

Editorial

The era of digital mental health interventions: we know they can be effective but are they also safe?

Urska Arnautovska and Alyssa Milton



Over the past two decades, digital mental health interventions (DMHIs) have seen a surge in studies with people experiencing mental ill-health, whether this be via web-based platforms, smartphone applications, text messages or other digital devices. Although DMHIs already demonstrate evidence of their acceptability and some of their effectiveness among different populations, the information about their safety is less clear. This Editorial reflects on a Delphi study by Taher and colleagues that explored the regulation of DMHIs and generated ten safety recommendations. We discuss these recommendations in the context of existing relevant literature and provide suggestions for further steps to advance research and policy on DMHIs in the UK and globally. Further dialog is needed, including the views and experiences of all key stakeholders, and particularly of people with lived experience, to ensure DMHIs are not only an

acceptable and potentially effective treatment approach, but also safe for those that use them.

Keywords

Adverse events; digital interventions; DMHI development; mental health services; safety.

Copyright and usage

© The Author(s), 2025. Published by Cambridge University Press on behalf of Royal College of Psychiatrists. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Urska Arnautovska (pictured) is a senior research fellow at the University of Queensland, Australia, and Senior Mental Health Clinician with the Metro South Addiction and Mental Health Service, Brisbane, Australia, with an interest in the physical comorbidities of people with schizophrenia.

Alyssa Milton is an associate professor at The University of Sydney, Australia, and Director of Knowledge Transfer with the Australian Research Council's Centre of Excellence for Children and Families over the Life Course, Sydney, Australia.

Over the past two decades, we have witnessed a rapid increase in the number of studies using digital mental health interventions (DMHIs), delivered remotely via computers, smartphone applications (apps), text messages or other digital devices such as wearables, aimed at improving various mental health conditions.¹ Only a decade ago, the acceptability of DMHIs among populations experiencing mental ill-health was still unclear, but systematic reviews² and individual empirical studies published since then have provided strong evidence of DMHIs being an acceptable and potentially effective treatment option for people with mental ill-health across a range of subpopulations.³

There is no doubt that such approaches can overcome some of the main limitations of traditional, face-to-face mental health treatments, such as costs associated with face-to-face delivery, restricted access and reach (often tied to clinical settings), along with challenges in supporting continuity of care during critical transition points (e.g. discharge from in-patient treatment).⁴ However, it is currently less clear if DMHIs present a safe treatment option. The safety of DMHIs, and psychological interventions in general, is typically defined by the occurrence of unwanted effects or negative outcomes, categorised as adverse events or serious adverse events (SAEs). SAEs, which include outcomes like death, disability or hospital admission, are consistently defined because of strict regulatory frameworks. In contrast, there is greater variability in conceptualising and measuring non-SAE, encompassing a broad range of less severe

negative effects such as psychological harm (e.g. distress caused by content) and exacerbation of symptoms (e.g. worsening of anxiety or depression). An additional consideration of safety includes risks related to data privacy and security, which are critical in DMHIs, particularly when dealing with sensitive information or in high-risk scenarios, such as suicide prevention apps, where a trade-off between monitoring for safety and ensuring user privacy must be carefully balanced.

One reason for the lack of clarity around the safety of DMHIs is likely related to the fact that, although a small proportion of DMHIs have been tested rigorously in sufficiently powered randomised clinical trials⁵ (e.g. the Horyzons trial:⁶ $n = 170$, found no significant differences compared with treatment as usual in the primary outcome (social functioning), but positive trends in vocational outcomes and use of emergency services; and the EMPOWER trial:⁷ $n = 73$, found lower fear of relapse in the EMPOWER group compared with treatment as usual), most DMHIs are available freely on the App store and lack proper evidence of safety and efficacy. As such, the issue is two-fold: first, there is a largely unregulated market supporting the existence of DMHIs, and second, we currently lack a regulatory method that is appropriate for the assessment of DMHIs. Currently, DMHIs are typically classified as 'medical devices' and, therefore, are considered to fall under regulatory processes that are responsible for regulating both pharmaceuticals and medical devices (e.g. the Medicines and Healthcare products Regulatory Agency in the UK, Therapeutic Goods Administration in Australia and Digital Services Act in the USA). Merging DMHIs with traditional medical interventions, however, risks overlooking critical aspects specific to mental health contexts.

