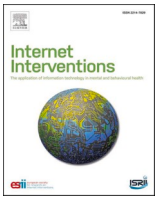
Internet Interventions 25 (2021) 100433

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent



Terms and conditions apply: Critical issues for readability and jargon in mental health depression apps

Sagar Jilka [a,b,*,1], Sara Simblett [a,b,1], Clarissa M. Odoi [a,b], Janet van Bilsen [a], Ania Wieczorek [a], Sinan Erturk [a], Emma Wilson [a], Magano Mutepua [a], Til Wykes [a,b]

[a] Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
[b] South London and Maudsley NHS Foundation Trust, London, UK

ARTICLE INFO

Keywords:
Mental health apps
Privacy policies
Readability
Jargon
Service users

ABSTRACT

Background: Mental health services are turning to technology to ease the resource burden, but privacy policies are hard to understand potentially compromising consent for people with mental health problems. The FDA recommends a reading grade of 8.
Objective: To investigate and improve the accessibility and acceptability of mental health depression app privacy policies.
Methods: A mixed methods study using quantitative and qualitative data to improve the accessibility of app privacy policies. Service users completed assessments and focus groups to provide information on ways to improve privacy policy accessibility, including identifying and rewording jargon. This was supplemented by comparisons of mental health depression apps with social media, music and finance apps using readability analyses and examining whether GDPR affected accessibility.
Results: Service users provided a detailed framework for increasing accessibility that emphasised having critical information for consent. Quantitatively, most app privacy policies were too long and complicated for ensuring informed consent (mental health apps mean reading grade = 13.1 (SD = 2.44)). Their reading grades were no different to those for other services. Only 3 mental health apps had a grade 8 or less and 99% contained service user identified jargon. Mental health app privacy policies produced for GDPR weren't more readable and were longer.
Conclusions: Apps specifically aimed at people with mental health difficulties are not accessible and even those that fulfilled the FDA's recommendation for reading grade contained jargon words. Developers and designers can increase accessibility by following a few rules and should, before launching, check whether the privacy policy can be understood.

1. Introduction

Resources for mental health services are scarce, and technology, such as smartphone apps, can be a way to use them efficiently (Krausz et al., 2019; Wykes, 2019). Thousands of mental health and wellbeing apps are available (Wykes and Schueller, 2019; Larsen et al., 2019) and they nearly always ask consumers to disclose personal data and to consent or assent to use these data (Razaghpanah et al., 2018). The privacy policy is usually presented at download and most apps cannot be used without first consenting to data being used by the app company. Privacy policies in Europe are governed by the General Data Protection Regulation (European Parliament and Council of European Union, 2016). GDPR states that privacy policies should be written "in a concise, transparent, intelligible and easily accessible form, using clear and plain language" (Article 12 of the GDPR (EU 2016/679)). The potential use of mental health apps has been advocated (Broughton, 2020) with some apps recording many more sessions and a spike in users during the pandemic. Apps for anxiety and depression have especially increased so it is now even more important that consent processes are appropriate (Lenahan, 2020; Herzog, 2020; Chowdhury et al., 2020).

Readability guidelines have been issued by the Food and Drug Administration (FDA) who suggest that an acceptable level is US school

* Corresponding author at: Institute of Psychiatry, Psychology & Neuroscience, King's College London, 16 De Crespigny Park, London SE5 8AF, UK.
E-mail address: sagar.jilka@kcl.ac.uk (S. Jilka).
[1] Joint first author.

https://doi.org/10.1016/j.invent.2021.100433
Received 5 January 2021; Received in revised form 20 May 2021; Accepted 14 July 2021
Available online 19 July 2021

8th grade (readable by someone aged 13) (Food and Drug Administration, 2014). This is measured using the Flesch reading grade (Flesch, 1948). This measure has been used extensively to analyse health care text (DuBay, 2004; Williamson and Martin, 2010) and app privacy policies (Powell et al., 2018; Robillard et al., 2019). To date, readability analyses of health apps' privacy policies have been concerned with the whole document, but this is problematic as long documents often have more difficult sections that are not identified from an overall readability score (Ennis and Wykes, 2016). Privacy policies are also notoriously long and littered with legal jargon that can affect accessibility and this has not yet been explored.

When privacy policies may be hard to read and understand, the ethics of this data sharing process has been questioned (Wykes and Schueller, 2019; Wykes et al., 2019), especially for individuals with mental health problems who are often excluded from using digital services (Ennis and Wykes, 2016; Wykes et al., 2019). The inaccessibility of privacy policies can add to this exclusion. We need to understand how difficult these policies are to read, to identify where jargon interferes with comprehension and be guided by service users on how they should change to improve their accessibility.

