Effective? Engaging? Secure? Applying the ORCHA-24 framework to evaluate apps for chronic insomnia disorder

Simon Leigh,[1] Jing Ouyang,[2] Chris Mimnagh[3]

[1]Lifecode Solutions, Liverpool, UK; [2]Innovation Hub, Alder Hey Children's Hospital, Liverpool, UK; [3]Wingate Medical Centre, Liverpool, UK

Correspondence to Simon Leigh, Lifecode Solutions, Liverpool L1 0AB, Merseyside, UK; simon@lifecode.co

ABSTRACT

Background Mobile health offers many opportunities; however, the 'side-effects' of health apps are often unclear. With no guarantee health apps first do no harm, their role as a viable, safe and effective therapeutic option is limited.

Objective To assess the quality of apps for chronic insomnia disorder, available on the Android Google Play Store, and determine whether a novel approach to app assessment could identify high-quality and low-risk health apps in the absence of indicators such as National Health Service (NHS) approval.

Methods The Organisation for the Review of Care and Health Applications- 24 Question Assessment (ORCHA-24), 24 app assessment criteria concerning data privacy, clinical efficacy and user experience, answered on a 'yes' or 'no' and evidence-driven basis, was applied to assess 18 insomnia apps identified via the Android Google Play Store, in addition to the NHS-approved iOS app Sleepio.

Findings 63.2% of apps (12/19) provided a privacy policy, with seven (36.8%) stating no user data would be shared without explicit consent. 10.5% (2/19) stated they had been shown to be of benefit to those with insomnia, with cognitive behavioural therapy apps outperforming hypnosis and meditation apps ($p=0.046$). Both the number of app downloads ($p=0.29$) and user-review scores ($p=0.23$) were unrelated to ORCHA-24 scores. The NHS-approved app Sleepio, consistently outperformed non-accredited apps across all domains of the ORCHA-24.

Conclusions Apps for chronic insomnia disorder exhibit substantial variation in adherence to published data privacy, user experience and clinical efficacy standards, which are not clearly correlated with app downloads or user-review scores.

Clinical implications In absence of formal app accreditation, the ORCHA-24 could feasibly be used to highlight the risk–benefit profiles of health apps prior to downloading.

INTRODUCTION

Mobile health (mHealth) is an emerging and rapidly developing opportunity, with the potential to play a significant role in transforming both the quality and efficiency of healthcare management. With approximately 1.7 billion worldwide users of health apps[1] estimated to account for around 4 million downloads of such apps every day, it is estimated that in 2017 alone, mHealth solutions could save a total of €99 billion in healthcare costs across the European Union.[2] Because this novel therapeutic medium can facilitate the effortless gathering and interpretation of vast medical data, health apps have emerged as an exciting opportunity to promote both pharmacovigilance and evidence-based practice. Furthermore, their flexible nature, which promotes user autonomy, may combat the negative impact that factors such as stigma have on healthcare seeking behaviour[3] while also enabling improved patient access to health information, anywhere and at any time.

Yet despite the significant potential for health apps to enhance the efficient and timely delivery of healthcare, there are currently numerous drawbacks when contemplating the use of these unregulated and largely unvalidated technologies. While it is mandatory for Medicines and Healthcare Products Regulatory Agency and US Food and Drug Administration (FDA)-approved medical devices and pharmaceuticals to provide a list of all known side effects, enabling users to weigh up the risks and benefits of use; unfortunately, the same cannot currently be said for health apps. This leaves considerable uncertainty as to whether health apps do in fact 'first do no harm', raising questions concerning whether health apps can rationally be chosen by users and clinicians alike as a viable, effective and safe therapeutic option.

One recent review of app-based psychological interventions available on the National Health Service (NHS) apps library demonstrated that fewer than 15% could demonstrate any evidence of clinical effectiveness.[4] Similarly, one in three (33%) certified apps on the NHS apps library were found to lack privacy policies, while 9 in 10 (89%) transmitted information over the internet without encryption.[5] While personal data protection is a fundamental right within Europe,[6] the potential 'side-effects' of using health apps can in fact be far greater than those associated with poor data governance, resulting in actual physical harm to users. A systematic assessment of iOS and Android apps calculating insulin dosage for those with type 1 diabetes established that 91% lacked numeric input validation, with 67% carrying a risk of inappropriate dosage recommendations.[7] Likewise, a review of smartphone applications assessing melanoma risk identified that 75% incorrectly classified at least 30% of melanomas as unconcerning, with one app displaying a sensitivity of just 6.8%.[8]

To date, attempts to evaluate the quality of health apps have faced numerous shortcomings and challenges, the first of which being a failure to acknowledge characteristics of apps that are important to clinicians and consumers, such as data governance[9] and user experience and aesthetics.[10] Second, applying any variation of a Likert scale, thereby relying on individual value judgements, leaves the inherent possibility that the experiences or expectations of the reviewer will impact the results of the assessment itself. Finally, app assessments including NHS accreditations or other one-off endorsements, which are typically not conducted at scale, are likely to result in information asymmetry. With much known about the characteristics of approved or accredited apps, but little known about how these characteristics compare with viable alternatives yet to be assessed and possibly currently in use, the subsequent inability to make informed choices about switching between apps is likely to result in opportunity costs.