To address this risk, a recent Delphi study by Taher and colleagues⁸ involved an international panel of 20 research experts suggesting solutions on how to adapt medical regulatory methods to better suit DMHIs. The strength of this research is in pursuing a regulatory solution that would be tailored to mental health

Table 1 Recommendations on how to improve the safety of digital mental health interventions and strategies for their implementation		
Theme	Ten recommendations that reached consensus	Strategies to implement recommendations
Improving adverse events data in DMHI trials	(a) Publish guidelines for professionals on how to assess and report adverse event data in digital mental health research (b) Standardise the optimum way to assess relatedness of an adverse event to the intervention, procedure or device, by using a variety of independent evidence sources (c) When evaluating safety, differentiate between self-guided and clinician guided interventions	Generate publication guidelines for reporting adverse events data that are required by journals to ensure that adverse event data are collected and reported consistently These guidelines should outline what adverse events data should be collected, when and how this should happen, and how these should be categorised, analysed and reported Determine the relatedness of the adverse event to the therapy
Broadening the definition of serious adverse events for DMHIs	(d) Broaden the scope of potential serious adverse events in digital studies to consider impact on mental health	The current definition of serious adverse events (hospitalisation, death, congenital abnormalities...) should be extended to be inclusive of: (a) Serious effects on mood, behaviour, general well-being and functioning, not just physical harms (b) Increase in harm toward others (c) Increase in intentional self-harm (d) Seizures induced by on-screen flashes (e) A suicide attempt
Short-term deterioration is not a safety concern in mental health therapy	(e) Monitor psychological deterioration to prevent potential harm, using post-treatment follow-ups (f) Utilise the suggestions outlined to support patients experiencing deterioration in any treatment context	Develop and maintain a strong therapeutic alliance Discuss a clear rationale for each therapy activity/exercise Patients need to be informed about short-term symptom deterioration/distress (through the process of informed consent). In a digital context this can be replicated through text, audio or video content It is important for therapists to normalise symptom deterioration. This can happen as part of a discussion with a trained mental health professional, or via the content of the digital therapy Help the patient think about sources of support that they can use between sessions if needed. In a digital context, provide users with alternative sources of support should they need it Provide patients/users with information on how to manage symptom deterioration
Adapting the yellow card scheme for DMHIs	(g) Improve the visibility of the yellow card scheme as the primary means of reporting adverse reactions in the digital mental health field (h) Change the yellow card ‘psychiatric conditions’ category label to ‘mental health adverse reactions’ (i) Expand the yellow card category that captures mental health data to incorporate informative subcategories and additional contextual information	Raise awareness among users that the yellow card system is used for reporting adverse reactions in digital mental health therapies Use a broader category such as ‘mental health adverse events’ is more appropriate. Include: (a) Severity/impact on the individual (b) Suicidal ideation/behaviour (c) Intentional self-harm (d) Threat to others Patient-reported outcomes automatically collected within the digital therapy itself
Using a harmonised approach to assess the safety of psychological therapies generally	(j) Develop an overarching, harmonised framework for assessing the safety of any psychological intervention (digital or otherwise)	Have a harmonised approach for safety reporting across face-to-face and digital interventions to increase our knowledge of the risks of mental health therapies and contribute to stronger safety standards and safer interventions
DMHI, digital mental health intervention.		

interventions delivered using various technology platforms, given that DMHIs are radically different compared with pharmaceutical interventions, and are characterised by the large heterogeneity of digital approaches and their common use as adjuncts to care, rather than as standalone treatments. Valuable outcomes of the Delphi study are a list of ten safety recommendations that reached consensus and the implementation strategies to support the assessment of the safety of DMHIs (Table 1).

Reflecting on these recommendations, our own experience with co-production and delivery of DMHIs,⁴ as well as previous initiatives aiming to capture the strengths and weaknesses regarding the methodology and safety of DMHIs,⁹ we point out several key elements and steps to advance this research and policy in this field. First, the implementation of a formalised regulatory method is essential for ensuring that DMHIs are both effective and safe, and do not inadvertently increase the risks of harm to their users, ultimately benefiting a broader range of individuals. Second, the safety of DMHIs should be assessed early in pilot feasibility studies, addressing any issues raised by key end users before scaling up. Involving end users throughout the development process of DMHIs early may also improve the implementation of digital treatment approaches, including user engagement with DMHIs, which, in turn, is likely to facilitate positive outcomes.¹⁰ Third, the need for assessment of the safety of DMHIs early in the research process underscores the importance of regulatory boards such as human research ethics committees being familiar with the guidelines for evaluation of DMHIs, such as the World Health Organization mHealth Evaluation, Reporting and Assessment (mERA) checklist.¹¹ Such formal regulatory processes would ensure that any risks associated with a specific DMHI to be used in a subsequent study, along with detection and mitigation of these risks, are clearly outlined in the risk evaluation and management plan within a study/trial protocol. Finally, we underscore the need for involvement of people with lived experience, especially those using DMHIs, to tailor these guidelines to specific conditions and risks.