This study investigates the accessibility of mental health depression app privacy policies in two phases. In phase 1, we sought the views of those using the apps – mental health service users - on general design principles for privacy policies that would make them more accessible, including identifying and rewording jargon. In phase 2, we built on these findings to investigate readability and length of mental health privacy polices compared to those from other services (finance, social media, and music). We expected mental health app privacy policies to be shorter and more readable than other service categories. But we also wanted to explore whether conforming to FDA guidelines would mean an absence of complicated text or jargon. We also wanted to discover whether the introduction of GDPR has made mental health app privacy policies more readable.

2. Methods

This was a mixed methods study in two phases. The first investigated service user views of the accessibility of two mental health app privacy policies and what made the policies difficult to understand. This included identifying jargon. The second phase built on these views and investigated the readability of mental health privacy policies and compared them to three other service categories: music, finance, and social media, identified jargon and whether GDPR made policies more readable.

2.1. Phase 1

2.1.1. App privacy policy selection

We selected two privacy policies from mental health therapy apps for depression. One app (SilverCloud), was available through the NHS on prescription and regularly used in primary mental healthcare services (NHS Apps, n.d.). The other app (MoodCalmer) was freely available for download. Both apps received a high usability rating on the mental health app review website, Psyberguide (Psyberguide, n.d.).

2.1.2. Service user sample

We recruited 31 service users with experience of a major depressive disorder and therefore likely to be users of the selected apps. All participants were recruited via the Maudsley Biomedical Research Centre's advisory groups (https://www.maudsleybrc.nihr.ac.uk/patients-public/support-for-researchers/), or via another mental health research study (RADAR CNS: https://www.radar-cns.org/). They took part in a rating exercise of both mental health app policies. Twelve from the entire group accepted an invitation to take part in focus groups on design solutions to make the policies more interpretable. A mental health service user advisory group was also consulted about alternative wordings for the service user identified jargon.

2.1.3. Measures

PHQ-8 (**Kroenke et al., 2009**) is a validated 8-item measure of the severity of symptoms associated with major depressive disorder, with a threshold of ≥10 indicating current clinical problems. Scores range from 0 to 24 with higher scores indicating a higher symptom severity.

Enlight evaluation tool (**Baumel et al., 2017**) is a validated 28 item quality assessment tool for mobile and web-based eHealth interventions. It covers a breadth of topics that have been used to assess the quality and therapeutic potential of apps and online therapies. Each item is scored from 1 to 5, with 5 representing higher quality. However, not all the items were appropriate to the assessment of the privacy policy alone so we engaged with a UK national young person's mental health advisory group (www.ypmhag.org) to define which were the most appropriate. They recommended eleven items: eight on usability (items 1–8) and three that were important to understand data journey, storage and use (items 9, 10 and 11) (see Table 1).

2.1.4. Procedure

2.1.4.1. Quantitative service user ratings. Following informed consent (ethics reference number: 18/NE/0340, North East - Newcastle & North Tyneside 2 Research Ethics Committee) participants completed the PHQ-8 and demographic, physical health and smartphone use information. Service users rated the two app policies using the Enlight evaluation tool and then highlighted jargon words or phrases in the two app policies that were difficult to understand.

Table 1
The eleven items from the Enlight questionnaire.

Item number	Feature	Description	Score range or value
1	*Navigation*	Is it easy/natural/frictionless to navigate through the document?	1–5
2	*Layout*	Does the document appear well-organized?	1–5
3	*Size*	Are the sizes of fonts/buttons/menus appropriate for the mental health app?	1–5
4	*Content Presentation*	Is the content presented in an engaging/interesting way (e.g., contains the right mix of video/audio/text/graphics)?	1–5
5	*Not Irritating*	Does the program avoid irritation in the user's experience (e.g., by controlling notifications/alerts/sounds or avoiding irritating colours/fonts/sounds/expressions)?	1–5
6	*Quality of Information*	Is the information provided in a clear and appropriate way for the target audience?	1–5
7	*Complete and Concise*	Is there sufficient information about what will happen to user's data without any omissions, over-explanations, or irrelevant text?	1–5
8	*Clarity about the App's Purpose*	Is there sufficient and accurate information about the target audience, the clinical aim (e.g., potential outcomes), and appropriate ways to use the app?	1–5
		Total score	8–40
9	*Data journey*	Are the risks around data transfer and exposure clear?	Yes/No
10	*Data storage*	Does the system notify users of how personal identifiable information will be kept confidential and secure?	Yes/No
11	*Data use*	Does the system notify users about how gathered data may be used (e.g. for commercial reasons)?	Yes/No

2.1.4.2. Focus groups. The focus groups explored user familiarity and willingness to use mental health apps, if and why they read privacy policies, how accessible they were, and what elements could be changed to make them more accessible (see supplementary material for topic guide). All participants were invited to provide feedback on the emerging themes in a second member checking group. Both groups took place in a university, lasted 90 min and were facilitated by mental health researchers, some of whom had personal experience of using mental health services.