The foundation of any reliable, consistent and meaningful mHealth assessment framework should be to establish a basis of safety, quality and effectiveness, in a way that assesses all relevant apps on an equal playing field. The aim of this study is to apply a novel framework for evaluating the quality and risks of health apps, based on observable indicators of data security, clinical efficacy and assurance and user experience and engagement. The Organisation for the Review of Care and Health Applications - 24 Question Assessment (ORCHA-24) framework, a total of 24 criteria derived from published mHealth,[11 12] data privacy[13–16] and user experience[17 18] standards and answered on an evidence-based 'yes' or 'no' basis without the requirement for scaling or value judgements, will

be applied to assess apps in the therapeutic area of chronic insomnia disorder presently available to download on the Android Google Play Store. In doing so, we aim to determine whether higher ORCHA-24 assessment scores, and thus improved compliance to published best-practice standards,[11–18] correlate with observable features of apps available at the point of download, including price, the total number of downloads or average user review scores. The ORCHA-24 will also be applied to the NHS-approved insomnia app 'Sleepio'[19] to determine any differences in scores between NHS-approved and non-approved apps. In doing so, we aim to determine the feasibility of using this short, scalable, evidence-driven framework as a proxy for app quality to potential downloaders of health apps, in the absence of time consuming and often highly expensive formal app evaluations.

OBJECTIVES

The aims of this study were to (1) assess the quality of apps for chronic insomnia disorder, available on the Android Google Play Store using the ORCHA-24, a combination of 24 'yes' or 'no' answered app assessment criteria regarding data privacy, clinical efficacy and user experience best practice standards; (2) determine whether ORCHA-24 assessment scores are related to observable factors at the point of download, including the number of app downloads to date, average user review scores or price; and (3) to obtain preliminary insights as to whether the ORCHA-24 could be used to identify high-quality and low-risk health apps at scale, and in the absence of observable quality indicators including NHS, FDA or National Institute for Health and Care Excellence (NICE) Medtech Innovation Briefings accreditations.

MATERIALS AND METHODS

Search strategy and inclusion criteria

In December 2016, we conducted a review of mobile apps available on the Android Google Play Store. To ensure valid comparisons within a single and well-defined therapeutic area, only apps dedicated to chronic insomnia disorder were considered. As a result, a single search term, 'insomnia', was used to identify relevant apps in the app store.

The preliminary screening of the apps identified was based on three factors: the app title, screenshots of the app and the description of the app in the Android Google Play Store. In each instance, apps were excluded if they were found to be games, unrelated to health, if they were for novelty purposes, including getting pets to sleep, or if they were not available in English language. Apps that were not excluded during the preliminary screening phase were subsequently excluded if (1) their primary purpose involved nothing more than providing an mp3 sound list, or (2) they were not intended for use by members of the general public. We also included the NHS-approved iOS app Sleepio, which can be found on the NHS Choices online mental health services website[19] and is available for social prescribing in selected Clinical Commissioning Groups.

All apps that met the inclusion criteria were downloaded and reviewed by two independent reviewers to ensure inter-rater reliability, the results of which are reported in terms of Cohen's kappa. Both reviewers received extensive training in the use of both the ORCHA-24 review and a more in-depth 120 question version prior to the research being conducted. Because the quality assessment of apps embodied a dedicated information governance component, once downloaded and subject to review, any apps that did not collect any user data, whether personally identifiable, sensitive or other, were excluded from the analysis.

App quality criteria and related standards

We assessed and ranked apps for the therapeutic relief of insomnia based on three overarching criteria: user experience and engagement, data privacy and clinical efficacy and assurance. Assessment criteria were chosen following a Delphi panel consisting of an information governance specialist, a health economist, an academic research fellow in mHealth and two clinicians (one primary and one secondary care). To be included within the assessment criteria, questions were required to be answered on a 'yes' or 'no' basis, on their ability to be answered with reference to evidence and without the opportunity for the opinions, expectations or value judgements of reviewers to affect outcomes.

Based on these criteria, and following a review of published best practice criteria, including the PAS-277: 2015,[11] AQuA accessibility testing criteria,[12 17] and legal standards, including the Data Protection Act 1998,[13] the Misrepresentation Act 1967,[20] and the European Data Protection Directive (95/46/EC),[14] the assessment criteria, listed in table 1, were selected.

Once the app review was completed, a report of the findings was emailed to the developer of each app. This allowed app developers the opportunity to scrutinise the assessment of their app and refute or provide evidence to the contrary of any conclusions made.

Analysis and outcomes

We assessed and ranked all apps based on the three subsections of the ORCHA-24 review, data privacy, clinical efficacy and assurance, and user experience and engagement, with all subsection scores combined to provide a total ORCHA-24 score. In the absence of preference elicitation exercises, assessment scores were obtained using a uniform weighting approach, with each question carrying equivalent weight as a constituent of the overall review score. As such, if an app received a score of 12/24, this was considered equivalent to any other app scoring 12/24, regardless of which questions were answered 'yes' or 'no' within each app assessment. All apps reviewed were subgrouped according to the 'method of action' of the application, namely (1) cognitive behavioural therapy (CBT), (2) hypnosis, (3) meditation and (4) multiple methods or other. We additionally collected data concerning the number of downloads of each app, the app price and the average user review score on the Android Google Play Store. These data were collectively used to perform subgroup analyses on the apps included and uncover any relationships between observable factors such as app price, or the number of downloads, and ORCHA-24 assessment score. One-way analysis of variance (ANOVA) was used to analyse variation between groups, with results reported at the standard 5% significance level. Levene's test of homoscedasticity was applied to ensure constant variance across independent samples. In the event of significant variations between groups, post hoc analyses using Tukey's honestly significant difference (HSD) were undertaken to uncover the origins of any statistically significant differences in ORCHA-24 assessment scores. All data analysis was conducted in Microsoft Office Excel 2013 (Microsoft, Redmond, Washington, USA).

FINDINGS

Two hundred and fifty-two apps were initially identified from the Android Google Play Store, with 215 excluded based solely on content provided in descriptions in the app store and within the app itself. Thirty-seven apps were downloaded with 19 apps excluded due to failure to collect any form of user data. This resulted in 18 Android apps plus one NHS-approved iOS app undergoing detailed review. Inter-rater reliability when applying the ORCHA-24 assessment was high, with an agreement rate of 94.5% and a Cohen's unweighted kappa coefficient of 0.88 (95% CI 0.84 to 0.93) (figure 1).

