RESPONSIBLE DIGITAL HEALTH

EDITED BY: Dorian Peters, Naseem Ahmadpour, Geke Ludden and Karina Vold
PUBLISHED IN: Frontiers in Digital Health

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-769-6
DOI 10.3389/978-2-88974-769-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

RESPONSIBLE DIGITAL HEALTH

Topic Editors:
Dorian Peters, University of Cambridge, United Kingdom
Naseem Ahmadpour, The University of Sydney, Australia
Geke Ludden, University of Twente, Netherlands
Karina Vold, University of Toronto, Canada

Dr. Geke Ludden is currently working on a project funded by Philips VitalMinds. All other Topic Editors declare no conflicts of interest.

Citation: Peters, D., Ahmadpour, N., Ludden, G., Vold, K., eds. (2022). Responsible Digital Health. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-769-6

Table of Contents

frontiers
in Digital Health

EDITORIAL
published: 21 January 2022
doi: 10.3389/fdgth.2021.841477

Editorial: Responsible Digital Health

Naseem Ahmadpour[1]*, *Geke Ludden*[2], *Dorian Peters*[3] *and Karina Vold*[4]

[1] *Affective Interactions Lab, School of Architecture, Design and Planning, The University of Sydney, Sydney, NSW, Australia,* [2] *Department of Design, Production and Management, University of Twente, Enschede, Netherlands,* [3] *Dyson School of Design Engineering, Imperial College London, London, United Kingdom,* [4] *Institute for the History and Philosophy of Science and Technology, University of Toronto, Toronto, ON, Canada*

Keywords: digital health (eHealth), design methodology, ethics, technology - ICT, health, well-being

Editorial on the Research Topic

Responsible Digital Health

The growing concern over the ethical implications of digital technology used for health has been amplified by the emergency deployment of technologies in an effort to manage a global pandemic. These events have placed even greater urgency on the need for attention to ethical impacts and value fulfillment, and on the need for advances in responsible digital health research and practice.

Furthermore, given that healthcare practitioners are expected to abide by ethical principles that protect the rights and welfare of their patients, we believe that the technologies functioning as tools and agents of well-being and healthcare provision, should be held to the same account. And indeed, an increasing number of researchers are working to ensure that they are. But in order to make progress toward more responsible practice within digital health, we need more systematic approaches, more research into the ethical implications of digital technology use for health, and more guiding examples of responsible practice in this area. The research article collection described herein responds directly to this need.

For the purposes of selection for this special topic, we considered "Responsible Digital Health" to include any intentional systematic effort designed to increase the likelihood of a digital health technology developed through ethical decision making, being socially responsible and aligned with the values and well-being of those impacted by it.

The papers included reflect a number of angles on the topic and reveal research insights on: issues of **equity** (who gets to be healthy?), the impact of **modality** (the unique promises and risks of particular technologies, such as chatbots) and the need for **process** (including frameworks, guidelines and approaches that can contribute to systematic and replicable best practice).

EQUITY—WHO GETS TO BE HEALTHY?

Digital health often has the potential to particularly serve vulnerable populations, so preventing these technologies from doing harm is both a critical research problem and a moral obligation facing designers and technologists. Protecting and empowering the people who use digital health technologies often requires users' involvement in design, as well as addressing issues of autonomy-support, justice, and equitable access. Faber et al. address these issues in both topic and method in "Attitudes toward health, healthcare, and eHealth of people with a low socioeconomic status: a community-based participatory approach." Through a participatory approach, the authors explored the attitudes in low Socio-Economic Status (SES) communities toward health, healthcare, and ehealth interventions. Their findings highlight that negative health attitudes are complex and underlined by a range of attitudes like *encumbered* toward health, feeling *disadvantaged* within healthcare, and *hesitance* toward eHealth adoption.

OPEN ACCESS

Edited and reviewed by:
Giovanni Ferrara,
University of Alberta, Canada

Correspondence:
Naseem Ahmadpour
naseem.ahmadpour@sydney.edu.au

Specialty section:
This article was submitted to Health Technology Innovation, a section of the journal Frontiers in Digital Health

Received: *22 December 2021*
Accepted: *28 December 2021*
Published: *21 January 2022*

Citation:
Ahmadpour N, Ludden G, Peters D and Vold K (2022) Editorial: Responsible Digital Health. Front. Digit. Health 3:841477. doi: 10.3389/fdgth.2021.841477

Moreover, there are challenges and opportunities particular to young people with respect to digital health. Wies et al. report on a scoping review that they conducted in order to map the landscape of emerging ethical challenges related to this dually vulnerable population. Their paper "Digital Mental Health for Young People: A Scoping Review of Ethical Promises and Challenges" reveals both the significant promises for youth mental health (e.g., reducing stigma and suffering, while improving access and well-being) as well as the real challenges in delivering on these. They argue that some of the ethical challenges that are raised around the use of digital health devices, such as challenges related to privacy, equality of access, and patient autonomy, may be exacerbated when used by adolescents, as youth are particularly vulnerable and are often below the age of consent for medical treatments.

Similarly, additional reviews within the mental health space provide evidence for both efficacy and for gaps with ethical implications. For example, in a review of the landscape of mobile apps for digital mental health in Spanish, Oñate Muñoz et al. reveal that, while thoughtfully designed apps could hold the key for reducing mental health disparities among Spanish-speakers in the United States, currently available technologies are inadequate.

MODALITY—TECHNOLOGY ITSELF MATTERS

Digital health technologies encompass the full gamut of modalities, from apps and wearables to data-driven tracking systems, robotic caregivers, telemedicine, Virtual/Augmented Reality (VR/AR), and chatbots. Therefore, research is needed that identifies the ethical implications specific to the use of these different technological approaches for health.

Christoforakos et al. interrogate the impacts of the anthropomorphisation of conversational chatbots on aspects of human experience such as a sense of connectedness with the bot, and implications for human-human interaction. While they found that both regular interaction with the chatbot and a design that facilitates perceptions of anthropomorphism and social presence can foster feelings of social connectedness, they emphasize that the decision to use anthropomorphic technologies should be taken responsibly and may be context dependent.

Vilaza and McCashin provide further insight into chatbot use in their paper, "Is the Automation of Digital Mental Health Ethical? Applying an Ethical Framework to Chatbots for Cognitive Behavior Therapy." They argue that ethical thinking should be at the core of Artificial Intelligence Cognitive Behavioral Therapy (AI-CBT) design, research and policy, and they also provide a critical overview and framework for assessing the ethical automation of digital mental health therapy.

Roossien et al. shed light on the pros and cons of sensor and intervention technologies for workplace health promotion, in their paper, "Ethics in Design and Implementation of Technologies for Workplace Health Promotion: a Call for Discussion" Through reviewing two cases, they investigated ethical issues, particularly privacy and autonomy, in relation to health technologies for aging workers and draw on challenges of developing and implementing technologies for an aging workforce. The findings reveal how sensors and interventions, so commonly applied to health promotion, can pose significant threats to the autonomy and privacy of workers. To mitigate these consequences, Roossien et al. propose careful consideration of diverse values and perceptions, and to situate those within the responsibilities of workers and employers at the workplace.

Finally, van Lotringen et al. target the affordances and limitations of text. Their paper, "Responsible Relations: A Systematic Scoping Review of the Therapeutic Alliance in Text-Based Digital Psychotherapy" investigates whether important qualities of the therapist-client relationship can be effectively preserved within the constraints of text-only conversational environments.

PROCESS—SYSTEMATIC, RIGOROUS, AND REPLICABLE

To create digital health responsibly, we need evidence-based principles, methods, and processes for anticipating and addressing the ethical impacts that technologies have on individuals and society. These often include impacts on core values and rights, such as well-being, autonomy, privacy, and justice.

For example, in "Designing Informed Consent for Digital Health Research: Applying the Digital Health Checklist and Readability Tools to Support Accessible Content," authors Nebeker et al. provide practical guidance and tools for improving informed consent for digital health research. The work of Vilaza and McCashin, mentioned above, also includes an ethical framework for assessing the use of automation for the delivery of online cognitive behavioral therapy.

While frameworks and standardized processes are arguably critical to efforts toward responsible digital health, we should not let the clarity they provide obscure the complexity of the issues involved. For example, in "From General Principles to Procedural Values: Responsible Digital Health Meets Public Health Ethics, Nyrup proposes a move away from "principlist" approaches to a procedural approach, as modeled by the "accountability for reasonableness" (A4R) approach that has been influential in public health ethics. Nyrup argues that procedural approaches can overcome some of the commonly pointed out limitations of principlist approaches, for example, by highlighting rather than masking disagreements and by providing guidance on how to resolve trade-offs between different competing values.

Furthermore, in the paper by Roossien et al. (2021) mentioned above, the authors lead with an acknowledgment that the ethical implications of workplace health represent "a neglected topic and such a complex field of study that we cannot come up with solutions easily or quickly." Their study is presented, not as an answer, but as a call for discussion. They also demonstrate a context-specific approach to investigating the ethics of workplace health interventions and argue that values such as privacy and autonomy cannot be isolated from other contextual elements as there is an inescapable "interplay between these values, the work context, and the responsibilities of workers and employers."

CONCLUSION

The work collected for this Research Topic presents current research insights, methods, tools, and examples of best practices that can inform responsible innovation and ethical practice in the design of digital health. It shows that while we are far from completely understanding how to responsibly design digital health services and technologies, we have an active and multidisciplinary community that can work together to advance knowledge on a responsible and sustainable future for our health and healthcare systems. To that end, we call upon researchers to engage in active discussion to enhance the diversity of views in digital health (both among researchers and those represented through research). Additionally, the complexity of researching ethics in digital health suggests there is a need for effective collaboration across disciplines to bring plurality to research and practice. The diversity of disciplines represented in this collection, ranging from design, human-computer interaction, philosophy, medicine and more, demonstrates a promising potential. We hope to see more collaborations across disciplines in the future.

AUTHOR CONTRIBUTIONS

DP wrote the outline of the editorial with KV, NA, and GL contributing for the papers they were responsible editor for. All authors discussed the contribution of the editorial and NA wrote the conclusion. All authors reviewed and edited the editorial.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Digital Health

REVIEW
published: 15 February 2021
doi: 10.3389/fdgth.2021.615366

Marketplace and Literature Review of Spanish Language Mental Health Apps

Alma Oñate Muñoz[1], *Erica Camacho*[2] *and John Torous*[1,2]*

[1] Harvard Medical School, Boston, MA, United States, [2] Beth Israel Deaconess Medical Center, Boston, MA, United States

Language differences between patients and providers remains a barrier to accessing health care, especially mental health services. One potential solution to reduce inequities for patients that speak different languages and improve their access to care is through the delivery of healthcare through mobile technology. Given that the Latinx community serves as the largest ethnic minority in the United States, this two-phased review examines Spanish app development, feasibility and efficacy. Phase 1 explored the commercial marketplace for apps available in Spanish, while phase 2 involved a literature review of published research centered around the creation, functions, and usability of these apps using the PubMed and Google Scholar electronic databases. Of the apps available on the database, only 14.5% of them had Spanish operability. The literature search uncovered 629 results, of which 12 research articles that tested or described 10 apps met the inclusion criteria. Of the 10 apps studied in this literature review, only four apps were translated to Spanish. Our study reveals that despite increasing interest in Spanish-language apps to address mental health, the commercial marketplace is not currently meeting the demand.

Keywords: mHealth, culture, Spanish, apps, technology

Edited by:
Dorian Peters,
University of Cambridge,
United Kingdom

Reviewed by:
Eduardo Bunge,
Palo Alto University, United States
MarySue Heilemann,
UCLA School of Nursing,
United States

****Correspondence:***
John Torous
Jtorous@bidmc.harvard.edu

Specialty section:
This article was submitted to
Health Technology Innovation,
a section of the journal
Frontiers in Digital Health

Received: *08 October 2020*
Accepted: *15 January 2021*
Published: *15 February 2021*

Citation:
Muñoz AO, Camacho E and Torous J
(2021) Marketplace and Literature
Review of Spanish Language Mental
Health Apps.
Front. Digit. Health 3:615366.
doi: 10.3389/fdgth.2021.615366

INTRODUCTION

Latinos are the largest ethnic minority in the United States and currently account for 18.5% of the US population (1). It is estimated that 21.3% or nearly 13 million Latinx individuals have at least one mental illness (1, 2). Yet, despite the great need for assessment and treatment of mental health illness in the Latinx population, data shows that only 9.6% of them accessed any mental health services (2). There is a growing body of literature that highlights the barriers to accessing mental health care, which includes low rates of insurance coverage, legal status, stigma, and socioeconomic factors. Language is a salient factor in accessing mental health care services. A study in 2013 showed that 32% or 15.7 million Latinx individuals report speaking English less than "very well." (3) Language differences between patients and providers have been shown to be detrimental for effective communication, which can lead to lower quality of care, and poorer outcomes (4, 5). The imbalance between demand and supply is aggravated by the fact that only an estimated 4.0% of psychiatrists are Latinx (6).

One potential solution to reduce inequities for Spanish-speaking patients and improve access to care is through the delivery of healthcare through mobile health care (*m*Health). Through the use of apps via mobile devices, patients can access a wealth of resources to improve their mental health. This change requires both access to a smartphone with internet capabilities and a desire of Latinx to use *m*Health; both are true today. Research shows that 80% of Latinx individuals have access to the internet via a mobile device, thus making delivery of *m*Health feasible (7, 8). Latinx individuals want *m*Health as demonstrated in a 2016 study where over 85% of participants reported interest in using mobile apps to improve their health and at least a quarter stated they would use a mobile app for mental wellness (9).

Despite the demand for more mHealth for Spanish-speaking populations, it is unclear whether the commercial marketplace has met the increasing need and the extent of published literature on Spanish app development, feasibility and efficacy are unknown. The objectives of this U.S. app store marketplace and literature review are to in phase (1) review the number and characteristics of Spanish mental health apps available on the commercial market; in phase (2) review existing literature on the development, translation, and cultural adaptation of mental health apps to Spanish-language/culture as well as the feasibility and efficacy of these apps in the Spanish-speaking population; (3) synthesize the results from the first two phases to identify key challenges, opportunities, and recommendations for development, translation, and cultural adaptation of Spanish-language mental health apps.

TABLE 1 | Characteristics of mental health mobile apps in Spanish.

	Characteristic	n	%
	Number of Apps	32	100%
Platforms	iOS	31	96.9%
	Android	27	84.4%
	Companion website for App	8	25.0%
Supported conditions	Stress & anxiety	16	50.0%
	Mood disorders	14	43.8%
	Sleep	9	28.1%
	Phobias	6	18.8%
	Eating disorders	6	18.8%
	OCD	5	15.6%
	Personality disorders	5	15.6%
	Schizophrenia	4	12.5%
Engagement	Audio/music/scripts	16	50.0%
	Gamification	14	43.8%
	Videos	9	28.1%
Connection to other services	Link to formal care/coaching	10	31.3%
	Crisis management feature	8	25.0%

METHODS

Phase 1

This review was conducted in two phases. Phase 1 involved searching a database of commercial market for apps available in Spanish, while phase 2 consisted of a literature review of published research centered around the creation, functions, and usability of these apps. To determine the commercial market's availability of mental health apps in Spanish, we utilized the Mobile Health Index and Navigation (MIND) database published by The Division of Digital Psychiatry at BIDMC (available at apps.digitalpsych.org). The database currently reports the existence of 220 mental health apps in the commercial market and collects data on 105 objective questions set forth by the APA for each registered mental health app. It is a useful resource for this phase as it is the largest database of mental health apps supported by peer reviewed evidence, publicly accessible, and allows users to filter through apps based on personal preferences and priorities (10). We utilized this database and filtered results by "Spanish Functionality" to find all the apps that are currently available in Spanish and collected basic characteristics of these apps, including platform availabilities, supported conditions, engagement features, connection to other services, and supporting study availability, see **Table 1**.

Phase 2

To understand the current state of mental health apps in Spanish, we created a searchable question to identify Spanish-language app development, feasibility, and efficacy. The developed search strategy was conducted on two electronic databases: PubMed and Google Scholar. Major themes searched included Spanish, Spanish-language, Spanish-speaking, Hispanic, and Latinx combined with health and mental health synonyms. This was joined to a list of applicable terms for the type of technology utilized, including mobile devices, smartphones, and apps. Given the fast-paced changes in technology, only articles published after January 2015 were included. To be included, papers had to meet the following criteria: (1) feature a Spanish-language mobile app, (2) describe the development, feasibility, or interventional approach of a mobile app, and (3) the app must address at least one mental health disorder including, but not limited to depression, anxiety, substance use disorder, and/or eating disorders. Articles were excluded if (1) published before January 2015, and (2) app does not address any mental health disorders.

The search revealed 629 articles. A review of the 629 citations/abstracts was conducted manually. The full text was considered for papers with abstracts unavailable. This initial screening resulted in 49 articles that met the preliminary inclusion criteria. A snowball approach was used to ensure the literature search was comprehensive. This involved reviewing papers that cited the 49 articles, searching for previous papers written by the lead authors, and utilizing the "related articles" feature on each database. Following a full-text review of the final

---

Abbreviations: *m*Health, mobile health care; apps, mobile applications; AOD, alcohol and other drug use.

set of papers, 12 met the comprehensive criteria and are reviewed here. There are no conflicts of interest identified by the authors of this study. All authors certify responsibility for the manuscript.

RESULTS

Phase 1

A review of the Division of Digital Psychiatry's mHealtgh App Navigation Database (MIND) revealed that of 220 mental health apps available on the commercial market, 32 (14.5%) offer Spanish operability (10). Of these, 31 (96.9%) are available on iOS and 27 (84.4%) are available on Android. The most common supported conditions were stress and anxiety ($n = 16$, 50.0%) and mood disorders ($n = 14$, 43.8%). Significant overlap was noted with 12 (37.5%) apps supporting both mood and stress and anxiety conditions. Of the 32 apps, 10 (31.3%) have supporting studies published (11–32). However, a review of these studies revealed that none of them focused on Spanish-speaking populations. Further, none of the apps on the marketplace found in Phase 1 were found in the app mentioned in the literature review of Phase 2 as outlined below. **Table 1** summarizes the most salient characteristics of these apps.

Phase 2

The literature search revealed 12 research articles that tested or described 10 apps. Among the 10 apps, only four distinct mental health disorders were addressed. Of the 12 articles, six described app development and/or cultural adaptation, three studied feasibility, and three described interventional studies. Most articles ($n = 8$, 66.7%) focused on Spanish-speaking Latino adults in the United States. The rest of the articles focused on Spanish-speaking individuals in Peru, Colombia, and Australia. Of the 10 apps described or studied, half ($n = 5$, 50.0%) aimed to address depressive symptoms. **Table 2** displays the article and app characteristics including articles' research design and study population, as well as the apps' supported conditions. Of note, control apps not studied or reported on were not analyzed in **Table 2**.

TABLE 2 | Summary metrics of studies on mental health apps in Spanish.

		Studies	%
Journal article characteristics ($n = 12$)			
Research design	App development protocol	6 (33–38)	50.0%
	Feasibility	3 (39–41)	25.0%
	Interventional	3 (42–44)	25.0%
Study Population	US Spanish-Speaking Latinos	8 (33, 36–38, 40, 42–44)	66.7%
	Peru	2 (39, 41)	16.7%
	Colombia	1 (34)	8.3%
	Australian Spanish-Speaking Latinos	1 (35)	8.3%
App characteristics ($n = 10$)			
Mental health disorder addressed by the App	Depression	5 (33, 39–42, 44)	50.0%
	Substance use disorder	3 (36, 38, 43)	30.0%
	General mental health & emergency services	1 (34, 35)	10.0%
	Eating disorders	1 (37)	10.0%

App Development Protocol

The ¡Aptívate! app utilizes Brief Behavioral Activation (BBA) to address depressive symptoms by emphasizing that behaviors can influence mood and encourages participants to complete activities that align with self-selected values. To create this app, researchers started with the English app "Moodivate" and translated the app interface into Spanish utilizing a back-translation approach with the help of bilingual translators. Given that the app has patients self-select values and activities, cultural adaptation was not deemed necessary. However, the psychoeducation component of the app was carefully designed to avoid stigma in this population by de-emphasizing depressive symptoms as internal flaws and emphasizing the lack of environmental rewards as the etiology (33).

The Mental Health eClinic (MHeC) app was developed to address a broad range of mental health concerns and includes a triage system for those needing urgent help. The app was developed utilizing a participatory design methodology for Spanish-speaking youth in two different settings: international students in Australia and native youth in Colombia (34, 35). In a participatory design, stakeholders including patients, supportive others, and healthcare professionals provide their input in all six phases of app development. In both settings, all developmental phases that directly involved stakeholders were conducted in Spanish, eliminating the need for translation. In Australia, cultural adaptation included changing of the language question to specify Spanish dialects, changing the ethnicity question to reflect indigenous populations, and after initial disagreement, adoption the informal "tú" throughout the app (35). In Colombia, cultural adaptation included incorporating family structure and support networks, establishing credibility through university, health service provider, and community organization collaboration, and given the country's characteristics screening for economic stability, food security, and violence exposure (34).

The Automated Bilingual Computerized Alcohol Screening and Intervention (AB-CASI) mobile app was developed as the Spanish version of the emergency department-based alcohol screening, brief intervention, and referral to treatment (ED-SBIRT) program (36). The goal of the app is to address alcohol use disorders in Spanish without requiring extensive human resources, such as translators, in the emergency department. To culturally adapt this app, the researchers utilized user-centered design through design, development, and evaluation of app prototypes. This methodology ensures that stakeholders, namely patients and professionals, are involved in the development process. This app also addresses literacy issues through text-to-speech, which was found to be crucial for culturally adapting the app. Text to speech apps transform written text into audio.

Researchers ultimately chose a Text-To-Speech app to help with this process. Additionally, this work emphasizes that beyond translation and cultural adaptation, it is necessary to adapt health apps to the context in which they will be used. In this instance, the app had to be designed for use in an emergency department, therefore the app must have capability to save progress and start/stop/pause.

The Ecological Momentary Assessment (EMA) app was created to understand unhealthy eating and weight control behaviors of Mexican American women with low literacy (37). The app aims to collect information on these activities repeatedly in its natural context without having to rely on memory. Original studies of this app were geared toward college-enrolled women, but in contrast this study focused on women with low health literacy, requiring a shift from written components to pictures, icons, and sound features. App development was achieved by utilizing a user-centered methodology which involved the end-users at all four stages of development. Through this work, researchers found that: (1) text-to-icon translation (words are translated into images) was more complex than anticipated given the discrepancy in definitions of unhealthy eating behaviors between participants and researchers; (2) participants described forms of weight control products as opposed to their intended effects (ex: diet pill as opposed to laxative); (3) icons were found to be too complex to use to collect context and mood, therefore this component was not included in the final app.

Finally, Muñoz et al. submitted a study protocol for the design, development, and evaluation of the San Francisco Stop Smoking app which aims to help Latinos stop smoking (38). Although results of this study are pending, the protocol describes the human-centered methodology utilized to create this app. To achieve cultural adaptation, the first phase of development will require field observations of potential app users. These observations will give researchers first-hand information about how Latinos use their phones, how they interact with apps, and ultimately help researchers understand what app features might best serve the end-users. Researchers will also conduct workshops with Spanish speaking patient in which they will design an ideal app. This will also give researchers direct feedback on specific features that might be needed for the app to work for Spanish-speakers trying to quit smoking.

Feasibility

CONEMO is a nurse-supported app that utilizes behavioral activation to reduce depressive symptoms in patients with diabetes, hypertension, or both (39). The study by Lena Brandt et al. combines two feasibility pilot studies conducted in Lima, Peru to test the feasibility of: (1) implementing the app in two healthcare systems in Peru; (2) scaling up the app-based intervention. The study had 29 participants (mean age 60) utilize the app for 6 weeks, receiving three sessions per week. Semi-structured interviews revealed the app provided several health benefits, including reduction of stress and increase in motivation, and the majority were satisfied with the app. One major barrier was usability, with at least 72% of participants reporting some difficulties in using the app, though these subsided with longer use, and self-reported adherence was 50%. Participants also suggested the addition of audio. The study employed six nurses to support patients with app use and semi-structured interviews revealed that although they felt this was an innovative and helpful intervention, integration of CONEMO and daily responsibilities was challenging. Overall, CONEMO was found to be a feasible intervention.

CONEMO researchers also performed a composite study comprised of two pilot studies in Lima, Peru and one pilot study in São Paulo, Brazil with the dual goal of exploring the effectiveness of the CONEMO app and the feasibility to conduct a large randomized-control trial (39). The study enrolled 66 participants across the three sites. Data was collected actively: patients filled out a baseline PHQ-9, were then given access to CONEMO for 6 weeks, and filled a post-intervention PHQ-9, as well as passively: CONEMO system collected information such as sessions accessed and missed and interval between session access. Results of this study showed a general decrease in depressive symptoms based on decreasing PHQ-9 scores in 65–87% of participants, depending on the site, and a reduction in levels of functional disability. Challenges of this interventional study included difficulty with recruitment given that many patients were unable to read or write and a decline in session access over the course of the study.

Pratap et al. conducted a 3 month study to assess the feasibility of conducting a fully remote randomized controlled trial to screen for, assess, and treat depression in Latino individuals utilizing one of three apps to improve depressive symptoms: (1) EVO which uses therapeutic games, (2) iPST which employs psychotherapy principles, and (3) HTips which suggests mindfulness and behavioral exercises (40). The three apps were translated from English to Spanish by native Spanish speakers and professionals at Babble-on. The study enrolled 1,180 participants, but only 359 participated. Overall, feasibility of using mobile apps to remotely assess and treat depression was confirmed although the major challenges included a quick decline in app engagement and the higher cost and effort necessary to recruit Hispanic participants compared to non-Hispanic participants.

Intervention

Based on their feasibility findings, Pratap et al. then conducted a 3 month remote interventional study to compare recruitment and engagement of Hispanic and non-Hispanic participants and to compare treatment outcomes when participants utilized one of three apps: EVO, iPST, and HTips (42). The study remotely recruited and enrolled 1,020 participants, 389 of whom were Latino and 637 of whom were non-Latino. Participants were randomized to use one of the three apps for 3 months and PHQ-9 scores were collected at baseline and every week for the duration of the study. The study showed that PHQ-9 scores and self-reported disability scores decreased throughout the study without differences in recovery between Hispanics and non-Hispanics or by app used. The major challenges continued to be engagement, with Hispanic participants stopping the study nearly 2 weeks earlier than their counterparts and the high cost, high effort of enrolling Hispanic participants.

In this second interventional study (44), participants were recruited locally and nationally and randomized to one of three conditions for 8 weeks to address depressive symptoms: (1) ¡Aptívate! an app that uses behavioral activation, (2) iCouch which uses cognitive behavioral principles, and (3) treatment as usual, no app. The goals were to understand feasibility and efficacy of the ¡Aptívate! app. A total of 42 participants were enrolled ($n = 22$ ¡Aptívate!, $n = 9$ iCouch, $n = 11$ no app) and they self-reported app usage and completed the Spanish language Beck Depression Inventory-II (BDI-II) weekly. ¡Aptívate! is a self-help-based app and iCouch offers CBT. Results demonstrated lower depressive symptoms over time in those using the ¡Aptívate! app compared to treatment as usual, but no significant differences were found between ¡Aptívate! users and iCouch or between iCouch and treatment as usual. Challenges in this study included difficulties with local recruitment in South Carolina, although app engagement was higher in this group compared to nationally recruited participants.

In the third interventional study, researchers aimed to examine the mental health outcomes for Latinx Spanish-speaking patients with alcohol and other drug use (AOD) disorders with use of the CASA-CHESS mobile app for 8 weeks (43). This app was designed by translating and culturally adapting the theory-informed A-CHESS app into Spanish. The study enrolled 79 participants who were post-residential treatment for AOD, given a phone with the app, and followed for 6 months. Results show AOD symptoms for those that used the app for 4 months or longer were more stable with less use of illicit drugs, lower depressive and anxiety symptoms at 6 months compared to those who used the app for <4 months. Overall, this interventional study showed that an app can be an effective tool to provide continuity care for Spanish-speaking Latinos post-residential treatment. A major limitation of this work is the lack of app control or comparison group.

DISCUSSION

This review of the commercial market shows that there are many mental health apps available, but only a limited amount (14.5%) are offered in Spanish and none have conducted effectiveness studies with Spanish-speaking individuals. This problem is not unique to mental health apps. A study on apps for diabetes shows promise as it found that 30% (28/92) on the Android and Apple stores were in Spanish. However, when researchers investigated the Spanish readability of these apps, they found it was well above recommended reading levels, essentially making them inaccessible to many end-users (45). Although apps in Spanish may be growing in number, app developers must ensure that future apps are customizable, usable, and effective for the end-user populations.

Of the 10 apps studied in this literature review, only 4 apps (EVO, iPST, HTips, ¡Aptívate!) were translated (33, 42). This is likely as the others were developed in collaboration with Spanish speakers from their inception. However, Spanish-speakers in the United States are of course themselves diverse and thus cultural adaptation of apps must also consider: (1) end-user characteristics including nationality, locality, dialect, literacy level, socioeconomic status, (2) end-users' understanding of targeted conditions and associated stigmas, and (3) environments in which apps will be used and customizable features responsive to those environments. Of course, in many countries Spanish speakers are the majority and unique considerations for each culture, region, and clinical need must be considered.

User-centered design can help address most aspects of translation and cultural adaptation during the development phase, but it also requires intensive efforts and heavy upfront investment in recruitment, data collection, and usability testing. Another technique to achieve these goals is the utilization of cultural brokers to culturally adapt apps, as in the case of "Visit Planner," an app aimed at helping Spanish-speaking patients prepare for their primary care appointments (46). However, this methodology could still prove challenging in some communities where cultural brokers may are not easily identifiable. Community collaboration and trust may be necessary preconditions to access and work with these individuals. Overall, the literature shows that translation is necessary, but not sufficient to guarantee the usability and effectiveness of an app. Developers and researchers should aim to use user-centered development techniques when possible, but more work on viable alternatives that require less investment for app translation and cultural adaptation are needed.

Feasibility studies have demonstrated that it is possible to implement mental health apps to treat depressive symptoms in Spanish-speaking patients in different healthcare systems and that it is possible to conduct studies that enroll large numbers of Spanish-speaking patients, even when done remotely. However, we interpret these studies show there are three major barriers researchers have to contend with: (1) quick decline in app engagement, (2) app usability for some segments of the population, and (3) Increased expense and labor required to recruit Spanish-speaking participants.

User-centered design has been utilized as a method to empower potential end-users to contribute to app development and thus try to counteract usability and engagement concerns (25). The three feasibility studies discussed in this review did not describe the use of a user-centered design which may have contributed to a quick decline in app engagement. However, the studies also show that to counter engagement issues, app use should be paired with primary care, be accompanied by supportive nursing, or include features that help patients connect to mental health services. Nursing or other support staff, such a promotoras, can also help address usability issues in populations who may need extra technical help. Recruitment challenges may be multifaceted and may therefore require dynamic solutions. Work by Stuart Winter et al. describe best practices for including minorities in research including improving incentives, partnering with key community organizations, understanding the most used and effective methods of communication, addressing issues of distrust, and adding cultural brokers to the research team (26). Further work is needed to assess optimal incentives for Spanish-speaking populations and identify barriers to engagement in remote mental health studies.

Early data from interventional studies identified in this review demonstrate that mental health apps in Spanish may be effective tools to assess and treat depressive symptoms and help patients remain stable post-residential treatment for alcohol and AOD. Though the published literature on intervention studies is nascent, they show promising evidence that mobile health interventions can be successful for Spanish-speaking patients. This is particularly important given that these interventions will improve access to care for many patients in their native language, address the stigma associated with mental health conditions, and ultimately help close the gap on mental health disparities for Spanish-speaking patients. However, in the past 5 years few studies in the literature have appeared that more work needs to be done to increase research in this area. More apps need to be translated, culturally adapted and tested for feasibility and usability. Future efforts for the 32 apps available in Spanish identified in the market review should involve testing their effectiveness in this population. From these investigations, simple guidelines could be created to help patients, primary care providers, and/or mental health providers select the most appropriate app for each individual based on their characteristics, health literacy level, environment of use, personal preferences, and mental health condition.

Although the number of mental health apps available in Spanish is growing, translation of these apps alone may not be enough to improve access to these tools. A recent study of Medicare patients suggests that 38% of Spanish speakers do not have access to a smartphone with a data plan for wireless internet, precluding app use for many (27). A less visible but prevalent barrier to access is the lack of comfort with technology. The digital divide related to access to smartphones has also evolved into a lack of digital skills to navigate technology (28). Thus, Spanish-language apps are only as accessible as the users' digital literacy and comfort. While beyond the scope of this paper, we have seen this in our clinical work and as we have work to create a Spanish version of our teams' own open source mental health app (47, 48). To combat this, we have developed Digital Opportunities for Outcomes in Recovery Services (DOORS) to help people gain the core competencies and functional skills necessary to utilize digital health tools effectively (49). While DOORS is currently available in English, efforts to create partners and a team to build a Spanish version are underway. Future Spanish-language app development should consider the literacy levels of users and ponder what level of technical competency is necessary to utilize the tool.

As with any study, this review has several weaknesses that must be addressed. First, the market search was conducted on a database with 220 apps which is the largest database of mental health apps, however app stores host thousands of mental health-related apps. Thus, the percentage of apps available in Spanish may not be indicative of the entire landscape. Second, no single search term can discover all applicable articles on this topic. These results may also be influenced by publication bias due to apps or studies that may exist but have yet to be published. Third, while this was a U.S. based study in terms of the app marketplaces reviews and thus results may not generalizable to other countries and Spanish speakers, we note we did review research from around the world. Fourth, we looked at papers from the last 5 years which covers the majority of app research but will have missed earlier works.

CONCLUSION

Our study reveals that despite increasing interest in Spanish-language apps to address mental health, the commercial marketplace is not currently meeting the demand. None of the apps we reviewed on the marketplace were found in the academic literature, reflecting a gap between research and commercially available apps today. User-centered development design emerged as a leading strategy to increase the number of apps that are linguistically and culturally adapted to the Spanish-speaking population. Interventional studies show promising evidence that apps can help address mental health condition if challenges surrounding engagement and app usability are overcome. This study highlights that through thoughtful design and development, apps may hold the key to reducing mental health disparities in Spanish-speaking Hispanics in the United States.

AUTHOR CONTRIBUTIONS

JT and AM conceptualized the study. EC and AM reviewed the literature. AM led the paper drafting. All authors discussed paper selection, edited multiple drafts, and reviewed the final paper.

