Comparing applets and oranges: barriers to evidence-based practice for app-based psychological interventions

Simon Leigh[1,2]

[1]Lifecode Solutions, Liverpool, UK; [2]Nexus Clinical Analytics, Euxton, UK
Correspondence to Simon Leigh, Lifecode Solutions, 54 St. James Street, Liverpool L1 0AB, UK; simon@lifecode.co

ABSTRACT

Poor-quality pharmaceuticals and medical devices rarely make it to market; however, the same cannot be said for app-based interventions. With a high availability but low evidence base for mHealth, apps are an increasingly uncertain prospect to users and healthcare professionals alike. Although in a first-best situation, the burden of proof concerning app safety, clinical and cost-effectiveness 'should' ultimately lie with app developers; a number of barriers to evidence generation, including the fact that 'acceptable evidence' itself is largely open to interpretation, mean that it may be folly to expect this paucity of real-world effectiveness research to improve. While the health technology assessment of established therapeutic modalities including pharmaceuticals and talking therapies benefits from the existence of approved evaluative guidelines, unfortunately the same cannot be said for app-based interventions, specifically with regard to outcomes measurement. As such, it would seem that in order to prevent the comparative assessment of apps simply becoming an exercise comparing apples and oranges, there is a clear need for consensus and guidance for app developers, as to which patient-reported outcome measures, among the hundreds available, are of clinical use to those making decisions, and should therefore be used when developing app-based interventions. By negating the fear that any evidence collected may be of poor quality, we can reincentivise developers to engage in evidence generation, and in doing so, maximise the likelihood of evidence-based decision-making taking a firm hold. However, only by dispelling the ambiguity around what acceptable evidence can and should look like, can we begin to do so.

BACKGROUND

Untreated mental health disorders are now the single largest cause of disability in the UK,[1] affecting one in four people and costing the English economy ∼£105 billion per year.[2] While waiting lists, demand[3] and financial pressures[4] for National Health Service (NHS) psychological interventions are on the rise, so is the use of apps and mHealth.[5] It is estimated that 71% of Britons own a smartphone,[6] 75% use smartphones or tablets to search for health information online[7] and 90% would use online services to contact healthcare professionals, were these services available.[6] When combined with the fact that the UK is the least expensive place in the world to engage with online solutions for digital health,[8] the potential patient and health service benefits that could be achieved through the wider use of high-quality evaluated apps could be considerable.

However, despite the potential for apps to play a valuable role within NHS-led mental healthcare, not all apps available to consumers are likely to be clinically effective, and of those that are, only a small number can demonstrate a clear picture of real-world effectiveness through the use of patient-reported outcome data.[9] Even with respect to NHS-accredited app-based psychological interventions, historically, as few as 15% have been backed by data to corroborate claims of effectiveness.[10] However, this paucity of high-quality effectiveness data is not a new phenomenon concerning electronic medical technologies, with the medical device industry historically suffering a similar shortage of evidence.[11] This is because unlike pharmaceuticals, which are required to undergo years of rigorous and controlled assessment concerning safety, dosing and effectiveness, regulators are often evaluating medical devices at a very early stage of their market life cycle.[11] Subsequently, the extent of product exposure, data collection and research is typically very sparse and particularly so if considering any longer term outcomes and the sustainability of treatment effects.

BARRIERS TO EVIDENCE-BASED PRACTICE WITH APPS

Although the majority of health apps are not currently classed as medical devices, this shortage of outcomes research is also observed within the market for app-based psychological interventions. Despite the apps industry quickly gathering momentum, with ∼165 000 health apps available online as of 2015,[12] an estimated 50% of such apps will receive fewer than 500 downloads across their entire product life cycle.[13] The result is that, if left to market forces, the rate of app uptake is likely to be prohibitively slow, thereby limiting the potential for app developers to gather sufficient data in order to power and detect meaningful treatment effects at conventional levels of statistical significance. This is likely to be particularly problematic if aiming to evaluate and publish data from apps within a time frame which is proportionate to the speed of app development, leaving a question regarding the value to app developers, of attempting to formally collect and analyse evidence of user outcomes at all.

