REPORT

The rise of digital mental health: why are we not talking more about safeguarding?

Many frameworks for digital mental health technologies set out the principles of risk assessment and patient safety that should be incorporated into services, but little guidance is offered on what to do when digital service users move into acute distress.

An exploration of the key considerations.

togetherall.com

togetherall

Digital mental health creates questions about protecting at-risk users

The last five years have seen an explosion in the digital health market. COVID-19 accelerated the growth and adoption of tech-delivered health products on the global stage. In some countries, the pandemic led to an easing of the barriers to entry[i] for some technologies amid the urgency created by lockdown and increasing numbers of people seeking support.

Mental health makes up a large share of the digital health market. Such technologies help to expand the reach of support, ease barriers of access to support services and improve the user experience for people who need or seek help.

The growth in the mental health segment is hard to quantify. The market, broadly defined, spans everything from consumer products that improve wellness to e-health services and disruptive digital mental health technologies that aim to identify, diagnose and treat[ii]. According to ORCHA, 5 million people download a digital health app every day. Different reports estimate that there are between 10-20,000 mental health apps and digital products available, with the total market value of the **digital mental health technology ('DMHT')** sector set to rise significantly this decade.

Responding to the proliferation of DMHTs, law makers, regulators, accreditation organisations and research bodies have been looking at appropriate standards and guidance needed to keep pace with a changing market. According to ORCHA, some 62.4% of digital mental health apps fall below clinical effectiveness requirements.[iii]

While there has been much progress since the start of the pandemic on how we evaluate and even regulate these technologies, the large and disparate landscape for DHMTs comes with challenging questions. In a market that spans everything from consumer-facing apps to tools approved for clinical use, who can provide a single measure of assurance that can be understood by consumers and clinicians, one that is dynamic to different tiers of intervention?

Amid those challenges, a consideration that ought to be a primary concern, but has received too little attention, is **safeguarding people at risk**.

With a wide choice of digital mental health products, and the complexity of available pathways, there is a high likelihood that some people will experience significant distress while using DMHTs not *intended* for at-risk users.

- Where does liability lie when an individual accessing a mental health service presents as being in distress or crisis?
- Who holds ultimate responsibility for duty of care in different scenarios?
- What can we expect of the organisations providing DMHTs?

Looking at the existing policy landscape, it becomes clear that risk assessment and safety are important principles of DHMT design. But how providers should keep users safe is less clearly defined.

Digital mental health technologies: a large and confusing landscape

The DMHT landscape is confusing to navigate for commissioners, providers, clinicians and users. All parties need reassurance about safety, quality and ethics. This is not only important for emerging technologies such as AI[iv] that present unique safeguarding challenges, but also where the *online* delivery of traditional support models, such as therapy or peer support, creates challenges to safeguard at-risk individuals.

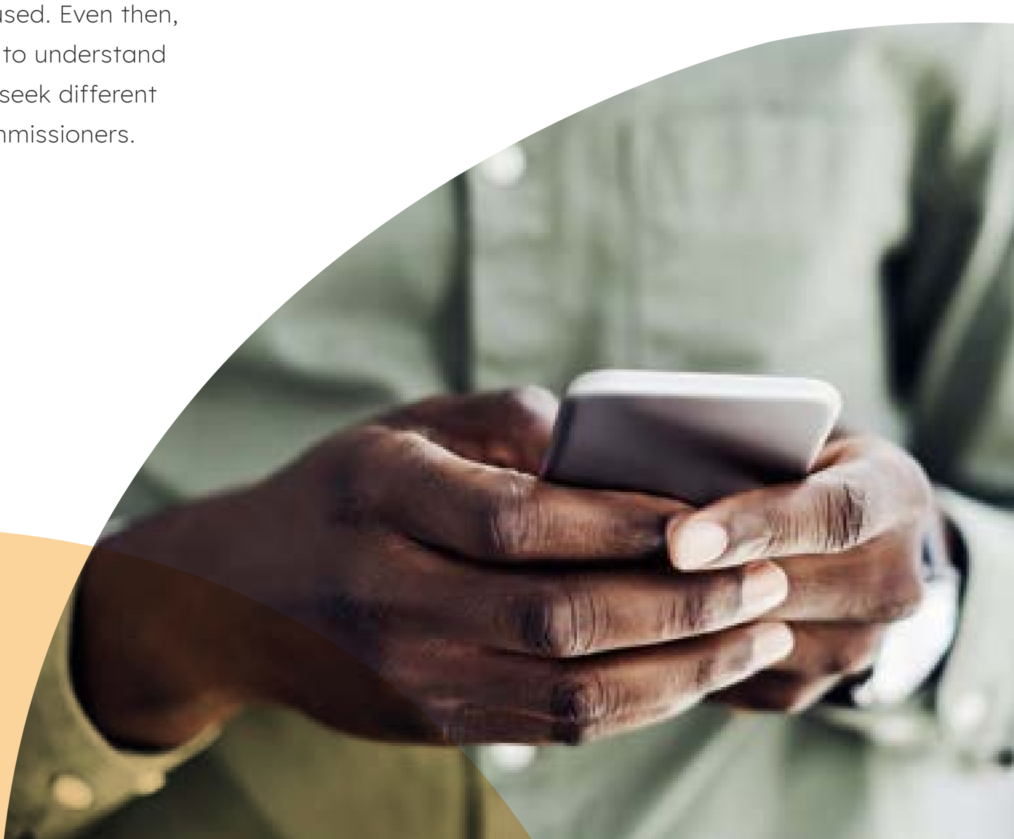
Part of the issue is the question of who pays for DMHTs. Creating trust and reassurance is a challenge when regulation is typically aligned to specific sectors (e.g. health), but the technology available covers everything from consumer products to clinically approved interventions.