2.1.4.3. Further service user advisory board consultation. We consulted a patient and public service user group called the 'feasibility and acceptability support team for researchers' (FAST-R; https://www.maudsleybrc.nihr.ac.uk/patients-public/support-for-researchers/) who provide expert consultation on mental health research. They generated alternative wording for jargon that had been identified by the participants.

2.1.5. Data analysis

2.1.5.1. Service user ratings. We characterised the demographic, clinical and smartphone use data of the participants. We investigated the adapted Enlight total usability score (items 1–8) for internal consistency (reliability) using Cronbach's alpha for both privacy policies using SPSS23 (IBM Corp, 2015), and if >0.7 we used it to investigate differences between the two privacy policies and to test whether demographic, clinical or smartphone characteristics affected participants' scores. Accessibility has several dimensions, so, we compared each of the 8 Enlight items between the two privacy policies using *t*-tests, and calculated proportions on the remaining 3 (items 9–11) that were scored as 'yes' or 'no'. The frequency of participant-identified jargon words and phrases in the two mental health privacy policies were also calculated.

2.1.5.2. Focus groups. The groups were audio recorded and transcribed verbatim. Two researchers independently coded the themes using a framework analysis ((Pope et al., 2000); details in supplementary methods titled 'thematic analysis') and the software package NVivo (version 10; QSR International, Melbourne, Australia).

2.1.5.3. Service user advisory board consultation. A descriptive list of alternative words or phrases generated by mental health service users were accumulated.

2.2. Phase 2

2.2.1. App privacy policy selection

A detailed protocol of data extraction is provided in supplementary methods (privacy policy extraction protocol: part 1). In February 2020 we used this protocol for the Google Play Android store and Apple app store for mental health apps, and three other service categories: music, finance and social media. Further searches for mental health app privacy policies were conducted on the websites Psyberguide (Psyberguide, n.d.), ORCHA (Orcha, n.d.) and the NHS apps library (NHS Apps, n.d.). After duplicates were removed, 699 app privacy policies remained: mental health ($n = 197$), social media ($n = 174$), music ($n = 161$) and finance ($n = 167$).

2.2.2. Measures

Readability was assessed using the Flesch reading grade (Flesch, 1948) for all privacy policies. A detailed protocol on how the readability was calculated is provided in supplementary method (privacy policy extraction protocol: part 2). This Flesch measure has been used in studies of health care text (DuBay, 2004; Williamson and Martin, 2010) and app privacy policies (Powell et al., 2018; Robillard et al., 2019). A high score indicates more complicated text. We also extracted information about whether the privacy policy was updated before or after GDPR (25th May 2016).

For every mental health app privacy policy, we also calculated the number of words and the number of service user identified jargon words highlighted in phase 1. For mental health app privacy policies that scored at or below reading grade 8, we extracted every 100-word iteration and calculated the Flesch reading grade of these extracts. So that the readability score was not inflated by a lack of punctuation, headings, or formatting, we used the reading grade at the 75th percentile.

2.2.3. Data analysis

Descriptive statistics are provided on the average length (word count) and readability (Flesch reading grade) of privacy policies and the frequency of apps privacy policies scoring at or below reading grade of 8. Differences between service category (mental health, social media, finance, music) in length and readability were explored using ANOVA and post hoc Tukey or *t*-tests.

For mental health apps, we used t-tests to investigate whether differences in length or readability were affected by the introduction of GDPR. We also report the number of jargon words identified in Phase 1.

For apps that fulfilled the FDA reading criterion, we investigated whether there were any other issues that affected their accessibility including the readability of 100-word sections and calculated how many word passages also fulfilled the FDA requirements.

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

3. Results

3.1. Phase 1

The thirty-one participants were on average 51 years old (SD = 14.73), with most identifying as women (71%), white (84%), having a concurrent physical disability (61%), and being educated to degree level or above (71%). Most had used a smartphone (87%), and currently reported at least mild symptoms of depression (71%) (Table 2). Those carrying out the ratings did not differ much from those taking part in the focus groups although they did become more representative with half

Table 2
The demographic and clinical characteristics of participants who took part in the two sections of the study.

