicians as patients who receive no benefit may view themselves as hopeless cases and not, as
in the BlueIce exemplar, just part of the quarter of patients who report no advantage from following the app.

Our approach has therefore been to provide information to patients at the point of the download that allows them to make an informed decision and which they can refer to later as part of their self-management plan.

**Responsibility in the Health App Marketplaces**

If these simple T4T principles are followed, then we will have gone some way towards protecting patients. Our view is that formal regulation is not needed unless the app is to be recommended to mental health systems and services. We just need information that will allow patients (and patient groups) to make informed choices. So whose job is it to monitor the T4T principles? Information that is not true can be picked up by Advertising standards authorities. Recent examples of this process are the US Federal Trade Commission fining of Lumosity for deceiving consumers with unfounded claims about cognitive benefits [37], and Carrot Neurotechnology for claiming that their app, Ultiemeeyes, can improve users vision [38]. Health apps are not a passing fad and the low barrier of entry into current app marketplaces has resulted in an environment that at best confuses and at worst delays effective treatment. The problems have been highlighted but rarely have clear solutions like ours been proposed. Developers may be encouraged to produce these answers by commercial advantages, as apps with T4T principles might increase consumer comfort and produce unique revenue streams through increased adoption, not only from direct to patient markets but also from health systems. They will also enjoy increased legitimacy among patient groups.

Confidence in the efficacy and safety of these health apps is the least that patients should expect in making a choice to buy or use them. It is now time that existing commercial app stores, specifically the Google Play and Apple iTunes stores, step back from their libertarian ideology and adopt some rules for health app marketing. They should tighten up the definition of health apps and adopt a system, ours hopefully, to allow patients to understand what to expect from a health app. Although some might believe that this proposal is “other worldly,” starting somewhere is important. Health app marketplaces have a duty, and health app developers a commercial advantage, from following our suggestions – we should not need to wait for another scandal or disaster before the Google Play or Apple iTunes stores step up to the plate and help prevent worthless products being pressed on those with mental health needs.

Contributors and Sources: SMS is the Executive Director of a mental health app review project of which TW serves on the Scientific Advisory Board. TW presented the outline and discussed the concept with SMS who wrote the initial draft. Both authors contributed to the revisions. TW is the guarantor.

Acknowledgements: TW acknowledges support from the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre (BRC) and the NIHR Senior Investigator Award. SMS acknowledges the support of One Mind as well as the Implementation Research Institute (IRI), at the George Washington University in St. Louis; through an award from the National Institute of Mental Health (5R25MH08091607) and the Department of Veterans Affairs, Health Services Research & Development Service, Quality Enhancement Research Initiative (QUERI). The views expressed are authors’ and not necessarily those of the NHS, the NIHR, the Department of Health or NIMH.

Conflicts of Interest: We have read BMJ policy on competing interests and declare the following: TW had developed a novel software intervention (CIRCuiTS) and sits on the PsyberGuide Scientific
Advisory Board. She has not received any funding from companies involved in this field. SMS has received funding from One Mind and serves as the Executive Director of PsyberGuide, a non-profit funded by One Mind. SMS has received funding payment for consulting to Annum Health and ORCAS, Inc. which develop mental health apps and has an interest in Optilife Technologies, Inc. to serve as a scientific advisor for Potentia Labs, Inc.
References