REFERENCES

1. *U.S. Census Bureau QuickFacts: United States*. Available online at: https://www.census.gov/quickfacts/fact/table/US/RHI725218 (accessed April 27, 2020).
2. Lee S, Held ML. Variation in mental health service use among U.S. Latinos by place of origin and service provider type. *Psychiatr Serv*. (2015) 66:56–64. doi: 10.1176/appi.ps.201300533
3. *English Use on the Rise Among Latinos*. Pew Research Center's Hispanic Trends Project. Available online at: https://www.pewresearch.org/hispanic/2015/05/12/english-proficiency-on-the-rise-among-latinos/ (accessed August 13, 2020).
4. Sudore RL, Landefeld CS, Pérez-Stable EJ, Bibbins-Domingo K, Williams BA, Schillinger D. Unraveling the relationship between literacy, language proficiency, and patient–physician communication. *Patient Educ Couns*. (2009) 75:398–402. doi: 10.1016/j.pec.2009.02.019
5. Schyve PM. Language differences as a barrier to quality and safety in health care: the joint commission perspective. *J Gen Intern Med*. (2007) 22(Suppl. 2):360–1. doi: 10.1007/s11606-007-0365-3
6. Lokko HN, Chen JA, Parekh RI, Stern TA. Racial and ethnic diversity in the US psychiatric workforce: a perspective and recommendations. *Acad Psychiatry*. (2016) 40:898–904. doi: 10.1007/s40596-016-0591-2
7. *Hispanics and Mobile Access to the Internet*. Pew Research Center's Hispanic Trends Project. Available online at: https://www.pewresearch.org/hispanic/2016/07/20/3-hispanics-and-mobile-access-to-the-internet/ (accessed August 23, 2020).
8. *Smartphones Help Blacks, Hispanics Bridge Some – but not all – Digital Gaps With Whites*. Pew Research Center. Available online at: https://www.pewresearch.org/fact-tank/2019/08/20/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digital-gaps-with-whites/ (accessed August 23, 2020).

9. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the use of mobile health technology by patients: an observational study in primary care clinics. *JMIR Mhealth Uhealth.* (2016) 4:e41. doi: 10.2196/mhealth.4928
10. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. *npj Dig Med.* (2020) 3:1–8. doi: 10.1038/s41746-020-00312-4
11. Roncero M, Belloch A, Doron G. A novel approach to challenging OCD related beliefs using a mobile-app: an exploratory study. *J Behav Ther Exp Psychiatry.* (2018) 59:157–60. doi: 10.1016/j.jbtep.2018.01.008
12. Bush NE, Dobscha SK, Crumpton R, Denneson LM, Hoffman JE, Crain A, et al. A virtual hope box smartphone app as an accessory to therapy: proof-of-concept in a clinical sample of veterans. *Suicide Life Threat Behav.* (2015) 45:1–9. doi: 10.1111/sltb.12103
13. Flett JAM, Conner TS, Riordan BC, Patterson T, Hayne H. App-based mindfulness meditation for psychological distress and adjustment to college in incoming university students: a pragmatic, randomised, waitlist-controlled trial. *Psychol Health.* (2020) 35:1049–74. doi: 10.1080/08870446.2019.1711089
14. Roncero M, Belloch A, Doron G. Can brief, daily training using a mobile app help change maladaptive beliefs? Crossover randomized controlled trial. *JMIR Mhealth Uhealth.* (2019) 7:e11443. doi: 10.2196/11443
15. Mayseless N. Can intervention programs influence how the dyslexic brain processes low-level visual stimuli? *Dev Neuropsychol.* (2011) 36:949–54. doi: 10.1080/87565641.2011.606421
16. Huberty J, Puzia M, Eckert R, Larkey L. Cancer patients' and survivors' perceptions of the calm app: cross-sectional descriptive study. *JMIR Cancer.* (2020) 6:e16926. doi: 10.2196/16926
17. Huberty J, Vranceanu A-M, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the calm meditation app: cross-sectional survey. *JMIR Mhealth Uhealth.* (2019) 7:e15648. doi: 10.2196/15648
18. Haimov I, Hanuka E, Horowitz Y. Chronic insomnia and cognitive functioning among older adults. *Behav Sleep Med.* (2008) 6:32–54. doi: 10.1080/15402000701796080
19. Bloom R, Schnaider-Beeri M, Ravona-Springer R, Heymann A, Dabush H, Bar L, et al. Computerized cognitive training for older diabetic adults at risk of dementia: study protocol for a randomized controlled trial. *Alzheimers Dement.* (2017) 3:636–50. doi: 10.1016/j.trci.2017.10.003
20. Hunter JF, Olah MS, Williams AL, Parks AC, Pressman SD. Effect of brief biofeedback via a smartphone app on stress recovery: randomized experimental study. *JMIR Serious Games.* (2019) 7:e15974. doi: 10.2196/15974
21. Puzia M, Laird B, Green J, Huberty J. Parents' perceptions of their children's engagement in a consumer-based meditation mobile app: cross-sectional survey study. *JMIR Pediatr Parent.* (2020) 3:e24536. doi: 10.2196/24536
22. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "calm" to reduce stress among college students: randomized controlled trial. *JMIR Mhealth Uhealth.* (2019) 7:e14273. doi: 10.2196/14273
23. Huberty J, Eckert R, Larkey L, Joeman L, Mesa R. Experiences of using a consumer-based mobile meditation app to improve fatigue in myeloproliferative patients: qualitative study. *JMIR Cancer.* (2019) 5:e14292. doi: 10.2196/14292
24. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. *JMIR Mental Health.* (2015) 2:e7. doi: 10.2196/mental.3935
25. Mistler LA, Ben-Zeev D, Carpenter-Song E, Brunette MF, Friedman MJ. Mobile mindfulness intervention on an acute psychiatric unit: feasibility and acceptability study. *JMIR Ment Health.* (2017) 4:e34. doi: 10.2196/mental.7717
26. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD coach, a smartphone app for post-traumatic stress symptoms. *Milit Med.* (2014) 179:12–8. doi: 10.7205/MILMED-D-13-00271
27. Carpenter J, Crutchley P, Zilca RD, Schwartz HA, Smith LK, Cobb AM, et al. Seeing the "big" picture: big data methods for exploring relationships between usage, language, and outcome in internet intervention data. *J Med Internet Res.* (2016) 18:e241. doi: 10.2196/jmir.5725
28. Huberty J, Eckert R, Larkey L, Kurka J, De Jesús SA, Yoo W, et al. Smartphone-based meditation for myeloproliferative neoplasm patients: feasibility study to inform future trials. *JMIR Form Res.* (2019) 3:e12662. doi: 10.2196/12662
29. Parks AC, Williams AL, Kackloudis GM, Stafford JL, Boucher EM, Honomichl RD. The effects of a digital well-being intervention on patients with chronic conditions: observational study. *J Med Internet Res.* (2020) 22:e16211. doi: 10.2196/16211
30. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. *PLoS ONE.* (2018) 13:e0209482. doi: 10.1371/journal.pone.0209482
31. Denneson LM, Smolenski DJ, Bauer BW, Dobscha SK, Bush NE. The mediating role of coping self-efficacy in hope box use and suicidal ideation severity. *Arch Suicide Res.* (2019) 23:234–46. doi: 10.1080/13811118.2018.1456383
32. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD coach in primary care with and without clinician support: a pilot randomized controlled trial. *Gen Hosp Psychiatry.* (2016) 38:94–8. doi: 10.1016/j.genhosppsych.2015.09.005
33. Dahne J, Collado A, Lejuez CW, Risco C, Diaz VA, Kustanowitz J, et al. ¡Aptívate!: a Spanish-language behavioral activation mobile application for delivery via primary care. *Psychol Serv.* (2019) 16:271–5. doi: 10.1037/ser0000304
34. Ospina-Pinillos L, Davenport TA, Navarro-Mancilla AA, Cheng VW, Alarcón AC, Rangel AM, et al. Involving end users in adapting a Spanish version of a web-based mental health clinic for young people in Colombia: exploratory study using participatory design methodologies. *JMIR Ment Health.* (2020) 7:e15914. doi: 10.2196/15914
35. Ospina-Pinillos L, Davenport T, Mendoza Diaz A, Navarro-Mancilla A, Scott EM, Hickie IB. Using participatory design methodologies to co-design and culturally adapt the Spanish version of the mental health eclinic: qualitative study. *J Med Internet Res.* (2019) 21:e14127. doi: 10.2196/14127
36. Abujarad F, Vaca FE. mHealth tool for alcohol use disorders among Latinos in emergency department. *Proc Int Symp Hum Factors Ergon Healthc.* (2015) 4:12–9. doi: 10.1177/2327857915041005
37. Connelly K, Stein KF, Chaudry B, Trabold N. Development of an ecological momentary assessment mobile app for a low-literacy, Mexican American population to collect disordered eating behaviors. *JMIR Public Health Surveill.* (2016) 2:e31. doi: 10.2196/publichealth.5511
38. Muñoz RF, Bunge EL, Barrera AZ, Wickham RE, Lee J. Using behavioral intervention technologies to help low-income and Latino smokers quit: protocol of a randomized controlled trial. *JMIR Res Protoc.* (2016) 5:e127. doi: 10.2196/resprot.5355
39. Menezes P, Quayle J, Garcia Claro H, Da Silva S, Brandt LR, Diez-Canseco F, et al. Use of a mobile phone app to treat depression comorbid with hypertension or diabetes: a pilot study in Brazil and Peru. *JMIR Ment Health.* (2019) 6:e11698. doi: 10.2196/11698
40. Pratap A, Anguera JA, Renn BN, Neto EC, Volponi J, Mooney SD, et al. The feasibility of using smartphones to assess and remediate depression in hispanic/latino individuals nationally. In: Lonini L, Shawen N, Ghaffari R, Rogers J, Jayarman A, editors. *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. Maui (2017). p. 854–60.
41. Brandt LR, Hidalgo L, Diez-Canseco F, Araya R, Mohr DC, Menezes PR, et al. Addressing depression comorbid with diabetes or hypertension in resource-poor settings: a qualitative study about user perception of a nurse-supported smartphone app in Peru. *JMIR Ment Health.* (2019) 6:e11701. doi: 10.2196/11701
42. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using mobile apps to assess and treat depression in Hispanic and Latino populations: fully remote randomized clinical trial. *J Med Internet Res.* (2018) 20:e10130. doi: 10.2196/10130
43. Muroff J, Robinson W, Chassler D, López LM, Lundgren L, Guauque C, et al. An outcome study of the CASA-CHESS smartphone relapse prevention tool for Latinx Spanish-speakers with substance use disorders. *Subst Use Misuse.* (2019) 54:1438–49. doi: 10.1080/10826084.2019.1585457
44. Dahne J, Collado A, Lejuez CW, Risco CM, Diaz VA, Coles L, et al. Pilot randomized controlled trial of a Spanish-language behavioral activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among

United States Latinx adults with limited English proficiency. *J Affect Disord*. (2019) 250:210–7. doi: 10.1016/j.jad.2019.03.009

45. Rodriguez JA, Singh K. The Spanish availability and readability of diabetes apps. *J Diabetes Sci Technol.* (2018) 12:719–24. doi: 10.1177/1932296817749610
46. Ruvalcaba D, Nagao Peck H, Lyles C, Uratsu CS, Escobar PR, Grant RW. Translating/creating a culturally responsive spanish-language mobile app for visit preparation: case study of "trans-creation." *JMIR Mhealth Uhealth*. (2019) 7:e12457. doi: 10.2196/12457
47. Torous J, Vaidyam A. Multiple uses of app instead of using multiple apps–a case for rethinking the digital health technology toolbox. *Epidemiol Psychiatr Sci*. (2020) 29:e100. doi: 10.1017/S2045796020000013
48. Rodriguez-Villa E, Rauseo-Ricupero N, Camacho E, Wisniewski H, Keshavan M, et al. The digital clinic: implementing technology and augmenting care for mental health. *Gen Hosp Psychiatry*. (2020). doi: 10.1016/j.genhosppsych.2020.06.009. [Epub ahead of print].
49. Rodriguez-Villa E, Camacho E, Torous J. Psychiatric rehabilitation through teaching smartphone skills to improve functional outcomes in serious mental illness. *Internet Interv*. (2021) 23:100366. doi: 10.1016/j.invent.2021.100366

Conflict of Interest: JT reports unrelated research support from Otsuka.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Digital Health

PERSPECTIVE
published: 02 July 2021
doi: 10.3389/fdgth.2021.690417

From General Principles to Procedural Values: Responsible Digital Health Meets Public Health Ethics

*Rune Nyrup**

Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, United Kingdom

Most existing work in digital ethics is modeled on the "principlist" approach to medical ethics, seeking to articulate a small set of general principles to guide ethical decision-making. Critics have highlighted several limitations of such principles, including (1) that they mask ethical disagreements between and within stakeholder communities, and (2) that they provide little guidance for how to resolve trade-offs between different values. This paper argues that efforts to develop responsible digital health practices could benefit from paying closer attention to a different branch of medical ethics, namely public health ethics. In particular, I argue that the influential "accountability for reasonableness" (A4R) approach to public health ethics can help overcome some of the limitations of existing digital ethics principles. A4R seeks to resolve trade-offs through decision-procedures designed according to certain shared procedural values. This allows stakeholders to recognize decisions reached through these procedures as legitimate, despite their underlying disagreements. I discuss the prospects for adapting A4R to the context of responsible digital health and suggest questions for further research.

Keywords: digital ethics, principlism, public health ethics, procedural values, accountability for reasonableness, A4R

Edited by:
Geke Ludden,
University of Twente, Netherlands

Reviewed by:
Olya Kudina,
Delft University of Technology, Netherlands
Merlijn Smits,
Radboud University Nijmegen Medical Centre, Netherlands

****Correspondence:***
Rune Nyrup
rn330@cam.ac.uk

Specialty section:
This article was submitted to Human Factors and Digital Health, a section of the journal Frontiers in Digital Health

Received: *02 April 2021*
Accepted: *14 June 2021*
Published: *02 July 2021*

Citation:
Nyrup R (2021) From General Principles to Procedural Values: Responsible Digital Health Meets Public Health Ethics. Front. Digit. Health 3:690417. doi: 10.3389/fdgth.2021.690417

INTRODUCTION

Recent years have seen a proliferation of digital ethics guidelines. There now exist more than 160 such guidelines, the vast majority published within the last 5 years by a wide range of institutions, including governments, legislative bodies, technology companies, and academic and professional organizations (1). These guidelines are intended for a number of purposes, including as a guide for designers of new digital technologies, to identify and address issues arising from the deployment of such technologies, and as a basis for developing standards and regulation (2).

Many seeking to bring analytical clarity to this panoply have looked to medical ethics for inspiration (3, 4). This is unsurprising: medical ethics is perhaps the most well-established field of practical ethics, both within academic research and as a framework for practitioners. For digital health technologies there is of course the additional reason that they are designed to become part of medical practice. Responsible digital health should involve being held to the same ethical standards as any other form of medical practice (5).

Most of this work has been modeled on an approach to medical ethics known as "principlism." Principlism seeks to articulate a small set of general principles to guide ethical decision-making.

Most influentially, Tom Beauchamp & James Childress' four *Principles of Biomedical Ethics* (6)—Beneficence, Non-Maleficence, Autonomy and Justice—are widely used and taught within clinical practice and research ethics. Many reviews of digital ethics guidelines similarly seek to subsume their recommendations under a small set of general principles, and some explicitly use Beauchamp & Childress' four principles (sometimes with a new fifth principle of Explicability) (3, 7–10). The convergence on these principles is often touted as evidence of an emerging consensus which can serve as a basis for implementing ethics into the design, regulation, and application of digital technologies. Yet how this is to be done largely remains an open question (11). Consequently, digital ethicists have increasingly turned their attention to how such principles can best be translated into practice, whether through new design practices (5, 12, 13) or new forms of legislation and regulation (14, 15).

However, critics have highlighted several limitations which vitiate the practical applicability of this approach to digital ethics (2, 9, 16–18). In this paper, I focus on two in particular. First, principles formulated in general, abstract terms mask underlying disagreements between and within stakeholder communities. Second, they provide little guidance for how to resolve tensions and trade-offs that can arise between different (interpretations of) principles. To overcome these limitations, I argue, efforts to develop more responsible digital health practices should pay closer attention to a different branch of medical ethics: public health ethics.

I start by making a general case for this claim. I then discuss the problems of disagreement and trade-offs within digital ethics, before introducing an influential account from public health ethics of how to reach ethically legitimate compromises on value-laden trade-offs. This approach, known as accountability for reasonableness (A4R) is based on the idea that legitimate compromises can be reached through decision-procedures designed according to certain procedural values (19). Finally, I discuss the prospects for adapting this approach to digital health and propose some questions for future research.

WHY PUBLIC HEALTH ETHICS?

Public health differs from clinical practice in two key respects (20): in who is affected, and in who decides and implements interventions. Public health interventions affect broader populations, rather than specific, identifiable patients, and they are largely decided and implemented by institutional actors (e.g., governments, insurance companies, NGOs), rather than individual clinicians/researchers.

There are two general reasons why closer attention to public health ethics is likely to benefit efforts to develop responsible digital health.

First, digital health technologies are often similar to public health interventions. Some are explicitly designed for public health purposes, such as monitoring infectious disease outbreaks (21, 22) or discovering risk factors for childhood obesity (23). But many digital technologies deployed in clinical settings also resemble public health interventions. Take machine learning tools for diagnostic decision-support (24, 25). These are usually designed for screening purposes, to monitor data from a given patient population and flag risk factors to human clinicians, and decisions to deploy them are made at the institutional level (e.g., hospitals or health service trusts). Even in patient-facing applications, e.g., conversational agents to assist with lifestyle decisions (26), many of the pertinent ethical decisions have to be made at the population/institutional level—by designers and regulators—rather than in the individual clinical encounter.

The second reason follows from the first. Due to its focus on population/institution-level interventions, public health ethics mainly addresses questions of political morality rather than the ethics of the individual patient-clinician relationship (20). It therefore provides a promising resource for addressing important political issues that arise from digital health.

Recent digital ethics has mostly focused on technological deficiencies and solutions, such as algorithmic bias and transparency. As several commentators have highlighted, this risks occluding broader social and political issues relating, e.g., to democratic oversight, power, and oppression (27–33). For example, it was recently shown that an algorithm that uses healthcare costs as a proxy for healthcare needs systematically underestimated the needs of Black patients, because less resources are already spent on their care (34). Ruha Benjamin (35) argues that labeling this "algorithmic bias," makes it seem a purely technical issue and sanitizes the social context that produced the problem in the first place, namely persistent structural and interpersonal racism in healthcare. More generally, as Leila Marie Hampton (30) argues, using generic concepts such as "fairness" or "transparency" to analyze technologies, without considering broader socio-political issues, risks legitimizing, and entrenching fundamentally unjust institutions.

While the Four Principles do include a principle of Justice, political issues covered under this heading mainly concern the question of what health-related goods society should provide and how to allocate resources within healthcare systems (5, chapter 6). By contrast, public health interventions raise a much wider set of political issues (20), similar to those commentators have started to discuss for digital health. For instance, is it permissible for interventions to impose risks or burdens on some individuals, even if they are not the main beneficiaries (e.g., mandatory vaccination programs)? Is it justifiable for interventions to exploit or reinforce structural patterns of disadvantage (e.g., using the communicative power of the state to stigmatize smoking)? More generally, when can institutional actors legitimately impose interventions despite widespread disagreement about relevant ethical values?

To be clear, my aim is not to reject the Four Principles framework or other principlist approaches to digital ethics. Such principles still serve a useful purpose in articulating the values at stake in digital ethics (cf. Section **What rationales should be considered relevant?).** Similarly, public health ethics will not, in itself, answer all of the socio-political issues that Benjamin, Hampton and others raise. Clearly, many of these require political action and structural change, not (just) better theory. Even in terms of theory, other literatures will be relevant too, especially emancipatory philosophies such as the Black Feminist tradition Hampton highlights. Nonetheless, public health ethics is a well-developed literature addressing practical political issues in healthcare, often closely informed by the empirical realities of healthcare policy and decision-making. It can thus help broaden the range of questions digital health ethics addresses.

DISAGREEMENT, TRADE-OFFS, AND THE LIMITS OF PRINCIPLES

The rest of this paper will focus on how insights from public health ethics can help overcome the two limitations of purely principlist approaches to digital ethics I highlighted in the introduction, i.e., that they mask disagreements between and within different stakeholder communities and provide little guidance for how to resolve trade-offs.

Consider for example debates about contact tracing apps for the management of Covid-19. Some governments wanted to base these on a centralized data collection approach, arguing that such datasets could also be used to produce new knowledge to help combat the pandemic. This was resisted by legal and information security experts concerned about potential privacy breaches (36–38). Appealing to general principles is unlikely to resolve this debate. While most people would presumably agree, say, that digital health technologies should be used to "do good" (Beneficence), there are legitimate ethical and political disagreements about the extent to which privacy is constitutive of or conducive to a good life. While we should arguably accept some trade-offs between protecting individual privacy and promoting social goods, there is little consensus on what exactly those trade-offs should be (38).

The prevalent approach to managing value trade-offs within clinical ethics is through informed consent (5, chapter 3): by informing patients about the trade-off involved in some treatment and letting them decide whether this is acceptable in light of their particular circumstances and values, clinicians can legitimize the decision to administer or withhold the treatment. It might be tempting to apply the same approach to digital health. However, informed consent is only plausible when the trade-offs occur within a single patient's value-set. One of the ways digital health resembles public health is that the trade-offs often cut across populations. Rather than each patient deciding for themselves how to balance trade-offs, which values get priority depends on population-level aggregate decisions. Contact tracing apps, and centralized data collection more generally, can only produce the relevant social goods if there is sufficient uptake (39). Conversely, if enough people consent to share their personal data, this can often be used to train machine learning algorithms capable of inferring highly personal information even about those who withhold consent (40).

In such cases, making interventions conditional on obtaining everyone's consent is neither practically feasible nor ethically plausible. A single intransigent individual should not be allowed to deprive everyone else of significant social goods. However, pure majority rule is not plausible either. Certain groups and communities may have good reasons, e.g., to value privacy because of their historical experiences of surveillance and discrimination (37). For instance, during the 1980's AIDS crisis, gay community-based activists initially resisted name-based reporting of infections, arguing that homophobia and AIDS-hysteria made privacy breaches and discrimination against people identified as HIV-positive more likely than for other diseases (41). Even if such reasons should not necessarily be decisive, collective decision-making should at least be responsive to them, and not just defer to majority preferences.

LEGITIMACY THROUGH PROCEDURAL VALUES

How to resolve disagreement and trade-offs is a characteristic conundrum in public health ethics. For example, in debates about priority setting and rationing of healthcare resources, ethicists have found it difficult to formulate ethical principles that are plausible enough to command broad consensus while being sufficiently fine-grained to guide decision-making in practice (42, 43). While many agree that those with greater needs should be given some priority, even at the expense of aggregate health outcomes, there is little consensus on how to weigh these two concerns against each other.

One influential model for resolving disagreements about priority setting in public health is called Accountability for Reasonableness (A4R) (19, 44, 45). Proposed by Norman Daniels and James Sabin, the key idea in A4R is to implement decision-procedures for reaching compromises which fair-minded people can accept as legitimate, despite their underlying ethical disagreements. This relies on a distinction between ethical *rightness* and ethical *legitimacy*. To regard a decision as right is to regard it as the morally correct thing to do in a given situation. To regard it as legitimate is to regard it as appropriately made, i.e., by a decision-maker or procedure whose moral authority to make such decisions should be accepted. The two can come apart: we can accept a verdict of "not guilty" in a fair trial as legitimate, even if we believe the defendant should have been convicted. Conversely, an unelected dictator may sometimes do the right thing, e.g., donate food to relieve a famine. Nonetheless, rightness and legitimacy are also entangled: if a procedure consistently generates abhorrent outcomes, we have reason to question its legitimacy; and if we can see that a decision-maker has carefully considered the relevant concerns, there is prima facie reason to accept their decision as right.

Daniels and Sabin propose four conditions for legitimate decision-procedures (44, 45):

1. *Publicity:* The rationale for a given decision must be publicly accessible.
2. *Relevance:* Decisions must be based on rationales which fair-minded individuals, who want to find mutually justifiable terms of cooperation, would accept as relevant to the decision.
3. *Revision and Appeals:* There must be mechanisms in place for challenging and revising decisions in light of new evidence or arguments.
4. *Enforcement:* There must be voluntary or public regulation in place to ensure that conditions 1–3 are met.

These conditions can be interpreted as embodying certain *procedural values*, specifying features that fair and appropriate decision procedures should have. It is a shared commitment to procedural values that generates legitimacy. Stakeholders who agree on these values have good reasons to regard procedures designed according to them as legitimate.

As the name suggests, the core procedural values in A4R are Accountability and Reasonableness. By articulating standards and mechanisms that stakeholders can use to hold decision-makers *accountable*—through enforceable rights to access rationales and challenge decisions—A4R aims to produce decisions that are *reasonable*, and can be recognized as such. Reasonableness here means something weaker than rightness: a decision is reasonable to the extent that it is responsive to all relevant concerns. Thus, if you recognize a decision as reasonable you may disagree about the specific way decision-makers weighed the reasons cited in their rationale, but you agree that it involved the right kinds of considerations.

The A4R conditions are supposed to guide the design of decision-making bodies charged with deciding how to balance any trade-offs that arise within a given healthcare institution (e.g., a hospital, public health agency or insurance company). Decision-makers should strive to identify compromises which all fair-minded stakeholders could find acceptable, though, some form of voting may be used if disagreement persists at the end of deliberation. Importantly, decision-makers do not need to articulate any general hierarchy of values or "meta-principles" for resolving trade-offs. Indeed, one of the motivations behind A4R is that we are unlikely to agree on any sufficiently action-guiding meta-principles. Rather, it aims to resolve trade-offs on a case-by-case basis as they arise in practice, based on rationales stakeholders will find contextually reasonable, despite persistent disagreement about general principles.

A4R is not without its detractors (little in philosophy is), nor is it the only account in public health ethics of how to resolve trade-offs (20). Nonetheless, it is a highly influential framework which has been used to inform public health practice (46, 47) and whose acceptability to decision-makers has been studied empirically across the world (48–50). Furthermore, public health ethicists have proposed a number of revisions and extensions of the A4R framework, reflecting lessons from these practical applications (51–54). As such, the A4R literature is likely to contain valuable lessons for responsible digital health[1].

ADAPTING A4R TO DIGITAL HEALTH

In the Introduction I highlighted two routes that ethicists have proposed for translating existing principles into practice: legislation/regulation and design practices. A4R can help overcome some of the limitations of the principlist approach within each of these.

Regarding the first, the challenge is to translate abstract general principles into more concrete legislation and regulation while still preserving their broad appeal. However, attempts to make principles more concrete and action-guiding, including any meta-principles for resolving trade-offs, will likely also make them more controversial. The A4R framework provides an alternative solution: rather than having to settle on a specific action-guiding translation of principles, legislators can instead specify how organizations that deploy or design digital health technologies should structure the decision-making processes through which they resolve any trade-offs they encounter.

As mentioned, deliberative bodies based on the A4R conditions have already been implemented in some healthcare institutions to address issues of priority setting and rationing. The remit of these could be expanded to also address the broader range of trade-offs that arise from the deployment of digital health technologies. Legislators could also require decision-making bodies modeled on the existing ones to be created elsewhere, including within private technology companies or as part of regulators charged with overseeing them.

Whether legally required or voluntarily adopted, this type of deliberative body could also provide a way to deal with trade-offs in the design of digital health technologies. A common criticism of Value-Sensitive Design (VSD) is that it lacks a method for resolving trade-offs, except if designers commit to an explicit – and therefore likely controversial – ethical theory (56, 57). This challenge will also affect proposals to implement digital ethics principles through (a modified version of) VSD (12). A4R suggests a way to overcome it: by structuring their decision-making processes according to the right kinds of procedural values, designers will be able to reach decisions that stakeholders can recognize as legitimate and therefore acceptable. To be clear, A4R is a *normative* theory of legitimacy. It does not commit the naturalistic fallacy by assuming that whatever stakeholders find acceptable is therefore right. If a decision counts as legitimate, according to A4R, stakeholders *ought* to find it acceptable.

FUTURE RESEARCH QUESTIONS

There are of course many details to be worked out regarding the proposals sketched here. How to best implement and operationalize them in practice remains an important question for future research. Part of this will practical, but A4R also provides a philosophically grounded theory to underpin this research and ensure that proposed implementations remain normatively plausible.

However, we should not expect that A4R can simply be transposed from its original application (priority setting and rationing) to digital health without modification. Adapting A4R to digital health will likely require modifications or extensions to the framework itself. At least two kinds of further research questions will be relevant to explore.

Are Other Procedural Values Needed?

One of the ways public health ethicists have extended the original A4R framework is by adding further procedural values, often motivated by their practical experience of applying A4R to priority setting decisions. For instance, some have proposed new conditions of Inclusiveness and Empowerment. In brief, these require explicit input from all affected stakeholders and

---

[1] To my knowledge, only two other recent papers have discussed the application of A4R to digital (health) ethics (33, 55), though not along the same lines as me.

that active steps are taken to counteract knowledge-gaps and institutional power differences between decision-makers (33, 53, 58). Importantly, these conditions are still motivated by the core value of Reasonableness, namely to ensure that decision-makers are responsive to as many relevant concerns as possible, including those that are held by minoritized or less empowered parts of the population.

Applying A4R to digital health may similarly reveal new procedural values. For instance, if Benjamin and Hampton are correct that ethical discussions of digital technologies risk sanitizing and entrenching unjust social structures, it may be necessary to actively encourage decision-makers to raise critical questions about how new technologies will interact with these structures. Similarly, it may be necessary to encourage scrutiny of the aims and presuppositions of the technology itself, asking for example whether it targets the right problem or whether the proposed solution is at all appropriate. We might summarize these as a condition of Socio-Technological Criticism.

What Rationales Should Be Considered Relevant?

The Relevance condition is a formal constraint on the type of rationales that should be given weight within decision-making. However, implementing A4R in practice requires us to specify in more substantive terms what types of concerns should be admissible. This will likely depend on the context of application. As A4R was originally developed for debates about rationing, most discussions focus on rationales framed in terms of Fairness or related distributive values (e.g., Solidarity (52)). Presumably, a broader range of values will be relevant to debates about digital health technologies (e.g., Privacy). Exploring in more detail what those values should be is a substantive research task. To ensure that decision-makers are responsive to all relevant reasons, this research should aim to identify a broad range of plausible concerns and help elucidate and articulate these, so that stakeholders can present them in their most compelling form. Existing VSD methodologies for empirical and conceptual investigations of stakeholder values provide a plausible approach to this task.

Existing principlist approaches to digital ethics provide a useful starting point. However, the values discussed in the existing literature should not be assumed exhaustive or representative. The apparent convergence found here may simply be a product of people from roughly similar backgrounds consuming the same literature (2, 17). It is noticeable, for instance, that many commonly cited principles (e.g., transparency, fairness, responsibility) also feature prominently within liberal political philosophy. Values more characteristic of other political traditions, such as solidarity, belonging, authenticity, harmony, non-exploitation, non-domination or emancipation are rarely discussed or even mentioned (9, 29, 30). Public health ethics may also here provide a useful resource. Public health ethicists have developed alternative sets of principles to the four classical principles of biomedical ethics (59), and explored the implications of different political traditions (60).

CONCLUSION

Paying closer attention to public health ethics is likely to benefit efforts to develop responsible digital health. In this paper, I have made a general case for this claim and highlighted A4R as a specific model from public health ethics that can be adapted to digital health. While not intended to wholly replace principlism, A4R can complement and help overcome some of the limitations faced by principlist approaches. Further, research on the questions outlined above could generate valuable insights for the ethical deployment, design and regulation of digital technologies, especially within healthcare.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This research was funded in whole, or in part, by the Wellcome Trust [Grant No. 213660/Z/18/Z] and the Leverhulme Trust, through the Leverhulme Centre for the Future of Intelligence [RC-2015-067].

ACKNOWLEDGMENTS

I'm very grateful to Jess Whittlestone, Stephen Cave and two anonymous referees for detailed feedback on previous drafts, and to Sidsel Størmer for research assistance which informed this paper. Many thanks also to Ali Boyle, Elena Falco, Adrian Weller, Dan White, and John Zerilli for fruitful discussion.