Data security

Thirteen apps (68.4%) collected 'personally identifiable' data with 63.2% of apps (12/19) providing a privacy policy to users, either within the app itself or as a link within the app's website. There was no significant difference in the presence of a privacy policy by app type, with 66.6% of CBT apps, 75% of meditation, 57.1% of hypnosis and 60% of apps using multiple methods providing a privacy policy to users. Each app that

Table 1 ORCHA-24 app assessment criteria and related standards

App assessment criteria	Related standards
Data governance	
(1) Does the app state that no data will be shared with other parties without explicit user consent?	► The Data Protection Act 1998: Principle 1[13] ► Data Protection Directive: Article 19[14]
(2) Does the app outline a process for managing data confidentiality breaches?	► Information Commissioner's Office: Guidance on data security breach management. Version 2.1. The Data Protection Act[15] ► The Charter of Fundamental Rights of the European Union: Article 7[24]
(3) Is there a data privacy policy, either within the app itself or on a website?	► PAS-277: 2015 6.2.3 .f (project documentation)[11] ► GSMA: Mobile and Privacy. Privacy Design Guidelines for Mobile Application Development (2012) (TCC2, TCC3, DRS4)[16]
(4) Does the data privacy policy, or statement, provide detail about what data is collected by the app?	► PAS-277: 2015 6.3.e(2)[11]: ► Data Protection Directive: Article 10[14]
(5) Does the data privacy policy, or statement, provide detail about what that data is used for by the app?	► The Data Protection Act 1998: Principle 2[13] ► Data Protection Directive: Article 10[14]
(6) Does the data privacy policy, or statement, state whether personal data are stored using recognised secure data storage technologies?	► The Data Protection Act 1998: Principles 6 and 7[13] ► The Data Protection Directive: Article 17[14]
(7) Does the data privacy policy, or statement, state that all personally identifiable data will be encrypted in transit between the device and any developer host storage? (eg, using FTP protocol)	► The Data Protection Directive: Article 17.[14] ► GSMA: Mobile and Privacy. Privacy Design Guidelines for Mobile Application Development (2012) (DRS2)[16] ► The Data Protection Act 1998: Principle 7[13]
(8) Does the data privacy policy state that only the minimum data items necessary for the app to function will be collected?	► The Data Protection Act 1998: Principle 3[13] ► The Data Protection Directive: Article 6 (data minimisation)[14]
Clinical efficacy and assurance	
(9) Is there a statement within the app itself, or the app store, about user feedback during design, development or testing?	► PAS-277: 2015 6.2.3 (c), 6.2.4, 6.5[11]
(10) Is there a statement either in the app or store about user involvement in testing?	► PAS-277: 2015 6.7[11]
(11) Is there a statement within the app that it has been tested and shown to be beneficial to someone with the relevant condition?	► Misrepresentation Act 1967[20] ► PAS 277: 2015 6.2.3(c) (Project documentation)[12]
(12) Is there a statement within the app, or app store, about the app having been through a clinical trial, or other form of testing to show real world effectiveness, and has received positive feedback?	► PAS-277: 2015, 6.2.3, 6.5[11]
(13) Is there a statement about how frequently any advice, guidance or content will be reviewed to ensure accuracy and clinical relevance?	► PAS-277: 2015, 6.7[11]
(14) Is there a statement within the app that it has been positively evaluated or validated by a clinical or other relevant expert?	► PAS-277: 2015, 6.2.3(b)[11]
(15) Is there any evidence within the app that the developer has attempted to validate any guidance or recommendations with academic expertise?	► No specific guidance. Result of Delphi panel discussion
(16) Is there a statement within the app identifying a list of review or accrediting bodies or individuals?	► No specific guidance. Result of Delphi panel discussion
User experience and engagement	
(17) Does the app provide support options for users with visual impairment? Including changing font sizes or colour?	► W3C: Accessibility Requirements for People with Low Vision Editor's Draft 6 June 2016, clause 3.3.1.[18] ► App Quality Alliance (AQuA): Accessibility Testing Criteria for Android Applications version 1.2: July 2015, sections 1.4.1 and 1.5.3[17]
(18) Does the app provide support options for users with hearing difficulties?	► AQuA: Accessibility Testing Criteria for Android Applications version 1.2: July 2015, section 3[17]
(19) Does the app contain a '?/HELP/ABOUT' function to aid user understanding?	► AQuA: Best Practice Guidelines for producing high-quality mobile applications version 2.3 – June 2013 page 23.[30]
(20) If clinical or technical terms are used, are they explained clearly to the user? (either within the content of the app or via a glossary)	► No specific guidance. Result of Delphi panel discussion
(21) Is there any statement within the app about how to report issues, bugs or errors to the developers?	► PAS-2772015 - Clause 10 (transparency)[11]
(22) Does the app set goals for users or allow them to set goals for themselves?	► No specific guidance. Result of Delphi panel discussion
(23) Is there a statement within the app about the developer's commitment to addressing problems reported to them? (eg, timescales to respond, commitment to eradicate reported bugs and faults)	► PAS-277: 2015–6.7 Accountability[11]
(24) Are there opportunities to link with other users of the app, including buddying, forums or group education?	► No specific guidance. Result of Delphi panel discussion

FTP, File Transfer Protocol; GSMA, Groupe Speciale Mobile Association; PAS, Publicly Available Standard.