	User ratings and jargon analysis ($n = 31$)	Improving privacy policies analysis ($n = 12$)
	Mean (SD) or % (n)	Mean (SD) or % (n)
Age (years)	51 (14.73) Range: 23–76	61 (8.80) Range: 48–76
Gender		
% Women	71 (22/31)	58 (7/12)
Ethnicity		
% White	84 (26/31)	92 (11/12)
% Other	16 (5/31)	8 (1/12)
Education		
Degree or above	71 (22/31)	50 (6/12)
A-Levels	6 (2/31)	8 (1/12)
GCSE	23 (7/31)	42 (5/12)
Smartphone Use		
% Yes	87 (27/31)	73 (8/11)
Physical disability		
% Yes	61 (19/31)	58 (7/12)
PHQ-8 score		
None	29 (9/31)	42 (5/12)
Mild	35.5 (11/31)	42 (5/12)
Moderate	12.9 (4/31)	0 (0/12)
Moderately severe	6.5 (2/31)	8 (1/12)
Severe	16.1 (5/31)	8 (1/12)

being women and fewer being educated to degree level.

3.1.1. Adapted Enlight ratings

Internal consistency for the adapted-Enlight total usability score (items 1–8) was high for both privacy policies (Cronbach's alpha = 0.928 (SilverCloud), and 0.935 (MoodCalmer)). SilverCloud scored significantly better on the total score (t(60) = 2.50, $P = 0.018$ [95% CI = 0.75–9.51]). Participant characteristics were not associated with the adapted-Enlight total scores, so we investigated the individual items. SilverCloud scored significantly better on: layout (t(60) = 2.19, $P = 0.037$ [95% CI = 0.10–1.20]), size of fonts/buttons/menus (t(60) = 2.70, $P = 0.009$ [95% CI = 0.24–1.63]), content presentation (t(60) = 2.75, $P = 0.008$ [95% CI = 0.26–1.67]) and quality of information (t(60) = 2.28, $P = 0.026$ [95% CI = 0.10–1.45]) (supplementary table 1). But there were no differences on the other items.

Participants rated both privacy policies as reasonable (>60% positive responses) for their explanation about the data journey, data storage and data use (items 9–11).

3.1.2. Identifying jargon

Participants identified 115 instances of jargon across the two mental health privacy policies. The most frequently identified are shown in Table 3 (full list in supplementary table 3). More instances of jargon were identified in the MoodCalmer app ($n = 263$) than the SilverCloud app ($n = 147$).

3.2. Improving privacy policies

3.2.1. Design issues

The twelve participants lacked familiarity with both mental health apps selected for our study but had a willingness to learn how to use similar apps. Their willingness was moderated by: trust in the providers; endorsement by professional bodies (e.g. NHS, universities); a strong evidence base; and positive user reviews. Some participants were wary of the motivations of private companies and cited data sharing with future employers, especially sensitive mental health information, as problematic, but not everyone was against sharing data. These contextual themes are summarised in supplementary Table 2.

Despite having to agree to a privacy policy, some participants spoke about ignoring the information as they were lengthy, difficult to navigate and in some cases, a 'waste of time.' Some participants suggested that familiarity and assumed trust in the provider, product or developer would mean there would be no need to read privacy policies. Other participants discussed what made them decide to read them (supplementary table 2). These included individual relevance, and the increased awareness of the potential risks to sharing ('selling') data. Where participants expressed a lack of understanding, one suggestion was to get someone with more expertise to help, however, participants also thought that there was a responsibility for the quality of information to be improved by the app provider: "*where people who are vulnerable in some sense, or have their health compromised by whatever they're going through – you know, if there's any kind of vulnerability, there should be a higher standard of communication in the terms and conditions.*"

An outline of practical guidelines for the design of future privacy policies was developed (Fig. 1) that included suggestions around layout, navigation, length and language of document, general accessibility, standardisation and being upfront and transparent.

One participant said that there is no choice but to agree, however "*at least if it's more accessible you feel you've got a choice rather than just, 'I have to tick because I need to use this app and I don't know what it's talking about'. At least you'd be given a choice and then, you make a decision accordingly.*"

Table 3
Jargon: Difficult to understand words or phrases (with five or more mentions), alongside suggested changes. *Indicates words that service users thought could not be changed but an explanation put in the glossary of the privacy policy.