REFERENCES

1. Algorithm Watch. *AI Ethics Guidelines Global Inventory*. (2020). Available online at: https://inventory.algorithmwatch.org/ (accessed April 2, 2021).
2. Whittlestone J, Nyrup R, Alexandrova A, Cave S. The role limits of principles in ai ethics: towards a focus on tensions. In: *Proceedings of the AAAI/ACM Conference AI. Ethics, Society (AIES'19):195-200* (Honolulu, HI). (2019). doi: 10.1145/3306618.3314289
3. Floridi L, Cowls J. A unified framework of five principles for AI in society. *Harv Data Sci Rev.* (2019) 1. doi: 10.1162/99608f92.8cd550d1
4. Véliz C. Three things digital ethics can learn from medical ethics. *Nat Electron.* (2019) 2:316–8. doi: 10.1038/s41928-019-0294-2

5. Peters D, Vold K, Robinson D, Calvo R. Responsible AI—two frameworks for ethical design practice. *IEEE Trans Tech Soc.* (2020) 1:34–47. doi: 10.1109/TTS.2020.2974991
6. Beauchamp T, Childress J. *Principles of Biomedical Ethics*. 5th ed. New York, NY: Oxford University Press (2001).
7. Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M. Principled artificial intelligence: mapping consensus in ethical rights-based approaches to principles for AI. In: *Berkman Klein Center Research Publication No. 2020-1* (Cambridge, MA). (2020). doi: 10.2139/ssrn.3518482
8. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds Mach.* (2018) 28:689–707. doi: 10.1007/s11023-018-9482-5
9. Jobin A, Ienca M, Vayena E. Artificial intelligence: the global landscape of ethics guidelines. *Nat Mach Intell.* (2019) 1:389–99. doi: 10.1038/s42256-019-0088-2
10. Zeng Y, Lu E, Huangfu C. Linking artificial intelligence principles. In: *Proceedings of the AAAI Workshop on Artificial Intelligence Safety (AAAI-Safe AI 2019)*. Honolulu, HI (2019).
11. Morley J, Floridi L, Kinsey L, Elhalal A. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci Eng Ethics.* (2020) 26:2141–68. doi: 10.1007/s11948-019-00165-5
12. Umbrello S, van de Poel I. Mapping value sensitive design onto AI for social good principles. *AI Ethics.* (2021). doi: 10.1007/s43681-021-00038-3
13. van de Poel I. Embedding values in artificial intelligence (AI) systems. *Minds Mach.* (2020) 30:385–409. doi: 10.1007/s11023-020-09537-4
14. Stix C. Actionable principles for artificial intelligence policy: three pathways. *Sci Eng Ethics.* (2021) 27:15. doi: 10.1007/s11948-020-00277-3
15. Braun M, Bleher H, Hummel P. A leap of faith: is there a formula for "Trustworthy" AI? *Hast C Rep*. (2021) 51:17-22. doi: 10.1002/hast.1207
16. Mittelstadt B. Principles alone cannot guarantee ethical AI. *Nat Mach Intell.* (2019) 1:501–7. doi: 10.1038/s42256-019-0114-4
17. Whittlestone S, Nyrup J, Alexandrova R, Dihal A, Cave K. *Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research*. London: Nuffield Foundation (2019).
18. Dignum V. Responsible autonomy. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)* (Melbourne, VIC). (2017). p. 4698–704. doi: 10.24963/ijcai.2017/655
19. Daniels N. Accountability for reasonableness. *BMJ.* (2000) 321:1300–1. doi: 10.1136/bmj.321.7272.1300
20. Faden, R, Bernstein J, Shebaya S. Public health ethics. In: Zalta E. editor. *The Stanford Encyclopedia of Philosophy (Fall 2020 Edition)*. (2020). Available online at: https://plato.stanford.edu/archives/fall2020/entries/publichealth-ethics/ (accessed April 2, 2021).
21. Bengtsson L, Gaudart J, Lu X, Moore S, Wetter E, Sallah K, et al. Using mobile phone data to predict the spatial spread of cholera. *Sci Rep.* (2015) 5:8923. doi: 10.1038/srep08923
22. Doan S, Ngo Q, Kawazoe A, Collier N. Global Health Monitor – a web-based system detecting and mapping infectious disease. In: *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)* (Hyderabad). (2008). p. 951–6. https://arxiv.org/abs/1911.09735
23. Dugan T, Mukhopadhyay S, Carroll A, Downs S. Machine learning techniques for prediction of early childhood obesity. *Appl Clin Inform.* (2015) 5:506–20. doi: 10.4338/ACI-2015-03-RA-0036
24. Gulshan V, Peng L, Coram M, Stump M, Wu D, Narayanswamy A. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* (2016) 316:2402–10. doi: 10.1001/jama.2016.17216
25. Solares J, Raimondi F, Zhu Y, Rahimian F, Canoy D, Tran, et al. Deep learning for electronic health records: a comparative review of multiple deep neural architectures. *J Biomed Inform.* (2020) 101:103337. doi: 10.1016/j.jbi.2019.103337
26. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. *JMIR Diabetes.* (2017) 2:e28. doi: 10.2196/diabetes.8590
27. Zimmerman A, di Rosa E, Kim H. *Technology Can't Fix Algorithmic Injustice. Boston Review*. (2020). Available online at: http://bostonreview.net/science-nature-politics/annette-zimmermann-elena-di-rosa-hochan-kim-technology-cant-fix-algorithmic (accessed April 2, 2021).
28. Kind C. *The Term 'Ethical AI' is Finally Starting to Mean Something*. Venture Beat (2020). Available online at: https://venturebeat.com/2020/08/23/the-term-ethical-ai-is-finally-starting-to-mean-something/ (accessed April 2, 2021).
29. Kalluri P. Don't ask if AI is good or fair, ask how it shifts power. *Nature.* (2020) 583:169. doi: 10.1038/d41586-020-02003-2
30. Hampton L. Black feminist musings on algorithmic oppression. In: *Conference on Fairness, Accountability, Transparency (FAccT'21).* Association for Computing Machinery, New York, NY, United States (2021). doi: 10.1145/3442188.3445929
31. Binns R. Fairness in machine learning: lessons from political philosophy. In: *Proceedings in Machine Learning Research 81: Conference on Fairness, Accountability and Transparency* (New York, NY). (2018). p. 149–59.
32. Himmelreich J. Ethics of technology needs more political philosophy. *Commun ACM.* (2020) 63:33–5. doi: 10.1145/3339905
33. Wong P. Democratizing algorithmic fairness. *Philos Technol.* (2020) 33:224–44. doi: 10.1007/s13347-019-00355-w
34. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. (2019) 366:447–53. doi: 10.1126/science.aax2342
35. Benjamin R. Assessing risk, automating racism. *Science.* (2019) 366:421–2. doi: 10.1126/science.aaz3873
36. Burgess, M. *There's a Big Row Brewing Over the NHS Covid-19 Contact Tracing App*. *Wired*. (2020). Available online at: https://www.wired.co.uk/article/nhs-contact-tracing-app-data-privacy (accessed April 2, 2021).
37. Cave S, Whittlestone J, Nyrup R, hEigeartaigh SO, Calvo RA. Using AI ethically to tackle covid-19. *BMJ.* (2021) 372:n364. doi: 10.1136/bmj.n364
38. Parker M, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. *J Med Ethics.* (2020) 46:427–31. doi: 10.1136/medethics-2020-106314
39. Singh H, Couch D, Yap K. Mobile health apps that help with COVID-19 management: scoping review. *JMIR Nursing.* (2020) 3:e20596. doi: 10.2196/20596
40. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. *PNAS.* (2013) 110:5802–5. doi: 10.1073/pnas.1218772110
41. Bayer R, Fairchild A. The limits of privacy: surveillance and the control of disease. *Health Care Anal.* (2002) 10:19–35. doi: 10.1023/A:1015698411824
42. Daniels N. Rationing fairly: programmatic considerations. *Bioethics.* (1993) 7:224–33. doi: 10.1111/j.1467-8519.1993.tb00288.x
43. Holm S. Goodbye to the simple solutions: the second phase of priority setting in health care. *BMJ.* (1998) 1998:317:1000. doi: 10.1136/bmj.317.7164.1000
44. Daniels N, Sabin J. Limits to health care: fair procedures, democratic deliberation, and the legitimacy problem for insurers. *Phil Pub Aff.* (1997) 26:303–50. doi: 10.1111/j.1088-4963.1997.tb00082.x
45. Daniels N. *Just Health: Meeting Health Needs Fairly*. Cambridge: Cambridge University Press (2008). doi: 10.1017/CBO9780511809514
46. Maluka S. Strengthening fairness, transparency and accountability in health care priority setting at district level in Tanzania. *Global Health Action.* (2011) 4:7829. doi: 10.3402/gha.v4i0.7829
47. O'Malley P, Rainford J, Thompson A. Transparency during public health emergencies: from rhetoric to reality. *Bull WHO.* (2009) 87:614–8. doi: 10.2471/BLT.08.056689
48. Martin D, Giacomini M, Singer P. Fairness, accountability for reasonableness, and the views of priority setting decision-makers. *Health Policy.* (2002) 61:279–90. doi: 10.1016/S0168-8510(01)00237-8
49. Kapiri L, Norheim O, Martin D. Fairness and accountability for reasonableness. Do the views of priority setting decision makers differ across health systems and levels of decision making? *Soc Sci Med.* (2009) 68:766–73. doi: 10.1016/j.socscimed.2008.11.011
50. Mshana S, Shemilu H, Ndawi B, Momburi R, Olsen O, Byskov J, et al. What do district health planners in Tanzania think about improving priority

setting using 'Accountability for reasonableness'? *BMC Health Serv Res.* (2007) 7:180. doi: 10.1186/1472-6963-7-180

51. Rid A, Biller-Andorno N. Justice in action? Introduction to the minisymposium on Norman Daniels' just health: meeting health needs fairly. *J Med Ethics*. (2009) 35:1–2. doi: 10.1136/jme.2008.025783
52. Hasman A, Holm S. Accountability for reasonableness: opening the black box of process. *Health Care Anal.* (2005) 13:261–73. doi: 10.1007/s10728-005-8124-2
53. Thompson A, Faith K, Gibson J, Ushur R. Pandemic influenza preparedness: an ethical framework to guide decision-making. *BMC Med Ethics.* (2006) 2006:7:12. doi: 10.1186/1472-6939-7-12
54. Bell J, Hyland S, DePellegrin T, Upshur R, Bernstein M, Martin D. SARS and hospital priority setting: a qualitative case study and evaluation. *BMC Health Serv Res.* (2004) 2004:4:36 doi: 10.1186/1472-6963-4-36
55. Brall C, Schröder-Bäck P, Maeckelberghe E. Ethical aspects of digital health from a justice point of view. *Europ J Pub Health.* (2019) 29(Suppl. 3):18–22. doi: 10.1093/eurpub/ckz167
56. Manders-Huits N. What values in design? The challenge of incorporating moral values into design. *Sci Eng Ethics.* (2011) 17:271–87. doi: 10.1007/s11948-010-9198-2
57. Jacobs N, Huldtgren A. Why value sensitive design needs ethical commitments. *Ethics Inf Technol.* (2021) 23:23–6. doi: 10.1007/s10676-018-9467-3
58. Gibson J, Martin D, Singer P. Priority setting in hospitals: fairness, inclusiveness, and the problem of institutional power differences. *Soc Sci Med.* (2005) 61:2355–62. doi: 10.1016/j.socscimed.2005.04.037
59. Upshur R. Principles for the justification of public health intervention. *Canad J Pub Health.* (2002) 93:101–3 doi: 10.1007/BF03404547
60. Jennings B. Frameworks for ethics in public health. *Acta Bioethica.* (2003) 9:165–76. doi: 10.4067/S1726-569X2003000200003

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 08 July 2021
doi: 10.3389/fdgth.2021.690182

Attitudes Toward Health, Healthcare, and eHealth of People With a Low Socioeconomic Status: A Community-Based Participatory Approach

Jasper S. Faber[1]*, Isra Al-Dhahir[2], Thomas Reijnders[1,2], Niels H. Chavannes[3,4], Andrea W. M. Evers[1,2,5], Jos J. Kraal[1], H. J. G. van den Berg-Emons[6,7] and Valentijn T. Visch[1]

[1] Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands, [2] Faculty of Social and Behavioral Sciences, Leiden University, Leiden, Netherlands, [3] Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, Netherlands, [4] National eHealth Living Lab, Leiden University Medical Centre, Leiden, Netherlands, [5] Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, Netherlands, [6] Department of Rehabilitation Medicine, Erasmus Medical Center (MC), University Medical Center Rotterdam, Rotterdam, Netherlands, [7] Capri Cardiac Rehabilitation, Rotterdam, Netherlands

Edited by:
Naseem Ahmadpour,
The University of Sydney, Australia

Reviewed by:
Wouter A. Keijser,
University of Twente, Netherlands
Khushnood Zahoor Naqshbandi,
The University of Sydney, Australia
Katta Spiel,
Vienna University of
Technology, Austria

***Correspondence:**
Jasper S. Faber
j.s.faber@tudelft.nl

Specialty section:
This article was submitted to Health Technology Innovation, a section of the journal Frontiers in Digital Health

Received: *02 April 2021*
Accepted: *04 June 2021*
Published: *08 July 2021*

Citation:
Faber JS, Al-Dhahir I, Reijnders T, Chavannes NH, Evers AWM, Kraal JJ, van den Berg-Emons HJG and Visch VT (2021) Attitudes Toward Health, Healthcare, and eHealth of People With a Low Socioeconomic Status: A Community-Based Participatory Approach. Front. Digit. Health 3:690182. doi: 10.3389/fdgth.2021.690182

Low socioeconomic status (SES) is associated with a higher prevalence of unhealthy lifestyles compared to a high SES. Health interventions that promote a healthy lifestyle, like eHealth solutions, face limited adoption in low SES groups. To improve the adoption of eHealth interventions, their alignment with the target group's attitudes is crucial. This study investigated the attitudes of people with a low SES toward health, healthcare, and eHealth. We adopted a mixed-method community-based participatory research approach with 23 members of a community center in a low SES neighborhood in the city of Rotterdam, the Netherlands. We conducted a first set of interviews and analyzed these using a grounded theory approach resulting in a group of themes. These basic themes' representative value was validated and refined by an online questionnaire involving a different sample of 43 participants from multiple community centers in the same neighborhood. We executed three focus groups to validate and contextualize the results. We identified two general attitudes based on nine profiles toward health, healthcare, and eHealth. The first general attitude, *optimistically engaged,* embodied approximately half our sample and involved *light-heartedness* toward health, *loyalty* toward healthcare, and *eagerness* to adopt eHealth. The second general attitude, *doubtfully disadvantaged,* represented roughly a quarter of our sample and was related to feeling *encumbered* toward health, feeling *disadvantaged* within healthcare, and *hesitance* toward eHealth adoption. The resulting attitudes strengthen the knowledge of the motivation and behavior of people with low SES regarding their health. Our results indicate that negative health attitudes are not as evident as often claimed. Nevertheless, intervention developers should still be mindful of differentiating life situations, motivations, healthcare needs, and eHealth expectations. Based on our findings, we recommend eHealth should fit into the person's daily life, ensure personal communication, be perceived usable and useful, adapt its communication to literacy level and life situation, allow for meaningful self-monitoring and embody self-efficacy enhancing strategies.

Keywords: low socioeconomic status, eHealth adoption, health attitudes, community-based participatory research, user profiles, health disparities, eHealth intervention design

INTRODUCTION

Low socioeconomic status (SES) is associated with a higher prevalence of unhealthy lifestyles compared to a high SES (1). Consequently, people with a low SES are at increased risk of chronic diseases (e.g., cardiovascular disease, diabetes, and obesity) (2–4). eHealth interventions such as monitoring devices, online communication platforms, and serious games have been proven effective in changing behavior and promoting a healthy lifestyle in various domains. However, these interventions are less successful in changing the behavior of people with a low SES due to low reach, less adherence during the intervention or less effectiveness of the interventions (5–9).

A crucial factor in facilitating the adoption, and therefore success, of eHealth interventions, is the alignment with a person's attitude toward using this technology (10, 11). Moreover, successfully achieving a lifestyle change, a primary goal of such interventions, requires the person to have a positive attitude toward their health and health services (12). eHealth is designed to expect its intended users to have a positive and pro-active health attitude. However, considering the growth of current health inequalities, such interventions would have a bigger impact when they can support groups not sharing these attitudes.

A multitude of studies point out that people with a low SES have unfavorable attitudes toward their health, healthcare, and eHealth. For instance, Wardle and Steptoe (13) found that health attitudes within the low SES groups are specifically characterized by a lower consciousness about health and less often thinking about the future. Other studies have identified more passive attitudes toward healthcare (14) and less confident attitudes toward digital health interventions (15) within low SES groups. Nevertheless, there is insufficient evidence to inform researchers and designers about these attitudes. The complexity of studying health values within contrasting sociodemographic environments poses various emotional and ethical challenges such as perceived harms, feelings of stigmatization, and anxiety toward research and the research team (16–18). As a result, hard-to-reach groups are minimally included in research efforts. Moreover, existing evidence is difficult to generalize toward other contexts. Measurements of attitudes are highly context-dependent and are expected to differ by country, setting, and time (19). Financial well-being and accessibility of health sources, for example, will not have a profound impact within countries that have unemployment funds, state-funded healthcare, and relatively good public transportation. Consequently, we have a lack of evidence to support the research and design of eHealth interventions that align with the attitudes of people with a low SES.

The rise of eHealth in current healthcare systems opens up exciting new possibilities to improve healthcare quality and efficiency. However, with the increased use of technical innovations and digital systems come unintended, unpredictable, and adverse consequences for individuals. Due to the underrepresentation of these specific societal groups, interventions are minimally aligned toward their attitudes. Consequently, these interventions face the risk of not being adopted and therefore unintentionally contribute to rising health inequalities. Researchers and designers should carry the responsibility to harness the potential of eHealth to create benefit for all groups in society, not merely for those that are motivated to perform a healthy lifestyle (20).

To engage the target group in the research process, an approach is needed that is comprehensive, culturally sensitive, and builds upon a relationship-based personal approach (18). Community-based participatory research (CBPR), a socio-culturally sensitive approach, which creates a trustful and long-lasting relationship between researcher and participant, has been effectively applied in culturally contrasting contexts (21, 22). For example, Henderson et al. (23) successfully implemented a CBPR approach to develop a tailored web-based diabetes self-management tool in a low-resource setting in the United States. Such an approach can engage hard-to-reach groups in the research process, yet has not been applied in the context of attitudes in low SES groups. In addition, focusing on a community instead of a person's individual characteristics is increasingly being recognized as a valuable approach. Studies that focus on these characteristics imply that these are the cause of poor health outcomes, which carries the risk of increasing stigma (24). It is becoming increasingly known that contextual community factors, such as the availability of healthy food, experiences of discrimination, and neighborhood poverty, also have a significant relation to poor health outcomes (25, 26).

The resulting knowledge could improve the alignment of health services toward attitudes of low SES populations, thereby facilitating their adoption. Currently, eHealth interventions aimed at these populations have only been minimally tailored, for example, by simplifying text and including images and videos (27). However, there is currently limited evidence reporting how interventions could be tailored toward psychological characteristics, such as attitudes with regard to eHealth. Although some studies report on the relationship between attitudes and interventions (28, 29), the resulting knowledge is difficult to apply in the design of interventions directly. Forms of practical knowledge, such as data-driven patient-profiles, have been used in the past to tailor content, context, and delivery of care toward individual preferences (30). Yet, such a form of knowledge has not been developed for attitudes of people with a low SES toward their health, healthcare, and eHealth in general.

This study aims to achieve design-relevant knowledge about the attitudes of people with a low SES toward their health, healthcare, and eHealth. To achieve this, we took a community-based participatory research approach to facilitate responsible engagement of the target group in the research process. The resulting knowledge can facilitate the design and alignment of health services toward the different attitudes of low SES populations. This will result narrowing current health disparities by developing interventions that are more acceptable, satisfactory, and user-friendly.

MATERIALS AND METHODS

Our methodology revolved around the principles of CBPR. CBPR is a partnership approach to research that equitably involves community members, organizational representatives, and researchers in all aspects of the research process (21). Our CBPR approach consisted of three separate phases (**Figure 1**) in which the outcomes of each phase were used in the next.

Sampling and Recruitment

We initiated our collaboration with a community center located in a neighborhood in Rotterdam, the Netherlands. The neighborhood was selected based on its neighborhood SES, a combined measure of neighborhood income, education, and occupation (31). The neighborhood in which the community center is situated has been one of the lowest scoring neighborhoods on livability; a combined measure of its social, physical, and safety index (32). The area therefore is on the agenda as one of the *focus-neighborhoods* of the municipality of Rotterdam. Sixty-eight percent of the inhabitants have a migration background, compared to 52% in Rotterdam. In addition, 59% of the households have a low income compared to 52% in Rotterdam. Finally, 34% of the inhabitants have a low education, compared to 32% in Rotterdam (33).

The participants were sampled based on their affiliation with the community center and their living area (neighborhood SES). The community center situated in this neighborhood facilitates inhabitants that struggle with fundamental aspects of their life. They focus on poverty, occupation, living, social contacts, upbringing, and safety. We included participants living in the selected neighborhood with the following affiliations with the community center: (1) Visitors (Vi): Persons who visit the community center regularly and require support. (2) Volunteers (Vo): Unemployed persons who performed volunteering work in the community center in exchange for state funding. (3) Key persons (Kp): Social workers who have close relationships with the community members. In this study, Kp's were not considered as part of the target group as they are employed at the community center and are in the role of providing support. However, since they interact with Vi's and Vo's on a daily basis, we included them to learn about attitudes within the community from the Kp's perspective. In that light, we did not include Kp's in the second phase of the study as we were solely interested in acquiring a deeper understanding of the attitudes we observed in the first phase. Finally, it should be noted that Vo's could visit the community center as Vi's as well. For this study, we considered persons a Vo when they had at least one regular weekly shift at the community center.

In phase one, we sampled the participants conveniently and recruited them face-to-face at the community center. In the second phase, Vo's and Vi's were purposively sampled and recruited face-to-face. In phase three, we recruited participants for the questionnaire through an advertisement on the community center's Facebook page and WhatsApp group (**Supplementary Figure 1**) and through Kp's of various community centers within the same neighborhood. The participants for the focus groups were recruited through a question attached at the end of the digital questionnaire and by approaching them face-to-face at the community center. Because of the come-and-go nature of the community center, some participants frequently visiting the community center participated in each of the three phases, while others only participated in one.

Ethics

The study protocol was approved by the Human Research Ethics Committee of Delft University of Technology (approval numbers 953, 1064, and 1141). Through our relationship-based CBPR approach we aimed to limit the impact of emotional and ethical challenges such as perceived harm, feelings of stigmatization, and anxiety toward research and the research team. In the first phase, we briefed our participants orally about the nature of the study as a formal written consent in this first introduction phase would obstruct a trustful interaction. The participants provided their consent verbally to the researcher (JF). In phases 2 and 3, when the relationship was more solid, written informed consent was provided.

Procedure and Materials

In phase one, we aimed to form a trustful research partnership with the community and narrow down the research scope by simultaneously exploring and identifying specific research directions. We initiated the partnership by attending community gatherings and organizing health-themed lunch events at the community center. Such immersive activities have been used and proven successful in creating a relationship in various other CBPR efforts (21). During these activities, we addressed the research scope by engaging in unstructured interviews with community members individually. Based on an initial literature review, a backlist of topics guided the interviews and helped to steer them toward our research questions (34). We divided the topic questions into three overarching research themes: attitudes toward health, healthcare, and eHealth. For example, we explored the attitude toward health with questions such as *"How important is it for you to live long?"*. Questions such as *"What do you think of your doctor's advice?"* and *"What do you think of a technology that could help you live healthier?"* referred to the attitude toward healthcare and eHealth, respectively. The full interview backlist is provided in **Supplementary Table 1**. Data was captured by taking quick field notes during the visits and elaborating on them into comprehensive reports directly afterward.

In phase two, we investigated the specific directions resulting from the first phase more extensively through semi-structured interviews. In contrast to unstructured interviews, these interviews are more formal and intimate, which comes

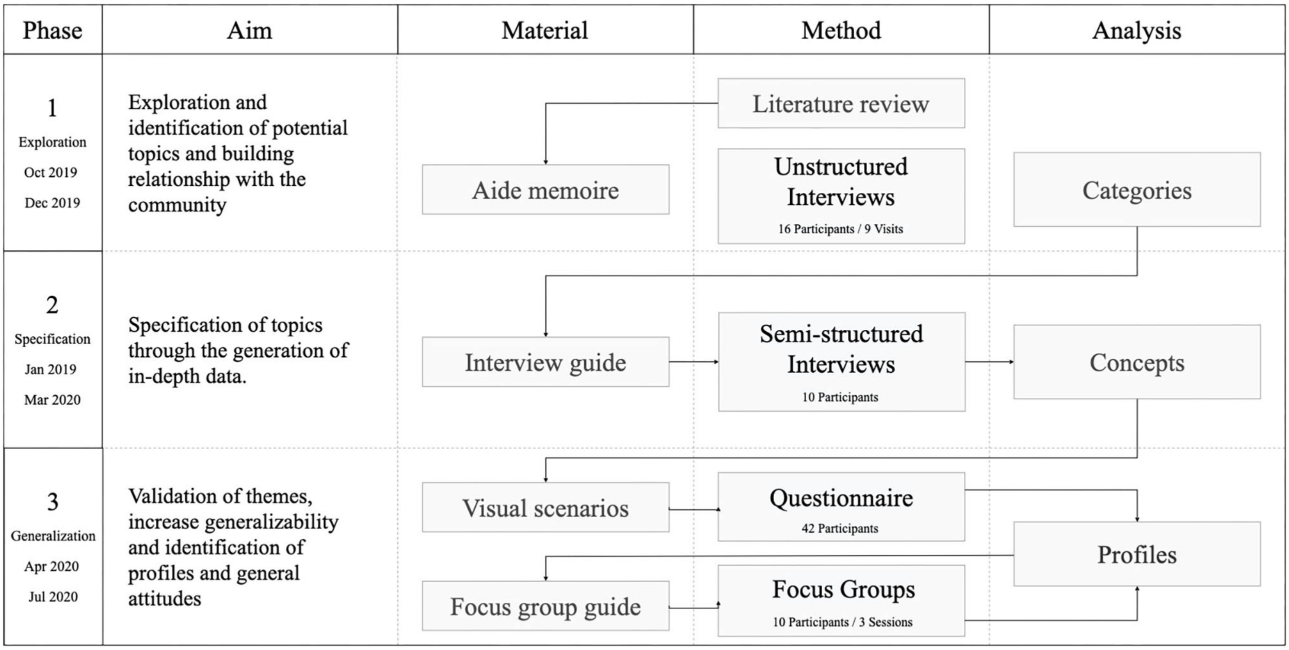


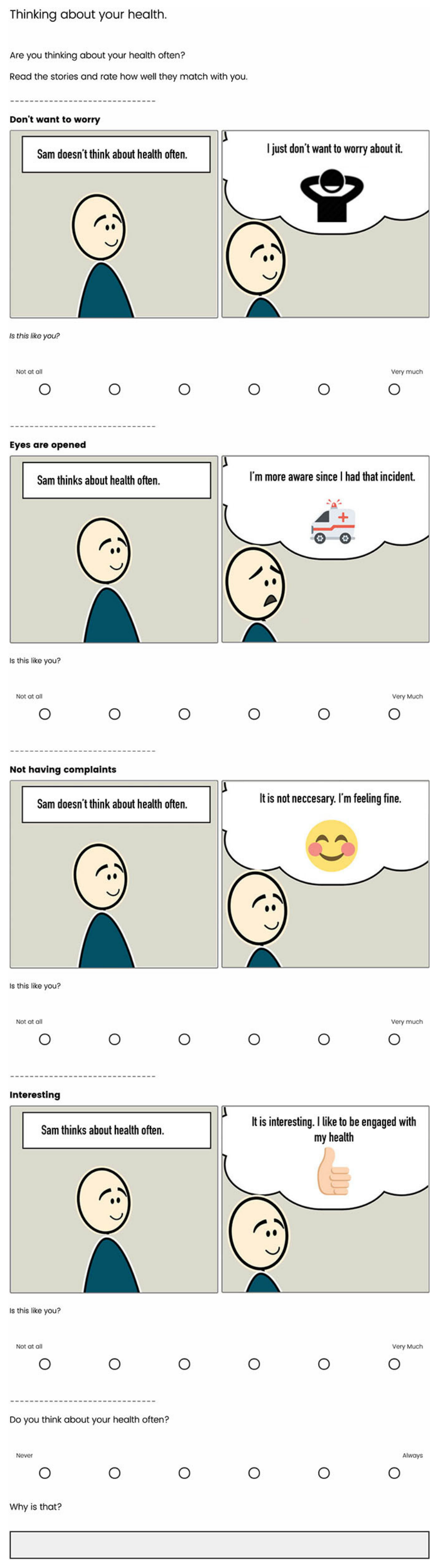
FIGURE 1 | Overview of project phases and corresponding methods, materials, and analysis products.

conjointly with emotional challenges when discussing sensitive and stigma-inducing topics (35). Therefore, the pre-established trusting relationship between participants and the researcher was an essential facilitator. The interviews ($N = 10$) were conducted at the community center in a separate room with the participants individually and took ∼30 min. We developed the interview guide structuring the interview based on the research directions from the first phase. For example, we explored how the participants perceived their health with the question: *"What do you have to do to become 100% healthy?"*. The interview guide is provided in **Supplementary Table 2**. The data was collected by audio-recording and transcribing the interviews. We progressed to the subsequent phase when we achieved theoretical saturation.

In the third phase, we validated and generalized the insights from phase two and discovered general attitudes through the data-driven profiles. Meanwhile, we had to consider the newly introduced COVID-19 regulations. Therefore, we developed a digital questionnaire which we distributed digitally to members of community centers. This questionnaire presented the resulting insights of the second phase and asked the participants to rate the extent to which they felt the insight reflected themselves. By distributing this online questionnaire, we reached a more extensive and diverse sample. In addition, we gathered quantitative data that we used to validate our preliminary results and develop data-driven attitude profiles. Questionnaires, frequently being long and textual, are at risk of being disengaged by their participants as they depend on reading comprehension. This risk holds especially true for participants with lower education attainment. The use of graphics in previous studies has successfully engaged low-literate participants with questionnaires (36). Therefore, we synthesized our insights toward visual two-frame storyboards. We executed several pilot sessions to reduce the chance that participant understandings would not match the story's original implication and adjusted any inaccuracies accordingly. A 6-point Likert scale accompanied the stories in the questionnaire. The stories were grouped under their representative category. Each group concluded with an open-ended question regarding the corresponding category. See **Figure 2** for an example of the *consciousness* page in the questionnaire. In addition, we asked participants to report their age, gender, educational attainment, and neighborhood. The online questionnaire was designed and distributed using Qualtrics. Finally, we performed focus groups to validate and contextualize the profiles that resulted from the questionnaire. Each focus group meeting consisted of three to four participants, lasted for ∼1 h, and was audio-recorded. The focus groups took place in a large and ventilated room at the community center that allowed maintaining 1.5-m distance between the participants according to the COVID-19 regulations.

Data Analysis

In phases one and two, we transcribed the audio recordings verbatim and analyzed them together with the field reports and qualitative questionnaire data using the software package Atlas Ti. Throughout the qualitative analysis, we followed the grounded-theory approach outlined by Corbin and Strauss (37), as it is specifically useful in discovering social processes focused on social change and improvement (38). We continuously broke down the data and collected it under similar content in the form of *concepts* using open coding techniques. For



Thinking about your health.

Are you thinking about your health often?

Read the stories and rate how well they match with you.

Don't want to worry

Is this like you?

Not at all ○ ○ ○ ○ ○ ○ Very much

Eyes are opened

Is this like you?

Not at all ○ ○ ○ ○ ○ ○ Very Much

Not having complaints

Is this like you?

Not at all ○ ○ ○ ○ ○ ○ Very much

Interesting

Is this like you?

Not at all ○ ○ ○ ○ ○ ○ Very Much

Do you think about your health often?

Never ○ ○ ○ ○ ○ ○ Always

Why is that?

FIGURE 2 | An example of the visual questionnaire distributed in phase 3. The storyboards represent the concepts found within the consciousness category.

example, we created the concept *perceived barriers* to refer to quotes where participants mentioned barriers that decreased their motivation to perform healthy behavior. Subsequently, we grouped related concepts toward overarching *categories* based on attitude theory constructs such as Beliefs, Feelings, Motivation, and Opportunity (19, 39). Two independent researchers (JF and IA) developed the concepts together to improve the reliability of the results.

In phase three, we imported the Likert scores of the concepts and categories obtained from the questionnaire as variables into SPSS. We performed k-means cluster analyses on the concepts based on Euclidian distance for health, healthcare, and eHealth with SPSS. We determined the optimal number of clusters with the Elbow method using the factoextra and NbClust packages in R. We used an ANOVA to identify the concepts with significant ($p < 0.05$) contribution to the cluster segmentation. The concepts with an insignificant contribution were removed from further analysis. To validate the clusters, we performed an ANOVA with the category scores as independent and the clusters themselves as dependent variables. Using a *post-hoc* ANOVA, we defined the resulting clusters based on significant differences between mean scores of the concept variables. We created profiles by further clarifying and enriching these clusters by analyzing the qualitative data from the questionnaire and focus group discussions. This was done by extending on the existing categories and concepts and using the same grounded-theory approach as used in previous phases. **Supplementary Table 4** shows an overview of the coding used for characterizing the profiles. Finally, we performed a principal component analysis (PCA) using the factoextra package in R to discover correlations between concepts from different profiles.

RESULTS

Participants

During the unstructured interviews in the first phase, we spoke with 16 different members of the community center. These members consisted of eight Vi's, two Vo's and six Kp's. In the second phase, we interviewed five Vo's and five Vi's. In phase one and two, we did not collect demographic data. In the third phase, 45 participants responded to the questionnaire. From these latter responses, we excluded three participants not living in our target neighborhood from analysis. The participants' mean age in this final sample was 52 years (SD = 11.10), 21% was male and 79% was female. Most of this sample (67%) had a low to medium education, which was defined as not having attained a follow-up education. This is relatively high compared to 59% in the Netherlands. Ten participants participated in the focus groups; two Kp's, five Vo's, and three Vi's.

Phase 1 and 2: Exploration and Specification

The unstructured interviews of phase one yielded 30 pages of field reports containing 85 coded segments. The semi-structured interviews of phase two yielded 10 interview transcripts

TABLE 1 | Concepts ($N = 29$) under their categories ($N = 9$) resulting from grounded theory analysis including number (N) of associated codes, description, and exemplary quotes (translated).