The latter point leads us to a key limitation of the Delphi study by Taher and colleagues: the lack of lived experience,¹² service user/patient and carer perspectives and an overrepresentation of views from those working in higher education. It is essential that all key end users, including people with lived experience of a specific mental health condition as well as clinicians and other therapists, provide their perspectives on what aspects of safety should be monitored and possibly mitigated when using DMHIs within real-life clinical and community settings. As such, a concept mapping approach may be better suited to capture the perspectives of all of these diverse users.

An example of a successful use of the concept mapping methodology to generate best practice in mental health is the Crisis Resolution Team Optimisation and Relapse Prevention (CORE) study.¹³ Through involvement of multiple stakeholder groups who prioritised and grouped 72 statements derived from various sources,¹⁴ their integrative approach enabled the development of a fidelity scale for crisis resolution teams, which was then piloted widely across UK,¹⁵ demonstrating feasibility and validity in its use. As such, using a concept mapping approach can ensure a more comprehensive understanding of what constitutes 'safety' from diverse perspectives and develop tailored, context-specific measures that reflect the unique features of DMHIs.

Further, involvement of people with lived experience is not only crucial during development, reporting and testing, but also in real-life implementation and monitoring of the intervention usage,¹² which would provide further reassurance of their safety, or identify needs for further improvement of safety. Additionally, exploration of the role of peer workers/coaches in the delivery and assessment of safety of DMHIs may be an avenue explored further in future

research. Adding these voices to the regulations around the safety of DMHIs may be especially important in the context of the increasing complexity of DMHIs, which may involve several end users (e.g. people experiencing mental ill-health, peer coaches, therapists, online peer community).

In line with this next step, it would also be important to include experts and end users from the Global South,¹⁶ as 85% of experts in the Taher et al study were based in the UK, with the remainder representing countries such as Canada, Australia and the Netherlands. Related to this point is also the need to make regulatory processes more accessible and affordable. As DMHIs developers, we are aware of the high costs and staffing demands associated with proper testing of such treatment approaches. Although national research funding bodies may fund such rigorous testing of a few lucky examples of DMHIs, to ensure the safety of a large majority of existing DMHIs, a different, stepped-level and potentially industry-driven, regulatory approach may be needed. Finally, it is essential to ensure that regulatory standards are proportionate and not unduly stringent for DMHIs, especially compared with pharmacological interventions. For pharmacological treatments, the risk of SAEs can be higher and conceptualised more clearly than in non-pharmacological trials. Although DMHIs do require high-quality safety monitoring, the potential for harm should not be overstated compared with pharmacological trials. Given that these two treatments options involve different risks, safety frameworks should reflect these differences to avoid creating an unfair imbalance in the quality standards between both treatment approaches.

In conclusion, it is evident that the issue of applying regulatory processes pertinent to pharmaceutical and medical devices to DMHIs warrants further dialog. To advance this dialog, considering many more views, including those using DMHIs, would be paramount. This may be especially critical, considering the fast-growing number of innovative DMHI that integrate artificial intelligence and related data collection methods (e.g. photos, posts on social media platforms). Further, although implementing the recommendations from the Delphi study has the potential to enhance the safety assessment processes for DMHIs, embedding these recommendations into existing frameworks such as World Health Organization mERA¹¹ checklist can present a useful step toward establishing formal regulatory methods for critical assessment and transparency in reporting of risks associated with digital technologies that aim to improve health outcomes across populations with mental ill-health. Improving the quality of mental health safety data, refining the risk mitigation processes and conducting more accurate risk-benefit analyses have the potential to lead to more effective and safer DMHIs, which, in the current digital era, have a strong potential to become a useful treatment partner across a wide range of mental health settings.

Urska Arnautovska , Health, Medicine and Behavioural Sciences, The University of Queensland, Brisbane, QLD, Australia; Metro South Addiction and Mental Health Service, Brisbane, QLD, Australia; and Queensland Centre for Mental Health Research, Wacol, QLD, Australia; **Alyssa Milton** , Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia; and Australian Research Council Centre of Excellence for Children and Families over the Life Course, Sydney, NSW, Australia

Correspondence: Urska Arnautovska. Email: u.arnautovska@uq.edu.au

First received 8 Jun 2024, final revision 21 Feb 2025, accepted 24 Feb 2025

Author contributions

U.A. drafted the original version of this editorial and provided references. A.M. provided substantial revisions and the addition of content and references. This editorial was finalised with contributions of both authors.