provided a privacy policy also provided a list of the data being collected, in addition to what the data collected was intended to be used for. However, only 7/12 apps with a privacy policy (36.8%) stated that no user data would be shared with other parties without explicit user consent. Additionally, just 25% (3/12) of apps with a privacy policy in place stated using recognised secure storage technologies, with 3/12 also stating that

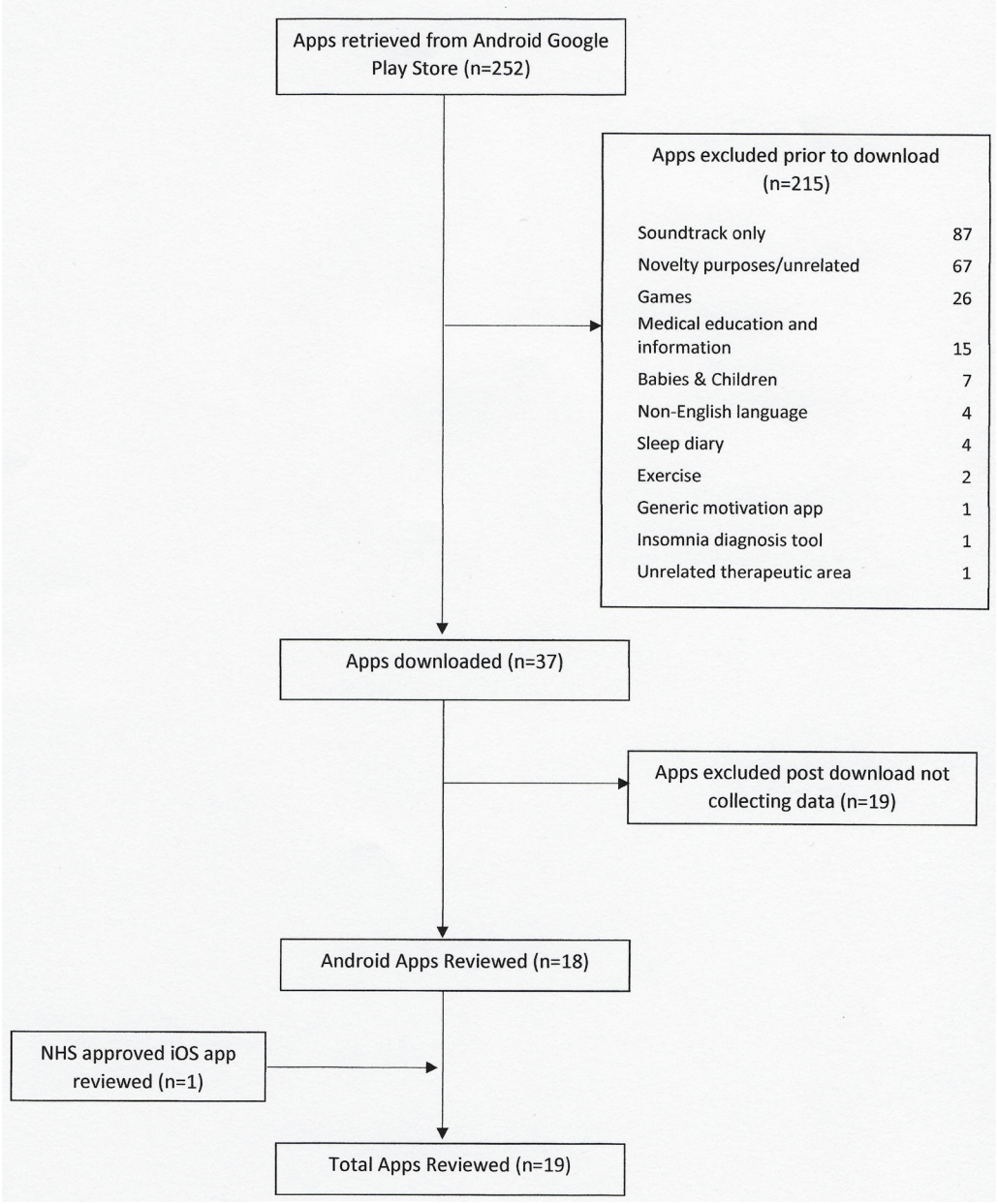

Figure 1 Details of apps retrieved.

all personally identifiable user data were encrypted in transit. Scores for data privacy varied considerably among the apps assessed, ranging from 0/8 to 7/8, with a mean score of 2.85 and a median of 3. While CBT apps (3.3) scored marginally higher than meditation (3.25), hypnosis (2.14) and apps using multiple methods (3.2), one-way ANOVA revealed no significant difference in adherence to data privacy guidelines, and subsequent ORCHA-24 data privacy scores, between app types ($p=0.86$).

Clinical efficacy and assurance

Approximately 68.4% of apps (13/19) contained a statement notifying users that the app had been positively evaluated by a clinical or other relevant expert, with 61.5% of this group (8/13) subsequently providing details of specific accrediting bodies or individuals. Furthermore, just 2/19 apps contained a statement asserting the app had been tested and shown to be of benefit to someone with the relevant condition, while only one app (5.3%) provided information to users regarding how often any advice or recommendations provided to users of the app would be reviewed to ensure clinical validity. ORCHA-24 scores for clinical efficacy and assurance also varied considerably on a per-app basis, ranging from 0/8 to 8/8, with a mean score of 2, and a median of 1. Again, results of ANOVA demonstrated that CBT apps scored significantly higher (4.3) than those applying meditation (1.25) and hypnosis methods (1) ($p=0.046$), while also demonstrating an improvement over those applying multiple methods (2.8) ($p=0.11$).

User experience

Almost all apps reviewed (17/19) provided support options for those with visual impairment, including changing the font size or colour scheme of the app, while 6/19 (31.6%) provided support options for those with hearing difficulties. Additionally, 16/19 apps provided a help/about section to aid understanding and use of the app, with only one app (5.3%) using medical or technical terms without clearly explaining them to the user.