Jargon words or phrases	Total mentions in the two privacy policies	Suggest change
Third parties/party	17	Other organisations/other people involved
Aggregate(d)	16	Combine(d)
Cryptographic security	12	Coded online security
Legitimate interests	12	Legally valid interest/lawful
Analytics	11	Investigation process/analysis
Ancillary	8	Secondary, subsidiary, extra
Affiliates/affiliated companies	7	Connected/associated
Demographic(s)	7	information about you/Population-statistics/population-related/ population structure
Legal basis*	7	Legal reason
Portability	7	Moveable/transferable
Vendors	7	Sellers
Subpoena(s)*	6	Court summons/call to court
Subsidiaries	5	Secondary bodies
without limitation	5	Unlimited

3.2.2. Alternative wording for jargon

Members of a service user advisory group provided suggested changes to the 115 instances of jargon (Table 3, and full list in supplementary table 3).

3.3. Phase 2

3.3.1. Privacy policy length and readability

Table 4 shows the means and standard deviations for the length and readability of the 699 app privacy policies. Very few privacy policies ($n = 6$) in any service category scored at or below the FDA's recommended reading grade 8 (mental health: 1.6%, music: 2%, social media and finance: 0%) (see Table 5 for list of mental health apps).

Privacy policies for the four service categories were significantly different in length ($F(3,673) = 6.73$, $P < 0.001$), with only mental health being shorter than finance ($P = 0.008$ [95% CI = −1569.98- -167.13]) and social media ($P < 0.001$ [95% CI = −1812.22- -440.21]).

Readability did not differ between service categories ($F(3, 673) = 2.15$, $P = 0.093$). When investigating reading grade further, the most accessible privacy policy belonged to a mental health app (5.8; 'MoodSpace – Stress, anxiety, & low mood self-help (Boundless Labs)'), but the most complicated privacy policy also belonged to a mental health app (30.4; Depression Quote Wallpapers HD (App Makerz)). There were 3 other outliers (i.e., cases with values above 1.5 times the interquartile range) in the mental health category (26.6, 21.1, and 19.5), 5 in the music category (21.4, 20.7, 21.1, 19.0, 21.9), 3 in social media (19.1, 19.1 and 20.0), and 2 in finance (29.3 and 20.8). After removing these outliers, we found a significant difference between service categories ($F(3, 664) = 4.123$, $P = 0.007$), with mental health apps more readable than music ($P = 0.034$, [95% CI = −1.074- -0.029]) and social media ($P = 0.015$, [95% CI = −1.118- -0.085]).

3.3.2. Mental health app privacy policies

3.3.2.1. The effect of GDPR. Only 48% of mental health apps' privacy policies were updated after GDPR laws came into effect but its introduction did not have a significant impact on readability (t(120) = −0.326, $p = 0.745$). However, they were on average longer when adapting GDPR regulations (t(120) = 2.297, $p = 0.023$) (see supplementary table 4 for descriptive data).

3.3.2.2. Jargon in mental health app privacy policies. All but one mental health app privacy policy (99%: 195/196; Depression Screening Test (Eddie Liu)) contained at least one example of service user identified