Funding

This study received no specific grant from any funding agency, commercial or not-for-profit sectors. A.M.'s research is supported partially by the Australian Government through the Australian Research Council's Centre of Excellence for Children and Families over the Life Course (project identifier: CE200100025), the Medical Research Future Fund Clinician Researchers Applied Research in Health (project identifier: MRF2032279) and the National Health and Medical Research Council (project identifier: 2025098).

Declaration of interest

None.

References

- 1 Torous J, Myrick K, Aguilera A. The need for a new generation of digital mental health tools to support more accessible, effective and equitable care. *World Psychiatry* 2023; **22**: 1.
- 2 Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone apps for the treatment of mental disorders: systematic review. *JMIR mHealth uHealth* 2020; **8**: e14897.
- 3 Fisher M, Etter K, Murray A, Ghiasi N, LaCross K, Ramsay I, et al. The effects of remote cognitive training combined with a mobile app intervention on psychosis: double-blind randomized controlled trial. *J Med Internet Res* 2023; **25**: e48634.
- 4 Milton A, Ozols AMI, Cassidy T, Jordan D, Brown E, Arnautovska U, et al. Co-production of a flexibly delivered relapse prevention tool to support self-management for long-term mental health conditions: a co-design and user-testing study. *JMIR Format Res* 2024; **8**: e49110.
- 5 Arnautovska U, Trott M, Vitangcol KJ, Milton A, Brown E, Warren N, et al. Efficacy of user self-led and human-supported digital health interventions for people with schizophrenia: a systematic review and meta-analysis. *Schizophr Bull* [Epub ahead of print] 27 Sep 2024. Available from: <https://doi.org/10.1093/schbul/sbae143>.
- 6 Alvarez-Jimenez M, Koval P, Schmaal L, Bendall S, O'Sullivan S, Cagliarini D, et al. The Horyzons project: a randomized controlled trial of a novel online social therapy to maintain treatment effects from specialist first-episode psychosis services. *World Psychiatry* 2021; **20**: 233–43.
- 7 Gumley AI, Bradstreet S, Ainsworth J, Allan S, Alvarez-Jimenez M, Aucott L, et al. The EMPOWER blended digital intervention for relapse prevention in schizophrenia: a feasibility cluster randomised controlled trial in Scotland and Australia. *Lancet Psychiatry* 2022; **9**: 477–86.
- 8 Taher R, Bhanushali P, Allan S, Alvarez-Jimenez M, Bolton H, Dennison L, et al. Bridging the gap from medical to psychological safety assessment: consensus study in a digital mental health. *BJPsych Open* 2024; **20**: e126.
- 9 Torous J, Firth J, Mueller N, Bendall S, Onnela JP, Baker JT. Methodology and reporting of mobile health and smartphone application studies for schizophrenia. *Harv Rev Psychiatry* 2017; **25**: 146.
- 10 Berry N, Machin M, Ainsworth J, Berry K, Edge D, Haddock G, et al. Developing a theory-informed smartphone app for early psychosis: learning points from a multidisciplinary collaboration. *Front Psychiatry* 2020; **11**: 602861.
- 11 Agarwal S, LeFevre AE, Lee J, L'engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ* 2016; **352**: i1174.
- 12 Palmer VJ, Banfield M. Omission of mental health lived-experience research in implementation research commission. *Lancet Psychiatry* 2024; **11**: 409.
- 13 Lloyd-Evans B, Christoforou M, Osborn D, Ambler G, Marston L, Lamb D, et al. Crisis resolution teams for people experiencing mental health crises: the CORE mixed-methods research programme including two RCTs. *Program Grants Appl Res* 2019; **7**: 1–102.
- 14 Lloyd-Evans B, Bond GR, Ruud T, Ivanecka A, Gray R, Osborn D, et al. Development of a measure of model fidelity for mental health Crisis Resolution Teams. *BMC Psychiatry* 2016; **16**: 1–12.
- 15 Lamb D, Milton A, Forsyth R, Lloyd-Evans B, Akther S, Fullarton K, et al. Implementation of a crisis resolution team service improvement programme: a qualitative study of the critical ingredients for success. *Int J Ment Health Syst* 2024; **18**: 18.
- 16 Kumar M. Championing equity, empowerment, and transformational leadership in (mental health) research partnerships: aligning collaborative work with the global development agenda. *Front Psychiatry* 2019; **10**: 421791.