This seemingly uncertain value of data collection in order to support any claims of effectiveness is likely compounded by a current absence of published guidelines for prospective app developers, as to how the merits of app-based interventions should be assessed. The result is that, unlike the structured and coordinated health technology assessment (HTA) of traditional health-generating technologies, including pharmaceuticals, talking therapies and medical devices, which benefit from the existence of approved guidelines[14] and a much clearer path from development to reimbursement, it is currently largely unclear what constitutes a minimum acceptable standard of evidence for app-based interventions. When combined with the ambiguity as to the form any evidence should take, whether prospective or retrospective, the preferred methodologies to be applied, including randomisation and blinding, and the follow-up, comparators and time horizon that should be considered, the ability of developers to provide meaningful data to inform the debate regarding the merits of app-based interventions seems a long way from realisation.

But perhaps most importantly, and regardless of the methodology applied, in order to prevent the evaluation and comparison of app-based psychological interventions simply becoming an exercise comparing apples and oranges, there is a clear need for consensus as to which patient-reported outcome measures (PROMs), among the hundreds that could potentially be deployed by prospective app developers,[15] should be incorporated when developing app-based interventions.

Some, including a recent perspective published in this journal,[16] have noted that the use of traditional quality indicators may be unrealistic in the context of apps, naturally leading to a discussion around a range of potential alternative indicators which may be more conducive to gauging app quality. Such indicators however, which include accessibility, user experience and technical quality, while useful from a general

assessment standpoint, have uncertain links to effectiveness and cost-effectiveness. While each of these measures will intrinsically impact on the overall clinical efficacy of an app, in the absence of clinical PROMs, their individual powers as a gauge of efficacy and value are limited, as it is largely unclear how much the NHS would be willing to pay for an X percentage point improvement in usability. On the a priori that the primary purpose of app-based psychological interventions is to alleviate psychological symptoms and actively manage mental health concerns, it seems vital that the elicitation of clinical efficacy, obtained through the use of PROMs, is given much greater consideration.

A consensus must be reached as to which PROMs actually provide utility to those making real-world treatment decisions, whether in line with existing minimum clinically important differences as used by the Improving Access to Psychological Therapies (IAPT) programme,[17] or the consistent application of alternative metrics which may be more conducive to use within apps. In the context of anxiety disorders, extensive questionnaires, including the 20-item and 18-item Beck Hopelessness Scale and Health Anxiety Inventory (HAI), designed to comprehensively assess mental well-being in routine clinical practice may be unsuitable for inclusion within app-based interventions. However, the less administratively burdensome and time-consuming 7-item Generalised Anxiety Disorder-7 (GAD-7) or short Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) may be better suited, especially when considering that meaningful data collection necessitates that such questionnaires be completed at baseline, post app use and ideally after a suitable period of follow-up.

In the absence of guidance to app developers as to which PROMs should be incorporated when building apps for the purpose of comparative assessment, the reality is that developers will simply continue to employ the metrics most likely to demonstrate the greatest efficacy for their product. Consequently, from the perspective of the clinician looking to provide high-quality support to patients, or the healthcare commissioner who may be considering the deployment of apps to supplement existing care pathways, applying a balanced, consistent and objective approach to the comparison of the costs and benefits, of the many app-based interventions currently available to consumers, both against one another and against existing NHS services, will be a significant challenge. Without the presence of a common denominator, it becomes almost impossible to compare the clinical and economic return on investment of a 10% improvement in self-belief from one app, a five-point reduction in the Penn State Worry from another and a three-point reduction in the Beck Depression Inventory from another, leaving a question as to which app is likely to deliver the greatest benefit to prospective users, and which, if any, should be recommended or funded in practice.