The mean ORCHA-24 assessment score for user experience and engagement ranged from a minimum of 2/8 to a maximum of 6/8, with a mean of 4.4 and a median score of 5. CBT apps again scored higher (5) than those applying meditation (4.25), hypnosis (4.6) and multiple

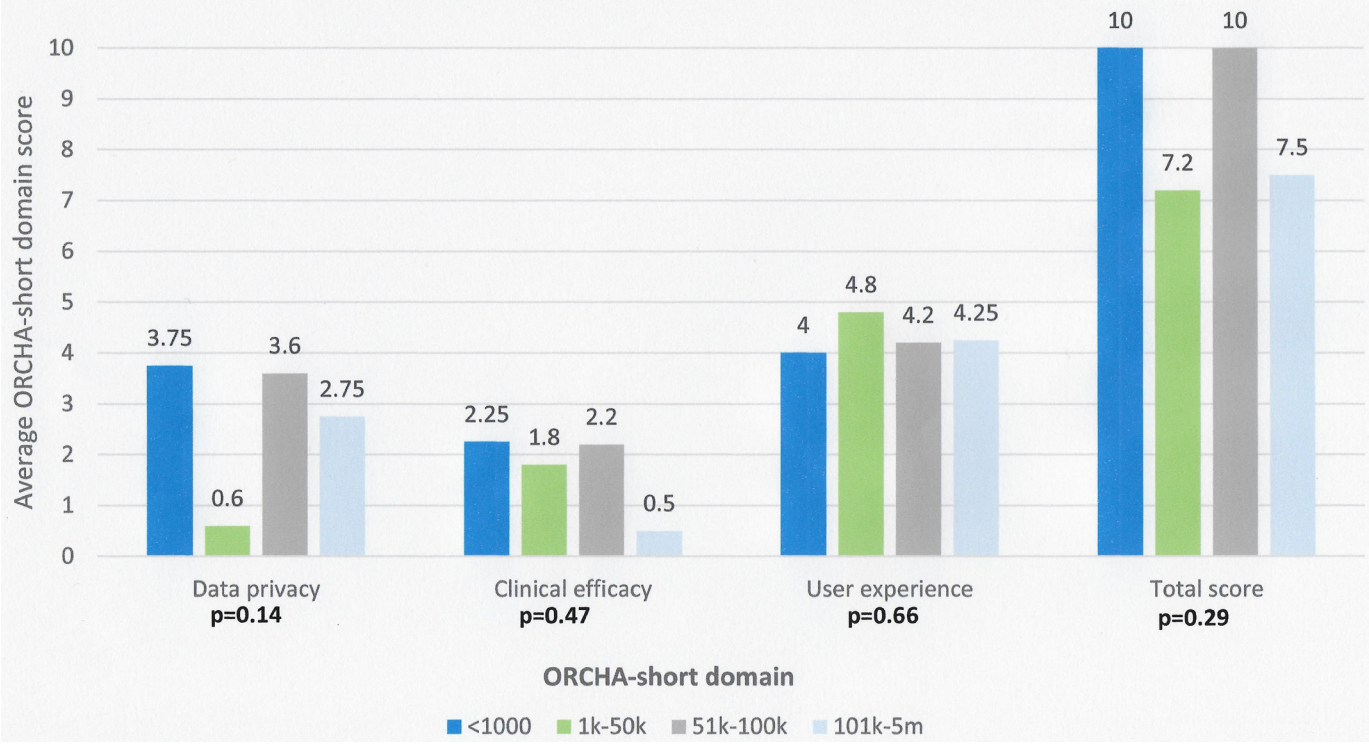

Figure 2 ORCHA-24 domain scores versus app downloads.

methods approaches (3.8); however, this was not statistically significant (p=0.347).

Downloads as an indicator of quality?

Apps with the lowest number of downloads (<1000) demonstrated the greatest ORCHA-24 data privacy score (3.75/8), compared with 0.6 in those with 1000–50 000 downloads, 3.6 in those with 51 000–100 000 downloads and 2.75 in those with >100 000 downloads (p=0.13). Additionally, apps with <1000 downloads also demonstrated the greatest clinical efficacy and assurance score (2.25) p=0.47, in addition to the joint highest overall ORCHA-24 score (10), p=0.29; however, this was not a statistically significant difference, as demonstrated in figure 2.

User review score as an indicator of quality?

Apps with the lowest user review score on the Android Google Play Store achieved the highest ORCHA-24 assessment scores with respect to data privacy (4.2) (p=0.06) and clinical efficacy and assurance (2.3) (p=0.6); however, this was not statistically significant. In the case of user experience, there were statistically significant differences in the mean ORCHA-24 score, with the results ranging from 3.7/8 (average review score of <3.5) to 5/8 (review score of 3.6–4.0) (p=0.05). Post hoc analysis using Tukey's HSD noted that a review score of 3.6–4.0 demonstrated significantly higher user experience scores than those with a review score of <3.5 and a numerically significant difference over those with a user review score of 4.1–4.5 (figure 3).