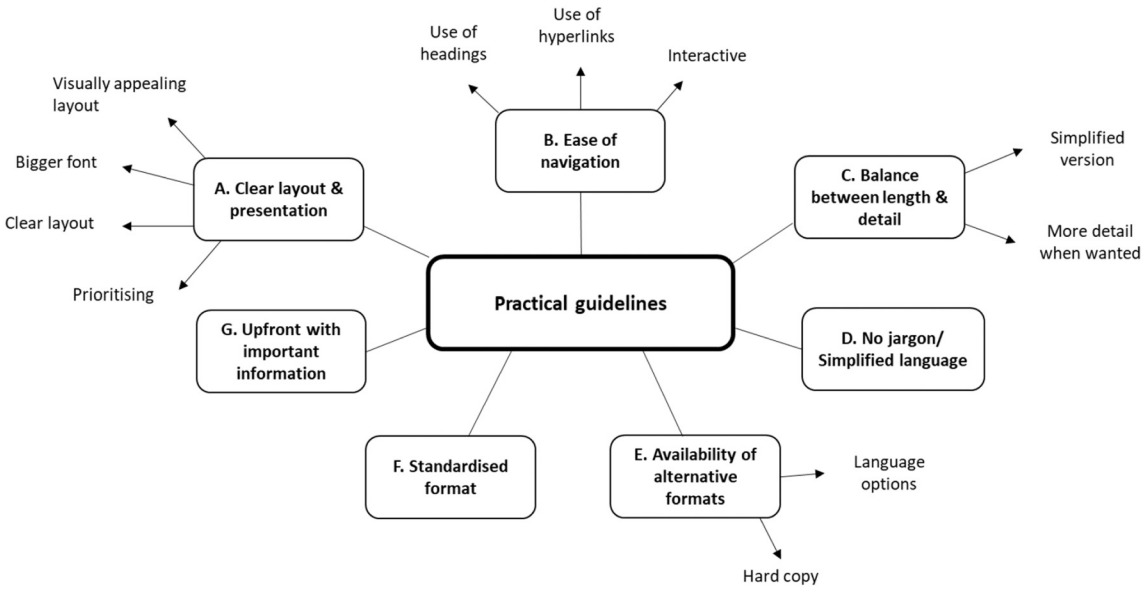


Fig. 1. Practical guidelines for the design of privacy policies developed by service users.

Table 4
Reading grade and length (word count) for each app service category.

	Flesch reading grade Mean (SD)	Length (word count) Mean (SD)
Mental health	13.06 (2.44)	1959.03 (1745.90)
Social media	13.58 (1.68)	3085.24 (2457.13)
Finance	13.17 (2.44)	2827.58 (2099.49)
Music	13.57 (2.27)	2464.73 (3559.67)

Table 5
Mental health apps whose reading grade is at or below 8, the reading grade of its 100 word most complicated section (at the 75th percentile), the length of their privacy policy, and the number of jargon words in its privacy policy.

App name (*developer name*)	Reading grade	Reading grade; most complicated 100 word section (at 75th percentile)	Word count	Number of jargon words
Depression medication manager (Early Access) (*MoodTools*)	7.30	13.35	236	9
Depression test (*MoodTools*)	7.30	7.70	236	9
MoodSpace - stress, anxiety, & low mood self-help (*Boundless Labs*)	5.80	8.20	328	5

jargon, with some privacy policies having 550 instances. There were no privacy policies which contained no jargon words and fulfilled the FDA's recommendation for reading grade.

3.3.3. Are app privacy policies with a reading grade ≤=8 accessible?

All three apps falling into this category contained service user identified jargon words (Table 5). We also calculated reading grades of every 100 word iteration of each policy. The grade at the 75th percentile was higher than the FDA recommendation in only 1 of the 3 (when rounded to the nearest grade) mental health apps (grade 13.35, Table 5). The percentage of 100 word passages that were at or below reading grade 8 varied considerably from 3% to 79%. An example of a complicated 100 word section taken from an FDA compliant privacy policy can be found in Text box 1.

4. Discussion

This was a novel mixed methods study expanding on previous readability analyses of privacy policies (Powell et al., 2018), incorporating mental health service user views. Few privacy policies comply with the FDA's recommended reading grade 8 level. The National Adult Literacy Survey revealed that about a quarter of US adults could not read or understand written materials above a fifth-grade level (Kirsch and United States, Educational Testing Service, and National Center for Education Statistics, 1993; Liu et al., 2011) and in the UK, "around 15 per cent, or 5.1 million adults have literacy levels at or below those expected of an 11-year-old" (Gilbert et al., 2018). Only three mental health privacy policies in our dataset were below these levels. Although they are more readable than other service categories, they are still too complicated for ensuring adequate consent processes. This was also true for social media, finance, and music apps. Although the introduction of GDPR has not improved readability, privacy policies are significantly longer when adapting GDPR regulations, and our service users highlighted problems with lengthy documents. Our analyses of 100-word chunks showed that although an app privacy policy can score less than grade 8 overall, they still contain sections of inaccessible text. These are increasingly important issues as a substantial proportion of users of such apps may not necessarily be English speakers.

Service users made recommendations for making privacy policies more accessible above and beyond what is captured in the adapted Enlight measure. This is important because recent evidence has shown that questionnaire ratings alone do not provide enough information for service users to make decisions about app usage (Zelmer et al., 2018). Novel findings included service users wanting privacy policies to prioritise critical information around consent at the beginning. The ease of navigation and use of hyperlinks and headings, in a standardised format, is important. Although the adapted Enlight measured quality of information in a clear and appropriate way, our service users expanded on this by highlighting the problems with jargon and wanted simplified versions of privacy policies. They identified a list of jargon words that developers should avoid, for example "third party", "aggregated" data, and "cryptographic security". In our sample of 196 mental health apps, only one mental health app had no jargon words, and no FDA compliant app had no jargon words.