THE APP DEVELOPER'S PERSPECTIVE

While it is clear that in a first-best situation, the burden of proof concerning app safety, clinical and cost-effectiveness 'should' ultimately lie with app developers, the barriers to effective and meaningful evidence generation that currently exist, including the fact that 'acceptable evidence' itself is largely open to interpretation, mean that it may be folly to expect the potential value of app-based interventions to be unlocked any time soon. Much like the NHS, app developers are faced with trade-offs and decisions regarding how best to allocate their limited resources; yet, unlike pharmaceutical and established medical device manufacturers, the majority of app developers are likely to be small and lacking adequate research and development funding and analytical expertise. The highly competitive nature of the market for app-based psychological interventions means that potentially expensive and time-consuming data collection and analytics will inevitably incur opportunity costs, that is, 'what benefit could have been achieved with these funds if used alternatively?' As such, app developers are currently likely to have little incentive to engage with existing regulatory frameworks,

which rely on time-consuming and often expensive randomised controlled trials (RCTs), with perceived returns on investment for competing business development activities, including advertising and app updates, likely to be far in excess of those associated with evidence generation. This is likely to be particularly true if developers fear that in the absence of guidance regarding what standard of evidence is acceptable, any evidence provided may be of poor quality, thereby negatively impacting sales or in some cases, even their reputation.[11]

REALISING THE POTENTIAL OF APP-BASED PSYCHOLOGICAL INTERVENTIONS

The high degree of competition and fast pace of development within the apps market, coupled with the minimal barriers to patients accessing apps, present a considerable opportunity for healthcare systems to benefit from the development of systems to improve the overall quality of app-based interventions. While poor-quality pharmaceuticals and medical devices rarely make it to market, the same cannot be said for app-based interventions, and it would seem that setting a high standard from the outset is vital to achieving long-term benefit for patients and the NHS. In certain therapeutic indications, apps could be deployed as a means of improving care quality and promoting efficiency, providing a temporarily sufficient 'bridging' treatment for those presenting with mild symptoms, and thereby allowing healthcare professionals to divert a greater amount of time to more challenging cases. Apps could be used as a relatively low-cost means of providing patient support and a continuity of care, and doing something when otherwise seen to be doing nothing, including providing coping strategies for those on waiting lists for talking therapies. Some app-based interventions may even turn out to be less effective and cost-effective than existing mental health services, and in some cases may even exacerbate mental health disorders or potentially widen existing health inequalities. Yet, before we can begin to address the many unknowns regarding the potential role and value of app-based psychological interventions within a 21st century NHS, and begin to maximise the potential of this infant therapeutic medium, we must first and foremost dispel the ambiguity around what 'acceptable evidence' to inform such decisions can look like.

Through acknowledging the current barriers to meaningful evidence generation that characterise the apps market, and adapting our approach to evidence generation accordingly, the NHS can begin to take full advantage of the current apps revolution, much the same way as the aviation, telecommunications and even taxi industries have done previously. A switch in emphasis, away from the traditional RCT and towards more pragmatic, less expensive and more widely available observational data, as suggested within this journal,[16] is likely to present a significant step towards circumventing a number of the current barriers to mHealth evidence generation.

However, not all evidence is equal, and prior to committing, en masse, to new alternative methodologies, it is essential that we first and foremost lay the groundwork as to what we are trying to answer with studies in mHealth. Only through clarifying what 'acceptable' evidence can and should look like, including guidance as to what additional observational data are necessary in order to negate the possibility of confounding and pooling bias, and providing sufficient support for the funding, collection and analysis of user data, can we expect the potential benefits of this therapeutic medium to be realised.