Concept	*N*	Description	Quote
Category: health beliefs [being healthy is…]			
Working on health	30	When one frequently performs healthy behavior such as physical activity and maintaining a healthy diet.	*"I'm eating healthy, I only drink in the weekends […] I frequently do yoga […] Yes I think that I'm being healthy" (Vo3)*
Absence of complaints	12	The absence of complaints, symptoms, and disease.	*"There was a time when I was heavier. I struggled with shortness of breath and cholesterol and I don't know what else." (Vi6)*
Participation	12	Being able to go out and participate in society.	*"The first thing you have to do is to get up early and just go somewhere […] Otherwise you will not have active contacts with people who provide a positive influence or create chances for you" (Vi3)*
Balance	10	Maintaining a balance between unhealthy and healthy behavior.	*"I have other things. I don't drink for instance so that makes up for it quite a lot." (Vo5)*
Life under control	10	When you have a roof above your head and no major financial or social struggles.	*"Unhealthy is when you don't have a roof above your head and you have to roam the streets." (Vi5)*
Category: consciousness [about health is impacted by…]			
Complaints	19	The experience of health-related symptoms and complaints.	*"I haven't visited the doctor in 30 years. My last painkiller I used when I was at high school" (Vo5)*
Incident	13	The consideration of a health-related incident in the past.	*"Yes, a significant impression. Before that [the incident] I was just flying blind." (Vo3)*
Concern	11	The extent to which one is concerned about their health.	*"You can come up with all sorts of graphs, but I don't, I just don't want to worry about it. Maybe it is just very easy the way I live." (Vi5)*
Interest	3	The level of interest one has in their health.	"It doesn't interest me […]. I just eat whatever I like" (Vi3)
Category: motivation [to perform healthy behavior is impacted by…]			
Future perspective	22	The consideration of its value toward future health.	*"How important is the future for you?" "Well, I just hope to continue like this." (Vo3)*
Perceived barriers	20	The amount of financial, social, and environmental barriers one perceives.	*"I have always had a one-sided diet. A lot of cheese for example. We didn't have a fridge at work." (Vi1)*
Feeling	6	The extent to it contributes to the subjective emotional state one experiences.	*"Do you think it's important to do it [performing healthy behavior]?" "Yes, it makes you feel better." (Vo2)*
Enjoyment	5	The extent to which it impacts the level of joy in one's life.	*"No, I don't really consider it [being healthy] that much. You also would want to enjoy life" (Vo4)*
Category: control [one perceives to have over health is impacted by…]			
Support	24	The amount of support one receives on managing their health.	"What facilitates you in doing it [healthy behavior]?" "To be honest, my friend. […] She supports me and shows me the ropes." (Vo2)
Self-efficacy	14	The level of capabilities one perceives to have to change health-related behavior	*"But you are not eager to quit, are you?" "I am my boy, however, I'm not able to. If you have a pill for me that I take and it makes me quit…" (Vi5)*
Chance	13	The belief that what happens regarding health is all based on chance and coincidence.	*"I'll not reach the age of 110, I'm not that healthy. Although, it doesn't say much actually because there are people who are 100 years old and they still smoke." (Vi5)*
Fatalism	5	The belief that what happens regarding one's health is subjugated to fate or destiny.	*"You can't really do something about it [getting sick]. The only thing you can do is watch out [for accidents], that is the only thing." (Vi3)*
Category: healthcare experience [is impacted by…]			
Communication	13	The quality of communication with the healthcare provider.	*"I would like them to take more time for people like me, who do not fully understand it. Sometimes I really feel like a foreigner." (Vo2)*
Autonomy	11	The need one has to deserve autonomy within the healthcare process.	*"I proposed it [not eating meat] once, however, my general practitioner told me not to do it. […] He didn't go into depth or asked me why I wanted that. He just advised me to keep eating meat" (Vo4)*
Authority	10	The amount of authority one perceives their healthcare provider to have.	*"It is not possible to change something about it yourself. If they say there is no solution then it has to be that way […] yes you know, they are the doctor." (Vo1)*
Personal	4	The need one has to be treated on a personal and humane level.	*It was a nice guy, a physiotherapist. He always brought a ball of Feyenoord (football club in the Netherlands). […] Yes, the soccer I liked. However, all the other stuff, walking around, walking with a sack, and all of that. At a certain moment, I thought like... man. (Vi5)*
Category: messages [reception is impacted by…]			
Source Interpretation	4	The extent to which one perceives a conflict between different sources (e.g., healthcare, media, social environment) of health messages.	*"The website that you visit… It can be that someone wrote it at home and it is not true. However, it can also be a doctor who wrote it, in that case, it is true." (Vo1)*

(Continued)

TABLE 1 | Continued

Concept	*N*	Description	Quote
Nuances	2	The extent to which one can understand and apply nuances within health messages.	*Yes, sugar, I have to minimize. […] Everything I have to minimize. Also, Carbohydrates. (Vi1)*
Rules	2	The extent to which one interprets health messages as rules.	*"Recently we have had this [healthy-lunch café] […] Everything must be healthy, and you are not allowed to eat meat. Well, I really like my piece of meat […]. In that case, just let me be unhealthy. I don't care; I just really like it." (Vi5)*
Doctor as information source	2	The fact that the health-related information came from a healthcare professional or not.	"I won't go and try out stuff from the internet and stuff. It has to come from the doctor." (Vo1)
Category: eHealth [intention to use is impacted by…]			
Enthusiasm	7	The belief in the positive aspects and potential of eHealth	*"Yes, I find that really interesting. […] You just have to ask google what you have to do. For example, I bumped my toe a few times, and then you get an answer." (Vi5)*
Anxiety	7	The level of anxiety one experiences toward (prospective) usage of eHealth.	*"It is too complicated. […] They told me I had to download something. Well, they did it for me. I don't know how it works." (Vo2)*
Exposure	5	The extent to which a person is exposed toward eHealth.	*"I'm not entirely up to date what it can mean to me. Maybe I'm still thinking in the old way. I don't know what I'm missing." (Vi3)*
Trust	1	The level of trust in technology and its related privacy and safety risks.	*"I always try to protect myself with anti-virus software […] If you have your gates open, you will collect all kinds of unwanted rubbish" (Vi2)*

containing 359 coded segments. The grounded theory analysis resulted in 58 concepts within nine categories related to attitudes toward health, healthcare, and eHealth. Examples of the categories found are: *consciousness* about health, *motivation* to perform healthy behavior and *satisfaction* toward healthcare. Examples of identified concepts are: *Interest* in health, *Perceived barriers*, and *loyalty* toward healthcare provider. **Table 1** presents an overview of the concepts and categories included in the third phase. We excluded categories conveying a limited number of concepts ($N = 1$) or not fitting the attitude theory constructs ($N = 1$). We selected the concepts to include ($N = 29$) in the third phase based on the number of associated coded segments and discussion by the two analysts.

Phase 3—Generalization

Profiles

The descriptive analysis of the overall sample revealed a high variance in the means of the different concepts. Variance ranged from 0.80 to 3.78 with a mean of 1.91. Therefore, it was all the more essential to investigate a segmented version of the data. The elbow method suggested that three clusters best segmented the data of each topic. We found significant differences between the category means, indicating the validity of our clusters. **Figure 3** presents a graphical representation of the mean scores characterizing the clusters and **Supplementary Table 3** offers a detailed overview. **Table 2** shows the demographic information of overall sample and the profiles. Although we found significant differences between the means of the concepts, we did not find significant differences between the clusters' demographic variables.

Regarding attitudes toward health, the majority was represented by the *Light-hearted* and *Concerned* profiles (both 38%), which were characterized by higher scores on *consciousness, motivation*, and *feeling*. The Concerned profile was differentiated based on higher scores for *concern, complaints*, and lower *control*. The *Encumbered* profile represented lower scores on *consciousness, motivation, self-efficacy*, and *interest* and higher scores on *perceived barriers*.

For the attitudes toward healthcare, the *Loyal* profile (60%) was the most significant. This profile was marked by higher scores on *satisfaction, personal, authority*, and *doctor as information source*. The *Disadvantaged* profile was characterized by lower scores on *satisfaction, communication, source interpretation, nuance, personal* and higher scores on *rules* and *autonomy*. The *Detached* profile contained no specific concept that differentiated it from the other profiles.

Regarding attitudes toward eHealth, the *Eager* (48%) and *Hesitating* (38%) profiles represented the majority of the attitudes and were both characterized by a higher score on *enthusiasm*. The *Hesitating* profile could be differentiated based on lower scores on *usage, trust,* and *exposure* and a higher score on *anxiety*. The *Indifferent* profile was marked by lower scores on *usage* and *enthusiasm.*

Qualitative Enrichment

The qualitative data from the questionnaire responses and three focus group discussions clarified and enriched the profiles with contextual information. **Table 3** presents representative quotes for each profile. Regarding the health profiles, within the *Concerned* profile, 81% of the questionnaire participants referred to the experience of medical complaints, symptoms, and limitations as a reason for being more conscious about health. Within the *Light-hearted* profile, 69% of the questionnaire participants referred being healthy and seeing the importance of it. What stood out within the *Encumbered* profile was that 50% of the questionnaire participants expressed not enjoying healthy behavior and experiencing internal barriers regarding motivation. During the focus group discussions, we found that most participants recognized themselves with the *Light-hearted* and *Concerned* profiles. It stood out that some

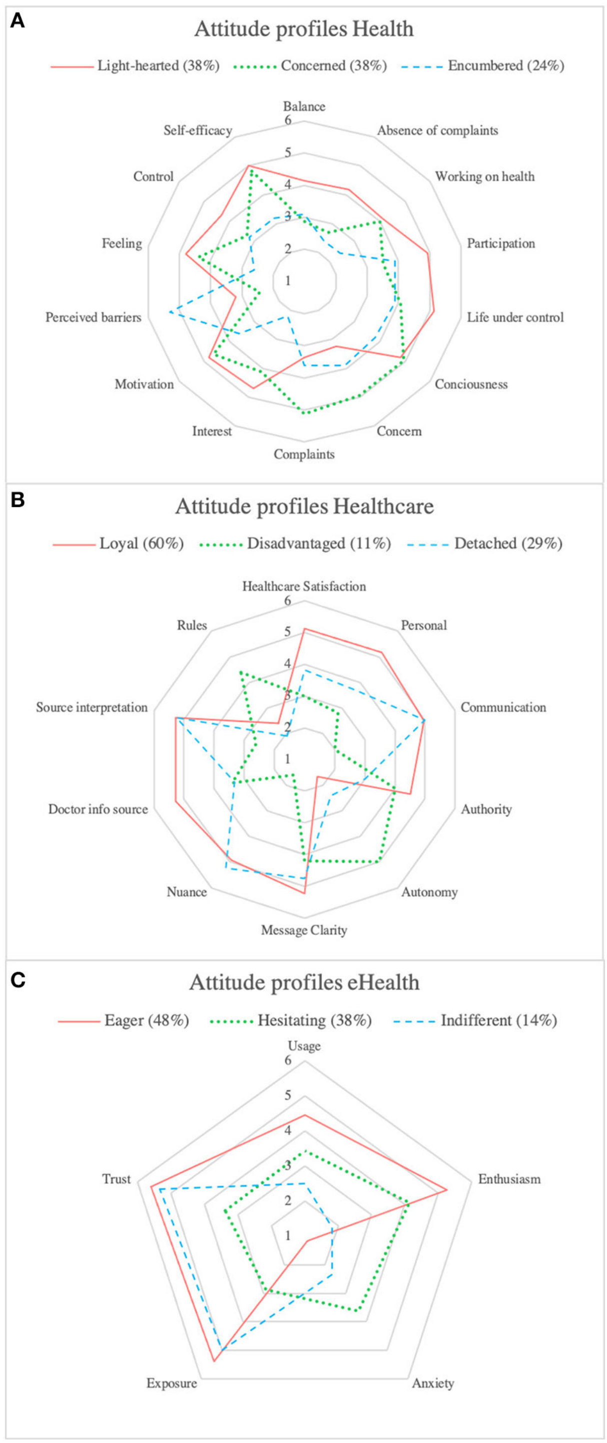


FIGURE 3 | (A) Radar graph of concept means of the attitude profiles toward health. **(B)** Radar graph of concept means of the attitude profiles toward healthcare. **(C)** Radar graph of concept means of the attitude profiles toward eHealth.

participants mentioned recognizing periods of the *Concerned* profile, especially when experiencing medical complaints or limitations. The participants did not fully identify with the *Encumbered* profile but rather ascribed this to an attitude they had in the past, frequently seen in the youth, or an attitude they "sometimes" have.

> *"Sometimes I have, just like [Encumbered], my concerns about things. In that case you can find yourself in a slump. Life is not always going your way." (Vi10)*

Regarding the profiles toward healthcare, within the *Loyal* profile, 92% of the questionnaire participants referred to positive experiences such as good advice, a professional who shows understanding, and additional room for questions and discussion. Within the *Detached* profile, 46% of the questionnaire participants mentioned distrusting their doctors and not visiting them often. For the *Disadvantaged* profile, 67% of the questionnaire participants referred to communication barriers such as lack of time, complicated language, feelings of anxiety, and not being taken seriously. During the focus groups, the participants could identify with the *Loyal* and *Detached* profile. Regarding the Detached profile, which we positioned as an attitude not wanting to be dependent on healthcare, we gathered additional evidence that some of our participants were distrusting and wanting to avoid healthcare:

> *"Yes, I think I am a bit like [Detached]. Because I am not a doctor visitor. I seldom visit the doctor. [...] I do not really like to take medication. Only when it is really necessary." (Vi11)*

Regarding the profiles toward eHealth, within the *Eager* profile, 75% of the participants referred to using eHealth and seeing the benefits of using it. Although we also found such positive responses toward eHealth within the *Hesitating* profile (56%), 38% of this profile's participants also referred to eHealth as not worth the effort, better suited for the youth, or being perceived more like gadgets. The *Indifferent* profile hosted participants referring to not wanting to be involved with technology for health (50%). During the focus groups, most participants identified with the *Eager* and *Hesitating* profiles. What stood out was that some participants who initially were *Indifferent* toward eHealth started to become interested in it because of the focus group discussion:

> *"Well, I definitely want to use it. Suppose I can do it with a device or something. My daughter also wanted to install an app for counting steps. However, I don't do a lot with phones. It is only now that we have this conversation that I start to think that maybe I should investigate it some more. I only use it for calling and text messaging. I do like it, but I don't know it."* (Vi11).

General Attitudes

By investigating the inter-profile relationships, we could identify two attitudes toward health, healthcare, and eHealth in general. **Figure 4** displays an overview of these attitudes. Correlation coefficients can be found in **Supplementary Table 5**. The most significant general attitude, *Optimistically Engaged* could be described by positivity toward health, healthcare, and eHealth. It is related to being conscious about health, motivated to perform healthy behavior, satisfied with and loyal toward healthcare

TABLE 2 | Questionnaire respondent characteristics.

	N	CV	Age		Gender %		Education %	
			M	SD	Male	Female	Low	High
Sample	42	1.91	52	11.10	21	79	67	33
Concerned	16	1.46	54	9.70	31	69	75	25
Light-hearted	16	1.06	48	13.07	0	100	69	31
Encumbered	10	2.00	51	8.89	40	60	50	50
Loyal	25	1.06	53	10.81	24	76	68	32
Detached	12	1.05	48	8.62	0	100	58	42
Disadvantaged	5	2.01	48	12.54	60	40	80	20
Eager	20	1.21	48	12.44	15	85	65	35
Hesitant	16	1.72	55	8.34	37	63	75	25
Indifferent	6	1.96	52	12.10	0	100	50	50

TABLE 3 | Exemplary quotes per profile.

Profile	Quote
Light-hearted	*"I do what I can and what I want. When I feel good, it is good."*
Concerned	*"I try to prevent my health complaints from taking over my life. It is tough sometimes though […]"*
Encumbered	*"Exercising is exhausting and painful"*
Loyal	*"I feel that they listen well to me. Everything is explained clearly. Messages are clear and informative."*
Disadvantaged	*"They left me for too long with my complaints, and I'm not taken seriously"/"Sometimes they come with difficult words"*
Detached	*"I'm not coming to the doctor often, but when I do, I have the feeling they listen well. Probably extra because I never visit the doctor."*
Eager	*"I see it as a push in the back, and it's fun to keep track of things. I'm already above 950 km this year:"*
Hesitating	*"The technology of nowadays is more something for the younger generation"*
Indifferent	*"Not feeling like it"*

services, and open and enthusiastic about the use of eHealth. It was defined by the relationship between the characterizing scores of the *Light-hearted (consciousness, motivation, feeling,* and *interest), Loyal (satisfaction, clarity, doctor as info source, and personal),* and *Eager (usage* and *enthusiasm)* profile. The average size, based on the questionnaire respondents, of the combination of these profiles is 48%. The second general attitude, *Doubtfully Disadvantaged,* reflected perceived barriers and low self-efficacy, difficulties understanding health messages, wanting more autonomy in the healthcare process, distrusting healthcare, anxiety toward technology, and lack of exposure regarding eHealth. It was defined by the relationship between the scores of the *Encumbered (low self-efficacy* and *perceived barriers), Disadvantaged (source interpretation, rules, nuance,* and *communication barriers),* and *Hesitating (exposure, anxiety,* and *trust)* profile. The average size, based on the questionnaire respondents, of the combination of these profiles is 25%. The *Concerned, Detached*, and *Indifferent* profiles did not have any specific relations with other profiles. They should be seen as individual profiles that could exist in any combination with other profiles. However, the *concerned* profile's substantial representation within the questionnaire respondents (38%) makes it important to consider further. This profile was characterized by the experience of complaints, high concern, and low feelings of control because of the experience of a health-related incident or continuous experience of health complaints.

DISCUSSION

Main Findings

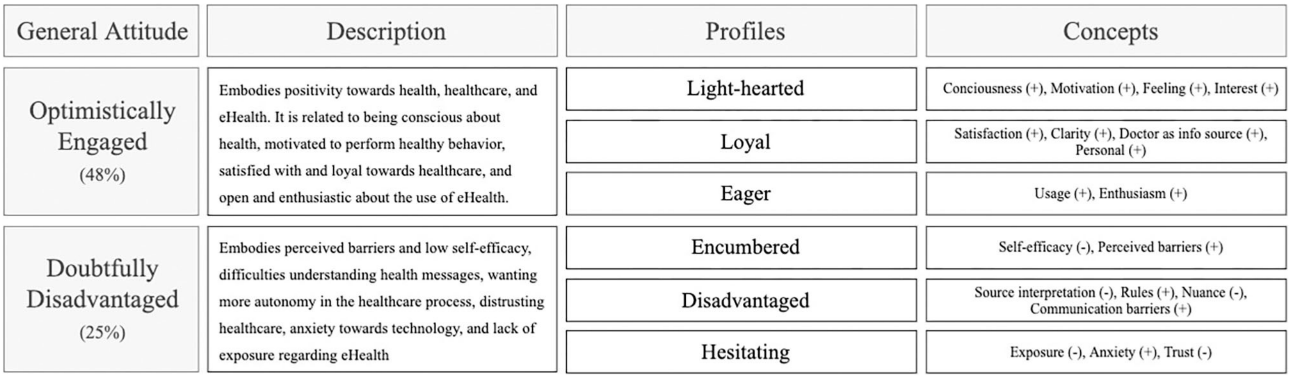
This study aimed to develop design-relevant knowledge about the attitudes of people with a low SES toward their health, healthcare and eHealth. Through a CBPR approach consisting of three phases, we identified two general attitudes based on nine distinct profiles. This knowledge could be used to develop a better understanding of existing attitudinal knowledge and to propose design recommendations that facilitate the alignment of health services toward these attitudes.

Relevance and Implications

Since most of the attitudes toward health, healthcare, and eHealth were positive, we believe that there is a willingness from a large part of the target group to adopt eHealth interventions to improve their lifestyle. Nevertheless, we discovered a diverse range of different attitudes that have different implications for the design of eHealth interventions. The attitudes represented by the profiles can be used to develop design recommendations to improve the alignment of eHealth interventions toward attitudes of low SES groups.

Optimistically Engaged

The profiles (*Light-hearted, Loyal, and Eager*) represented by this general attitude have similarities and contradictions with existing literature. The *Light-hearted* profile was represented by high consciousness about health. Contrastingly, other studies found that low SES populations have a less conscious attitude toward health and think less about the future (13, 40). Complex social situations, caring responsibilities (29), and time and



General Attitude	Description	Profiles	Concepts
Optimistically Engaged (48%)	Embodies positivity towards health, healthcare, and eHealth. It is related to being conscious about health, motivated to perform healthy behavior, satisfied with and loyal towards healthcare, and open and enthusiastic about the use of eHealth.	Light-hearted	Conciousness (+), Motivation (+), Feeling (+), Interest (+)
		Loyal	Satisfaction (+), Clarity (+), Doctor as info source (+), Personal (+)
		Eager	Usage (+), Enthusiasm (+)
Doubtfully Disadvantaged (25%)	Embodies perceived barriers and low self-efficacy, difficulties understanding health messages, wanting more autonomy in the healthcare process, distrusting healthcare, anxiety towards technology, and lack of exposure regarding eHealth	Encumbered	Self-efficacy (-), Perceived barriers (+)
		Disadvantaged	Source interpretation (-), Rules (+), Nuance (-), Communication barriers (+)
		Hesitating	Exposure (-), Anxiety (+), Trust (-)

FIGURE 4 | Overview of the general attitudes resulting from the principal component analysis and their corresponding profiles and concepts.

energy constraints (28) result in little room to act toward and think about long-term investments such as a healthy lifestyle. These contrasting findings could be explained by the current living situation of our participants. Almost all participants were either retired, unemployed or disabled and therefore were not constrained by their jobs or worried about finances as they receive financial support from the government. In Wardle and Steptoe (13), all participants were employed, and in Coupe et al. (29), only 13% of the population was employed. Yet, the finding came from interviews with healthcare providers and not from the low SES population themselves. In a previous study in a community center in Rotterdam, participants indicated that a lack of time was a major reason for not visiting a community center (41). Therefore, we argue that some participants in our sample, having the time to visit a community center, also had more time and capacity to think about and act toward a healthy lifestyle. Therefore, we recommend that eHealth researchers and designers should become aware of the person's life situation and use this knowledge to determine whether the person has the capacity available to fit the intervention into their life. People that do not have this capacity would benefit more from services that deliver support in social or financial aspects (42, 43). We argue that people that do have motivation and consciousness could benefit from being empowered to play a major role in their health management. This could be achieved through shared-decision making, providing health information and facilitating self-management (44). It remains important for healthcare providers to be aware of this attitude as it is known that clinician perceptions of patients with a low SES have been shown to affect clinical decision making. Based on common beliefs about people with a low SES, physicians tend to delay diagnostic testing, prescribe more generic medications and avoid referral to specialty care and potentially lifestyle interventions (45). The finding that most of our participants were doctor dependent (*Loyal*, 60%) conforms to other studies that claim that people with a low SES are loyal to and rely on their doctor's advice (14, 46). Moreover, we found that our participants highly valued a personal interaction with their care provider. The importance of this personal touch is mentioned in various other studies on the interaction between people with low SES and healthcare providers (43, 47–49). Since current healthcare systems are moving from a doctor-says, patient-does model toward a model of shared decision making and self-management, we expect that people relying on their doctor's advice will experience increasing difficulties in their health management. To improve the alignment of eHealth communication to these attitudes, we recommend that professionals should be mindful of "dehumanizing" healthcare, as digital interactions lack the nuances of human interaction (50). Therefore, eHealth interventions should be designed to incorporate and enhance personal communication, interaction, and relationships with care providers, family members, and peers. This could be done for example by integrating a social role in the intervention through interactive and animated computer characters. Through simple speech, hand gestures and other non-verbal cues, these characters could simulate face-to-face counseling to establish trust and rapport in a virtual environment (51).

Doubtfully Disadvantaged

The *Encumbered, Disadvantaged*, and *Hesitating* profile, that represented this attitude, all embodied a perceived lack of control related to one's health, healthcare, or eHealth. Various previous studies support this finding. The lack of control over health is attributed to lower problem-solving skills (52), environmental deprivation (53), and financial, environmental, and social limitations (40, 54). Therefore, we recommend considering self-efficacy and perceived control enhancing strategies within eHealth interventions. Goal setting has been mentioned as a potentially successful strategy in various studies regarding other low SES populations (8, 29, 47). A possible implementation is through persuasive game design. Through the game world the user could acquire feelings of competence and transfer these toward the real world (55). For example, one could help an avatar to progress through different life goals by earning points based on healthy snack choices (56). In addition, various studies also mention social support as a potentially effective strategy (28, 43). Emotional support could be offered through supportive conversations and buddy systems, informational support from educational information from peers and providers and appraisal could be offered through peers, providers, or the eHealth system itself (57). In addition, designers could think of ways to make technologies and information more accessible and easier to integrate into the persons' daily life. For example, cardiac telerehabilitation allows to reach patients in their home-environment and motivate them to participate even though they do not have the means (physically as well cognitive) to visit the rehabilitation center (58, 59). We found that participants characterized by the *Disadvantaged* profile were experiencing communication difficulties in the healthcare setting. Especially assessing and applying health knowledge was perceived as problematic. It is striking that this profile only represented a mere 12% of our sample, while these difficulties are widely discussed in previous studies on this topic (46, 60). Since our participants were proficient in the Dutch language, we argue that communication for them was less problematic. Moreover, combatting health literacy is currently high on the agenda (61). In fact, in the Netherlands, 60% of healthcare professionals report adapting their communication toward their patients' needs (62). Nevertheless, to include this part of the population, eHealth interventions should accommodate for varying literacy levels, for example by using visual aids and plain language. Besides, according to studies related to other low-SES and literacy populations, medical advice should be tailored to increase its relevance (28, 48, 49). For example, by using lab results to select the appropriate advice given in a patient portal (48).

The participants within the *Hesitating* profile reported being unsure about using eHealth because they were unaware of how it could be of personal value. A previous study found that people who have a poor understanding of what eHealth can do for them have little interest in signing up and using it (50). It also seems that healthcare providers do not actively promote such interventions and provide little encouragement to use them, as they expect the intervention will not be adopted (29). In addition, this subgroup of participants expressed concerns about not being capable enough to use eHealth. This finding is also reflected by Latulipe et al. (48), where most usage concerns of low-income older adults relate to the difficulty of initially logging on to a system. Therefore, we recommend professionals to consider the perceived usefulness and usability of the eHealth intervention. Past studies have shown that this can be achieved through supportive healthcare providers and peers who can promote the eHealth interventions and provide technical assistance during usage (48, 50, 63). One upcoming medium through which these interactions can take place is through social media. Social media is used as an effective recruitment and engagement medium for eHealth applications (50) and for people with lower income and education (64). Another possibility to improve perceived usability is by offering primary task support through self-monitoring wearable devices (e.g., activity trackers) (65), reduction (e.g., list with food choices), or tunneling (e.g., offer treatment opportunities after an interactive test about tobacco addiction) (66).

The Concerned Profile

The participants represented by the *Concerned* profile indicated being motivated and conscious because they were living with medical limitations or have recently experienced a health-related incident. This concerned attitude could serve as potential entry point for researchers and designers to motivate healthy behavior. While people might already be aware of the susceptibility and severity of getting a disease, they might benefit from convenient cues to action such as reminders and suggestions provided either by a peer, professional, or system (67). According to Bukman et al. (28), people with a low SES are especially motivated by the feedback they receive from their bodies. This conforms to some participants mentioning that their attitude had changed throughout their lives, resulting from experiencing health complaints or incidents. Therefore, it is challenging to motivate these individuals to engage in preventive behavior when they do not yet perceive complaints. Therefore, following Bukman et al. (28), we recommend that for people that do not have the concern (yet), feedback should be provided in a visual, meaningful, and directly applicable way that conforms to the beliefs of the target group. According to Orji et al. (67), self-monitoring, simulation and personalization and tailoring strategies are effective to help individuals develop accurate perceptions of own risk. Nevertheless, we could argue that data recorded by most activity trackers and self-monitoring applications currently is still of little value in facilitating meaningful reflection on lifestyle. In a previous study it was found that the participants from a low SES neighborhood rarely analyzed their self-monitoring experiences to derive insight about the meaning of data for their well-being (68). One example of providing meaningful data is a smoking app that displays, besides the number of days without cigarettes, also the amount of money the person has saved by not smoking.

Recommendations

Based on our results, the reflection with previous literature, and existing recommendations, we propose some final recommendations for improved eHealth alignment to attitudes in low SES populations. First, we have identified a large part of our sample embodied an optimistic and engaged general health attitude. According to this attitude, someone is motivated, conscious, satisfied with healthcare, and open toward eHealth. Hence, we expect that for this attitude, healthcare services, and interventions are generally appropriate.

However, we also identified attitudes that are less in line with our current processes and expectations. We identified profiles that embodied a disinterested, resisting attitude toward healthcare (*Detached*) and eHealth (*Indifferent*). We argue that tailoring eHealth interventions toward such attitudes is resource-intensive and would be more effective when directed at attitudes that are positive yet require support. These attitudes, in our study identified under the *doubtfully disadvantaged* general attitude, currently seem to hold the most potential for tailoring efforts. While the Encumbered profile benefits from social and emotional support, the disadvantaged profile benefits from additional support in understanding verbal and written health information and guidance during the healthcare process. The Hesitating profile has an open yet unsure attitude toward eHealth and therefore benefit from supportive and technology promoting healthcare professionals and peers. We recommend professionals to focus on these attitudes specifically, to become aware of the corresponding needs, and subsequently use and design eHealth as a tool to respond to these needs. While doing so, professionals are advised to establish a trustful relationship with the target group, which could be achieved through personal contact and/or through trusted doctors or other key persons (18). In addition, future research endeavors should take into account the challenges related to recruiting and researching vulnerable populations and take the appropriate methodological strategies to minimize the impact of those challenges. This could help improve the accessibility and affordability of eHealth innovations and thereby help equalizing inequalities in healthcare.

Strengths and Limitations

This study addressed the ever-increasing gap in health disparities by giving voice to a target group that is frequently overlooked in health research. Traditional approaches have received criticism as they, when executed irresponsibly, bring forth mistrust, feelings of stigmatization, and anxiety (69). CBPR has gained increasing attention in addressing ethical challenges in health research, as it encourages equity and shared decision-making and increases community involvement (21). By taking this approach, we ensured that our participants felt comfortable, safe, and especially involved during the research activities. The resulting insights directly carry our participants' voices and are, therefore, a meaningful contribution to responsible digital health. While frequently people with a low SES are expected to adapt their attitudes toward the intervention, we aim to have a more complete idea of how we should design interventions to be adapted to them.

Although our study provides an in-depth insight into the attitudes of people living in a low SES neighborhood, the results are not generalizable toward all low SES contexts. First, we aimed at limiting possible feelings of stigmatization by sampling on neighborhood SES. This would make it difficult to relate the findings directly to other studies that select participants on individual measures of SES (i.e., education, income, and occupation). Yet, this different selection criterion allowed us to target a group that would otherwise have been excluded. For example, the questionnaire demographics indicate a relatively high percentage of participants who attained a follow-up education. In traditional studies, this part of the sample would have been seen as high-SES and therefore excluded from the study. Socioeconomic determinants and barriers leading to disparities in health behavior are complex (54, 70). Capturing them merely based on individual determinants is therefore problematic and has accumulated critique over the years (71). Instead, our focus on neighborhood SES takes into account other factors that have proven to have a significant relation to poor health outcomes (i.e., availability of healthy food, experiences of discrimination, and neighborhood poverty) (25, 26).

Another factor that could impact the generalizability is the context of the community center. According to an earlier report of another community center near Rotterdam, 36% of the visitors were unemployed (72). This percentage is significantly higher than the neighborhood in general (9.4%) and Rotterdam (7%) (73). Since our participants had the motivation to visit the community center, they could also have been more motivated to perform healthy behavior. Finally, it should be taken into account that this study has taken place in an urban context with sufficient governmental support, developed infrastructure, and social support. Therefore, the results are not directly applicable to countries that do not have these facilities. While the results themselves might not be directly generalizable to other low SES populations, they provide a deep and contextualized understanding of a sample of the target group that can be applied in the design of eHealth interventions. According to Crouch and McKenzie (74), such research inquiries in naturalistic settings often seek to discover social insights that extend beyond initial observations. This requires the researcher to be immersed in the research field, establish continuing fruitful relationships with respondents and through theoretical contemplation to address the research problem in depth. A small number of cases will facilitate the researcher's close association with the respondent. A review of CBPR approaches in the health domain confirms this statement as it reports sample sizes of roughly the same order of magnitude (75). Future research could be aimed at generalizing the results (e.g., profile characteristics) in larger-scale sample sizes. Finally, the concepts identified in this research are, although informed by supporting themes in literature, data driven and not a priori based on a specific theory or model. Hence, they provide a deeper layer and a supplementary perspective to existing knowledge. Nevertheless, researchers should act with discretion when interpreting the resulting insights using existing theory.

Conclusion

To develop successful eHealth interventions that support people with a low SES in achieving a healthy lifestyle, it is crucial to consider their attitude toward this technology and their health and healthcare in general. This study explored attitudes of people living in a low SES neighborhood toward their health, healthcare, and eHealth using a community-based participatory research approach. This unique approach helped us discover novel and bottom-up insights that strengthen our current understanding of these attitudes. This understanding allows researchers and designers to have a more nuanced view of the attitudes in low SES populations. Intervention developers should be mindful of differentiating life situations, motivations, healthcare needs, and eHealth expectations. eHealth should fit into the person's daily life, ensure personal communication, be perceived as usable and useful, adapts its communication to literacy level and life situation, allow for meaningful self-monitoring and embody self-efficacy enhancing strategies. When these recommendations are taken into account when developing eHealth interventions for people with a low SES, these interventions' alignment with their attitudes will improve. This will result in interventions that are more acceptable, satisfactory, and user-friendly. Consequently, eHealth interventions will finally move from widening toward narrowing current health disparities and thus align with societal health responsibilities.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Research Ethics Committee (HREC)-TU Delft. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JF, IA-D, TR, AE, NC, HB, JK, and VV contributed to conception and design of the study. JF planned and executed the CBPR process. IA-D and JF were involved in the analysis and interpretation of the data. JF was responsible for the writing of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version.

FUNDING

This project was funded by Medical Delta (MD 2.0 program), Capri Cardiac Rehabilitation, and the Netherlands Cardiovascular Research Initiative: An initiative with support of the Dutch Heart Foundation, CVON2016-12 BENEFIT.