Tools that are recognised and prescribed by clinicians naturally go through more stringent vetting, but the process of accreditation may differ depending on who commissions them, how they are funded and where in the country they are used. Even then, healthcare professionals who want to understand the suitability of a digital tool may seek different assurances to those sought by commissioners.

Outside of healthcare, many apps and tools are made available by educators, public health initiatives, local government, charities and insurance products. These organisations receive little in the way of standardised guidance on what reassurances they should seek and how to choose products that suit their population's needs.

Finally, there is the vast consumer-facing market with millions of apps available to the public through their phones. *"Five million people download a health app every single day,"* according to Liz Ashall-Payne, Chief Executive of ORCHA. *"Most people then go on and delete it within 24 hours. Why? Because they pick the wrong technology for them, and they didn't know how to find the right one."*[v]

Standards and assessments to create assurance are emerging, but the picture looks very different from country to country.



What's going on in the world of DMHT policy?

When it comes to digital mental health products, the last two years have produced no shortage of reading material. Many documents, from voluntary standards to evaluative frameworks and even regulation have been created to make the landscape clearer and easier to navigate.

Much of the policy accelerated since the COVID-19 pandemic has sought to embed common principles for digital mental health tools entering the market. While the detail and the enforceability of such principles differs between countries, the common themes include **governance, privacy, security, compliance, co-production, safety, efficacy and ethics.**

North America

The Food and Drug Administration (FDA) is the US federal regulator with a "public health responsibility to oversee the safety and effectiveness of medical devices – including mobile medical apps". Guidance in September 2022 was issued to clarify definitions over which mobile apps are a focus for the FDA. As is the case elsewhere in the world, *"the intended use of a mobile app"* is key to determining whether it falls under FDA oversight, and even then, enforcement is discretionary, based on level of the understood risk to the public[vi]. In short, many general wellbeing and DMHTs fall outside of either the scope or the focus for the FDA.

Meanwhile, The Amercian Psychiatric Association's App Evaluation Model allows psychiatrists to rate DMHTs as part of a five-part evaluation process. The criteria are consistent with other common principles of digital health standards and cover access, privacy/security, evidence, usability and data integration. The model makes it easier for psychiatrists to make more informed choices.

In addition, US-based One Mind PsyberGuide publishes assessments of mental health apps based on publicly available (non-verified) information. The criteria reviewed cover credibility, transparency, user experience and 'professional reviews'. It aims to build transparency into the process for different stakeholders but especially users.

In Canada, eMHIC has helped to develop an assessment framework for mental health apps to ensure quality across *"over 450 criteria based on seven standards: 1) data and privacy; 2) clinic evidence; 3) clinical safety; 4) usability and accessibility; 5) security and technical stability; 6) cultural safety, social responsibility and equity; and 7) enhanced data sovereignty."* [vii]

EU and continental Europe

Despite the relative ease of cross-border regulatory join up in Europe, the landscape of guidance and regulation differs from country to country.