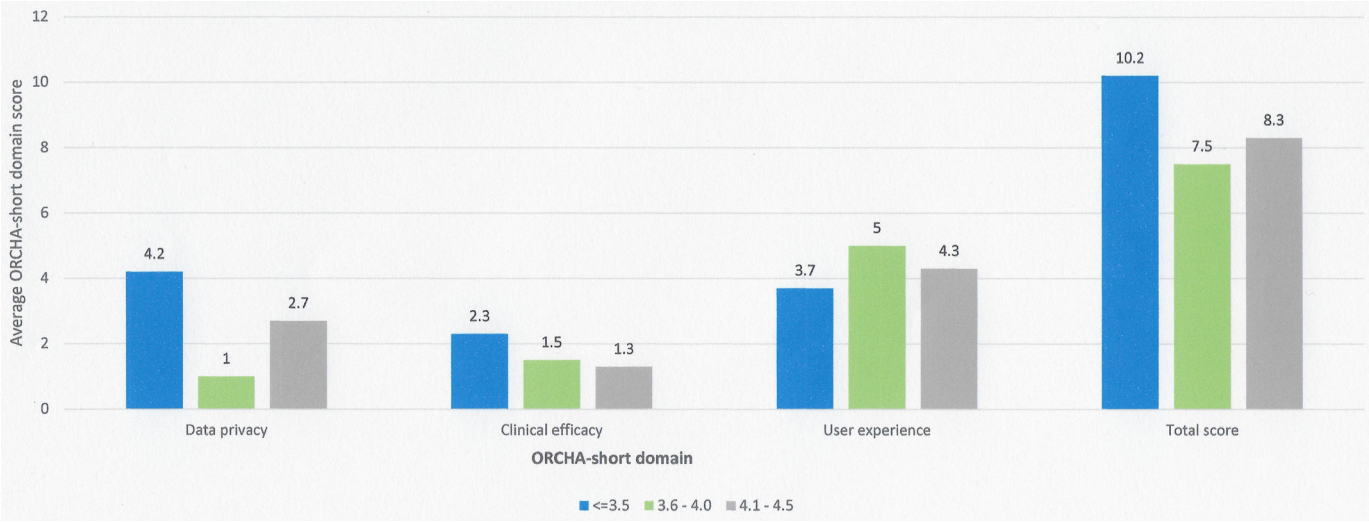

Figure 3 ORCHA-24 domain score versus average user review score.

Price as an indicator of quality?

Paid for apps demonstrated improved ORCHA-24 assessment scores for data security (3.4 vs 2.4, p=0.4), and clinical efficacy and assurance (2.6 vs 1.4, p=0.19), but a slightly lower score for user experience (4.0 vs 4.5, p=0.38). Overall, paid for apps scored a total ORCHA-24 score of 10/24, compared with 8.2 for alternatives that were free at the point of download (p=0.24).

Using the ORCHA-24 as a horizon scanning tool

The NHS-approved app Sleepio consistently outscored the comparator apps on all domains of the ORCHA-24. The most numerically significant differences were observed in the domains of clinical efficacy and assurance (8 vs 1.7) and data privacy (7 vs 2.6). On combining all elements of the ORCHA-24 review, Sleepio achieved a total score of 20/24, comparing to an average 8.6 among non-accredited apps (range 3/24 to 14/24).

DISCUSSION

This study represents a review and analytical assessment of apps dedicated to the alleviation of chronic insomnia disorder and available to download via the Android Google Play Store. Apps reviewed varied considerably in their approaches to data privacy, user experience and documenting evidence of clinical effectiveness. Apps with CBT as a mechanism of action consistently outperformed apps marketed as hypnosis, meditation or other methods such as neurolinguistics programming, across all domains of the ORCHA-24; with the largest difference occurring when examining evidence of clinical efficacy (p=0.11). A review of app and web-based psychological interventions, conducted in the context of the NHS apps library, also recently demonstrated that technologies applying CBT are most likely to be backed by evidence of clinical effectiveness.[4] While both studies rely on relatively small numbers of apps included, this suggests that apps using methodologies with a sufficient evidence-base, such as CBT, may be a less uncertain prospect, even if evidence in the context of the app itself is lacking.

Comparing the average results across all domains of the ORCHA-24, apps scored much higher in the domains of user experience (4.4/8) and data privacy (2.8/8) than clinical efficacy and assurance (2.1/8). This resulted in a statistically significant difference when comparing the former with the latter (p=0.03), suggesting that demonstrating evidence of clinical benefit remains a significant obstacle for app developers. Given that many app development companies are likely small and lacking adequate funding for research and development, this is perhaps unsurprising. Although ideally the burden of-proof of clinical effectiveness should ultimately lie with app developers, previous research has suggested that the absence of technical support for those unfamiliar with clinical research makes it unlikely that evidence generation will become a priority for app developers.[21]

While evidence of effectiveness is paramount to clinicians and policy makers alike, it is also currently not clear how important evidence of clinical efficacy, including accreditation or RCTs, is to the end users of apps. Subsequently, it remains unclear to what extent such factors are likely to influence the decision to download and engage with apps, when compared with usability or data privacy concerns for example.

A large-scale analysis of 18 000 popular free apps available on the Google Play Store recently highlighted that ~50% were missing a privacy policy, despite ~70% processing personally identifiable information,[22] synonymous with the 63.2% and 68.4% observed respectively in this study. Additionally, just 36.8% of apps with a data privacy policy in this study stated that no user data would be shared without explicit user consent, a finding also synonymous with existing estimates of 17%[19] to 50%.[23] One app user opinion poll conducted in 2014 by the Pew Research Center highlighted that 9 in 10 prospective app downloaders say that having clear information about how their data will be used is 'very' or 'somewhat' important when choosing whether to download an app.[24] The same research highlighted that 6 in 10 have chosen not to install an app when learning exactly how much personal information the app required to use it, while 43% admitted to uninstalling an app for the same reason, after initially downloading it.[25]

Given that apps with high user reviews are known to be downloaded significantly more than those with low user reviews,[26] our borderline significant finding that the apps in this study with the lowest user review scores demonstrated superior achievement of data privacy endpoints versus those with higher review scores of 3.6–4.0 and 4.1–4.5, respectively (p=0.06), may be of some concern. Similarly, apps with the lowest user review scores also demonstrated improved ORCHA-24 scores in the domain of clinical efficacy and assurance; however, this was a relatively minor and insignificant improvement (p=0.6).