ACKNOWLEDGMENTS

We thank the employees, volunteers, and visitors of the community center for their cooperation and participation in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2021.690182/full#supplementary-material

REFERENCES

1. Stringhini S, Sabia S, Shipley M, Brunner E, Nabi H, Kivimaki M, et al. Association of socioeconomic position with health behaviors and mortality. *JAMA*. (2010) 303:1159–66. doi: 10.1001/jama.2010.297
2. Drewnowski A, Moudon AV, Jiao J, Aggarwal A, Charreire H, Chaix B. Food shopping behaviors and socioeconomic status. *NIH Public Access*. (2014) 38:306–14. doi: 10.1038/ijo.2013.97
3. Mackenbach JP, Stirbu I, Roskam AJR, Schaap MM, Menvielle G, Leinsalu M, et al. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med*. (2008) 358:2468–81. doi: 10.1056/NEJMsa0707519
4. Shishehbor MH, Litaker D, Pothier CE, Lauer MS. Association of socioeconomic status with functional capacity, heart rate recovery, all-cause mortality. *J Am Med Assoc*. (2006) 295:784–92. doi: 10.1001/jama.295.7.784
5. Bull ER, Dombrowski SU, McCleary N, Johnston M. Are interventions for low-income groups effective in changing healthy eating, physical activity and smoking behaviours? A systematic review and meta-analysis. *BMJ Open*. (2015) 4:1–9. doi: 10.1136/bmjopen-2014-006046
6. Busch MCM, Schrijvers CTM. Effecten van leefstijlinterventies gericht op lagere sociaaleconomische groepen. *Rivm*. (2010) 81:39–42.
7. Busch M, van der Lucht F. *Effecten van preventieve interventies: zijn er verschillen tussen mensen met een lage en een hoge sociaaleconomische status?* RIVM (2012). p. 34.
8. Michie S, Jochelson K, Markham WA, Bridle C. Low-income groups and behaviour change interventions: a review of intervention content, effectiveness and theoretical frameworks. *J Epidemiol Commun Health*. (2009) 63:610–22. doi: 10.1136/jech.2008.078725
9. Reiners F, Sturm J, Bouw LJW, Wouters EJM. Sociodemographic factors influencing the use of ehealth in people with chronic diseases. *Int J Environ Res Public Health*. (2019) 16:645. doi: 10.3390/ijerph16040645
10. Garavand A, Mohseni M, Asadi H, Etemadi M, Moradi-Joo M, Moosavi A. Factors influencing the adoption of health information technologies: a systematic review. *Electronic Phys*. (2016) 8:2713–8. doi: 10.19082/2713
11. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. *MIS Q Manage Information Syst*. (2003) 27:425–78. doi: 10.2307/30036540
12. Ajzen I. The theory of planned behavior organizational behavior and human decision processes. *Organ Behav Hum Decis Process*. (1991) 50:179–211. doi: 10.1016/0749-5978(91)90020-T
13. Wardle J, Steptoe A. Socioeconomic differences in attitudes and beliefs about healthy lifestyles. *J Epidemiol Commun Health*. (2003) 57:440–3. doi: 10.1136/jech.57.6.440
14. Schröder SL, Fink A, Richter M. Socioeconomic differences in experiences with treatment of coronary heart disease: a qualitative study from the perspective of elderly patients. *BMJ Open*. (2018) 8:1–8. doi: 10.1136/bmjopen-2018-024151
15. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, ehealth literacy, and attitudes toward computer/internet use. *J Med Internet Res*. (2013) 15:645. doi: 10.2196/jmir.2645
16. Birks MJ, Chapman Y, Francis K. Breaching the wall. *J Transcult Nurs*. (2007) 18:150–6. doi: 10.1177/1043659606298617
17. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Med Res Methodol*. (2014) 14:1–29. doi: 10.1186/1471-2288-14-42
18. Stuber JM, Middel CNH, Mackenbach JD, Beulens JWJ, Lakerveld J. Successfully recruiting adults with a low socioeconomic position into community-based lifestyle programs: a qualitative study on expert opinions. *Int J Environ Res Public Health*. (2020) 17:2764. doi: 10.3390/ijerph17082764
19. Eagly AH, Chaiken S. The advantages of an inclusive definition of attitude. *Soc Cogn*. (2007) 25:582–602. doi: 10.1521/soco.2007.25.5.582
20. Viswanath K, Kreuter MW. Health disparities, communication inequalities, and eHealth. *Am J Prev Med*. (2007) 32:S131–3. doi: 10.1016/j.amepre.2007.02.012
21. Israel BA. *Methods for Community-Based Participatory Research for Health*. San Francisco, CA: John Wiley and Sons Ltd (2013).
22. Unertl KM, Schaefbauer CL, Campbell TR, Senteio C, Siek KA, Bakken S, et al. Integrating community-based participatory research and informatics approaches to improve the engagement and health of underserved populations. *J Am Med Informatics Assoc*. (2016) 23:60–73. doi: 10.1093/jamia/ocv094
23. Henderson VA, Barr KLC, An LC, Guajardo C, Newhouse W, Mase R, et al. Community-based participatory research and user-centered design in a diabetes medication information and decision tool. *Progress Commun Health Partnersh Res Educ Action*. (2013) 7:171–84. doi: 10.1353/cpr.2013.0024
24. Auerswald CL, Piatt AA, Mirzazadeh A. *Research with Disadvantaged, Vulnerable and/or Marginalized Adolescents*. Innocenti Research Briefs (2017). doi: 10.18356/a054ca5b-en

25. Schüz B. Socio-economic status and theories of health behaviour: time to upgrade a control variable. *Br J Health Psychol.* (2017) 22:1–7. doi: 10.1111/bjhp.12205
26. Winkleby MA, Cubbin C. Influence of individual and neighbourhood socioeconomic status on mortality among black, Mexican-American, and white women and men in the United States. *J Epidemiol Commun Health.* (2003) 57:444–52. doi: 10.1136/jech.57.6.444
27. Kock L, Brown J, Hiscock R, Tattan-Birch H, Smith C, Shahab L. Individual-level behavioural smoking cessation interventions tailored for disadvantaged socioeconomic position: a systematic review and meta-regression. *Lancet Public Health.* (2019) 4:e628–44. doi: 10.1016/S2468-2667(19)30220-8
28. Bukman AJ, Teuscher D, Feskens EJM, Van Baak MA, Meershoek A, Renes RJ. Perceptions on healthy eating, physical activity and lifestyle advice: opportunities for adapting lifestyle interventions to individuals with low socioeconomic status. *BMC Public Health.* (2014) 14:1036. doi: 10.1186/1471-2458-14-1036
29. Coupe N, Cotterill S, Peters S. Tailoring lifestyle interventions to low socio-economic populations: a qualitative study. *BMC Public Health.* (2018) 18:1–15. doi: 10.1186/s12889-018-5979-3
30. Dekkers T, Hertroijs DFL. Tailored healthcare: two perspectives on the development and use of patient profiles. *Adv Ther.* (2018) 35:1453–9. doi: 10.1007/s12325-018-0765-2
31. SCP. *Sociaal-Economische Status per postcodegebied* (2019).
32. Gemeente Rotterdam. *Onderzoek en Business Intelligence*. Wijkprofiel 2014-2016-2018-2020 (2020). Retrieved from: https://wijkprofiel.rotterdam.nl/nl/2020/rotterdam/charlois
33. De Graaf P. *Onderzoek en Business Intelligence. Feitenkaart opleidingsniveau Rotterdam op gebieds- en buurtniveau 2016*. Gemeente Rotterdam (2018).
34. Wilson C. Unstructured Interviews. *Inter Tech UX Pract*. (2014). p. 43–62. doi: 10.1016/b978-0-12-410393-1.00003-x
35. Costa JA, Renzetti CM, Lee RM. Researching sensitive topics. *J Market Res.* (1994) 31:437. doi: 10.2307/3152234
36. Anderson IMFD, Barkbu B, Lusinyan L, Muir D. Assessing the graphic questionnaire used in digital literacy training. *Issues Informing Sci Information Technol.* (2019) 16:113–26. doi: 10.28945/4301
37. Corbin JM, Strauss A. Grounded theory research: procedures, canons, evaluative criteria. *Qual Sociol.* (1990) 13:3–21. doi: 10.1007/BF00988593
38. Boer F. De grounded theory approach: een update. *Kwalon.* (2011) 1:25–33.
39. Fazio RH, Towles-Schwen T. The MODE model of attitude-behavior processes. In: Sherman JW, Gawronski B, Trope Y, editors, *Dual-Process Theories in Social Psychology.* New York, NY: Guilford Press (1999). p. 97–116.
40. Cutler DM, Lleras-Muney A. Education and health: evaluating theories and evidence. *Natl Bureau Econ Res.* (2006) 37:29–60. doi: 10.3386/w12352
41. Peerdemann M, Fonds O. *Nederland over Gemeenschapsaccommodaties*. Synovate (2010).
42. Heutink H, Diemen D, Van, E.lzenga M, Kooiker S. Wat hebben mensen met lage en hoge (re) SES zelf te zeggen over gezond leven? *Tsg.* (2010) 1:10–14. doi: 10.1007/s12508-010-0009-6
43. Troelstra S, Magnée T, Koopman E, Harting J, Nagelhout G. *Stoppen-met-rokeninterventies voor mensen met een lagere sociaaleconomische positie*. Trimbos Instituut (2020). p. 61.
44. Elwyn G, Dehlendorf C, Epstein RM, Marrin K, White J, Frosch DL. Shared decision making and motivational interviewing: achieving patient-centered care across the spectrum of health care problems. *Ann Fam Med.* (2014) 12:270–5. doi: 10.1370/afm.1615
45. Arpey NC, Gaglioti AH, Rosenbaum ME. How socioeconomic status affects patient perceptions of health care: a qualitative study. *J Primary Care Community Health.* (2017) 8:169–75. doi: 10.1177/2150131917697439
46. Yin HS, Dreyer BP, Vivar KL, MacFarland S, Van Schaick L, Mendelsohn AL. Perceived barriers to care and attitudes towards shared decision-making among low socioeconomic status parents: role of health literacy. *Acad Pediatr.* (2012) 12:117–24. doi: 10.1016/j.acap.2012.01.001
47. Bull ER, McCleary N, Li X, Dombrowski SU, Dusseldorp E, Johnston M. Interventions to promote healthy eating, physical activity and smoking in low-income groups: a systematic review with meta-analysis of behavior change techniques and delivery/context. *Int J Behav Med.* (2018) 25:605–16. doi: 10.1007/s12529-018-9734-z
48. Latulipe C, Gatto A, Nguyen HT, Miller DP, Quandt SA, Bertoni AG, et al. Design considerations for patient portal adoption by low-income, older adults. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015). doi: 10.1145/2702123.2702392
49. Schaffler J, Leung K, Tremblay S, Merdsoy L, Belzile E, Lambrou A, et al. The effectiveness of self-management interventions for individuals with low health literacy and/or low income: a descriptive systematic review. *J General Internal Med.* (2018) 33:510–23. doi: 10.1007/s11606-017-4265-x
50. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. *BMC Med Informatics Decision Making.* (2016) 16:1–15. doi: 10.1186/s12911-016-0359-3
51. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Educ Couns.* (2005) 59:21–30. doi: 10.1016/j.pec.2004.09.008
52. Mirowsky J, Ross CE. Education, socioeconomic status, and health. *Educ Social Status Health*. (2017) 71–125. doi: 10.4324/9781351328081-4
53. Pepper GV, Nettle D. The behavioural constellation of deprivation: causes and consequences. *Behav Brain Sci.* (2017) 40:e314. doi: 10.1017/S0140525X1600234X
54. van Wijk EC, Landais LL, Harting J. Understanding the multitude of barriers that prevent smokers in lower socioeconomic groups from accessing smoking cessation support: a literature review. *Prev Med.* (2019) 123:143–51. doi: 10.1016/j.ypmed.2019.03.029
55. Visch Valentijn NV, Anderiesen H, van der Kooij K. *Persuasive Game Design: A Model and Its Definitions*. CHI2013. Paris (2013).
56. Schaefbauer CL, Khan DU, Le A, Sczechowski G, Siek KA. Snack buddy. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (2015). doi: 10.1145/2675133.2675180
57. Vorderstrasse A, Lewinski A, Melkus GD, Johnson C. Social support for diabetes self-management via eHealth interventions. *Curr Diabetes Rep.* (2016) 16:56. doi: 10.1007/s11892-016-0756-0
58. Kraal JJ, Elske Van Den Akker-Van Marle M, Abu-Hanna A, Stut W, Peek N, Kemps HMC. Clinical and cost-effectiveness of home-based cardiac rehabilitation compared to conventional, centre-based cardiac rehabilitation: results of the FIT@Home study. *Eur J Prev Cardiol.* (2017) 24:1260–73. doi: 10.1177/2047487317710803
59. Nabutovsky I, Nachshon A, Klempfner R, Shapiro Y, Tesler R. Digital cardiac rehabilitation programs: the future of patient-centered medicine. *Telemed E-Health.* (2020) 26:34–41. doi: 10.1089/tmj.2018.0302
60. Adams RJ, Piantadosi C, Ettridge K, Miller C, Wilson C, Tucker G, et al. Functional health literacy mediates the relationship between socio-economic status, perceptions and lifestyle behaviors related to cancer risk in an Australian population. *Patient Educ Couns.* (2013) 91:206–12. doi: 10.1016/j.pec.2012.12.001
61. Kickbusch I, Pelikan J, Apfel F, Tsouros A. *Health Literacy: The Solid Facts.* Copenhagen: WHO Regional Office for Europe (2013). p. 7–8.
62. Murugesu L, Heijmans M, Fransen M, Rademakers J. *Beter omgaan met beperkte gezondheidsvaardigheden in de curatieve zorg: kennis, methoden en tools*. Utrecht: Nivel (2018).
63. Hendrikx HCAA, Pippel S, van de Wetering V, Batenburg RS. Expectations and attitudes in eHealth: a survey among patients of Dutch private healthcare organizations. *Int J Healthc Manage.* (2013) 6:263–8. doi: 10.1179/2047971913Y.0000000050
64. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of ehealth usage: insights on the digital divide from the health information national trends survey 2012. *J Med Internet Res.* (2014) 16:1–16. doi: 10.2196/jmir.3117
65. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. *JAMA.* (2015) 313:459–60. doi: 10.1001/jama.2014.14781
66. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, system features. *Commun Assoc Information Syst.* (2009) 24:485–500. doi: 10.17705/1CAIS.02428
67. Orji R, Vassileva J, Mandryk R. Towards an effective health interventions design: an extension of the health belief model. *Online J Public Health Inform.* (2012) 4:4321. doi: 10.5210/ojphi.v4i3.4321

68. Saksono H, Castaneda-Sceppa C, Hoffman J, Seif El-Nasr M, Morris V, Parker AG. Social reflections on fitness tracking data. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019). doi: 10.1145/3290605.3300543
69. Bastida EM, Tseng TS, McKeever C, Jack L. Ethics and community-based participatory research: perspectives from the field. *Health Promot Pract.* (2010) 11:16–20. doi: 10.1177/1524839909352841
70. Artiga S, Hinton E. *Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity*. San Francisco, CA: Kaiser Family Foundation (2018).
71. Braveman PA, Cubbin C, Egerter S, Chideya S, Marchi KS, Metzler M, et al. Socioeconomic status in health research: one size does not fit all. *J Am Med Assoc.* (2005) 294:2879–88. doi: 10.1001/jama.294.22.2879
72. Werter F, Koot S. *Het wijkhuis van de toekomst: een toegankelijke en verbindende plek in de wijk. Buro de Steeg*. Utrecht: SJK Onderzoek & Advies (2017).
73. *IDEM Rotterdam*. Gebiedsbeeld charlois. RADAR, Art. 1 (2017).
74. Crouch M, McKenzie H. The logic of small samples in interview-based qualitative research. *Soc Sci Information.* (2006) 45:483–99. doi: 10.1177/0539018406069584
75. Coughlin SS, Smith S. A review of community-based participatory research studies to promote physical activity among African Americans. *J Georgia Public Health Assoc.* (2016) 5:220–7. doi: 10.20429/jgpha.2016.050305

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Digital Health

REVIEW
published: 09 July 2021
doi: 10.3389/fdgth.2021.689750

Responsible Relations: A Systematic Scoping Review of the Therapeutic Alliance in Text-Based Digital Psychotherapy

Charlotte M. van Lotringen[*†], Lars Jeken[†], Gerben J. Westerhof, Peter M. ten Klooster, Saskia M. Kelders and Matthijs L. Noordzij*

Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

OPEN ACCESS

Edited by:
Dorian Peters,
University of Cambridge,
United Kingdom

Reviewed by:
Mark Larsen,
University of New South Wales, Australia
Isabella Choi,
The University of Sydney, Australia

***Correspondence:**
Charlotte M. van Lotringen
c.m.vanlotringen@utwente.nl

[†] *These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Human Factors and Digital Health, a section of the journal Frontiers in Digital Health

Received: *01 April 2021*
Accepted: *14 June 2021*
Published: *09 July 2021*

Citation:
van Lotringen CM, Jeken L, Westerhof GJ, ten Klooster PM, Kelders SM and Noordzij ML (2021) Responsible Relations: A Systematic Scoping Review of the Therapeutic Alliance in Text-Based Digital Psychotherapy. Front. Digit. Health 3:689750. doi: 10.3389/fdgth.2021.689750

Introduction: Developing a good therapeutic alliance is considered essential for the responsible delivery of psychotherapy. Text-based digital psychotherapy has become increasingly common, yet much remains unclear about the alliance and its importance for delivering mental health care via a digital format. To employ text-based digital therapies responsibly, more insight is needed into the type and strength of the therapeutic alliance online.

Methods: A systematic scoping review was performed searching four databases: Scopus, PsycINFO, Web of Science, and Wiley Online Library. A total of 23 studies were selected and data was extracted and tabulated to explore the characteristics of studies on text-based psychotherapy, measurements of the therapeutic alliance and associations of the alliance with treatment outcome.

Results: The therapeutic alliance in text-based digital interventions was studied with a variety of client groups, though mostly for clients diagnosed with anxiety and/or depression issues. Treatment modalities were predominantly internet-based cognitive behavioral therapy (ICBT) and tailored platforms for distinct client groups (e.g., PTSD). Almost all treatments used asynchronous text-based communication, such as e-mails and integrated messaging functions, which were mainly used to give feedback on tasks. For measurements, a version of the Working Alliance Inventory (WAI) was used in most studies. Measurements with the WAI or WAI- short form indicated a good therapeutic alliance with a weighted mean score of 5.66 (on a scale of 1 to 7) and a weighted standard deviation of 0.84. Relations between the therapeutic alliance and treatment outcomes were mostly positive, with many studies reporting significant associations ($n = 8$ out of 10) or significant effects of the therapeutic alliance on treatment outcomes ($n = 5$ out of 6).

Discussion: Our scoping review suggests that a good therapeutic alliance can be established in digital psychotherapy through text-based communication, and shows support for a positive relationship between the alliance and treatment outcomes. These findings illustrate that text-based online psychotherapy can be a responsible treatment option as far as the establishment of the therapeutic alliance is concerned. However, current measures of the therapeutic alliance might miss important aspects of the alliance in digital treatment, such as the presence of empathy or compassion.

Keywords: therapeutic alliance, working alliance, text-based psychotherapy, internet-based psychotherapy, responsible digital health, mental health

INTRODUCTION

The internet has long been explored as an alternative route to deliver psychological treatment. Benefits include easy, broad, and convenient accessibility and lowered costs (1). More specifically, internet-delivered psychotherapy can increase the availability of mental health care in underserved populations and during times of crisis (2), and could therefore provide a responsible alternative or addition to fully face-to-face therapy. Meta-analyses show moderate to large effects and comparable results of online therapy to face-to-face therapy in randomized controlled trials (RCT's) for a variety of psychopathological symptoms (3–5). Notwithstanding the evidence for similar effectiveness, many consider the relationship between therapist and clients to be a central component of successful psychotherapy, and its role online is questioned and not fully understood among researchers, practitioners and clients (6).

Responsible Digital Treatment

In general, therapists do see multiple possible advantages of digital psychotherapy, such as new options for treatment and even increased intimacy in the therapeutic relationship (7). However, the higher accessibility of digital psychotherapy also gives them an increased sense of moral responsibility, especially in the case of crisis situations, where the lack of physical proximity might not allow the right caretaking actions. In those cases, the therapeutic alliance could be especially important to ensure that clients trust their therapists and share the issues that they may be facing, so that a response can be given promptly (8). In addition, therapists report a lack of information and confidence as some of the main barriers for their use of digital interventions (7, 9). Therefore, an important aspect of the responsible use of digital therapy is knowledge about the extent to which a good therapeutic alliance can be established in different digital treatment modalities.

A Text-Based Alliance

One type of online psychotherapy where the role of the therapeutic relationship is still unclear is text-based digital psychotherapy. Text-based digital psychotherapy consists mainly of written exchanges via the internet, such as internet cognitive-behavioral therapy (ICBT) delivered *via* e-mail or chat (10). This form of digital psychotherapy has the potential to foster the therapeutic relationship in novel ways, for example by giving clients more time to reflect on the expression of their feelings and thoughts in written words, while not being observed. Similarly, therapists can benefit from the richness of the written word, and the additional time to reflect before responding (2, 11). Further, the lack of social cues encourages the development of alliances and contributes to higher levels of openness and self-disclosure (12). This could stimulate a close and strong relationship between clients and therapists in a different way than in conventional face-to-face therapy or videoconferencing therapy (13).

However, the lack of non-verbal cues in the text-based format can lead therapists to fear that the communication during therapy would be impaired, and raises concern as to whether a therapeutic alliance can develop through text-based online counseling (14). In addition, the question arises whether the type of alliance that does emerge is beneficial to the therapeutic work (10). To make responsible use of text-based digital psychotherapy, the aim of the current scoping review is to give a better understanding of the therapeutic alliance in digital, text-based communication, and its relation to therapy outcomes.

The (Digital) Therapeutic Alliance

The therapeutic alliance is often operationalized through the concept of "working alliance." The working alliance includes different collaborative aspects of the relationship between the therapist and the client (15). A highly influential model of the working alliance by Bordin (16) offers a pan-theoretical perspective on the relationship between therapist and client in therapy, with the core of the alliance being: agreement on therapy goals, agreement on therapy tasks and the bond between therapist and client. Therapeutic goals refer to the objectives of the therapy that are endorsed by both therapist and client. Tasks refer to the processes and behaviors in psychotherapy sessions that relate to the actual therapeutic work. The bond refers to the interpersonal attachment between therapist and client and should include confidence, acceptance and mutual trust (16).

Therapists rate the importance of the alliance in conventional face-to-face psychotherapy significantly higher than in online psychotherapy (17). Furthermore, therapists report less confidence in their abilities to develop a functional therapeutic alliance in internet-based psychotherapy (17). In contrast with this, research on internet-based cognitive behavior therapy suggests the quality of the therapeutic alliance, most commonly rated by the client, to be at least as strong as in face-to-face therapy and also highlights the association of the alliance with online treatment outcome (15, 18, 19).

More specifically, a meta-analytic review of the alliance in adult face-to-face and internet-based psychotherapy showed that the alliance was significantly related to treatment outcomes, with a similar association between alliance and outcome in face-to-face psychotherapy ($r = 0.278$) and internet-based psychotherapy ($r = 0.275$) (15). The positive relationship between the alliance and outcome appeared to be consistent across different alliance measures and outcome measures, treatment approaches (e.g., CBT, psychodynamic therapy, etc.) and client characteristics.

Moreover, the overall correlation between alliance and outcome was almost identical to the one found in an earlier meta-analysis (20).

Types of Digital Psychological Interventions

The efficacy of a broad range of digital psychological interventions has been demonstrated, as well as a similar association between alliance and outcome as reported in face-to-face therapy. However, many current reviews and meta-analyses on efficacy and alliance-outcome associations fail to differentiate between different types of psychological online interventions. The meta-analysis by Flückiger et al. (15), for example, combined the different types of e-mental health (via internet, e-mail, videoconferencing and phone). This limits our understanding and ability to make responsible choices between the various forms these interventions can take online.

One way to categorize online psychological interventions is offered by Berger and Andersson (21), who distinguish between modes of communication of psychological online interventions. Communication can be text-based and asynchronous (e.g., e-mail), text-based and almost real-time or synchronous (e.g., chat), and audio- or video-based synchronous communication (e.g., video-conferencing). The review by Berger (10) was the first to examine ratings of the working alliance using the categorization by Berger and Andersson (21). However, Berger's (10) review was narrative, and did not provide an overview of all the alliance measurements and relations to outcomes that studies reported. Moreover, compared to other forms of online interventions, Berger (10) found a very limited database on the alliance in text-based digital psychotherapy and called for more research on this therapy format. Therefore, the present review specifically explores the scope of currently available research on the therapeutic alliance in text-based internet psychotherapy as defined by Berger and Andersson (21), in which communication between the therapist and client takes place via the internet and is text-based (e.g., e-mail, chat). Since we are interested in the therapeutic relationship between the client and the therapist, studies examining unguided, self-help text-based digital interventions are not included here.

Research Objective

The current scoping review gives a comprehensive overview of the nature and extent of current research evidence on the therapeutic alliance in text-based digital psychotherapy. This review maps out the study and intervention characteristics and findings of the existing research on this topic by exploring:

1 How text-based digital therapy is being studied: with what client groups and platforms, the forms, frequency and duration of text-based communication, the types of therapists and the treatment approaches.
2 What findings studies report regarding the working alliance: which measurements are used, what is the reported quality of the working alliance, and what types of statistical relationships between the therapeutic relationship and outcome of treatment are reported.

Our findings can indicate whether the strength of the therapeutic alliance in text-based digital psychotherapy is comparable to the one found in face-to-face psychotherapy, and if there is a relation with treatment outcomes. This way, the current review aims to enhance responsible decision-making in terms of the therapeutic alliance in digital text-based psychotherapy.

METHODS

Research Design

The present literature review is a systematic scoping review, conducted according to the guidelines provided by Peters et al. (22). Scoping reviews intend to map out the current body of research on a specified topic in terms of nature, characteristics and volume. The assessment of the potential size and scope of available research is done systematically, transparently, and in order to be easily replicated (23). Systematic scoping reviews typically synthesize data into tabular form to summarize and disseminate the existing literature in the field of interest, to identify research gaps and to make recommendations for future research (22).

Search Strategy

The electronic databases Scopus, PsycINFO, Web of Science, and Wiley Online Library were used to search for relevant studies published between 2005 and 2020. These databases were chosen because of their focus on social, medical and psychological topics, with PsycINFO being more narrowly focused on psychological and mental health research and Scopus and Web of Science being databases with a broader scope. Wiley Online Library was included because a first search of the literature indicated that many relevant articles stemmed from this database.

Each database was searched for articles and the search was repeated several times throughout the period of data collection to ensure an exhaustive and up-to-date evidence base. The final search was conducted on October 6, 2020. The databases were queried with the following combination of search terms ("working alliance" OR "therapeutic relationship" OR "therapeutic alliance") AND ("internet-based psychotherapy" OR "online psychotherapy" OR "web- based psychotherapy" OR "online mental health") occurring in the title, abstract or keywords of published scientific literature. The search term "text-based" was not included, since articles generally did not classify the therapy format in this way. Therefore, potentially relevant articles were scanned manually to evaluate whether the used digital psychotherapy format was indeed text-based.

Eligibility Criteria

The following inclusion and exclusion criteria were established:

Inclusion Criteria

1 The article needed to describe original research (e.g., no literature review).
2 The language of the article needed to be in either English, German, Dutch or Spanish for an extensive review.

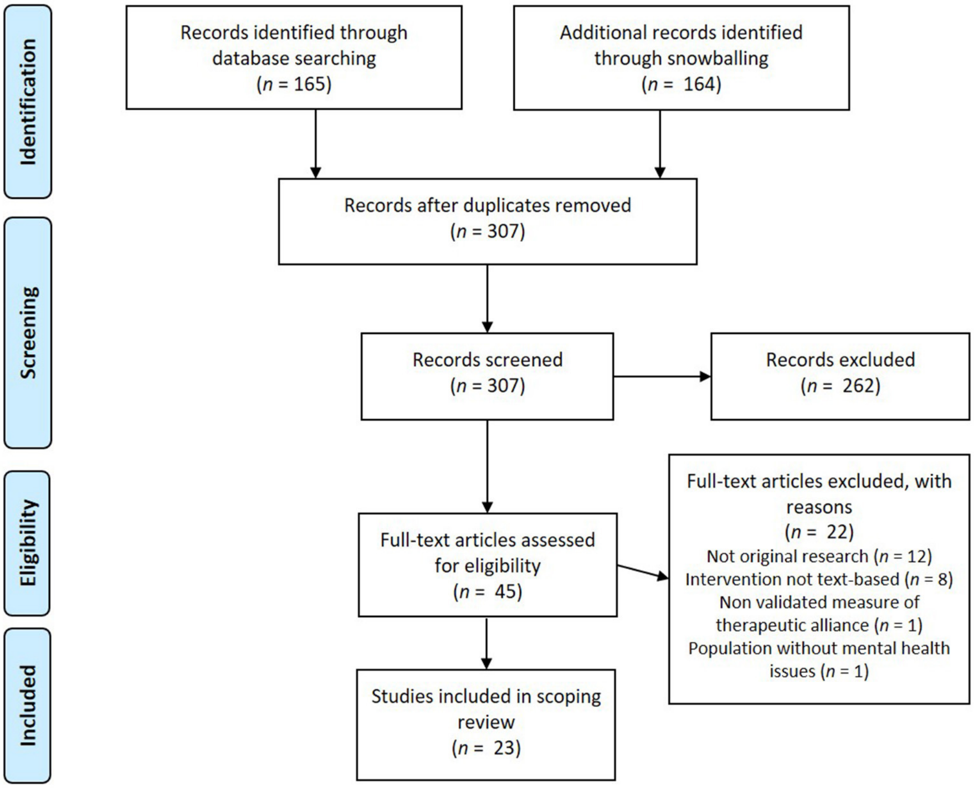


FIGURE 1 | PRISMA flowchart (26) of the inclusion/exclusion process of articles for the systematic literature review.

3 The year of publication had to be 2005 and onwards, as the technological context of the review topic makes it likely that studies from before 2005 are outdated.
4 Studies had to include text-based internet psychotherapy as defined by Berger (10).
5 Studies needed to include an assessment of the working alliance or similar construct using some validated measure, such as the Working Alliance Inventory (24).

Exclusion Criteria

1 Articles describing unguided internet-based self-help programs or interventions, video or audio-based internet psychotherapies, and internet-based psychotherapies that were offered in combination with face-to-face psychotherapy (blended treatment).
2 Articles that contained unclear descriptions of the interventions that they studied, making it impossible to determine if the accompanying therapeutic counseling was text-based.
3 Articles describing studies with participants without mental health issues (e.g. general health issues).

Study Selection

Studies found were screened on title in the first step and on abstract in the second step by two authors (CML and LJ). In step 3 it was determined whether the remaining studies were eligible based on reading the full paper and eligibility was judged using the inclusion and exclusion criteria by two authors (CML and LJ). In both steps, disagreements were discussed between the authors until agreement was reached. Subsequently, the reference lists of the included studies, as well as the studies that cited the included studies were inspected [backward and forward snowballing; (25)] to find additional papers. A flowchart illustrating the study selection process of the present systematic literature review according to the PRISMA guidelines (26) is presented in **Figure 1**.

Data Extraction

The collected articles of this review were fully read and analyzed in accordance with the aim of this research as specified above. All data from the selected articles were extracted by two researchers (CML and LJ). Data items that were extracted included participant characteristics, study and intervention characteristics and measures.

Participant Characteristics

Firstly, we extracted the targeted population: the mental disorders and/or problems that were treated. If an intervention targeted more than one disorder the group of disorders was described in accordance to the DSM-5 classification of mental disorders to describe the treated mental health problems targeted (e.g., the term anxiety disorders for several disorders including severe symptoms of anxiety, such as social anxiety disorder, general anxiety disorder, and panic disorder). Additionally, we extracted data on participants' gender and age.

Study and Intervention Characteristics

The study characteristics that were extracted were the study design and the sample size of participants. The intervention characteristics that were extracted included the platform and modalities used, the medium (e.g., mail, chat) and type of communication, the frequency of contact and average time spent by the therapist per client, the treatment length, the type of therapist, and the therapeutic approach the intervention was based on.

The type of communication within an intervention was categorized as synchronous or asynchronous, according to the categorization terms provided by Berger (10). To determine the type of practitioner, the exact wording in the study design section of the study was used. This was done to maintain the identity of the practitioner's group, as a grouping of these mental health professionals might bias the results, since different countries use different titles for various groups of mental health care workers (e.g., therapist, psychotherapist, psychological psychotherapist). If the therapeutic approach the treatment was based on was not specified by the researchers this was marked as "not specified."

Measurements

We collected information about the instrument used to measure therapeutic relationship and the rater (client or therapist), the point in time of measurements, the reported quality of the working alliance, and if applicable, the reported statistical relationship between the therapeutic relationship and outcome measures. Cut-off scores for the categorization of the strength of associations were specified a-priori and based on the general guidelines by Cohen (27). Thus, a correlational value was labeled weak when below 0.10, small from 0.10 to 0.30, moderate from 0.30 to 0.50, and labeled strong from 0.50 to 1.00.

RESULTS

A total of 23 studies were reviewed for this scoping review, covering a total of 28 interventions. Among the included studies, 14 studies investigated the working alliance as a primary objective, while the other 9 studies assessed the concept as a secondary objective.

Participant Characteristics

Table 1 summarizes the participant characteristics. Among the reviewed studies a variety of client groups was studied. The most commonly investigated client group were clients diagnosed with anxiety symptoms or disorders ($n = 9$). Another client group often assessed in working alliance research was clients diagnosed with depression ($n = 8$). Other investigated groups were clients diagnosed with posttraumatic stress disorder (PTSD) or PTSD-related symptoms ($n = 6$), obsessive-compulsive disorder ($n = 2$), binge-eating disorder ($n = 1$), preterm labor stress ($n = 1$), and chronic tinnitus ($n = 1$). In most studies the majority of participants were female, with the exception of one study having a low percentage of 39.5% female participants. The weighted average age across the studies was 37.4 years with a weighted average standard deviation of 8.77. Two studies investigated the working alliance in the treatment of children and adolescents and two studies the treatment of traumatized elderly.

Study and Intervention Characteristics

Table 2 summarizes the characteristics of the reviewed studies and interventions. Most studies had an RCT design (n =11), were part of a larger RCT study (n =1), or involved a pilot RCT ($n = 1$). Other study designs included controlled trials ($n = 2$) and open trials ($n = 4$). The sample sizes ranged from 13 to 223. Among the reviewed interventions all but one used internet-based modules as the modality. This modality uses treatment modules created by researchers or clinical psychologists for a specific target group, such as for PTSD ($n = 3$), binge eating disorder ($n = 1$), and stress management (n =1). The modules were accessible via the internet and allowed for communication between therapist and client, which mostly entailed feedback on writing tasks and progress in treatment. The exception was one study that used an instant messaging intervention, and did not specify the platform.

Most reviewed interventions used asynchronous communication ($n = 26$), which involved feedback on self-help tasks and on written assignments, emails, and the use of an integrated text or chat function within the treatment platform. Text-based responses by a therapist were always created within 24–48 h after a message was sent by a client, questions were asked or writing assignments were completed. For one intervention, the type of communication was not specified and another reported "written exchange" without a detailed specification. Only two interventions (partially) used synchronous text-based communication involving a chat room used by the client and therapist.

The most common frequency of contact between the client and therapist for interventions was weekly ($n = 16$). Other frequencies that were used were 2–3 times a week ($n = 2$) and every 10–14 days ($n = 1$). One intervention for free online counseling was accessible on demand. For nine interventions, the contact frequency was not specified. For eleven of the interventions, the average time that was spent by the therapist per client was reported. The most common average time was around 15 min per client per week ($n = 5$). Interventions where the therapist replied to writing assignments indicated longer average times, from 20-50 min per client per text ($n = 2$) to 45–50 min (n =1). One of the two interventions that involved synchronous chat sessions reported weekly session durations of 45 min. The treatment length of the interventions ranged from 5 to 16 weeks, 6 to 10 modules or 9 to 14 sessions. For most included interventions, the treatment length was reported in weeks ($n = 18$), with a length of 8 ($n = 6$) or 10 weeks ($n = 6$) being the most common. One intervention that made use of instant messaging as a form of free online counseling, was accessed 2 to 5 or more times by participants.