An EY report in 2022 for the French government, assessing digital health standards across EU member states, Scotland and Norway, analysed the maturity of different areas of regulation for DHTs. It found after assessing 29 countries that issues of security, data privacy, data management and interoperability were covered by relatively mature and advanced standards, while **issues of liability, and ethics were far less likely to be included and covered in European regulations.**

Scandinavia, which aims to integrate its healthcare systems this decade, has recently launched NordDEC, a single digital health evaluation framework covering Sweden, Denmark, Finland, Norway and Iceland. It aims to *"evaluate and identify trusted digital health technologies within healthcare and preventive care. It also aims to provide product developers and owners with clear visibility on what good looks like in order to inform product development, market access strategies and commercial positioning."* [viii]

Germany has become the first country to create a federal-level system to approve digital health (including mental health) apps for prescription and reimbursement by insurance. The *Digitale Gesundheitsanwendung*, or 'DiGA', means that the 90% of the German population covered by insurance – and their healthcare professionals – can now use the DiGA website to search for DHTs assessed and pre-approved for clinical prescription. The key criteria covered by the assessment process, focuses on security and functional assurance, data security, data privacy, data interoperability. References to patient safety falls under a section titled 'Additional Quality Conditions' and regarding risk, references the need to set expectations for the insured patient.

England and Wales

Two standards (DCB 0129 and DCB 0160) have been in place in the NHS for some time. Recognised by English law, they outline processes to mitigate the potential for harm when digital products are used in healthcare and they set out the process and responsibilities for assessing risk. Building on this, the NHS' Digital Technology Assessment Criteria (DTAC) is designed to give everyone – staff, patients and commissioners – assurance that DHTs meet national standards in five key areas – the first of which is clinical safety.

NICE continually works to update its Evidence Standards Framework (ESF), developed before the pandemic for use by both commissioners. It is designed to cover a broad spectrum of DHTs for use in healthcare, arranged in three tiers, which can broadly be summarised as a) technologies that create efficiencies b) technologies for patient self-care c) technologies relating to treatment or diagnosis or monitoring. Building on NICE's role in Health Technology Assessment (HTA), the institute is also trialling a new approach to help DTAC-compliant technologies generate evidence earlier and more easily through an initiative called the Early Value Assessment.

UK-headquartered, but increasingly international, The Organisation for the Review of Care and Health Apps (ORCHA) specialises in DHT evaluation and helps drive awareness of accredited tools through pre-assessed app libraries that help build transparency, trust and assurance. Its core assessment builds on DTAC standards and other international frameworks.

Despite all this work, the regulatory landscape for digital mental health remains unclear. The government has acknowledged the regulatory challenges and the MHRA and NICE are now working together to produce guidance on 'risk-proportionate regulation of digital mental health products'[ix].

Australia and New Zealand

In New Zealand, the Digital Mental Health & Addiction Tool (DMHAT)[x] is an assessment framework to support DMHTs meet acceptable quality standards. As with Canada, this has been developed in collaboration with eMHIC. The guidelines are specific to mental health and focus on clinical safety as a core principle.

The Australian Commission on Safety and Quality in Health Care developed the National Safety and Quality Digital Mental Health Standards (NSQDMH), which was launched in November 2020. Accreditation of service providers begins in late 2022. As well as having established some of the most advanced accreditation processes for digital mental health, specifically in relation to at-risk users, the Australian standards appear to set out the clearest language and expectations for DMHTs to recognise and respond to users experiencing acute distress:

> **"Serious adverse events may be preceded by changes in a person's behaviour or mood that can indicate a deterioration in their mental state. Early identification of deterioration may improve outcomes but can be more difficult in a digital setting. However, digital services should not mean a higher level of risk. A systematic approach to recognising deterioration early and responding to it appropriately is therefore required, noting that the response may include calling for emergency assistance internally or via external emergency response systems."**
>
> **Australian Commission on Safety and Quality in Health Care, 'National Safety and Quality Digital Mental Health Standards' 2020**

Focusing our attention on safeguarding

Risk monitoring and management is a core component of good practice in mental health services, whether provided in person, online, within healthcare or not. Distress is dynamic and can be affected by circumstances that can change over the briefest of timeframes. Policy makers, and even private companies, are moving quickly to address the need for clearer standards in digital mental health. We have seen that this is challenged by a complex ecosystem of different technologies, sectors, buyers and support needs.