The discovery that the NHS-approved app Sleepio convincingly and consistently outperformed all non-accredited apps across all domains of the ORCHA-24 may also be a significant finding, particularly when considering the poor visibility of app quality and risk indicators that characterises the ever-growing health app market. 'One-off' endorsements by bodies including the NHS, FDA or NICE's health app briefings are rare, and often time consuming, with an NHS digital assessment expected to take an average 3–4 weeks per app.[27] The majority of questions covered by the ORCHA-24 (21/24), with the exception of those concerning accessibility (Q17 and18) and determining whether the app provides a list of review or accrediting bodies (Q16), are now also included within the beta NHS digital assessment question set. However, unlike the more rigorous and inclusive NHS digital assessment, which also takes account of economic evidence and assesses the quality of the code within apps, the short and simple ORCHA-24 is designed to be conducted at scale. Because assessments and approvals by bodies such as the NHS are currently an exception as opposed to a rule, this leaves consumers with relatively few indicators regarding which of the many available health apps to use. At present, 26 NHS-approved apps are provided through the updated beta NHS apps library[28]; however, upwards of 165 000 are available to download.[29] Therefore, in therapeutic areas yet to benefit from the existence of approved or recommended health apps, the ORCHA-24, a short and scalable assessment, designed to be interpreted by end users of apps, may provide a feasible, timely and cost-effective means of informing potential users of the risks and benefits associated with the many competing apps available for download.

There are a number of limitations to this study, both in terms of research design and the limitations of the ORCHA-24 assessment itself. First, limiting the inclusion of apps almost exclusively to those available on the Android Google Play Store may have limited the generalisability of the study findings. However, this is not the first analysis of mHealth solutions to focus predominantly on Android compatible health apps.[23] Because the number of app downloads were believed a priori to be potentially correlated with app quality, Android apps were the only option, of the mainstream app platforms available, that could provide such data.

Second, the uniform weighting applied to each question, such that proof of a positive result from an RCT is considered of equal value to the presence of a privacy policy within the app, may not be an accurate representation of what users, and healthcare professionals alike, value when deciding on which mHealth solutions to use. In the absence of preference elicitation data, however, we deemed this to be the only fair and objective approach to the respective weighting of responses in the assessment, in order to prevent value judgements concerning the relative values of each of the 24 criteria subsequently affecting the results of the study. It is important that future research with relevant stakeholders, including app users, clinicians and policy makers, highlights the relative importance and appropriate utility weights of the numerous criteria within the ORCHA-24 app assessment. Such research would add considerable value in a time when the well-being and preferences of both healthcare users and providers alike are increasingly being considered within the context of transformational change and healthcare reimbursement.

Third, while the ORCHA-24 app assessment was developed as an objective and evidence-based tool, the requirement for the assessment to be operable at scale necessitated assumptions, some of which may have limited the precision of the tool with respect to a small number of the assessment criteria. Primarily, if an app failed to state the existence of clinical efficacy data, in either the app store description, or within the app itself, it was assumed that such information did not exist. Similarly, if an app developer stated user involvement during design and testing, in the absence of means of verification, it was assumed that this information was correct. As a result, some aspects of the tool, primarily questions 9–11, were open to gamesmanship, and future research may look to reduce the respective weighting of such questions, and thereby improve the precision of the tool.

Finally, while the tool aimed to capture as much relevant and important information as possible, with respect to user experience, clinical efficacy and data privacy, these questions were deemed of relevance following a Delphi process in a single group of mHealth stakeholders, and as a result, we appreciate the possibility for potentially important criteria to have been omitted. While this is a concern, it is our collective view that this tool should not be seen as a replacement for more in-depth assessments, such as those using NHS digital assessment criteria, but rather as an adjunct, with the possibility of being used in the time that it takes to perform such assessments, or as a horizon-scanning and prioritisation tool beforehand. As such, we do not view this as a major limitation but rather a necessity of the assessment being able to be applied at scale.

While the ORCHA-24 has proved a reliable indicator of app quality within the area of chronic insomnia disorder, it is imperative that future research addresses whether the ORCHA-24 yields similar results in alternative therapeutic areas, prior to considering wider use of the tool for the assessment of health apps.

CLINICAL IMPLICATIONS

While mHealth represents significant opportunities to health systems, the potential 'side-effects' of using health apps are often unclear, giving no guarantee of 'first doing no harm' and limiting the viability of this therapeutic medium. This study has uncovered substantial variation in the quality of apps dedicated to chronic insomnia disorder, which was not clearly correlated with observable factors such as app downloads or user review scores. The NHS-approved insomnia app Sleepio consistently and convincingly outperformed non-accredited apps across all domains of the ORCHA-24 assessment, suggesting that in the absence of formal app accreditation, this tool may feasibly be applied to highlight the risk–benefit profile of health apps to prospective users prior to downloading. However, further research is required to determine whether this outcome will also be observed in other therapeutic areas.

Acknowledgements The authors wish to acknowledge the contributions of Matthew Leahy and Sam Leahy, who assisted in the design of the questionnaire and the reviewing of apps.

Funding This analysis was funded by a research grant from ORCHA Healthcare Ltd.

Competing interests SL reports grants from ORCHA Healthcare Limited, from null, during the conduct of the study.