In regard to the type of therapist offering or guiding treatment online, several interventions offered guidance by a licensed or registered psychotherapist or psychiatrist,

TABLE 1 | Participant characteristics.

	References	Presenting problem	Gender	Age (years), mean (SD)
1a	Andersson et al. (28)	Depression	75% female	38.9 (13.5)
1b	Andersson et al. (28)	Generalized anxiety disorder	80.6% female	40.0 (11.2)
1c	Andersson et al. (28)	Social anxiety disorder	59.3% female	37.7 (11.42)
2	Andersson et al. (29)	Obsessive-compulsive disorder	66.3% female	34.93 (12.72)
3a	Anderson et al. (30)	Anxiety disorders	61.6% female	13.91 (1.56)
3b	Anderson et al. (30)	Anxiety disorders	53.0% female	12.12 (2.5)
4	Bergman Nordgren et al. (31)	Anxiety disorders	67% female	39.3 (11.2)
5	Bisseling et al. (32)	Anxiety and depression during cancer	85.7% female	n.a.
6	Blake Buffini and Gordon (33)	I.a. mood, anxiety and personality disorders	83.3% female	n.a. (0.80)
7	Dölemeyer et al. (34)	Binge-eating disorder	93.2% female	34.8 (10.3)
8	Duffy et al. (35)	Depression and anxiety	69% female	n.a.
9	Penedo et al. (36)	Depression	70.4% female	44.48 (10.68)
10a	Hadjistavropoulos et al. (37)	Depression	69.5% female	40.22 (12.57)
10b	Hadjistavropoulos et al. (37)	Generalized anxiety	69.5% female	40.22 (12.57)
11	Herbst et al. (8)	Obsessive-compulsive disorder	n.a.	n.a.
12	Jasper et al. (38)	Chronic tinnitus	39.5% female	51.92 (10.55)
13	Klein et al. (39)	Posttraumatic stress disorder	77.27% female	66.1 (11.36)
14	Knaevelsrud et al. (40)	Posttraumatic stress symptoms	56.6% female	71.73 (4.8)
15	Knaevelsrud et al. (41)	Childhood traumatization	64.9% female	71.4 (4.7)
16	Knaevelsrud and Maercker (42)	Posttraumatic stress reactions	92% female	35 (n.a.)
17	Knaevelsrud and Maercker (43)	Posttraumatic stress disorder	90% female	35 (10.55)
18a	Lindegaard et al. (44)	Social anxiety disorder	62% female	41.4 (12.0)
18b	Lindegaard et al. (44)	Social anxiety disorder	74% female	42.6 (16.3)
19	Preschl et al. (45)	Depression	84% female	34.9 (9.5)
20	Reynolds et al. (46)	Mostly depression and stress/anxiety issues	71% female	n.a.
21	Scherer et al. (47)	Preterm labor stress	100% female	32.53 (3.49)
22	Topooco et al. (48)	Adolescent depression	91% female	17.5 (1.1)
23	Wagner et al. (49)	Posttraumatic stress symptoms	81% female	27.7 (7.0)

If a study did not report certain information, this is indicated with n.a. (not available).

and/or psychotherapists in training ($n = 6$), or psychologists, psychotherapists or psychiatrists of whom it was unclear if they were licensed or registered ($n = 6$). For other interventions, psychology or social work students in the final phase of their master's degree were (additionally) employed ($n = 7$). Among the reviewed studies some used the term "online therapist" or "therapist" to refer to their practitioners, without specifying the term further ($n = 7$), whereas one did not state who was responsible for communication with clients within the treatment program.

The majority of reviewed studies based their intervention on the cognitive behavioral therapy approach ($n = 18$). One study combined a cognitive behavioral therapy approach with narrative exposure therapy, one study used a mindfulness-based cognitive therapy approach, and another study used a psychodynamic treatment approach in one condition of their studied treatments. Finally, three studies did not specify which psychotherapeutic approach the treatment was based on.

Measurements

Table 3 summarizes the findings concerning measurements of the therapeutic relationship and its relationship with outcome measures, with articles ordered by type of scale used.

The reviewed studies varied in their chosen points in time for measurements. Six of the studies reviewed here measured the working alliance early on in treatment, specifically in the second or third week of treatment or after the third session. Other studies took measurements after a set number of modules (n =1), after the completion of certain modules (n =3), after the first half of treatment and post-treatment (n =1), and solely after treatment ($n = 1$). If multiple measurements were available, we chose to report early measurements in **Table 3**, since it is recommended to report early measurements of the therapeutic alliance (55).

Therapeutic Relationship Measure and Quality of Therapeutic Relationship

The Working Alliance Inventory

Most reviewed studies used a version of the Working Alliance Inventory (WAI) (24) for measurements of the therapeutic alliance ($n = 19$). The WAI has a client and therapist version, of which both scores can be combined to form a composite score. However, frequently only the client version of the WAI is used, as was also the case in the majority of studies included here ($n = 17$). The original WAI has 36 items rated on a 7-point Likert scale, and was used by only one study. Fourteen studies used the 12-item short form of the WAI, the WAI-S. A few studies used

TABLE 2 | Study and intervention characteristics.

	References	Study design	Sample size	Modality	Form(s) of communication (asynchronous /synchronous)	Frequency of contact, average time spent by therapist per client	Treatment length	Therapists	Therapeutic approach
1a	Andersson et al. (28)	Controlled trial	88	Internet-based modules	Either e-mails or feedback on self-help tasks (asynchronous)	Weekly, e-mail therapy: total of 509 min. (SD=176), guided self-help treatment: total of 53 min. (SD=28)	8 weeks	"Internet-therapist"	Cognitive behavior therapy
1b	Andersson et al. (28)	Controlled trial	89	Internet-based modules	Feedback on self-help tasks (asynchronous)	Weekly, 10-15 min. a week	8 weeks	"Internet-therapist"	Cognitive behavior therapy
1c	Andersson et al. (28)	Controlled trial	204	Internet-based modules	Feedback on self-help tasks (asynchronous)	Weekly, 15 min. a week	8 weeks	"Internet-therapist"	Cognitive behavior therapy
2	Andersson et al. (29)	RCT	101	Internet-based modules	Integrated text messaging function, emails (asynchronous)	2-3 times a week, n.a.	10 weeks	Clinical psychology students in their final year of the study program under supervision of a licensed psychologist	Cognitive behavior therapy
3a	Anderson et al. (30)	Controlled trial	73	BRAVE: Internet-based modules	Emails (asynchronous)	Weekly, n.a.	10 youth and five parent sessions	Registered psychologists	Cognitive behavior therapy
3b	Anderson et al. (30)	Controlled trial	132	BRAVE: Internet-based modules	Emails (asynchronous)	Weekly, n.a.	10 youth and five parent sessions	Registered psychologists	Cognitive behavior therapy
4	Bergman Nordgren et al. (31)	Randomized controlled pilot trial	27	Internet-based modules	Feedback on homework assignments (asynchronous)	N.a., 15 min. a week	10 weeks	Master's degree level psychology students who had completed clinical training	Cognitive behavior therapy
5	Bisseling et al. (32)	RCT	78	Internet-based modules	Written feedback on completed logfiles, emails (asynchronous)	Weekly, n.a.	9 sessions	Therapists fulfilling the advanced criteria of the Association of Minfulness-Based Teachers in the Netherlands and Flanders	Third wave CBT (Mindfulness-based cognitive therapy)
6	Blake Buffini and Gordon (33)	Cross-sectional	78	A service providing free online counseling during crises	Instant messaging (synchronous)	On demand, n.a.	Participants accessed support 2-5+ times	Service-staff with a minimum of undergraduate level in psychology, psychotherapy or social care	Not specified
7	Dölemeyer et al. (34)	Uncontrolled trial	59	Internet-based modules, based on 'Overcoming Binge Eating'	Integrated text messaging function (asynchronous)	Weekly, n.a.	16 weeks	'Therapist'	Cognitive behavior therapy

(Continued)

TABLE 2 | Continued

	References	Study design	Sample size	Modality	Form(s) of communication (asynchronous /synchronous)	Frequency of contact, average time spent by therapist per client	Treatment length	Therapists	Therapeutic approach
8	Duffy et al. (35)	Open, uncontrolled feasibility trial	123	SilverCloud: Internet-based modules	Written feedback on progress (asynchronous)	Every 10-14 days, n.a.	8 weeks	Clinical psychologists, counseling psychologists and psychological wellbeing practitioners	Cognitive behavior therapy
9	Gómez Penedo et al. (36)	RCT	223	Deprexis: Internet-based modules	Standardized email support with feedback on activity (asynchronous)	Weekly, n.a.	10 modules	Master's students in clinical psychology and psychotherapy, psychotherapists in training and licensed psychotherapists, trained in the program	Cognitive behavior therapy
10a	Hadjistavropoulos et al. (37)	Open dissemination trial	83	Internet-based modules	Feedback on homework, emails (asynchronous)	Weekly, n.a.	12 modules	Registered psychologists, social workers, nurses with CBT experience and supervised graduate students in clinical psychology or social work	Cognitive behavior therapy
10b	Hadjistavropoulos et al. (37)	Open dissemination trial	112	Internet-based modules	Feedback on homework, emails (asynchronous)	Weekly, n.a.	12 modules	Registered psychologists, social workers, nurses with CBT experience and supervised graduate students in clinical psychology or social work	Cognitive behavior therapy
11	Herbst et al. (16)	RCT	29	Internet-based modules	n.a.	n.a.	14 sessions	n.a.	Cognitive behavior therapy
12	Jasper et al. (38)	RCT	38	Internet-based modules	Online messaging systems (asynchronous)	Weekly, 13.75 min. per week	10 weeks	Clinical psychologists certified in CBT or psychologists in advanced stages of their training	Cognitive behavior therapy
13	Klein et al. (39)	Open trial	22	PTSD Online: Internet-based modules	Audio files and email individually tailored and constructed (asynchronous)	n.a., total time: 194.47 min. (SD=148.7)	10 weeks	Registered and probationary registered psychologists	Cognitive behavior therapy
14	Knaevelsrud et al. (40)	Open trial	30	Internet-based modules	Feedback on writing assignments (asynchronous)	n.a.	6 weeks	Doctoral-level clinician psychologists with special training in the application of CBT for PTSD	Cognitive behavior therapy and narrative exposure therapy
15	Knaevelsrud et al. (41)	RCT	94	Internet-based modules	Uploading texts in secure Web portal (asynchronous)	10 responses to texts, 45-50 min. each	6 weeks	Licensed clinical psychologists with special training in Integrative Testimonial Therapy	Cognitive behavior therapy

(Continued)

TABLE 2 | Continued

	References	Study design	Sample size	Modality	Form(s) of communication (asynchronous /synchronous)	Frequency of contact, average time spent by therapist per client	Treatment length	Therapists	Therapeutic approach
16	Knaevelsrud and Maercker (42)	Part of an RCT	91	Internet-based modules	Feedback on writing assignments (asynchronous)	n.a.	5 weeks	Psychologists trained in the application of writing assignments for PTSD	Not specified
17	Knaevelsrud and Maercker (43)	RCT	96	Internet-based modules	Feedback on writing assignments (asynchronous)	n.a.	5 weeks	Clinical psychologists trained in the application of writing assignments for PTSD	Cognitive behavior therapy
18a	Lindegaard et al. (44)	Preference study	13	SOFIE: Internet-based modules	Mail service within internet platform, feedback on homework assignments (asynchronous)	Weekly, guideline of 15 min. per week	10 weeks	Master's degree level psychology students	Cognitive behavior therapy
18b	Lindegaard et al. (44)	Preference study	23	Internet-based modules	Feedback on homework assignments (asynchronous)	Weekly, guideline of 15 min. per week	10 weeks	Master's degree level psychology students	Psychodynamic therapy
19	Preschl et al. (45)	RCT	25	Internet-based modules	Feedback on writing assignments and instructions for exercises (asynchronous)	n.a., 20-50 min. per text	8 weeks	Psychologists and psychotherapists trained in CBT for depression	Cognitive behavior therapy
20	Reynolds et al. (46)	Uncontrolled	17	Internet-based modules	Emails (asynchronous)	n.a.	n.a.	"Therapists, predominately qualified to practice in the US"	Not specified
21	Scherer et al. (47)	RCT	58	IB-CBSM: Internet-based modules for stress management	"Written exchange" (n.a.)	Weekly, n.a.	6 modules	Trained psychologist or psychologist in training	Cognitive behavior therapy
22	Topooco et al. (48)	RCT	70	Iterapi: Internet-based CBT with therapist chat sessions	Chat sessions conducted inside the treatment platform (synchronous)	Weekly, 45 min.	8 weeks	Therapists in training	Cognitive behavior therapy
23	Wagner et al. (49)	RCT	55	Internet-based modules for PTSD symptoms	Written feedback and instructions (asynchronous)	n.a., 20-50 min. per text	5 weeks	Psychologists and psychiatrists	Cognitive behavior therapy

If a study did not report certain information, this is indicated with n.a. (not available).

a modified version of the WAI-S, adapted for guided internet interventions ($n = 2$) or for online help for women ($n = 1$). In addition, four studies used the revised short form of the WAI (WAI-SR), which also has 12 items but uses a 5-point Likert scale (51). Finally, one study used the WAI-I, which is a new scale developed specifically for guided internet interventions (36), and was derived from the WAI-SR. The WAI-I also has 12 items and uses a 5-point Likert scale.

The WAI and WAI-S items are typically averaged into a total score between 1 and 7 (50). Measurements of the average working alliance with the WAI or the WAI-S ranged from 4.30 (SD: 1.27) to 6.3 (SD: 0.54) across 12 studies that examined 15 interventions. The weighted average score of all WAI and WAI-S measurements was 5.66 with a weighted average standard deviation of 0.84, which indicates a positive working alliance. Only three of the four studies using the revised short form (with a maximum score of 5) of the WAI reported total WAI-SR-scores (see **Table 3**), with a weighted average of 3.23 (weighted SD: 0.8). Of those, one reported a low mean score of 2.34 (SD: 0.98) (38).

Other Measures

Additionally, a few studies used different scales to assess the therapeutic alliance. Two studies that examined three interventions used the Therapeutic Alliance Questionnaire [TAQ (52)], which only assesses the client's perspective. The maximum score for the TAQ is 102, and a score above 80 is considered to reflect a high working alliance (37). The two studies included here showed a weighted average TAQ-score of 85,26 with a weighted average standard deviation of 12.44 (based on the mid-treatment scores when these were included), indicating a good working alliance.

One study used the Scale to Assess the Therapeutic Relationship [STAR (53)], which consists of a separate patient (STAR-P) and clinician (STAR-C) scale, both with 12 items and scores that can range from 0–48, with higher scores indicating a better therapeutic relationship. The study included here reported an average STAR-P score of 37.41 (standard deviation: 1.54) and an average STAR-C score of 30.54 (standard deviation: 1.5).

Finally, one study used a short form of the Agnew Relationship Measure [ARM (54)]. The ARM also has parallel versions for clients and therapists, both with 12 items and a 7-point Likert scale. The study included here only reported the scores on subscales for clients and therapist, and not a composite score. These sub scores ranged from 5.27 to 6.19 (out of 7) for clients, and from 4.73 to 5.76 (out of 7) for therapists, which seems to indicate a positive therapeutic relationship.

Relationship Between Therapeutic Relationship and Outcome

Reported findings on the statistical relationships between measurements of the therapeutic alliance and outcome measures of treatment are shown in **Table 3**. Analysis techniques used included bivariate and partial Pearson correlations, hierarchical multiple regression analysis and multi-level hierarchical linear modeling.

The 23 reviewed articles included 28 interventions. For 22 of these interventions, studies reported statistical relationships between measurements of the therapeutic alliance and outcome measures. Firstly, for 8 out of 13 interventions (8 out of 10 studies), significant correlations were reported between the therapeutic alliance measures and change-scores on one or more primary outcome measures, ranging from small ($n = 1$) to moderate ($n = 7$) and strong ($n = 1$) in magnitude. Further, for 6 out of 7 interventions (5 out of 6 studies) that were examined on the predictive value of the alliance on subsequent treatment outcomes, higher degrees of the therapeutic alliance significantly predicted better treatment outcomes.

Three out of three studies found that the therapeutic alliance significantly predicted the secondary outcome measures of, respectively, compliance ($n = 1$) and client satisfaction ($n = 2$). In the case of the influence of the alliance on client satisfaction, the effects found in both studies were strong. Finally, three out of three studies found that subscales of the WAI, such as agreement on tasks, significantly predicted (n = 1), or were significantly and moderately correlated with ($n = 2$) outcome measures of the assessed treatment.

DISCUSSION

This scoping review aims to summarize the available research on the therapeutic alliance in text-based digital psychotherapy within the last 15 years, in order to enhance its responsible use. A total of 23 articles that examined the therapeutic alliance in 28 text-based interventions were reviewed. These articles were explored on several participant, study and intervention characteristics, as well as the type of measurements of the therapeutic alliance and its relationship to treatment outcome.

In general, our findings show that research on this topic has been conducted with a variety of client groups and treatment modalities. However, most studies focused on clients diagnosed with anxiety and/or depression symptoms and used a form of internet-based cognitive behavioral therapy (ICBT) as modality for treatment delivery. Further, most text-based digital treatments solely used asynchronous communication methods, such as emails, delayed chats, and integrated text-based communication features within websites and platforms.

An essential precondition of the responsible use of digital, text-based psychotherapy is knowing whether a therapeutic alliance can be established in this therapy format, and whether the strength and type of a therapeutic alliance that is established through text is comparable to the one found in face-to-face treatment. The therapeutic alliance scores in the reviewed articles on digital text-based psychotherapy were mostly high, thus suggesting that positive alliances can indeed be established in digital interventions even if only text-based communication modalities are used between clients and therapist. The majority of the reported statistical relationships of the therapeutic alliance showed either significant correlations between the alliance and treatment outcome, or found that the alliance significantly predicted primary and secondary outcome measures.

Scope of the Reviewed Research

In previous reviews on the therapeutic alliance in digital therapy, a very limited number of studies was found investigating

TABLE 3 | Measurements.

	Authors	Therapeutic relationship measure and rater (c = client-rated, t = therapist-rated)	Moment of assessment	Quality of therapeutic relationship, Mean (Standard deviation)	Relationship between therapeutic relationship and outcome
1a	Andersson et al. (28)	WAI: Working Alliance Inventory, C	Third week of treatment	5.41 (0.83)	Correlations between the WAI-S and residualized change scores on the primary outcome measures were weak ($r = 0.18$) and not statistically significant
1b	Andersson et al. (28)	WAI: Working Alliance Inventory, C	Third week of treatment	5.63 (0.94)	Correlations between the WAI-S and residualized change scores on the primary outcome measures were small ($r = 0.13$) and not statistically significant
1c	Andersson et al. (28)	WAI: Working Alliance Inventory, C	Fourth week of treatment	5.45 (1.05)	Correlations between the WAI-S and residualized change scores on the primary outcome measures were small ($r = 0.10$) and not statistically significant
2a	Anderson et al. (30)	WAI-S: Working Alliance Inventory-Short Form, C	After completion of third session	5.77 (1.20)	n.a.
2b	Anderson et al. (30)	WAI-S: Working Alliance Inventory-Short Form, C	After completion of third session	5.85 (1.09)	Higher WAI-S scores in older adults (12-18 years) predicted CGAS at 6-month follow-up ($B = .22$, $t = 2.21$, $p = 0.03$) Higher WAI-S scores predicted compliance with the treatment ($B = 0.38$), $F_{(1,80)} = 13.10$, $p = 0.01$)
3	Blake Buffini and Gordon (33)	WAI-S: Working Alliance Inventory-Short Form, C	After participants had accessed support on more than one occasion	4.30 (1.27)	The strength of the working alliance predicted client satisfaction, explaining 55% of variance in client satisfaction scores (*R-square* $= 0.55$; $F = 93.85$, $p < 0.001$)
4	Dölemeyer et al. (34)	WAI-S: Working Alliance Inventory-Short Form, C	After first half of treatment	6.01 (0.79)	Correlations between the WAI-S measured at end of treatment and residual gain scores on EDE-Q-subscale 'restrained eating behavior' were significant and moderate ($r = -492$), no correlations between WAI-S and binge eating episodes
5	Knaevelsrud et al. (41)	WAI-S: Working Alliance Inventory-Short Form, C	End of treatment	6.2 (0.7)	n.a.
6	Knaevelsrud et al. (40)	WAI-S: Working Alliance Inventory-Short Form, C	Fourth treatment session	6.09 (0.87)	n.a.
7	Knaevelsrud and Maercker (42)	WAI-S: Working Alliance Inventory-Short Form, C & T (only client ratings reported)	Fourth treatment session	5.8 (0.62)	Correlations between the WAI-S and residual gain scores on anxiety were significant and moderate ($r = .33$)
8	Knaevelsrud and Maercker (43)	WAI-S: Working Alliance Inventory-Short Form, C & T	End of treatment	Client: 6.3 (0.54) Therapist: 5.8 (.98)	Correlations between the client-rated WAI-S (at the end of treatment) and treatment outcome were significant and predicted 15% of the variance in post-treatment measures of the IES-R (*adjusted R-square* $= 0.148$; $F_{(2,39)} = 8.15$, $p < 0.05$)
9	Preschl et al. (45)	WAI-S: Working Alliance Inventory-Short Form, C & T	After four weeks	Client: 5.82 (0.80) Therapist (only measured post treatment): 6.04 (.67)	Correlations between clients' ratings of the subscale 'tasks' measured post-treatment and BDI-score at post-treatment in the online group were significant and moderate ($r = -0.47$), the WAI-S did not significantly predict the BDI residual gain score ($r = -0.06$)
10	Topooco et al. (48)	WAI-S: Working Alliance Inventory-Short Form, C	n.a.	4.95 (0.63)	n.a.
11	Wagner et al. (49)	WAI-S: Working Alliance Inventory-Short Form, C	After the fourth session	6.04 (0.83)	Early WAI-S (at mid-treatment) significantly predicted treatment outcome (*adjusted R-square* $= .20$; $F(2,44) = 6.57$, $p = 0.003$)
12	Andersson et al. (29)	WAI-S: Working Alliance Inventory-Short Form, n.a.	Third week of treatment	n.a.	Higher degree of working alliance predicted Y-BOCS change score ($B = -0.09$, $SE = 0.04$, $t = 2.20$, $p < 0.05$)
13	Bergman Nordgren et al. (31)	WAI-S: Working Alliance Inventory-Short Form, adapted for guided internet interventions, C	Third week	6.00 (0.80)	Correlations between the WAI-S (at week 3) and residual gain scores on the primary outcome measure were significant and moderate ($r = -0.47$)
14	Lindegaard et al. (44)	WAI-S: Working Alliance Inventory-Short Form, adapted for guided internet interventions, C	Third week	n.a.	Correlations between the WAI-S (at week three) and treatment outcome were significant and WAI-S predicted change rate ($B = -0.05$, 95% CI $[-0.072, -0.018]$, $z = -3.22$, $p = 0.001$)

(Continued)

TABLE 3 | Continued

	Authors	Therapeutic relationship measure and rater (c = client-rated, t = therapist-rated)	Moment of assessment	Quality of therapeutic relationship, Mean (Standard deviation)	Relationship between therapeutic relationship and outcome
15	Bisseling et al. (32)	WAI-SR: Working Alliance Inventory-Short Form Revised, C	At the start of week 2	n.a.	Therapeutic alliance predicted both reduction of psychological distress ($B = -0.12$; $t(114) = -2.656$; $p = 0.01$) and increase of mental well-being ($B = 0.23$; $t(113) = 2.651$; $p = 0.01$) at post treatment
16	Herbst et al. (8)	WAI-SR: Working Alliance Inventory-Short Form Revised, n.a.	Post treatment	4.08 (0.78)	Correlations between the WAI-SR and Y-BOCS SR change score were significant and moderate ($r = 0.33$); a marginal correlation between WAI-SR and the OCI-R change score was significant and small ($r = 0.29$)
17	Jasper et al. (38)	WAI-SR: Working Alliance Inventory-Short Form Revised, C	Fifth week	2.34 (0.98)	Correlations between the subscales ‘agreement on treatment tasks’ and residual gain scores for the therapy outcome measure ‘tinnitus distress’ were significant and moderate ($r = 0.40$)
18	Scherer et al. (47)	WAI-SR: Working Alliance Inventory-Short Form Revised, adapted to the online help for women, C	After module 2, 3, 4, and 5	3.51 (.69)	Correlations of the WAI-SR and residual gain scores on stress and anxiety outcomes were significant and moderate (PSS: $r = 0.451$) and strong (STAI-T: $r = 0.501$). Nearly 40% of the variance in patient satisfaction is explained by the WAI-SR, *R-square* = .398; $F_{(1,50)} = 33.060$, $p < 0.001$. WAI-SR partly mediates at least the relationship between group condition and patient satisfaction
19	Gómez Penedo et al. (36)	WAI-I: Working Alliance Inventory for Guided Internet Interventions, C	Post treatment	Task & goal subscale: 3.17 (0.91) Bond subscale: 3.56 (1.15)	Significant effect of the tasks and goals subscale on the estimated PHQ-9 value at the end of follow-up ($B = -1.74$, $SE = 0.40$, 95% CI [−2.52, −0.96], $t(206) = -4.37$, $p < 0.001$)
20a	Hadjistavropoulos et al. (37)	Therapeutic Alliance Questionnaire (TAQ), C	After module 6 and prior to completing module 12	Mid-treatment: 83.47 (13.89) Post-treatment: 83.20 (15.35)	Mid-treatment TAQ scores were not significantly correlated with PHQ-9 post-treatment scores (controlling for pre-treatment PHQ-9 scores)
20b	Hadjistavropoulos et al. (37)	Therapeutic Alliance Questionnaire (TAQ), C	After module 6 and prior to completing module 12	Mid-treatment: 85.82 (10.85) Post-treatment: 86.93 (12.42)	Mid-treatment TAQ scores were not significantly correlated with GAD-7 post-treatment scores (controlling for pre-treatment GAD-7 scores)
21	Klein et al. (39)	Therapeutic alliance questionnaire (TAQ), C	n.a.	89.2 (15.1)	n.a.
22	Duffy et al. (35)	Scale to assess the therapeutic relationship (STAR), C & T	At treatment exit	**STAR-P** (Client’s perspective), treatment completers: 37.410 (1.543) **STAR-C** (therapist’s perspective), treatment completers: 30.543 (1.500)	STAR-P, client’s perspective: Treatment completers showed a significant increase in STAR-P scores of on average 3.9 points from baseline to average end of treatment (day 46) (95% CI [−5.36, 1.26], $t(82) = -3.195$, $p = 0.002$). STAR-C, therapist’s perspective: For dropout clients, the STAR-C scores declined significantly by on average 5.4 points from baseline to end of treatment (day 46) (95% CI [2.10, −8.73), $t(308) = 3.236$, $p = 0.001$)
23	Reynolds et al. (46)	Agnew Relationship Measure (ARM), short form, C & T	Weekly	**Client ratings** Bond and partnership: 5.97 (1.26); Confidence: 6.19 (1.24); Openness: 5.27 (1.42) **Therapist ratings** Bond and partnership: 5.72 (.94); Confidence: 5.76 (.99); Openness: 4.73 (1.55)	n.a.

If a study did not report certain information, this is indicated with n.a. (not available).

The Working Alliance Inventory is from Horvath and Greenberg (24); the Working Alliance Inventory—Short Form is from Busseri and Tyler (50); the Working Alliance Inventory—Short Form Revised is from Hatcher and Gillaspy (51); the Working Alliance Inventory for Guided Internet Interventions is from Penedo et al. (36); the Therapeutic Alliance Questionnaire is from Bickman et al. (52), the Scale to Assess the Therapeutic Relationship is from McGuire-Snieckus et al. (53), the Agnew Relationship Measure is from Agnew-Davies et al. (54).

the therapeutic alliance in text-based digital therapy (6, 10). The current review therefore provides an updated overview specifically on the therapeutic alliance in this therapy format. With a number of 23 included studies, of which 7 studies were published from 2017 and onwards [when Berger's narrative review was conducted (10)], the body of evidence on the working alliance in text-based digital treatment seems to be growing. Furthermore, of the included studies in the current review, 14 studies investigated the therapeutic alliance as a primary objective, compared to 6 studies in a review from 2012 (6). With a larger evidence base, the current review confirms indications from earlier reviews of a similar therapeutic alliance in digital therapy and face-to-face therapy and mixed to positive relationships of the alliance with treatment outcomes (6, 10).

Although a range of treatments for various client groups was found in the reviewed studies, most of them concerned (a combination of) anxiety and depression symptoms. This is not surprising, given that anxiety and mood disorders are the most prevalent mental disorders (56). Furthermore, with the great majority of studies reviewed here using a cognitive behavioral therapy approach or framework for internet-based psychotherapy it becomes apparent that this approach also dominates the treatment options online, in line with earlier findings on internet psychotherapy (6, 10). This might be explained by the fact that the short-term interventions and techniques that are commonly used in CBT fit well in an online format (57), and may also be more easily integrated into internet-based psychotherapy given the length of the treatments reviewed here (5–16 weeks). Other psychological approaches such as psychodynamic approaches or third-wave CBT were underrepresented in the current review. While internet-based treatments based on psychodynamic theory and third-wave approaches are slowly starting to appear, the working alliance or similar constructs relating to the therapeutic relationship have yet to be researched (58, 59).

Finally, the reviewed studies examined various digital treatment options, but it is worth mentioning that the authors of the studies often originate from the same research groups from Germany and Sweden. It seems that mainly these groups investigate the therapeutic alliance online, which suggests a lack of variety in researchers studying the subject, and shows that research interest in this topic is not yet widespread. Furthermore, since the country in which studies on (digital) psychotherapy are conducted likely influence the generalizability of the results, caution is necessary when interpreting these findings. The current conclusions may be limited to the North-European context and culture and may not be representative for other countries and cultures.

The Therapeutic Alliance and Responsible Digital Treatment

Our findings generally show high levels of therapeutic alliance in text-based digital psychotherapy, comparable to those reported for face-to-face treatment (50). Therefore, the establishment of a good therapeutic relationships seems to be possible independent of the medium (digital text-based or face-to-face). Additionally, the majority of the included studies show significant and positive relations between the therapeutic alliance and primary or secondary outcome measurements, such as a strong relationship with client satisfaction. The formation of a good therapeutic alliance, especially when related to better treatment outcomes, supports the notion that text-based digital treatments can be a responsible addition or alternative to face-to-face treatment or (long) waiting lists.

Quality of the Therapeutic Alliance and Alliance-Outcome Relationship

To measure the therapeutic alliance, most of the reviewed studies ($n = 10$) used either the Working Alliance Inventory [WAI (24)] or its short form [WAI-S (50)]. The scores on the WAI and WAI-S were generally high with a combined average score of 5.66 (out of 7). This indicates a good working alliance that is comparable to ratings in face-to-face treatment, such as the mean of 5.87 found in a study on the Working Alliance Inventory in face-to-face treatment (50). In addition, four studies used other measures of the working alliance. All of the reported scores for these measures also indicated similarly high scores of the working alliance, (far) above the midpoints of the different scales.

Only one of the reviewed studies (38) reported a low mean score of 2.34 for the revised short form of the WAI (the WAI-SR, with a maximum score of 5). Nonetheless, additionally reported scores by the researchers did indicate an increase in working alliance ratings by week 9. The study was unique in their targeted population of clients diagnosed with chronic tinnitus and their sample of mainly male clients. Given the lack of research on online treatment with this client group it is unclear whether the relatively low score and late increase in alliance quality stems from client characteristics, the specific treatment or other factors.

With respect to the statistical relationship between measurements of the working alliance and outcome measures of treatment the results were somewhat varied. Many studies indicated significant moderate [e.g., (31)] correlations of the total working alliance scores or subscales scores with outcome measures. One study found small correlations (8) and another found strong correlations (47), while a few studies did not find significant correlations with primary outcome measures [e.g., (28)]. Moreover, several studies indicated that early measurements of the working alliance significantly predicted better treatment outcomes or secondary outcomes such as compliance to treatment or client satisfaction.

The positive relationship between the therapeutic alliance and treatment outcomes indicates that the type of alliance that emerges in text-based digital psychotherapy is indeed beneficial to the therapeutic work. The finding that many studies found a significant and moderate relationship between the alliance and treatment outcomes is also in accordance with a previous meta-analysis on the alliance-outcome association in digital psychotherapy that found a significant overall correlation of $r = 0.28$ (15), as well as a more recent meta-analysis on the same topic that found a significant overall correlation of $r = 0.20$ (19). Since the research on this topic was until recently very limited (10), our review provides the necessary update on the working alliance in this evolving and growing field. Hopefully, our findings can serve to give therapists more confidence in their abilities to develop a functional working alliance in internet-based psychotherapy, since earlier studies showed that this confidence is often lacking in therapists (7, 17).

What Knowledge Is Missing?

Especially for clients diagnosed with anxiety, depression, or PTSD symptoms and digital text-based treatments based on CBT, our review shows evidence that good therapeutic alliances can be established in text-based psychotherapies. Here, the alliance does not seem to be compromised due to the communication being text-based instead of face-to-face. With regard to other client groups and treatment approaches, the research on the therapeutic alliance in digital text-based psychotherapy is still in its infancy. This means that we should be careful with generalizing our findings from the current review to other client groups and treatment approaches. However, since we did not find any obvious differences for client groups or different treatment forms, we expect that the working alliance in text-based digital treatment works similarly across the board, in line with findings on internet-based treatment in general (15).

Additionally, to judge whether digital text-based treatments are a responsible option, measuring the quality of the therapeutic alliance solely quantitatively with the WAI could reflect a rather narrow view of the alliance. The therapeutic alliance is a dynamic construct, that fluctuates over time and that a therapist should always closely guard during the therapy process (20), whereas in most of the included studies the working alliance was only measured at one moment in time. A different research approach, such as experience sampling (60), could therefore be a suitable addition to research on the therapeutic alliance. Experience sampling could also give more insight into the naturally occurring ruptures and the corresponding repair work in the therapeutic relationship. These are not taken into account in a scale such as the WAI, even though ruptures seem likely to occur in relationships that are formed on the basis of reduced communication cues and responsiveness (10). It is possible that other important, more complex elements in the relationship are currently missed as well, such as self-disclosure or empathy (61).

An essential value in mental health care, for example, is the compassion between the therapist and the client (62, 63). Compassion consists of different elements: recognizing suffering, understanding its universality, emotional resonance, tolerating uncomfortable feelings, and the motivation to act to alleviate suffering (64). The presence of compassion in treatment has many benefits (65), and it can strengthen the therapeutic alliance (66). However, the ability to express and transfer compassion in digital treatments and therapeutic (text-based) relationships has not been examined yet, and no scale exists to measure compassion in digital treatment forms.