A 'Global Governance Toolkit for Digital Mental Health' developed by the World Economic Forum and Deloitte in 2021 identified 'harm to patient though malfunction' and 'misleading content' as important safety risk factors to consider in DMHT design. The NHS's 'Digital Clinical Safety Strategy'[xii] discusses risk management and issues around data interoperability.

There is, however, a difference between the concept of 'do no harm' in your service and actively intervening based on best practice when it is appropriate and risk has been identified.

Few reports and frameworks spell out the issues relating to the duty of care that DMHTs or their commissioners have in relation to service users at risk. We have seen above that the Australian NSQDMH goes furthest to highlight these issues. In the UK, NICE guidelines (for certain DHTs, where communication with other individuals is facilitated via that DHT), say that "appropriate safeguarding measures are in place around peer-support and other communication functions within the platform."

It asks providers to:

> **"Describe who has access to the platform and their roles within the platform. Describe why these people or groups are suitable and qualified to have access. Describe any measures in place to ensure safety in peer-to-peer communication, for example through user agreements or moderation."**
>
> **(NICE, Evidence Standards Framework for digital health technologies, 2019)**

For digital mental health providers, predicting, identifying and responding to risk episodes presents challenges that are unique to digital delivery. Users can be anonymous, their location is not always known, and presentation 'cues' that are easier to detect face-to-face are not present online. That said, as Martinez and Farahan have pointed out[xiii], advances in research of using 'digital phenotyping' may offer ways to more accurately spot risk factors before people reach a crisis point.

Much of the guidance we have is based on the *intended* design of the DMHT: who is the technology meant to be used by and for what purpose? While certain DHTs may be *intended* for use with low-risk groups, the complex landscape and many pathways through which apps and services can be accessed, means that access cannot always be gatekept by an appropriately qualified person. Indeed, to have access narrowly restricted this way would also be counter to the goals of widening access and population-level support that many digital tools are designed to address.

And then there are apps available via app stores. Research conducted before the pandemic on a search of 2690 relevant apps and a systematic assessment of 69 of them found that:

"Non-existent or inaccurate suicide crisis helpline phone numbers were provided by mental health apps downloaded more than 2 million times. Only five out of 69 depression and suicide prevention apps offered all six evidence-based suicide prevention strategies. This demonstrates a failure of Apple and Google app stores, and the health app industry in self-governance, and quality and safety assurance." [xiv]

Two years later, Parrish *et al.* (2021) found in a review of 35 relevant and widely used apps that only a third of them provided in-app crisis resources. The research concluded that to address the inconsistency, **crisis language should be included as part of app evaluation frameworks** and internationally accessible, vetted resources should be provided to app users. [xv]

Since the guidance and frameworks, which steer how newer DMHTs might address individuals in crisis, remains underdeveloped, what best practice can we look at for managing risk episodes?

Let's look at how we approach this at Togetherall.

Togetherall – a case study in online safeguarding

Togetherall is not new. It is one of the longest serving digital mental health services that is widely available. The idea is simple. Among other helpful tools, we provide an online space for safe and anonymous peer support, 24/7. We've been doing this since our inception in 2007 and, in that time, not only have we seen the rapid entry of DMHTs into the market, and the establishment of new policy in response, but we have also learned a great deal about keeping people safe.

Our core activity – to facilitate a place where people can gain and give support with peers, instantly anonymously and safely – is not subject to regulation. However, we do follow all the guidance created in recent years and have put our service through evaluations including ORCHA. That said, long ago, we developed our own safeguarding framework and protocols. We did this based on a 'duty of care,' an established principle and obligation in clinical practice to avoid foreseeable harm to service users. Long before the digital mental health wave, our clinical team designed our way of working to reflect what they had learned in clinical practice: to identify, monitor, intervene appropriately and resolve.

Fortunately, the overall percentage of our users who experience a crisis episode is relatively low. Indeed, our service is not *intended* for people experiencing or likely to move into acute distress. But what we know, both from our clinical experience and from working with the users of Togetherall – our community members – is that DMH providers/ services must plan for the potential of risk.

Here are some examples of cases where we monitored and then worked with the members to offer support.

WARNING:
The following contains content and themes that may be distressing

1. Safeguarding

- A platform post mentioning a relative who was intoxicated and unable to attend to their child.
- A young person reporting that their mother was being emotionally abusive, leading to suicidal thoughts.