Provenance and peer review Not commissioned; externally peer reviewed.

doi:10.1136/eb-2017-102751

Received 12 June 2017; Revised 9 September 2017; Accepted 13 September 2017

REFERENCES

1. **Silva BM,** Rodrigues JJ, de la Torre Díez I, *et al*. Mobile-health: a review of current state in 2015. *J Biomed Inform* 2015;**56**:265–72.
2. **Price Waterhouse Coopers (PWC): socio-economic impact of mHealth: an assessment report for the European Union.** 2013 https://www.gsma.com/iot/wp-content/uploads/2013/06/Socio-economic_impact-of-mHealth_EU_14062013V2.pdf (accessed 9 Mar 2017).
3. **World Health Organisation World Health Report.** Mental disorders affect one in four people. http://www.who.int/whr/2001/media_centre/press_release/en/ (accessed 18 Apr 2017).
4. **Leigh S,** Flatt S. App-based psychological interventions: friend or foe? *Evid Based Ment Health* 2015;**18**:97–9.
5. **Huckvale K,** Prieto JT, Tilney M, *et al*. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. *BMC Med* 2015;**13**:214.
6. **Charter of fundamental rights of the European union (2000/c 364/01).** article 8. http://www.europarl.europa.eu/charter/pdf/text_en.pdf (accessed 7 apr 2017).
7. **Huckvale K,** Adomaviciute S, Prieto JT, *et al*. Smartphone apps for calculating insulin dose: a systematic assessment. *BMC Med* 2015;**13**:106.
8. **Wolf JA,** Moreau JF, Akilov O, *et al*. Diagnostic inaccuracy of smartphone applications for melanoma detection. *JAMA Dermatol* 2013;**149**:422–6.
9. **Stoyanov SR,** Hides L, Kavanagh DJ, *et al*. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 2015;**3**:e27.
10. **Khoja S,** Durrani H, Scott RE, *et al*. Conceptual framework for development of comprehensive e-health evaluation tool. *Telemed J E Health* 2013;**19**:48–53.
11. **PAS 277:2015 health and wellness apps.** Quality criteria across the life cycle. Code of practice. http://shop.bsigroup.com/forms/PASs/PAS-2772015/ (accessed 4 Feb 2017).
12. **App Quality Alliance.** Baseline testing criteria for android applications. updated with v1.6,30 Nov 2014. http://www.appqualityalliance.org/AQuA-test-criteria-for-android-apps (accessed 5 Feb 2017).
13. **The data protection act 1998.** http://www.legislation.gov.uk/ukpga/1998/29/contents (accessed 4 Jun 2017).
14. **The European Data Protection Directive (95/46/EC).** http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31995L0046 (accessed 5 May 2017).
15. **Information Commissioner's Office.** Guidance on data security breach management. version 2.1. the data protection act. (2012). https://ico.org.uk/media/for-organisations/documents/1562/guidance_on_data_security_breach_management.pdf (accessed 6 Apr 2017).
16. **GSMA: Mobile and Privacy.** Privacy design guidelines for mobile application development 2012. https://www.gsma.com/publicpolicy/wp-content/uploads/2012/03/gsmaprivacydesignguidelinesformobileapplicationdevelopmentv1.pdf (accessed 18 May 2017).
17. **App Quality Alliance.** Accessibility testing criteria for windows applications version 1.0 October 2015. http://www.appqualityalliance.org/Accessibility_Testing_Criteria (accessed 7 Feb 2017).
18. **W3C.** Accessibility requirements for people with low vision editor's draft 06 Jun 2016, clause 3.3.1. https://www.w3.org/TR/low-vision-needs/ (accessed 5 Apr 2017).
19. **NHS.** Online Mental Health Services: Sleepio. http://www.nhs.uk/Conditions/online-mental-health-services/Pages/sleepio.aspx (accessed 21 May 2017).
20. **Misrepresentation act 1967.** http://www.legislation.gov.uk/ukpga/1967/7/section/3 (accessed 18 Apr 2017).
21. **Leigh S.** Comparing applets and oranges: barriers to evidence-based practice for app-based psychological interventions. *Evid Based Ment Health* 2016;**19**:90–2.
22. **Zimmeck S,** Wang Z, Zou L, *et al*. Automated analysis of privacy requirements for mobile apps. 2016 http://shomir.net/pdf/publications/plt_2016s.pdf (accessed 23 Apr 2017).
23. **Blenner SR,** Köllmer M, Rouse AJ, *et al*. Privacy policies of android diabetes apps and sharing of health information. *JAMA* 2016;**315**:1051–2.
24. **EU Commission.** EU charter of fundamental rights (2009). http://ec.europa.eu/justice/fundamental-rights/charter/index_en.htm (accessed 11 Feb 2017).
25. **Olmstead K,** Atkinson M. Pew Research Center: Internet & Technology. Apps permissions in the google play store. 2015 http://www.pewinternet.org/2015/11/10/apps-permissions-in-the-google-play-store/ (accessed 28 May 2017).
26. **Khalid H.** On the link between mobile app quality and user reviews. M.Sc Thesis. 2014 https://qspace.library.queensu.ca/bitstream/handle/1974/12596/Khalid_Hammad_201410_MSC.pdf;jsessionid=25F5D563137E7802DBF77A2F0AB3F84E?sequence=1 (accessed 28 May 2017).
27. **Apps Information for Developers about the Apps Library (Beta).** NHS digital assessment questions – Beta. frequently asked questions. https://developer.nhs.uk/digital-tools/frequently-asked-questions/ (accessed 8 Sep 2017).
28. **NHS apps.** https://apps.beta.nhs.uk (accessed 2 May 2017).
29. **IMS institute for health informatics.** Medicines use and spending in the US—a review of 2015 and outlook to 2020. http://www.imshealth.com/en/ thought-leadership/ims-institute (accessed 2 Apr 2016).
30. **AQuA.** Best practice guidelines for producing high quality mobile applications version 2.3. http://www.appqualityalliance.org/AQuA-best-practice-guidelines (accessed 5 May 2017).