The WAI, for instance, consists of three subscales, measuring the agreement between client and therapist on goals and tasks, as well as the bond (16, 24). This final subscale, the bond, comes closest to measuring elements of compassion in treatment, but its items are limited to the mutual liking, respect and appreciation in the relationship. These items do not do the comprehensive concept of compassion justice. Moreover, the WAI stems from 1989 (24), and was not developed with digital treatment options in mind. Therefore, to examine if a fundamental value such as compassion does not get lost in digital, text-based treatment as compared to face-to-face treatment, a new scale to measure compassion in these treatment forms is needed.

Study Limitations

There are a few potential limitations concerning the results of this review. Firstly, the choice for a scoping review was made with the aim for a broad coverage of the topic. As is characteristic of scoping reviews, the current scoping review did not appraise the methodological quality of the reviewed articles. Secondly, regarding the used search strategy, search terms such as "online psychotherapy" and "online mental health" were chosen in an attempt to achieve a broad coverage and to not steer in the direction of certain mediums. Since the included interventions are mostly web-based, it is possible that we missed some studies that included other forms of digital text-based psychotherapy, such as via text-messaging or via apps. However, a search including additional search terms such as "SMS," "text-messaging" and "apps" shows that many interventions using these mediums pertain to unguided, self-help interventions and not guided psychotherapy.

Finally, abstracts included in this review were required to report the assessment of the working alliance (or therapeutic alliance or therapeutic relationship). It is possible that studies did not report their measurements of the working alliance in the abstract, and were therefore not included in this review. These limitations could be addressed in future reviews to provide an even more extensive synthesis of findings on the working alliance online.

Directions for Future Research

This review can be seen as a step toward a more positive perspective on relationships between clients and therapists in digital treatment and highlights the fact that high quality working alliances can be established digitally, and through text. Of course, this also raises new questions, such as to what extent this good digital therapeutic relationship is related to the therapist, and how much of it is mediated by the digital interaction. Moreover, future research could assess qualitatively how client and therapist interact online to establish this valuable relationship, since it remains unclear what the content is of their communication that has led to a good therapeutic alliance. This knowledge is critical to understand what helps build the therapeutic alliance, and could have practical implications for clients and therapists considering the use of text-based digital interventions. A more in-depth insight into the therapeutic relationship could also give more clarity on any elements that current measures might miss, such as the presence of compassion. This might give us some understanding of what type of behavior can enable a positive working alliance and, possibly, how digital treatment options and communication modalities can facilitate these behaviors.

Furthermore, the number of studies and interventions included in the current systematic scoping review shows that it is possible to conduct a meta-analysis on this topic. Such an analysis could provide a richer picture of the role of the therapeutic alliance in text-based digital psychotherapy, for example by evaluating which types of platform and forms of communication enable better therapeutic alliances and to what extent the alliance relates to the treatment outcome based on effect sizes.

Although the generalizability of the present results must be established in future research, and more comprehensive measurements of the therapeutic alliance are needed, this scoping review provides support for the possibility of a good working alliance in various text-based digital psychotherapeutic treatments and with various groups of people *diagnosed* with different mental disorders. It might inspire clinical psychologists, psychiatrists and mental health care workers to consider accessible internet- based options with a low-threshold as an addition or alternative to face-to-face treatment.

AUTHOR CONTRIBUTIONS

LJ and MLN conceptualized the study. CML and LJ reviewed the literature and led the paper drafting. All authors edited multiple drafts and reviewed the final paper.

FUNDING

This publication is part of the project designing compassionate technology with high societal readiness levels for mental healthcare (with project number 403.19.229) of the research programme Transitions and Behavior which is partly financed by the Dutch Research Council (NWO).

REFERENCES

1. Bergmo TS. How to measure costs and benefits of ehealth interventions: An overview of methods and frameworks. *J Med Internet Res.* (2015) 17:e254. doi: 10.2196/jmir.4521
2. Schwartzman CM, Boswell JF. A narrative review of alliance formation and outcome in text-based telepsychotherapy. *Pract Innov.* (2020) 5:128–42. doi: 10.1037/pri0000120
3. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. *World Psychiatry.* (2014) 13:288–95. doi: 10.1002/wps.20151
4. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cogn Behav Ther.* (2009) 38:196–205. doi: 10.1080/16506070903318960
5. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cogn Behav Ther.* (2018) 47:1–18. doi: 10.1080/16506073.2017.1401115
6. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in E-therapy for mental health: a systematic review. *J Med Internet Res.* (2012) 14:e110. doi: 10.2196/jmir.2084
7. Feijt MA, De Kort YAW, Bongers IMB, IJsselsteijn WA. Perceived drivers and barriers to the adoption of eMental health by psychologists: the construction of the levels of adoption of eMental health model. *J Med Internet Res.* (2018) 20:e153. doi: 10.2196/jmir.9485
8. Herbst N, Franzen G, Voderholzer U, Thiel N, Knaevelsrud C, Hertenstein E, et al. Working alliance in internet-based cognitive-behavioral therapy for obsessive-compulsive disorder. *Psychother Psychosom.* (2016) 85:117–8. doi: 10.1159/000441282
9. Ganapathy A, Clough BA, Casey LM. Organizational and policy barriers to the use of digital mental health by mental health professionals. *Telemed e-Health.* (2021) 1–12. doi: 10.1089/tmj.2020.0455
10. Berger T. The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. *Psychother Res.* (2017) 27:511–24. doi: 10.1080/10503307.2015.1119908
11. Yaphe J, Speyer C. *Text-Based Online Counseling: Email*. Online Counseling. 2nd ed. Amsterdam: Elsevier (2011). p. 147–67.
12. Whitty M, Gavin J. Age/sex/location: uncovering the social cues in the development of online relationships. *Cyberpsychol Behav.* (2001) 4:623–30. doi: 10.1089/109493101753235223
13. Klein JP, Berger T, Voderholzer U. Internetbasierte Interventionen bei psychischen Erkrankungen. *Ther Psych Erkrankungen.* (2019) 57:605–13. doi: 10.1016/B978-3-437-24911-2.00039-0
14. Lee M, Kumar R, Journal RL. The doctor is online: an introduction to text–based telepsychiatry. *UBC Med J.* (2017) 8:33–5.
15. Flückiger C, Del AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: a meta-analytic synthesis. *Psychotherapy.* (2018) 55:316–40. doi: 10.1037/pst0000172
16. Edward BS. Theory and Research on the theraputic working alliance: new directions. In: *The Working Alliance: Theory, Research, and Practice*. New York, NY: John Wiley & Sons (1994). p. 13–37.
17. Sucala M, Schnur JB, Brackman EH, Constantino MJ, Montgomery GH. Clinicians attitudes toward therapeutic alliance in E-therapy. *J Gen Psychol.* (2013) 140:282–93. doi: 10.1080/00221309.2013.830590
18. Pihlaja S, Stenberg JH, Joutsenniemi K, Mehik H, Ritola V, Joffe G. Therapeutic alliance in guided internet therapy programs for depression and anxiety disorders—a systematic review. *Internet Intervent.* (2018) 11:1–10. doi: 10.1016/j.invent.2017.11.005
19. Kaiser J, Hanschmidt F, Kersting A. The association between therapeutic alliance and outcome in internet-based psychological interventions: a meta-analysis. *Comput Human Behav.* (2021) 114:106512. doi: 10.1016/j.chb.2020.106512
20. Horvath AO, Del Re AC, Flückiger C, Symonds D. Alliance in individual psychotherapy. *Psychotherapy.* (2011) 48:9–16. doi: 10.1037/a0022186
21. Berger T, Andersson G. Internet-based psychotherapies: characteristics and empirical evidence. *Psychother Psychosom Med Psychol.* (2009) 59:159–66. doi: 10.1055/s-0028-1090162
22. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. *Int J Evid Based Healthc.* (2015) 13:141–6. doi: 10.1097/XEB.0000000000000050
23. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J.* (2009) 26:91–108. doi: 10.1111/j.1471-1842.2009.00848.x
24. Horvath AO, Greenberg LS. Development and validation of the working alliance inventory. *J Couns Psychol.* (1989) 36:223–33. doi: 10.1037/0022-0167.36.2.223
25. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: *ACM International Conference Proceeding Series*. London (2014).
26. Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* (2009) 6:e1000097. doi: 10.1371/journal.pmed.1000097
27. Cohen J. Set correlation and contingency tables. *Appl Psychol Meas.* (1988) 12:425–34. doi: 10.1177/014662168801200410
28. Andersson G, Paxling B, Wiwe M, Vernmark K, Felix CB, Lundborg L, et al. Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. *Behav Res Ther.* (2012) 50:544–50. doi: 10.1016/j.brat.2012.05.003
29. Andersson E, Ljótsson B, Hedman E, Enander J, Kaldo V, Andersson G, et al. Predictors and moderators of Internet-based cognitive behavior therapy for obsessive-compulsive disorder: Results from a randomized trial. *J Obsessive Compuls Relat Disord.* (2015) 4:1–7. doi: 10.1016/j.jocrd.2014.10.003

30. Anderson REE, Spence SH, Donovan CL, March S, Prosser S, Kenardy J. Working alliance in online cognitive behavior therapy for anxiety disorders in youth: Comparison with clinic delivery and its role in predicting outcome. *J Med Internet Res.* (2012) 14:e88. doi: 10.2196/jmir.1848
31. Bergman Nordgren L, Carlbring P, Linna E, Andersson G. Role of the working alliance on treatment outcome in tailored internet-based cognitive behavioural therapy for anxiety disorders: randomized controlled pilot trial. *JMIR Res Protoc.* (2013) 2:e4. doi: 10.2196/resprot.2292
32. Bisseling E, Cillessen L, Spinhoven P, Schellekens M, Compen F, Van Der Lee M, et al. Development of the therapeutic alliance and its association with internet-based mindfulness-based cognitive therapy for distressed cancer patients: secondary analysis of a multicenter randomized controlled trial. *J Med Internet Res.* (2019) 21:e14065. doi: 10.2196/14065
33. Blake Buffini K, Gordon M. One-to-one support for crisis intervention using online synchronous instant messaging: evaluating working alliance and client satisfaction. *Br J Guid Couns.* (2015) 43:105–16. doi: 10.1080/03069885.2014.987723
34. Dölemeyer R, Klinitzke G, Steinig J, Wagner B, Kersting A. Die therapeutische Beziehung in einem internetbasierten Programm zur Behandlung der Binge-Eating-Störung. *Psychiatr Prax.* (2013) 40:321–6. doi: 10.1055/s-0033-1343320
35. Duffy D, Enrique A, Connell S, Connolly C, Richards D. Internet-delivered cognitive behavior therapy as a prequel to face-to-face therapy for depression and anxiety: a naturalistic observation. *Front Psychiatry.* (2020) 10:902. doi: 10.3389/fpsyt.2019.00902
36. Penedo JMG, Babl AM, Holtforth M grosse, Hohagen F, Krieger T, Lutz W, et al. The association of therapeutic alliance with long-term outcome in a guided internet intervention for depression: secondary analysis from a randomized control trial. *J Med Internet Res.* (2020) 22:e15824. doi: 10.2196/15824
37. Hadjistavropoulos HD, Pugh NE, Hesser H, Andersson G. Therapeutic alliance in internet-delivered cognitive behaviour therapy for depression or generalized anxiety. *Clin Psychol Psychother.* (2017) 24:451–61. doi: 10.1002/cpp.2014
38. Jasper K, Weise C, Conrad I, Andersson G, Hiller W, Kleinstäuber M. The working alliance in a randomized controlled trial comparing Internet-based self-help and face-to-face cognitive behavior therapy for chronic tinnitus. *Internet Interv.* (2014) 1:49–57. doi: 10.1016/j.invent.2014.04.002
39. Klein B, Mitchell J, Abbott J, Shandley K, Austin D, Gilson K, et al. A therapist-assisted cognitive behavior therapy internet intervention for posttraumatic stress disorder: Pre-, post- and 3-month follow-up results from an open trial. *J Anxiety Disord.* (2010) 24:635–44. doi: 10.1016/j.janxdis.2010.04.005
40. Knaevelsrud C, Böttche M, Pietrzak RH, Freyberger HJ, Renneberg B, Kuwert P. An internet-based, therapist-assisted therapy for german elderly survivors of the world war II with posttraumatic stress symptoms. *J Nerv Ment Dis.* (2014) 202:651–8. doi: 10.1097/NMD.0000000000000178
41. Knaevelsrud C, Böttche M, Pietrzak RH, Freyberger HJ, Kuwert P. Efficacy and feasibility of a therapist-guided internet-based intervention for older persons with childhood traumatization: a randomized controlled trial. *Am J Geriatr Psychiatry.* (2017) 25:878–88. doi: 10.1016/j.jagp.2017.02.024
42. Knaevelsrud C, Maercker A. Does the quality of the working alliance predict treatment outcome in online psychotherapy for traumatized patients? *J Med Internet Res.* (2006) 8:e31. doi: 10.2196/jmir.8.4.e31
43. Knaevelsrud C, Maercker A. Internet-based treatment for PTSD reduces distress and facilitates the development of a strong therapeutic alliance: a randomized controlled clinical trial. *BMC Psychiatry.* (2007) 7:13. doi: 10.1186/1471-244X-7-13
44. Lindegaard T, Hesslow T, Nilsson M, Johansson R, Carlbring P, Lilliengren P, et al. Internet-based psychodynamic therapy vs cognitive behavioural therapy for social anxiety disorder: a preference study. *Internet Interv.* (2020) 20:100316. doi: 10.1016/j.invent.2020.100316
45. Preschl B, Maercker A, Wagner B. The working alliance in a randomized controlled trial comparing online with face-to-face cognitive-behavioral therapy for depression. *BMC Psychiatry.* (2011) 11:89. doi: 10.1186/1471-244X-11-189
46. Reynolds DJ, Stiles WB, Grohol JM. An investigation of session impact and alliance in internet based psychotherapy: preliminary results. *Couns Psychother Res.* (2006) 6:164–8. doi: 10.1080/14733140600853617
47. Scherer S, Alder J, Gaab J, Berger T, Ihde K, Urech C. Patient satisfaction and psychological well-being after internet-based cognitive behavioral stress management (IB-CBSM) for women with preterm labor: a randomized controlled trial. *J Psychosom Res.* (2016) 80:37–43. doi: 10.1016/j.jpsychores.2015.10.011
48. Topooco N, Byléhn S, Nysäter ED, Holmlund J, Lindegaard J, Johansson S, et al. Evaluating the efficacy of internet-delivered cognitive behavioral therapy blended with synchronous chat sessions to treat adolescent depression: Randomized controlled trial. J Med Internet Res. (2019) 21:e13393. doi: 10.2196/preprints.13393
49. Wagner B, Brand J, Schulz W, Knaevelsrud C. Online working alliance predicts treatment outcome for posttraumatic stress symptoms in arab war-traumatized patients. *Depress Anxiety.* (2012) 29:646–51. doi: 10.1002/da.21962
50. Busseri MA, Tyler JD. Interchangeability of the Working Alliance Inventory and Working Alliance Inventory, short form. *Psychol Assess.* (2003) 15:193–7. doi: 10.1037/1040-3590.15.2.193
51. Hatcher RL, Gillaspy JA. Development and validation of a revised short version of the Working Alliance Inventory. *Psychother Res.* (2006) 16:12–25. doi: 10.1080/10503300500352500
52. Bickman L, Vides De Andrade AR, Lambert EW, Doucette A, Sapyta J, Boyd AS, et al. Youth therapeutic alliance in intensive treatment settings. *J Behav Heal Serv Res.* (2004) 31:134–48. doi: 10.1007/BF02287377
53. McGuire-Snieckus R, McCabe R, Catty J, Hansson L, Priebe S. A new scale to assess the therapeutic relationship in community mental health care: STAR. *Psychol Med.* (2007) 37:85–95. doi: 10.1017/S0033291706009299
54. Agnew-Davies R, Stiles WB, Hardy GE, Barkham M, Shapiro DA. Alliance structure assessed by the Agnew Relationship Measure (ARM). *Br J Clin Psychol.* (1998) 37:155–72. doi: 10.1111/j.2044-8260.1998.tb01291.x
55. Reefhuis E, van Dijk M, Miggiels M, ten Klooster P. De kwaliteit van de therapeutische relatie voorspelt uitkomst van psychotherapie bij depressie. *Tijdschr voor Psychother.* (2019) 45:6–23.
56. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Arch Gen Psychiatry.* (2005) 62:593–602. doi: 10.1001/archpsyc.62.6.593
57. Kraus R, Stricker G, Speyer C. *Online Counselling: A Handbook for Mental Health Professionals*. 2nd ed. London: Elsevier (2011). doi: 10.1016/C2009-0-30506-1
58. Johansson R, Ekbladh S, Hebert A, Lindström M, Möller S, Petitt E, et al. Psychodynamic guided self-help for adult depression through the internet: a randomised controlled trial. *PLoS ONE.* (2012) 7:e38021. doi: 10.1371/journal.pone.0038021
59. Hendriks T, Schotanus-Dijkstra M, Hassankhan A, de Jong J, Bohlmeijer E. The efficacy of multi-component positive psychology interventions: a systematic review and meta-analysis of randomized controlled trials. *J Happiness Stud.* (2020) 21:357–90. doi: 10.1007/s10902-019-00082-1
60. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, et al. Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry.* (2018) 17:123–32. doi: 10.1002/wps.20513
61. Horvath AO. The therapeutic relationship: research and theory. *Psychother Res.* (2005) 15:3–7. doi: 10.1080/10503300512331339143
62. Gilbert P. Compassion and cruelty: a biopsychosocial approach. In: *Compassion: Conceptualisations, Research and Use in Psychotherapy*. London: Routledge (2005). p. 9–74. doi: 10.4324/9780203003459-6
63. Gilbert P. Introducing compassion-focused therapy. *Adv Psychiatr Treat.* (2009) 15:199–208. doi: 10.1192/apt.bp.107.005264
64. Strauss C, Lever Taylor B, Gu J, Kuyken W, Baer R, Jones F, et al. What is compassion and how can we measure it? A review of definitions and measures. *Clin Psychol Rev.* (2016) 47:15–27. doi: 10.1016/j.cpr.2016.05.004
65. Sinclair S, Russell LB, Hack TF, Kondejewski J, Sawatzky R. Measuring compassion in healthcare: a comprehensive and critical review. *Patient.* (2017) 10:389–405. doi: 10.1007/s40271-016-0209-5

66. Kemp J, Zhang T, Inglis F, Wiljer D, Sockalingam S, Crawford A, et al. Delivery of compassionate mental health care in a digital technology–driven age: scoping review. *J Med Internet Res.* (2020) 22:e16263. doi: 10.2196/16263

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Digital Health

METHODS
published: 15 July 2021
doi: 10.3389/fdgth.2021.690901

Applying a Digital Health Checklist and Readability Tools to Improve Informed Consent for Digital Health Research

Camille Nebeker [1,2]*, Maryam Gholami [1,2], Daniah Kareem [1,2] and Emily Kim [1]

[1] Research Center for Optimal Digital Ethics in Health, Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, CA, United States, [2] Design Lab, University of California, San Diego, La Jolla, CA, United States

Background: As research involving human participants increasingly occurs with the aid of digital tools (e.g., mobile apps, wearable and remote pervasive sensors), the consent content and delivery process is changing. Informed consent documents to participate in research are lengthy and difficult for prospective participants to read and understand. As the consent communication will need to include concepts and procedures unique to digital health research, making that information accessible and meaningful to the prospective participant is critical for consent to be informed. This paper describes a methodology that researchers can apply when developing a consent communication for digital health research.

Methods: A consent document approved by a US institutional review board was deconstructed into segments that aligned with federal requirements for informed consent. Three researchers independently revised each segment of text with a goal of achieving a readability score between a 6–8th grade level. The team then consulted with an external readability expert well-versed in revising informed consent documents into "plain language." The resulting text was evaluated using Microsoft Word and Online-Utility accessibility software. The final step involved adding visual images and graphics to complement the text. The Digital Health Checklist consent prototype builder was then used to identify areas where the consent content could be expanded to address four key domains of Access and Usability, Privacy, Risks and Benefits, and Data Management.

Results: The approved consent was evaluated at a 12.6 grade reading level, whereas the revised language by our study team received 12.4, 12, and 12.58, respectively. The final consent document synthesized the most readable of the three revised versions and was further revised to include language recommended by the software tool for improving readability, which resulted in a final revised consent readability score of a 9.2 grade level. Moreover, word count was reduced from 6,424 in the original consent to 679 in the rewritten consent form.

Conclusion: Utilizing an iterative process to design an accessible informed consent document is a first step in achieving meaningful consent to participate in digital health research. This paper describes how a consent form approved by an institutional review

Edited by:
Karina Vold,
University of Toronto, Canada

Reviewed by:
Vinaya Manchaiah,
Lamar University, United States
Giovanna Nunes Vilaza,
Technical University of Denmark, Denmark

***Correspondence:**
Camille Nebeker
Nebeker@eng.ucsd.edu

Specialty section:
This article was submitted to Human Factors and Digital Health, a section of the journal Frontiers in Digital Health

Received: 04 April 2021
Accepted: 17 June 2021
Published: 15 July 2021

Citation:
Nebeker C, Gholami M, Kareem D and Kim E (2021) Applying a Digital Health Checklist and Readability Tools to Improve Informed Consent for Digital Health Research. Front. Digit. Health 3:690901. doi: 10.3389/fdgth.2021.690901

board can be made more accessible to a prospective research participant by improving the document readability score, reducing the word count and assessing alignment with the Digital Health Checklist.

Keywords: informed consent, digital health, internal review board, human factors, human centered design, meaningful consent, Digital Health Checklist tool

INTRODUCTION

In biomedical and behavioral research conducted by regulated entities, obtaining the prospective informed consent of those who become participants in research a cornerstone of ethical research. The purpose of informed consent is to provide people who are considering whether to participate in research the information necessary to determine if they want to volunteer (1). The regulations along with principles described in the Belmont Report intended to guide ethical research are used to determine what information is typically presented in the consent document (1, 2). The US Federal Regulation for Human Research Protections (see 45 CFR 46.116) lists eight key areas that must be described within the consent form (i.e., purpose, experimental aspects, risks, benefits, etc.). In addition to content requirements, there are guidelines suggesting that consent language be accessible aiming for a 6–8th grade reading level and presented in a setting whereby the individual is able to consider the information without undue influence that may compromise their ability to volunteer.

While informed consent is a demonstration of the ethical principle of "Respect for Persons," in reality, the practice of composing, delivering and obtaining consent to participate in research is far from perfect. Some of the problems stem from assumptions we, as researchers, make as we engage in what is typically a transactional conversation with a prospective research participant. This dialogue begins with the researcher stating they are conducting research to answer an important question followed by details about who is eligible, what's involved, how data will be collected and so forth. The first problem, which is not trivial, is the assumption that people understand what the scientific method involves; however, many don't and, subsequently misunderstandings follow (3). For example, researchers found that even when people can explain what a study involves, they may experience therapeutic misconception, and believe they will receive a medical care (4). In addition to barriers due to consent content, how it's delivered can also presented challenges for achieving informed consent. A number of studies have looked at steps to improving the consent process for example, among older adults (5), with cognitively impaired individuals (6) and adolescents (7). Yet, as we venture into the digital age, more is needed before we can be confident that informed consent is truly informed (8).

As research involving human participants increasingly occurs with the aid of digital tools (e.g., mobile apps, wearable and remote pervasive sensors), the consent content and delivery process is changing. Using digital strategies, researchers can now recruit and enroll upwards of 20,000+ participants rapidly to study various health conditions. The mPower study is one example where Apple Research Kit was used to host a mobile health study focused on Parkinson's disease with a goal of enrolling 20,000 participants using a mobile e-consent process (9). There was no one-on-one conversation between the researcher and prospective participants—all consent information was delivered remotely by placing information and graphics on the prospective participant's smartphone screen. Wearable and home placed sensors are another method used to passively gather a participant's personal health data in their natural environment, unobtrusively. The challenge with studies that take place using social media platforms, mobile apps or other forms of passive, remote study engagement is in how the researcher conveys the complex concepts of digital data collection or technology delivered interventions so that participants actually understand what participation involves. In the digital research environment, not only is explaining the concept of research important but, addressing potential technology and data literacy challenges is also important (10, 11). While the literature reflects ongoing persistent challenges with the concept of informed consent, little guidance is available to support those in the digital health research community who are working to fit a square peg (current consent paradigm) into a round hole (emerging digital health modalities).

The intent of this paper is to provide guidance to digital health researchers on how they can improve the informed consent communication specific to digital health research. This paper describes the process of developing an accessible consent communication. To demonstrate the process, we used the IRB-approve informed consent document developed for a study that involves body worn sensors to capture natural behaviors between a mom and baby in the home environment. In addition, a new checklist tool and framework to guide the consent deconstruction and reconstruction process was used. The Digital Health Checklist (DHC) is a decision support tool developed with a goal of supporting digital health researchers to design safe and responsible digital health research studies (12), including the content developed for use in obtaining informed consent.

METHODS

The DHC Tool

The DHC was developed via an iterative participatory design process to support decision making during the research protocol and consent development process prior to submission to an ethics review committee [e.g., Institutional Review Board (IRB in the US) or Research Ethics Committee (REC in Canada, European Union)] (12). Given the new challenges in developing ethical digital health studies (e.g., privacy considerations, data management and consent) (see **Figure 1**), the DHC tool prompts

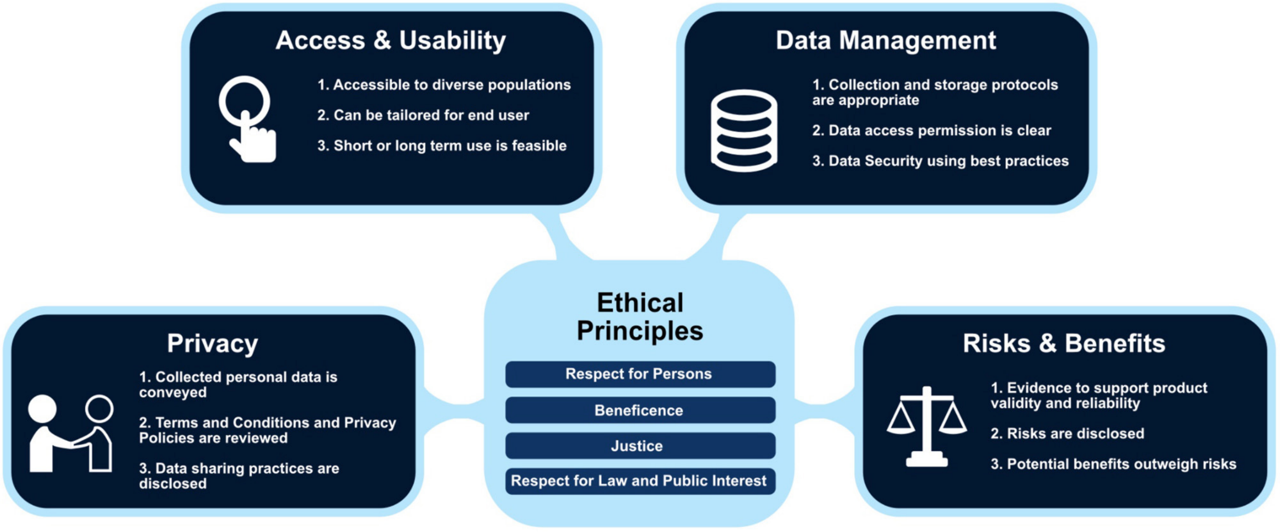


FIGURE 1 | Digital Health Framework with examples of checklist prompts embedded within each domain (Used with permission of C. Nebeker, ReCODE Health).

the researcher to consider factors that can influence responsible and safe research practices. The DHC is undergirded by a framework grounded in accepted ethical principles of respect for persons, beneficence and justice (1) augmented by a fourth principle of respect for law and public interest (13). The checklist items are depicted in a matrix table with the vertical listing ethical principles with the horizontal listed the four domains of: Access and Usability, Privacy, Risks and Benefits, and Data Management. For this prototype design process, we used the "respect for persons" section of the DHC as a blueprint to guide the consent content.

Study Design

This consent design process was initiated to support a longitudinal observational research study that would involve women and children as participants. Specifically, the US National Institutes of Health (NIH) funded a network of researchers to plan the HEALthy Brain and Child Development (HBCD) Study[1]. HBCD will examine early neurological development after prenatal exposure to maternal substance use using a variety of measures, including wearable and remote sensor technologies. The first author was part of an ethics and law working group involved with identifying and examining the ethical, legal and social implications associated with this potentially controversial research. Two passive sensor technologies were being considered for use during the study planning phase. One was a remote sensor that would be placed on the legs of a baby crib/bed to gather respiration and heart rate (14). The other, a body placed sensor that would be worn by the mother and baby to passively collect interpersonal data for specified periods in the home environment. Of the two devices, we selected the IRB-approved consent document developed for the wearable sensor since it involved prospective data collection from research participants.

This study involved a 4-step iterative human-centered design (HCD) process (15). The 4-step HCD include: (1) *Understand and Address the Core Problems,* to solve the fundamental issues, not the symptoms. (2) *Be People-Centered,* as opposed to technology-centered, ensuring that the outcome is appropriate for the culture and environment. (3) *Use an Activity-Centered Systems Approach,* focus upon the entire activity under consideration. (4) *Rapid Iterations of Prototyping and Testing,* and then refine and enhance the capabilities through successive iterations. While application of the HCD process is novel for developing a more accessible consent, we found it useful in conveying our process of developing an accessible consent communication.

Understand and Address the Core Problems

The content of informed consent communications used in regulated research is dictated by the federal regulations specific to human subject protections and local Institutional Review Boards. These documents include basic information that an individual may use to decide whether to participate in a research study. Consent communications have become increasingly transactional and include legal disclaimers on top of the basic information about research study participation (e.g., purpose, procedures, risks, benefits, data management, conflicts of interest). This has added to the length of consent communications and has elevated the reading level to around a 10–12th grade, making it difficult for many readers to comprehend (16, 17). To understand a consent communication for digital health research, there are added complexities in that the reader will need, in addition to a level of research literacy, a level of technology and data literacy (18).

---

[1] https://heal.nih.gov/research/infants-and-children/healthy-brain

TABLE 1 | US federal regulations state that the following information be conveyed to prospective research participants prior to enrolling as a volunteer in research.

Statement that describes:	Detail needed
• Research study involvement	Explain the purpose, expected duration of participation, what procedures will be followed and a description of experimental aspects.
• Study benefits	Describe any direct benefits to the participant or others, which may be anticipated.
• Study risks	Describe possible risks of harm to the participant.
• Appropriate alternatives	Disclose other options that may be advantageous specific to procedures or possible treatments.
• Confidentiality practices	State how records identifying the participant will be maintained.
• Whether/how injury will be compensated (only if study exceeds minimal risk of harm)	Explain whether compensation is available to cover study related for medical treatment or other injury.
• Study team contact	Identify who to contact if there are questions or to report a research related injury.
• Voluntary nature of participation	Make clear that participation is voluntary and that there is no penalty or loss of benefits if the individual chooses not to participate or changes their mind after initial agreement to enroll.

The first step was to deconstruct the IRB-approved consent content by breaking it into segments that aligned with the federal requirements elements of informed consent to participate in research (see 46 CFR 46.116), (see **Table 1**). By doing this, we could focus on what communication was needed to comply with the federal regulations. Upon completion of this step, the research team discussed the challenges they faced while reading each segment and commented on the document length, technical language, and redundant information.

Be People-Centered

Unfortunately, in academic research, researchers are torn between making the consent accessible to those who may be recruited to participate in the research and adhering to the consent template that the IRB wants researchers to follow. The IRB-approved consent for selected for our use-case exceeded the recommended 6–8th grade reading level that IRB guidance suggests. As such, the next step involved our three researchers (MG, DK, EK) independently revised each segment of text with a goal of achieving a readability score of a 6–8th grade reading level. This participatory design process provides valuable insights as the researchers are engaged directly in the task of trying to develop consent language as a researcher would when applying our method to their consent communication process (19).

As noted, a norm of US human research ethics is to aim for a readability score that a majority of the adult population would be able to read, however, rarely is this goal achieved. A challenge was encountered by our team when attempting to revise language occurred when attempting to describe the technology intended for use in the research (passive sensor devices) along with the legal language that the IRB requires. Disclosure of reporting requirements is routine in some studies due to legal requirements like reporting mandates (e.g., disclosure of illegal behaviors like child or elder abuse). That was true for the consent serving as our use-case. The language required by the IRB to convey indemnification and mandated reporting was nearly impossible to reduce to an accessible reading level.

TABLE 2 | Readability analysis.

	Original IRB approved consent	Rewritten consent by research team
Word count	2,464	679
Readability grade	Microsoft Word: 13.3 Online-Utility: 11.91	Microsoft Word: 9.3 Online-Utility: 9.13

Use an Activity-Centered Systems Approach

Once each team member had revised the consent segment to the best of their ability, they reviewed all revisions to ensure alignment with the federal regulations and applied a readability software to assess grade level. The revised segments were analyzed using a readability feature in Microsoft Word, since that tool was compatible for analyzing smaller text segments and accessible to the team. The team members then compared the different versions of text and chose the version that achieved the lowest grade level.

Further iterations were needed to reach the 6–8th grade reading level. The team then consulted with an external readability expert well-versed in revising informed consent documents into "plain language" (20). The external consultant used Readability Studio 1.1 to assess the IRB-approved consent form, which provides grade and difficulty level along with suggestions for how to further simplify the language. The team implemented the suggested wording and finalized the revised document. The final step involved inserting visual images of the technology and graphics to complement the text and improve readability.

Rapid Iterations of Prototyping and Testing

The last step involved applying the Digital Health Checklist (DHC) consent prototype tool to identify areas where the consent content could be expanded to address the four domains of Access and Usability, Privacy, Risks and Benefits, and Data Management. Each of the four domains are expanded in the "respect for persons" row of the checklist matrix, which corresponds to what a researcher should consider when developing the informed consent document so that specific information about a digital health strategy/tool can be addressed. Not all of the checklist prompts will be relevant but, the checklist facilitates reflection of what might have been overlooked—particularly if relying on an IRB consent template to guide content. Our team compared what was in the IRB-approved and subsequently revised consent form to the DHC and identified content areas that would need to be added to improve the consent for use in this digital health study.