2. Self-harm or risk of suicide

- A member posting from a location where they were at high risk of suicide.
- A member posting about having cut their wrists.
- A member saying they have taken an extremely high quantity of painkillers. Another who had a bottle of bleach by their bed.

3. Domestic violence

- A post mentioning violence against an elderly relative suffering with dementia.
- A member posting that their partner is hitting them, and they feel trapped but can't leave.

Clinical philosophy

Togetherall can be thought of as ‘people helping people, scaled by technology, monitored by clinicians’. Our clinical staff monitor, nurture, facilitate and intervene with the goal of fostering a safe community of peers who are focused on helping each other with mental health concerns.

Because we believe in the power of peer support, we aim to monitor, observe and intervene only when needed, and as lightly as possible, to keep the focus on the community while also ensuring safeguarding capacity when needed. We seek to empower our members to give and take in a healthy, supportive, honest and anonymous manner. We manifest this philosophy by carefully training our clinical staff to act as a unified shaping/caring force in the community.

Clinical structure

Togetherall employs a large multidisciplinary clinical team of licensed/registered/professionally accredited mental health professionals including social workers, counsellors, nurses, psychologists and psychiatrists. Each staff member is painstakingly recruited, evaluated, and trained; they receive months of hands-on guidance and supervision to ensure consistent and high-quality practice.

- **Our ‘Wall Guides’** interact with members anonymously, routine low-risk community management, monitoring, and signposting to customised resources.
- **Lead ‘Wall Guides’** are responsible for delegating work and reviewing/actioning potential risk on the platform.
- **Senior Clinical Team** provide overall leadership per shift, guidance to LWGs/WGs, and are responsible for handling external communications related to member safety and crisis management.

Clinical practices

Togetherall's clinical philosophy and structure provide the foundation for our clinical practices in the community. Using their professional training, experience and ethics drawn from mental health practice, Togetherall clinical staff shape the community using a range of tools including:

- Removing/editing identifying information to ensure anonymity.
- Adjusting language in community content.
- Hiding content for review and editing.
- Monitoring all activity surrounding a particular post.
- Contributing to the community by sharing information, sparking discussion, or contributing to discussions. This activity contributes to a vibrant community and lets members know that "we've got this" in the event that they are concerned about another member.
- Messaging with the member to clarify or adjust content to balance their intent and community rules.
- Messaging with a member with concern and/or resources related to risk.

When it is determined that a member is in crisis, at risk, in need of more intensive support, or unable to follow the community rules, clinical staff have a range of interventions they can use. One of which is to 'escalate' the case.

Risk escalation

Partners who work with Togetherall – in health, education and other sectors – expect us to be able to take care of their populations who choose to join Togetherall for peer support. We have an established pathway to identify and refer those in imminent risk scenarios. Our philosophy is to make access to Togetherall as easy as possible while also ensuring we gather enough information to safeguard members.

In the event that we have identified a situation of serious risk, our clinical team can:

- work with the individual and de-escalate the situation
- externally escalate off platform based on agreed and locally specific protocols

While just less than 1% members on Togetherall experience a risk episode requiring clinical team escalation, all members are monitored 24/7. In 2022, the content of 40% of members was actively reviewed for moderation based on content which may have indicated potential risk.

Embedding the principle of safeguarding by design

A diverse marketplace of accessible DMHTs effectively guarantees that a percentage of users will be or become 'in-crisis', regardless of a DMHT's intended application. Therefore, Togetherall believes that greater attention must be paid to ensuring professional safeguarding of those at risk when using digital mental health services.

While efforts are made to ensure that DMHTs cause no unintended harm, these services must also be proactively engineered to integrate robust, responsive and professional safeguarding to attend to the percentage of users at risk, as well as the percentage of users whose mental health will deteriorate while using the application. These facts are well known to mental health providers and they must be fully considered by the developers of DMHTs.

While DMHTs vary in their design, complexity and operational capacity, we believe that DMHTs can, and must, do more to integrate the principles and practices of crisis intervention in order to protect all users. We offer the following recommendations for those considering a DMHT.

DMHT providers should transparently, precisely and clearly describe how they provide monitoring, oversight and safeguarding.
Use of vague terms, such as 'clinical monitoring' or 'monitoring with AI' without specific details, should be viewed as unacceptable. DMHT providers should provide a comprehensive and specific description of how users are monitored for risk, how those users are supported through intervention, and details regarding the mental health professionals and practices used to ensure that risk is fully managed.