Text Box 1
This 100 word passage is from an app with an acceptable reading grade (8.2), but this particular passage has a reading grade of 11.4.

"*MoodSpace app data is automatically backed up to your Google Drive - which you have control of. Here, it's safely encrypted, and protected by your Google account credentials. If you don't trust Google, you can delete your backup and disable backup from your device settings (just type 'backup' into the settings search box). Anonymous information we collect To make the app work well at all we collect the following anonymous data: Crash reports If you've never seen the app crashing, it's because as soon as one happens, we get a crash report. A little red light flashes in our office, a*"

4.1. Strengths and limitations

The strength of this study is the combination of a rigorous qualitative exploration of service user views of privacy policies with an in-depth quantitative analysis to explore inaccessibility and so provides a richer analysis of the problems. We also compared the readability of mental health apps with other service apps to detect if they were service category specific.

A potential limitation is the two mental health policies reviewed by service users in phase 1 were both therapy apps, and therefore not wholly representative of a wide range of mental health apps available. However the jargon was not therapy specific and was also present in almost every mental health privacy policy. These are also the types of app that have been recommended to service users as helping with mental health difficulties. Our participants were mostly women and well educated (although these demographics were more balanced for the focus groups). They also had few mental health symptoms and were taking part in a study of depression and smartphone technology and so may not be representative. Despite this, they found the privacy policies difficult to understand, so our findings are likely to be conservative with accessibility problems likely to increase with lower educational attainment and a wider range and severity of symptoms.

5. Conclusions

We have demonstrated that nearly all privacy policies are inaccessible and that mental health apps are no different to other services. More recent publicised concerns about data sharing of seemingly innocuous location data (Thompson and Warzel, 2019) will make individuals wary of apps and potentially be a barrier to their use, making clearly specified app privacy policies essential hopefully following our framework. For mental health apps it is crucial for developers to understand the needs of people with mental health difficulties which may differ from the general population. A reduction in this one barrier to using digital products might reduce mental health service user concerns and improve access and use of technology that might help mental health recovery.

Acknowledgements and contributors

SJ, SS & TW designed the study. All authors carried out the literature search. SJ, SS, CO, JvB, AW, SE, EW & MM collected the data and contributed to data analysis. SJ, SS & TW interpreted the data and wrote the manuscript. The authors would like to thank all service user advisory group members for the advice on this study.

Funding

This work was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London (IS-BRC-1215-20018).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.invent.2021.100433.

References

NHS Apps (n.d.). Retrieved from: https://www.nhs.uk/apps-library/. Accessed in August 2018 and in February 2020.

Baumel, A., Faber, K., Mathur, N., Kane, J.M., Muench, F., 2017. Enlight: a comprehensive quality and therapeutic potential evaluation tool for Mobile and web-based eHealth interventions. J. Med. Internet Res. 19 (3), e82 https://doi.org/10.2196/jmir.7270.

Broughton, E., 2020. Can apps help you manage your mental health during the coronavirus pandemic? [online] Patient.info. Available at: https://patient.info/news-and-features/can-apps-help-you-manage-your-mental-health-in-lockdown.

Chowdhury, B., Dodds, B., Durrant, B., Boland, B., Quetteville, B., Dodds, B., Field, B., 2020. Mental Health Apps Downloaded More Than 1M Times Since Start of Virus Outbreak. [online]. Available at:. The Telegraph [Accessed 4 September 2020]. https://www.telegraph.co.uk/technology/2020/05/16/mental-health-apps-downloaded-1m-times-since-start-virus-outbreak/.

DuBay, W.H., 2004. The principles of readability. Retrieved from. http://www.impactinformation.com/impactinfo/readability02.pdf. Accessed 18 Dec 2018.

Ennis, L., Wykes, T., 2016. Sense and readability: participant information sheets for research studies. Br. J. Psychiatry 208 (2), 189–194. https://doi.org/10.1192/bjp.bp.114.156687.

European Parliament and Council of European Union, 2016. Regulation (EU) 2016/679. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN (Accessed: 21 August 2020).

Flesch, R., 1948. A new readability yardstick. J. Appl. Psychol. 32 (3), 221–233. https://doi.org/10.1037/h0057532.

Food and Drug Administration, 2014. Informed consent – guidance for IRBs, Clinical Investigators and Sponsors. Retrieved from. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/informed-consent. Accessed 21 Oct 2019.

Gilbert, L., Teravainen, A., Clark, C., Shaw, S., 2018. A National Literacy Trust research report: literacy and life expectancy - an evidence review exploring the link between literacy and life expectancy in England through health and socioeconomic factors. Retrieved from. https://literacytrust.org.uk/documents. Accessed 21 Oct 2019.