Through raising the perceived importance and informative value of evidence generation, we can maximise the likelihood of evidence-based decision-making taking a firm hold, and as a result, benefit from timely and rigorous assessment of app-based interventions, rather than the current reality of a trade-off between the two. In doing so, we can begin to generate meaningful clinical and economic insights that can help shape and improve the standard of care with respect to mental health services, and highlight which of the thousands of app-based interventions currently available to consumers are likely to result in

measurable clinical benefit and at a reasonable price (this applies to the NHS in the UK, but can be extended to other countries as well). However, only by providing sufficient incentives for app developers to collect patient-reported outcomes, providing a clear means of navigating the currently complex and uncertain regulatory landscape, and making it clear exactly what form of evidence is required, can we begin to do so.

REFERENCES

1. **The Kings Fund.** Long-term conditions and mental health: the cost of co-morbidities. 2012. http://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/long-term-conditions-mental-health-cost-comorbidities-naylor-feb12.pdf (accessed 28 Mar 2016).
2. **The health foundation.** Is mental health care improving? http://www.health.org.uk/sites/default/files/IsMentalHealthCareImproving.pdf (accessed 26 Mar 2016).
3. **The Kings Fund.** Mental health under pressure. 2015. http://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/mental-health-under-pressure-nov15_0.pdf (accessed 28 Mar 2016).
4. **The Telegraph.** Mental health patients 'at risk' due to budget cuts, leading think tank warns. 2015. http://www.telegraph.co.uk/news/health/11990289/Mental-health-patients-at-risk-due-to-budget-cuts-leading-think-tank-warns.html (accessed 23 Mar 2016).
5. **Boulos MN**, Brewer AC, Karimkhani C, et al. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online J Public Health Inform* 2014;**5**:229.
6. **Nuffield trust.** Delivering the benefits of digital health care. 2016. http://www.nuffieldtrust.org.uk/sites/files/nuffield/nutj4099_healthtechreport_16.2.16_web.pdf (accessed 21 Mar 2016).
7. **Department of Health and UK Trade and Investment.** The UK: your partner for digital health solutions. 2015. https://www.gov.uk/government/publications/theuk-your-partner-for-digital-health-solutions (accessed 5 Apr 2016).
8. **NIHR MindTech Healthcare Technology Co-operative**. Technologies for remote therapy and management—multiple criteria, multiple stakeholders. 2014. http://www.rcpsych.ac.uk/pdf/Craven%20Michael%20P%20CAP14.pdf (accessed 18 Mar 2016).
9. **Martínez-Pérez B**, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. *J Med Internet Res* 2013;**15**:e120.
10. **Leigh S**, Flatt S. App-based psychological interventions: friend or foe? *Evid Based Ment Health* 2015;**18**:97–9.
11. **Campbell B.** NICE medical technologies guidance: aims for clinical practice. *Perioper Med (Lond)* 2013;**2**:15.
12. **IMS institute for health informatics.** Medicines Use and Spending in the US—A Review of 2015 and Outlook to 2020. 2015. http://www.imshealth.com/en/thought-leadership/ims-institute (accessed 2 Apr 2016).
13. **IMS Institute for Healthcare Informatics.** Patient Apps for Improved Healthcare from Novelty to Mainstream. 2013. http://obroncology.com/imshealth/content/IIHI%20Apps%20report%20231013F_interactive.pdf (accessed 3 Apr 2016).
14. **National Institute for Health and Care Excellence (NICE)**. Guide to the methods of technology appraisal. 2013. https://www.nice.org.uk/process/pmg9 (accessed 8 Apr 2016).
15. **National Institute for Mental Health in England.** Outcomes compendium. http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_093677.pdf (accessed 5 Apr 2016).
16. **Nicholas J**, Boydell K, Christensen H. mHealth in psychiatry: time for methodological change. *Evid Based Ment Health* 2016;**19**:33–4.
17. **NHS England.** Improving Access to Psychological Therapies Measuring Improvement and Recovery Adult Services. Version 2. 2014. http://www.iapt.nhs.uk/silo/files/measuring-recovery-2014.pdf (accessed 29 Mar 2016).