Buyers of digital mental health services and products from all sectors (health, government, education, insurance and private companies) should ask detailed questions if a seller does not outline the above.

Because the DMHT marketplace is largely unregulated, buyers should not assume that commonly used terms carry universal meaning and they should carry out due diligence on the process and policies for monitoring and safeguarding. This can include the following:

- A recent job advertisement for those providing clinical monitoring. We recommend looking for baseline requirements such as education, experience, and licensure/registration.
- Evaluate whether routine work is being done by volunteers or non-professionals.
- The credentials of staff members providing coverage or oversight.
- A schedule of coverage provided by mental-health professionals.
- A mental-health leadership structure that provides supervision, training, and oversight.
- Details about how AI and technology is implemented in risk identification and management. Explore whether AI and automated tools may be implemented without professional checks and balances.
- Enquire whether the DMHT assumes a duty of care when a user is identified as 'at risk'.
- Enquire whether the DMHT provides passive resourcing to members (e.g., a message that relies on their action) or whether the DMHT professionals directly engage with the member.
- Enquire how quickly the DMHT can detect risk and how quickly can a professional support the individual at risk?
- Enquire how quickly mental health professionals working for the DMHT are able to respond to requests from users.
- If a service makes promises regarding multiple languages, ask for evidence demonstrating how complex technologies (e.g., AI) have been translated, the schedules/availability/response times of mental health providers in each language, and reliance on translation lines/technologies in lieu of bilingual individuals.

References

i) Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry. 2021;20(3):318-335. doi:10.1002/wps.20883

ii) World Economic Forum (2022). Governance Frameworks in Digital Mental Health. [online] WEF. Available at: https://www.weforum.org/whitepapers/global-governance-frameworks-in-digital-mental-health.

iii) ORCHA (2022). The people's view of digital in NHS mental health support: UK population attitudes and behaviour report. [online] ORCHA. Available at: https://info.orchahealth.com/digital-for-mental-health-attitudes-and-behaviour-report.

iv) Koutsouleris, N., Hauser, T.U., Skvortsova, V. and Choudhury, M.D. (2022) From promise to practice: towards the realisation of AI-informed mental health care. The Lancet Digital Health. [online]. 4 (11), pp.e829–e840. Available from: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(22)00153-4/fulltext [Accessed 16 November 2022].

v) Cellan-Jones, R. Health apps - helpful or harmful? [online] rorycellanjones.substack.com. Available at: https://rorycellanjones.substack.com/p/health-apps-helpful-or-harmful [Accessed 22 Nov. 2022].

vi) Center for Devices and Radiological Health (2019). Policy for Device Software Functions and Mobile Medical Applications. [online] U.S. Food and Drug Administration. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications.

vii) World Economic Forum (2022). Governance Frameworks in Digital Mental Health. [online] WEF. Available at: https://www.weforum.org/whitepapers/global-governance-frameworks-in-digital-mental-health.

viii) NordDEC. https://norddec.org/. [online] Available at: https://norddec.org/ [Accessed 22 Nov. 2022].

ix) GOV.UK. Mental health funding of £1.8m welcomed by MHRA and NICE to explore regulation of digital mental health tools. [online] Available at: https://www.gov.uk/government/news/mental-health-funding-of-18m-welcomed-by-mhra-and-nice-to-explore-regulation-of-digital-mental-health-tools.

x) emhf.emhicglobal.com. The Guide – DMHAT. [online] Available at: https://emhf.emhicglobal.com/guide/ [Accessed 22 Nov. 2022].

xi) Assessment and management of risk of patients causing harm RCPsych https://www.rcpsych.ac.uk/members/supporting-you/assessing-and- managing-risk-of-patients-causing-harm Accessed June 2021

xii) NHS Transformation Directorate. Digital Clinical Safety Strategy. [online] Available at: https://transform.england.nhs.uk/key-tools-and-info/digital-clinical-safety-strategy/.

xiii) Martinez, C. & Farhan, I. Making the right choices, Reform, July 2019

xiv) Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. BMC Med. 2019;17(1):231. Published 2019 Dec 19. doi:10.1186/s12916-019-1461-z

xv) Parrish EM, Filip TF, Torous J, Nebeker C, Moore RC, Depp CA. Are Mental Health Apps Adequately Equipped to Handle Users in Crisis? Crisis. 2021 May 27. doi: 10.1027/0227-5910/a000785.

togetherall

CONTACT US

info@togetherall.com

togetherall.com