Herzog, K., 2020. Mental health apps draw wave of new users as experts call for more oversight. Available at:. CNBC [Accessed 4 September 2020]. https://www.cnbc.com/2020/05/24/mental-health-apps-draw-wave-of-users-as-experts-call-for-oversight.html.

IBM Corp, 2015. Released. In: IBM SPSS Statistics for Windows, Version 23.0. IBM Corp, Armonk, NY.

Kirsch, I.S., United States, Educational Testing Service, & National Center for Education Statistics, 1993. Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey, 2nd ed. Office of Educational Research and Improvement, U.S. Dept. of Education, Washington, D.C.

Krausz, R.M., Ramsey, D., Wetterlin, F., Tabiova, K., Thapliyal, A., 2019. Accessible and cost-effective mental health care using E-Mental Health (EMH). In: Javed, A., Fountoulakis, K. (Eds.), Advances in Psychiatry. Springer, Cham.

Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B.W., Berry, J.T., Mokdad, A.H., 2009. The PHQ-8 as a measure of current depression in the general population. J. Affect. Disord. 114 (1), 163–173. https://doi.org/10.1016/j.jad.2008.06.026.

Larsen, M.E., Huckvale, K., Nicholas, J., Torous, J., Birrell, L., Li, E., Reda, B., 2019. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit. Med. 2 (1), 18. https://doi.org/10.1038/s41746-019-0093-1.

Lenahan, M., 2020. Calm records 120% more app sessions than headspace during quarantine [online] Blog.apptopia.com. Available at: https://blog.apptopia.com/calm-app-outperforms-headspace-during-quarantine [Accessed 4 September 2020].
Liu, E., Lo, T.-F., Hsieh, H.-L., 2011. In: Literacy and Health: Evidence From the 2003 National Assessment of Adult Literacy, p. 19.
Orcha (n.d.). Retrieved from: https://www.orcha.co.uk/. Accessed in February 2020.
Pope, C., Ziebland, S., Mays, N., 2000. Analysing qualitative data. Br. Med. J. 320 (7227), 114–116.
Powell, A., Singh, P., Torous, J., 2018. The complexity of mental health app privacy policies: a potential barrier to privacy. JMIR MHealth UHealth 6 (7), e158. https://doi.org/10.2196/mhealth.9871.
Psyberguide, (n.d.). Retrieved from https://psyberguide.org/. Accessed in August 2018 and in February 2020.
Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., Kreibich, C., Gill, P., 2018. Apps, trackers, privacy, and regulators: a global study of the mobile tracking ecosystem. In: Proceedings 2018 Network and Distributed System Security Symposium. Presented at the Network and Distributed System Security Symposium, San Diego, CA. https://doi.org/10.14722/ndss.2018.23353.
Robillard, J.M., Feng, T.L., Sporn, A.B., Lai, J.A., Lo, C., Ta, M., Nadler, R., 2019. Availability, readability, and content of privacy policies and terms of agreements of mental health apps. Internet Interv. 17, 100243 https://doi.org/10.1016/j.invent.2019.100243.
Thompson, S., Warzel, C., 2019. Opinion | Twelve Million Phones, One Dataset, Zero Privacy. [online]. Available at:. Nytimes.com [Accessed 3 Jan. 2020]. https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html.
Williamson, J.M.L., Martin, A.G., 2010. Assessing the readability statistics of national consent forms in the UK. Int. J. Clin. Pract. 64 (3), 322–329. https://doi.org/10.1111/j.1742-1241.2009.02245.x.
Wykes, T., 2019. Racing towards a digital paradise or a digital hell? J. Ment. Health 28 (1), 1–3. https://doi.org/10.1080/09638237.2019.1581360.
Wykes, T., Schueller, S., 2019. Why reviewing apps is not enough: transparency for trust (T4T) principles of responsible health app marketplaces. J. Med. Internet Res. 21 (5) https://doi.org/10.2196/12390.
Wykes, T., Lipshitz, J., Schueller, S.M., 2019. Towards the design of ethical standards related to digital mental health and all its applications. Curr. Treat. Options Psychiatry 6 (3), 232–242. https://doi.org/10.1007/s40501-019-00180-0.
Zelmer, J., van Hoof, K., Notarianni, M., van Mierlo, T., Schellenberg, M., Tannenbaum, C., 2018. An assessment framework for e-mental health apps in Canada: results of a modified Delphi process. JMIR Mhealth Uhealth 6 (7), e10016.