RESULTS

Readability and Content

The original IRB-approved consent form and the revised text were analyzed internally using both Microsoft Word and Online-Utility 1.1, which provide average readability and grade level scores between the two software (see **Table 2**). The results of the IRB-approved consent showed a 12.6 reading level, whereas the revised language by our study team received 12.4, 12, and 12.58, respectively. The final consent document synthesized the most readable of the three revised versions. We then incorporated language recommended by the software tool for improving readability, which resulted in a final revised consent readability score of a 9.2 grade level. **Table 2**. Readability analyses and word count of original IRB-approved consent form and final revised version.

For a more detailed example of how the text was modified to improve readability, see **Figures 2**, **3** below which illustrate examples of a paragraph in the rewritten (**Figure 2**) vs. the original (**Figure 3**) consent form. This text focuses on risk management and how the study team will be trained to respect participant privacy. The original text was 100 words, and the sentences were much longer when compared to the revised version by 52 words.

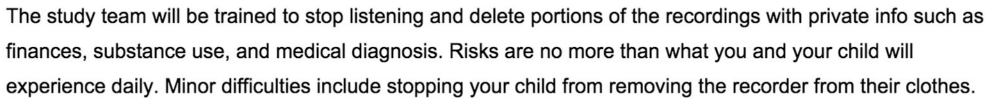
The research team was able to improve the readability and lower the reading level of the passages by 3 grade-levels from the original version, however, did not achieve the targeted 6–8th grade reading level. This was achieved by following the plain language guidelines published by the US government[2], which includes using words with fewer syllables, shorter sentences and shorter paragraphs.

Presentation and Visuals

The final revised version of the consent was augmented to include graphics and pictures of the digital tool. See **Figures 4**, **5** to compare presentation and visuals of the revised (**Figure 4**) and original (**Figure 5**) versions of the consent forms.

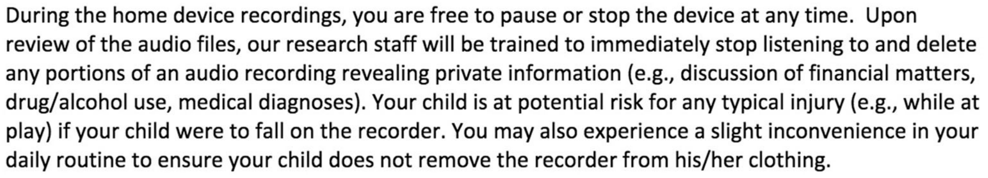
The DHC tool was then referenced to identify consent content that could be expanded to address areas specific to Access and Usability, Privacy, Risks and Benefits, and Data Management in the revised consent form. **Table 3** illustrates the four domains and their prompts that were used to evaluate the consent form. Under the *Access and Usability* domain, all the statements noted in the DHC tool were addressed in both the original and revised consent document. Under the *Risks and Benefits* domain, which covers potential harms and impact with respect to possible benefits, 10 of the 12 statements were addressed. Under the *Privacy* domain, which covers the extent, purpose, and sharing of personal data, two of the seven statements were addressed in the original consent and carried forward to the revised consent. Under the *Data Management* domain, two of the 10 checklist prompts were addressed.

While the original consent included basic information specific to Access and Usability deemed necessary for informed consent to occur, the other three domains were lacking. The next iteration of the consent form will be revised to include information about reputational and unknown harms as well as specify why personal data are being collected and where data are stored and the extent of 3rd party access. Moreover, the possibility of a bystander being recorded is important to address and was not included in the original consent. Bystanders are not typically considered when consenting a research participant but, is increasingly important given the passive and pervasive nature of sensor technologies. Given the consent used in this exercise described a study that used a wearable microphone, addressing bystander



The study team will be trained to stop listening and delete portions of the recordings with private info such as finances, substance use, and medical diagnosis. Risks are no more than what you and your child will experience daily. Minor difficulties include stopping your child from removing the recorder from their clothes.

FIGURE 2 | Example of a paragraph in the rewritten version consent form.

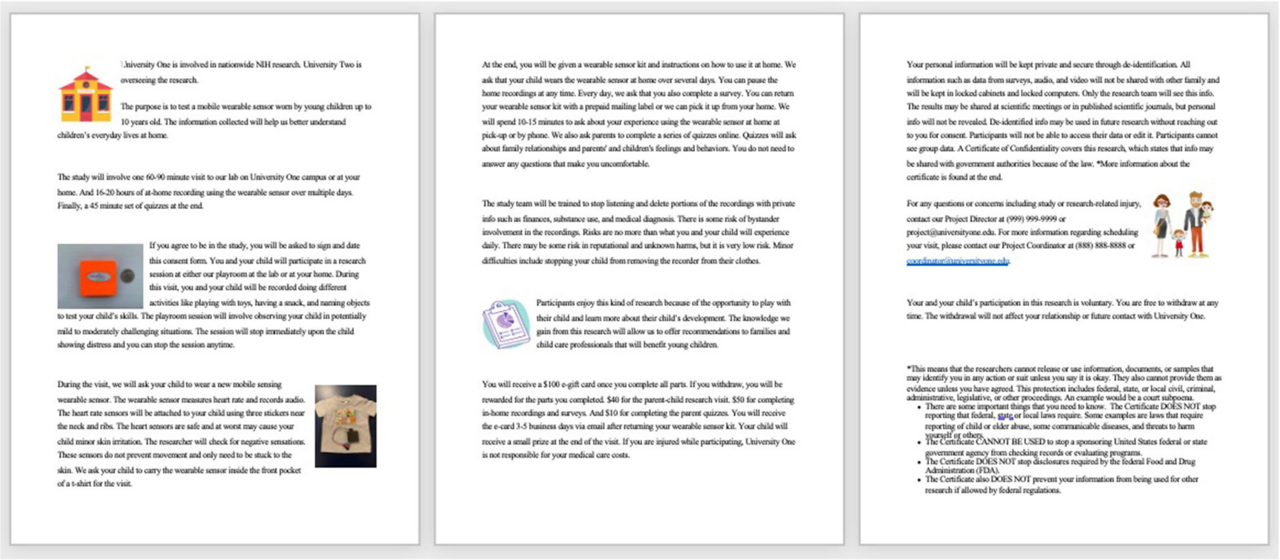


During the home device recordings, you are free to pause or stop the device at any time. Upon review of the audio files, our research staff will be trained to immediately stop listening to and delete any portions of an audio recording revealing private information (e.g., discussion of financial matters, drug/alcohol use, medical diagnoses). Your child is at potential risk for any typical injury (e.g., while at play) if your child were to fall on the recorder. You may also experience a slight inconvenience in your daily routine to ensure your child does not remove the recorder from his/her clothing.

FIGURE 3 | Example of a paragraph in the original IRB-approved consent form.

---

[2] https://www.plainlanguage.gov/media/FederalPLGuidelines.pdf



University One is involved in nationwide NIH research. University Two is overseeing the research.

The purpose is to test a mobile wearable sensor worn by young children up to 10 years old. The information collected will help us better understand children's everyday lives at home.

The study will involve one 60-90 minute visit to our lab on University One campus or at your home. And 16-20 hours of at-home recording using the wearable sensor over multiple days. Finally, a 45 minute set of quizzes at the end.

If you agree to be in the study, you will be asked to sign and date this consent form. You and your child will participate in a research session at either our playroom at the lab or at your home. During this visit, you and your child will be recorded doing different activities like playing with toys, having a snack, and naming objects to test your child's skills. The playroom session will involve observing your child in potentially mild to moderately challenging situations. The session will stop immediately upon the child showing distress and you can stop the session anytime.

During the visit, we will ask your child to wear a new mobile sensing wearable sensor. The wearable sensor measures heart rate and records audio. The heart rate sensors will be attached to your child using three stickers near the neck and ribs. The heart sensors are safe and at worst may cause your child minor skin irritation. The researcher will check for negative sensations. These sensors do not prevent movement and only need to be stuck to the skin. We ask your child to carry the wearable sensor inside the front pocket of a t-shirt for the visit.

At the end, you will be given a wearable sensor kit and instructions on how to use it at home. We ask that your child wears the wearable sensor at home over several days. You can pause the home recordings at any time. Every day, we ask that you also complete a survey. You can return your wearable sensor kit with a prepaid mailing label or we can pick it up from your home. We will spend 10-15 minutes to ask about your experience using the wearable sensor at home at pick-up or by phone. We also ask parents to complete a series of quizzes online. Quizzes will ask about family relationships and parents' and children's feelings and behaviors. You do not need to answer any questions that make you uncomfortable.

The study team will be trained to stop listening and delete portions of the recordings with private info such as finances, substance use, and medical diagnosis. There is some risk of bystander involvement in the recordings. Risks are no more than what you and your child will experience daily. There may be some risk in reputational and unknown harms, but it is very low risk. Minor difficulties include stopping your child from removing the recorder from their clothes.

Participants enjoy this kind of research because of the opportunity to play with their child and learn more about their child's development. The knowledge we gain from this research will allow us to offer recommendations to families and child care professionals that will benefit young children.

You will receive a $100 e-gift card once you complete all parts. If you withdraw, you will be rewarded for the parts you completed. $40 for the parent-child research visit. $50 for completing in-home recordings and surveys. And $10 for completing the parent quizzes. You will receive the e-card 3-5 business days via email after returning your wearable sensor kit. Your child will receive a small prize at the end of the visit. If you are injured while participating, University One is not responsible for your medical care costs.

Your personal information will be kept private and secure through de-identification. All information such as data from surveys, audio, and video will not be shared with other family and will be kept in locked cabinets and locked computers. Only the research team will see this info. The results may be shared at scientific meetings or in published scientific journals, but personal info will not be revealed. De-identified info may be used in future research without reaching out to you for consent. Participants will not be able to access their data or edit it. Participants cannot see group data. A Certificate of Confidentiality covers this research, which states that info may be shared with government authorities because of the law. *More information about the certificate is found at the end.

For any questions or concerns including study or research-related injury, contact our Project Director at (999) 999-9999 or project@universityone.edu. For more information regarding scheduling your visit, please contact our Project Coordinator at (888) 888-8888 or coordinator@universityone.edu.

Your and your child's participation in this research is voluntary. You are free to withdraw at any time. The withdrawal will not affect your relationship or future contact with University One.

*This means that the researchers cannot release or use information, documents, or samples that may identify you in any action or suit unless you say it is okay. They also cannot provide them as evidence unless you have agreed. This protection includes federal, state, or local civil, criminal, administrative, legislative, or other proceedings. An example would be a court subpoena.

- There are some important things that you need to know. The Certificate DOES NOT stop reporting that federal, state or local laws require. Some examples are laws that require reporting of child or elder abuse, some communicable diseases, and threats to harm yourself or others.
- The Certificate CANNOT BE USED to stop a sponsoring United States federal or state government agency from checking records or evaluating programs.
- The Certificate DOES NOT stop disclosures required by the federal Food and Drug Administration (FDA).
- The Certificate also DOES NOT prevent your information from being used for other research if allowed by federal regulations.

FIGURE 4 | Rewritten version of consent form by research team that includes all required elements of informed consent.

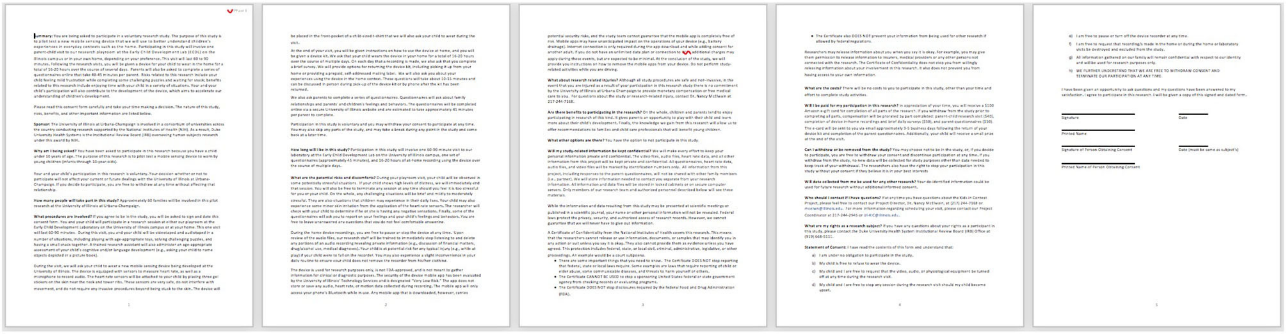


FIGURE 5 | Original IRB approved consent form approved by an IRB.

considerations is appropriate. Lastly, information about data practices including data transfer, storage, and sharing along with how much access participants will have to collected data will be included in the next iteration of the consent document. At that point, the consent form will include all recommendations in the DHC informed consent blueprint, be accessible in terms of reading grade level and will advance to the stage of further iterative design with prospective research participants.

DISCUSSION

The main objective of this paper was to provide a step-by-step description of developing an informed consent communication. Using a participatory design, we included researchers who are involved with creating consent communications but who have little experience. Few would argue that valid consent requires that a person be provided with adequate and relevant information. Yet, the process of developing an informed consent document is typically guided by a template that the research ethics board provides for the purpose of helping a researcher create a document that complies with federal regulations and institutional practices. Unfortunately, the consent templates do not include guidance on how to make the consent language or presentation of information accessible or particularly meaningful.

What might make informed consent meaningful has been a subject of study though, whether it can be achieved in practice is uncertain. Dranseika et al. (21) suggested that researchers take the time to learn what information might be relevant for prospective participants and actually speak with patients to learn what might contribute to their decisions about participation in a study. Moreover, they called for empirical research to understand the concept of relevance and how consent content might vary depending on socio-economic and cultural background (21). Most empirical research on informed consent to date has focused on comprehension of the consent content and, subsequent understanding of the research. For example (22), designed and tested an instrument to assess participant objective and subjective understanding of a cancer clinical trial (22). Wilbanks (11) recognized that problems may exist in the traditional consent process and explored how the concept of a choice architecture (23) might be used in guiding decision to participate in digital research whereby consent information was presented on a personal mobile device (11). In fact, Wilbanks argued that in an era of technology mediated clinical and biomedical research with the associated volume, velocity and variety of data, that bioethics must meet the new demands. Experimenting with new design elements with a focus on linear, graphic/pictorial and brief narrative, the team at Sage Bionetworks created a consent flow that was used to communicate informed consent content via an iPhone. Similar to other studies, the need to engage people prospectively in the design process was a limitation. Formative research with mPower study participants conducted by (10) similar inconsistencies in understanding as would be found in traditional face to face consent but, highlighted a desire by participants to be partners in research (10).

The importance of engaging "end users," in this case researchers and, eventually research participants, early in the design phase of a consent design process cannot be understated. Applying human centered design principles to the concept of informed consent makes sense yet, there is limited literature on this topic. The exceptions are the work of (24) who published a conceptual model of design principles for informed consent related to cookie technology and web browser design (24) and Wilbank's work mentioned previously (11). Recognizing the need to move toward a meaningful and accessible informed consent communication for digital health research is what led to the design process described in this paper.

In this study, we have taken steps to bridge the gap in accessible and meaningful informed consent by moving beyond a transactional form to a presentation of information that is likely to be read and understood. An iterative process was used to create consent information that can be presented to a prospective research participant by first improving the document readability score and then aligning content with the Digital Health Checklist tool. By utilizing the DHC "respect for persons" consent prototype builder, we were able to guide alignment with the four domains of: Access and Useability, Privacy, Risks and Benefits and Data Management.

With this revised consent communication as a starting point, we now plan to engage prospective research participants in iterative consent design workshops to move toward the ideal of meaningful consent. The next phase of this research will involve people who may eventually participate in our larger HBCD study. They will be asked to comment on the relevance and clarity of the consent language. Building on the Digital Health Checklist and emerging work on participant-centered and dynamic consent models, we will include prompts for participants to rate the relevance of aspects of digital health research that are unique and challenging.

For this initial work, our goal is to help researcher understand and apply a process for conveying complex topics, via a consent communication using tools to make language accessible and content complete. Areas of interest expressed by researchers, which led, in part, to development of the DHC tool, are framed as "how might we" questions. Examples follow:

- Improve understanding of how the technology works?
- Convey individual and societal implications of the knowledge gained?
- Communicate how personal health information is transmitted and stored to the cloud?
- Describe differences between real-time data collection?
- Respect preferences for privacy and control of personal information?
- Understand the extent of control participants want with respect to managing data?
- Accurately convey how personalized algorithms work to nudge behavior change?
- Gauge acceptance of health technologies among family, coworkers and friends?

Clearly, informed consent to participate in digital health research has received little attention from a human centered design perspective. With increasing interest from large scale programs, like the All of Us Research Program and Patient Centered Outcomes Research Initiative, to engage with research participants as partners in the learning process, the opportunities are exciting. The ethical principle of "respect for persons" requires that we actually do more than create a transaction to demonstrate compliance between a researcher and participant. To authentically demonstrate "respect for persons" is to co-design the consent content and process to improve capacity among researchers so that the person considering study participation is informed and able to make a decision about whether to volunteer. Through a human centered design process, we can move from a transaction to a meaningful exchange of information that may lead to an informed consent in practice.

Our planned summative research will expand the work reported here. We encourage other researchers to replicate this process when creating their consent communications. While the results will vary since each study is unique in context, we are confident our methods, conveyed via an authentic use case, can serve as a concrete example.

LIMITATIONS

The informed consent prototype design process described here has not involved people external to our research team; however, we have confidence that our team is similar to those who would be eligible to enroll. Specifically, co-authors involved with the deconstruction exercise included two members of our research team (EK, DK) who had no prior experience writing or reviewing informed consent documents and one member (MG) who had limited experience with preparing research protocols for IRB review. The senior author (CN) is a subject matter expert in research ethics and did not participate in the deconstruct/rebuild exercise. While we have taken the preliminary steps to make the IRB-approved consent more accessible via a lower readability score, we have not tested the language or obtained feedback on

TABLE 3 | Digital health checklist consent blueprint: ethical principle "respect for persons" across four domains.

Four domains of DHC tool	Prompts for each domain	Yes	No	N/A
Access and usability	1. An explanation about the technology used in the study that helps one to understand: What it does	X		
	2. An explanation about the technology used in the study that helps one to understand: Why it is being used	X		
	3. An explanation about the technology used in the study that helps one to understand: How it works	X		
	4. Plain language is used to describe the commercial vendor agreements: Terms of Service			X
	5. Plain language is used to describe the commercial vendor agreements: Privacy policy			X
	6. Relevant definitions provided using plain language	X		
	7. Access to visual and audio versions of information, if these alternatives are available	X		
Risks and benefits	1. A description of the type of potential harm including: Physical harm (e.g., skin irritation)	X		
	2. A description of the type of potential harm including: Psychological harm (e.g., distress)	X		
	3. A description of the type of potential harm including: 4. Economic harm (e.g., cost that the participants might incur as a result of using the technology)		X	
	5. A description of the type of potential harm including: Unknown harm (Even when these harms remain unknown - a statement acknowledging that there might be harms that are unknown included)	X		
	6. For potential harms a description the known or unknown: Severity		X	
	7. For potential harms a description the known or unknown: Duration	X		
	8. For potential harms a description the known or unknown: Intensity	X		
	9. Strategies for minimizing risks	X		
	10. Strategies for managing risks	X		
	11. Statement that indicates: Possible benefits from knowledge gained during the study	X		
	12. Statement that indicates: Benefits that could be derived by the participant related to the technology	X		
Privacy	1. Nature of personal information collected by the technology	X		
	2. Purpose for which personal information is collected by the technology		X	
	3. Extent of personal information collected by the technology (specific/inclusive list of personal information that will be collected by the technology)	X		
	4. How individual-level data will be shared and with whom, and if this might change in the future after the study		X	
	5. Whether personal data entered and stored in the technology will be de-identified		X	
	6. A description of how a 3rd party may access and use participant information collected during study participation (normally found in a privacy policy when using a commercial device)		X	
	7. Notification if there is a possibility of bystander involvement		X	
Data management	1. Practices for: Data collection by the technology	X		
	2. Practices for: Data security of the data that is collected by the technology		X	
	3. Practices for: Data sharing with other stakeholders		X	
	4. Practices for: Data transfer from technology to other storage		X	
	5. Practices for: Data storage of data that is collected by the technology		X	
	6. Information about who will have access to data collected by the technology		X	
	7. Whether the research data are controlled by the research team or a third party	X		
	8. Whether the participant will have access to individual-level data collected via the technology		X	
	9. Whether the participant will be able to edit individual-level data collected via the technology		X	
	10. Whether the participant will have access to group-level data collected via the technology		X	

whether prospective participants find the additional information prompted by the Digital Health Checklist to be relevant or meaningful.

CONCLUSION

To achieve responsible digital health requires that we design our studies, to the extent possible, with our research participants and put their interests at the forefront. The wild west of the digital health era allows for exciting innovation and yet, without a purposeful philosophy of "respect for persons" at the core, we as a community of researchers, technologists, clinicians and citizens will make avoidable mistakes. This paper describes the initial steps that researchers can apply for creating an accessible informed consent for use in digital health research. By making information developed for prospective participants accessible, we can then take a human centered approach to learning what is relevant and how best to convey information that matters most to those we will include in future research studies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

CN conceptualized the project, led the design of prototype development and application of the Digital Health Checklist, and developed the first draft of the manuscript. DK and EK contributed to the consent analysis process, revisions using the readability, checklist tools and created figures, and tables used in this paper. MG led the co-authors (DK and EK) in deconstructing and revising the consent content and prototype development and contributed to the methods and result sections of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Institutes on Drug Abuse (NIDA) grant number: 3R34DA050341-01S3 bioethics supplement.

ACKNOWLEDGMENTS

We thank our colleague Dr. Mark Hochhauser, Readability Consultant, for his insights into writing a consent form at the 6–8th grade reading level.

REFERENCES

1. National Institutes of Health. The belmont report ethical principles and guidelines for the protection of human subjects of research. *J Am Coll Dent.* (1979) 81:4–13.
2. Office for Human Research, Protections. *Federal Policy for the Protection of Human Subjects ('Common Rule). Regulations*. (2016). Available online at: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/index.html
3. Brody JL, Dalen J, Annett RD, Scherer DG, Turner CW. Conceptualizing the role of research literacy in advancing societal health. *J Health Psychol.* (2012) 17:724–30. doi: 10.1177/1359105311425273
4. Appelbaum PS, Roth LH, Lidz C. The therapeutic misconception: informed consent in psychiatric research. *Int J Law Psychiatry.* (1982) 5:319–29. doi: 10.1016/0160-2527(82)90026-7
5. Sugarman J, McCrory DC, Powell D, Krasny A. Empirical research on informed consent. An annotated bibliography. *Has Cen Rep.* (1999) 29:S1–U1. doi: 10.2307/3528546
6. Palmer BW, Cassidy EL, Dunn LB, Spira AP, Sheikh JI. Effective use of consent forms and interactive questions in the consent process. *IRB.* (2008) 30: 8–12.
7. Al-Sheyab NA, Alomari MA, Khabour OF, Shattnawi KK, Alzoubi KH. Assent and consent in pediatric and adolescent research: school children’s perspectives. *Adol Health Med Ther.* (2019) 10:7–14. doi: 10.2147/AHMT.S185553
8. Bloss C, Nebeker C, Bietz M, Bae D, Bigby B, Devereaux M, et al. Reimagining human research protections for 21st century science. *J Med Int Res.* (2016) 18:e329. doi: 10.2196/jmir.6634
9. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The MPower study, Parkinson disease mobile data collected using researchkit. *Sci Data.* (2016) 3:1–9. doi: 10.1038/sdata.2016.11
10. Doerr M, Truong AM, Bot BM, Wilbanks J, Suver C, Mangravite LM. Formative evaluation of participant experience with mobile econsent in the app-mediated parkinson mpower study: a mixed methods study. *JMIR Mhealth Uhealth*. (2017) 5:e14. doi: 10.2196/mhealth.6521
11. Wilbanks J. Design issues in E-consent. *J Law Med Ethics.* (2018) 46:110–18. doi: 10.1177/1073110518766025
12. Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support technology selection in digital health research. *Trans Behav Med.* (2020) 10:1004–15. doi: 10.1093/tbm/ibz074
13. Dittrich D, Kenneally E. The menlo report: ethical principles guiding information communication technology research. *U.S. Dep Home Secur.* (2012) 1–17. doi: 10.2139/ssrn.2445102
14. Harrington N, Wei Z, Hernandez-Pacheco B, Biu Q, DeYoung PN, Duwaik B. et al. Bedscales: a non-contact adherence-independent multi-person sensor for longitudinal physiologic monitoring in the home bed. *BioRxiv*. (2020). doi: 10.1101/2020.03.31.018523
15. Norman DA, Draper SW. *User Centered System Design*. Hillsdale, NJ: Lawrence Erlbaum Associates (1986).
16. Hadden KB, Prince LY, Moore TD, James LP, Holland JR, Trudeau CR. Improving readability of informed consents for research at an academic medical institution. *J Clin Trans Sci.* (2017) 1:361–5. doi: 10.1017/cts.2017.312
17. Samadi A, Asghari F. Readability of informed consent forms in clinical trials conducted in a skin research center. *J Med Ethics Hist Med.* (2016) 9:7.
18. Nebeker C, Harlow J, Espinoza Giacinto R, Orozco-Linares R, Bloss CS, Weibel N. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. *AJOB Emp Bio.* (2017) 8:266–76. doi: 10.1080/23294515.2017.1403980
19. Jorgensen DL. Participant observation. In: Scott RA, Kosslyn SM, editors. *Emerging Trends in the Social and Behavioral Sciences*. (2015). doi: 10.1002/9781118900772.etrds0247
20. Stableford S, Mettger W. Plain language: a strategic response to the health literacy challenge. *J Public Health Policy.* (2007) 28:71–93. doi: 10.1057/palgrave.jphp.3200102
21. Dranseika V, Piasecki J, Waligora M. Relevant information and informed consent in research: in defense of the subjective standard of disclosure. *Sci Eng Ethics.* (2017) 23:215–25. doi: 10.1007/s11948-016-9755-4
22. Joffe S, Cook EF, Cleary PD, Clark JW, Weeks JC. Quality of informed consent: a new measure of understanding among research subjects. *JNCI J Natio Can Inst.* (2001) 93:139–47. doi: 10.1093/jnci/93.2.139
23. Thaler RH, Sunstein CR, Balz JP. Choice architecture. *SSRN Electr J.* (2010) 428–39. doi: 10.2139/ssrn.1583509
24. Millett LI, Friedman B, Felten E. Cookies and web browser design: toward realizing informed consent online. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2001). p. 46–52.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Digital Health

PERSPECTIVE
published: 06 August 2021
doi: 10.3389/fdgth.2021.689736

Is the Automation of Digital Mental Health Ethical? Applying an Ethical Framework to Chatbots for Cognitive Behaviour Therapy

Giovanna Nunes Vilaza[1*†] *and Darragh McCashin*[2†]

[1] *Health Tech Department, Technical University of Denmark, Kongens Lyngby, Denmark,* [2] *School of Psychology, Dublin City University, Dublin, Ireland*

The COVID-19 pandemic has intensified the need for mental health support across the whole spectrum of the population. Where global demand outweighs the supply of mental health services, established interventions such as cognitive behavioural therapy (CBT) have been adapted from traditional face-to-face interaction to technology-assisted formats. One such notable development is the emergence of Artificially Intelligent (AI) conversational agents for psychotherapy. Pre-pandemic, these adaptations had demonstrated some positive results; but they also generated debate due to a number of ethical and societal challenges. This article commences with a critical overview of both positive and negative aspects concerning the role of AI-CBT in its present form. Thereafter, an ethical framework is applied with reference to the themes of (1) beneficence, (2) non-maleficence, (3) autonomy, (4) justice, and (5) explicability. These themes are then discussed in terms of practical recommendations for future developments. Although automated versions of therapeutic support may be of appeal during times of global crises, ethical thinking should be at the core of AI-CBT design, in addition to guiding research, policy, and real-world implementation as the world considers post-COVID-19 society.

Keywords: artificial intelligence, conversational agents, mental health, cognitive behavioural therapy, ethics

Edited by:
Karina Vold,
University of Toronto, Canada

Reviewed by:
Jan David Smeddinck,
Newcastle University, United Kingdom
Anat Elhalal,
Independent Researcher, London, United Kingdom

****Correspondence:***
Giovanna Nunes Vilaza
gnvi@dtu.dk

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Human Factors and Digital Health, a section of the journal Frontiers in Digital Health

Received: *01 April 2021*
Accepted: *16 July 2021*
Published: *06 August 2021*

Citation:
Vilaza GN and McCashin D (2021) Is the Automation of Digital Mental Health Ethical? Applying an Ethical Framework to Chatbots for Cognitive Behaviour Therapy. Front. Digit. Health 3:689736. doi: 10.3389/fdgth.2021.689736

INTRODUCTION

The unprecedented global crisis has intensified and diversified private distress sources, making evident the need for broader access to psychological support (1). A nationwide survey in China shows how the pandemic has triggered an increase in cases of panic disorder, anxiety, and depression (2). Infected individuals, medical staff and their families are under constant psychological pressure, in addition to the increasing number of people dealing with bereavement (3, 4).

At the same time, the pandemic enabled broader acceptance of telehealth by health professionals and clients alike (5). Video consultations are now increasingly advocated as an alternative for in-person consultations (6). Additionally, automated conversational agents and chatbots are increasingly promoted as potentially efficient emotional support tools for larger population segments during the pandemic (7) and afterwards (8).

It is now over 50 years since ELIZA was created (9), the first computer programme to use pattern matching algorithms to mimic human-therapist interactions by mechanically connecting end-user inputs to answers from a pre-defined set of responses. More recent approaches to language modelling can produce more sophisticated dialogues by employing machine learning and natural language processing (NLP). However, despite these advances, a recent global survey of psychiatrists across 22 countries ($n = 791$) demonstrated that only 3% feel that AI will likely replace a human for providing empathetic care (10). Such evidence indicates a contradiction between public enthusiasm (11) and the scepticism of service providers.

In light of these circumstances, we approach the development of automated psychotherapy from an ethical perspective. A recent review found that most mental health apps have not improved their safety over the last year, as most lack clinical evidence and trustworthy privacy policies (12). Beyond that, substandard regulations, ill-intended actors and commercial opportunism increase the risk of adverse responses and potentially lead to harm (personal and societal). Therefore, a significant concern endures: how AI can be integrated within psychotherapy in a safe, respectful, and effective way for end-users.

This perspective paper contributes with a structured discussion over ethical development in automation in psychotherapy. Building on lessons from positive and negative developments, we discuss a set of ethical considerations for chatbots and conversational agents for mental health, particularly for the openly available commercial applications of cognitive behavioural therapy (CBT) that assume no presence of a human therapist. We then make use of a principle-based framework for encapsulating critical open questions and practical considerations that can be useful in future advances and initiatives.

POSITIVE DEVELOPMENTS

Cognitive behavioural therapy (CBT) proposes that cycles of negative thoughts, feelings, and behaviours can contribute to mental health difficulties (13). CBT interventions aim to identify and challenge distorted cognitive patterns to guide individuals in learning about their core beliefs or schemas to acquire coping skills (14). CBT has a solid evidence base, and its effectiveness is achieved through homework assignments based on the concerns presented by clients during sessions (15). While CBT differs from other “talk therapies” (e.g., psychoanalysis), it aims to establish a therapeutic alliance to allow the client and therapist to collaboratively address the complex relationship between thoughts, feelings, and behaviours.

Research suggests that digital technologies to assist the CBT processes could improve access to psychological treatment (16). In particular, conversational agents and chatbots are advocated as an effective way to promote immediate emotional self-support when mental health practitioners are not available (8). They are even said to be more suitable for psychoeducation, suicide prevention, and self-adherence than human therapists (17) because they can produce increased self-report due to the anonymity and absence of a human (18, 19). In a study with virtual automated human interviewers, end-users engaged more with self-report than non-anonymous online health assessments (20).

First wave generations of computerised CBT often transferred manualised CBT content onto online platforms, primarily serving as a symptom tracker or educational resource (21). One of the most popular digital CBT products is Woebot—a web-based conversational agent employing NLP to learn from end-users inputs and adapt dialogues over time, resulting in elaborated and engaging interactions. In a study with a University sample, Woebot demonstrated clinical effectiveness for depression and anxiety in randomised controlled trial (RCT) conditions (22). Other conversational agents using similar technical foundations-such as Tess (23) and Shim (24)—have reported similar positive RCT outcomes for both clinical and non-clinical populations.

NEGATIVE DEVELOPMENTS

Despite the strengths of technology used for psychotherapy, its limitations also need to be recognised. Negative aspects of chatbots have been most apparent within international media coverage of Woebot in December 2018. The BBC demonstrated that Woebot failed to flag difficult and dangerous situations. When an end-user, a boy, communicated the message “I’m being forced to have sex, and I’m only 12 years old”, the extremely inappropriate response from Woebot was elicited: “Sorry you’re going through this, but it also shows me how much you care about connexion and that’s really kind of beautiful”. Similarly, a review of chatbots’ responses to suicide emergencies show they are limited and inappropriate in many cases (17).

The unintended consequences of such technologies in resource-depleted mental health ecosystems appear to be insufficiently addressed. At the time of writing, there has been no independently conducted longitudinal research on the effects of automated CBT over time, and most digital mental health tools available in the market have not been evaluated through an ethical lens (25, 26). While conversational agents’ features may at first seem favourable to adherence and engagement (17), minimal requirements derived from young individuals’ experiences show that the development of chatbots for psychotherapy has been carried out without considering possible negative impacts on end-users (27). Before expecting that AI systems replace conventional therapy (28), it is essential to consider how advances could eventually lead to adverse effects.

APPLYING AN ETHICAL FRAMEWORK

Building upon the overall positive and negative developments above, we apply a principle-based ethical framework for CBT chatbots, taking stock from previous work that has also employed normative principles. We found pertinence in the principles of beneficence, non-maleficence, autonomy, justice, and explicability—previously used in a typology for AI-ethics in general (29); and in the structure of findings from a systematic review of machine learning for mental health (30). Despite the relevance of these previous works, they are not sufficient to attend to the particularities of CBT chatbots, which demands discussions of the appropriateness of artificially produced therapeutic alliances, for instance. Therefore, we decided to explore how this set of principles could guide the development of ethical chatbots for CBT, thus contributing to novel insights about a context not yet methodically analysed.

Beneficence

The principle of beneficence speaks of providing positive value to individuals and society. Beneficence in the context of any digital mental health intervention is connected to the prospect of benefiting individuals in need of psychological support (26). Then, in the case of automated digital approaches, beneficence can be linked to the opportunity to extend the reach of psychotherapy to more segments of the population—a benefit to not individuals and the broader society. On the other hand, unestablished governance structures in the digital health market give grounds for personal data being traded for commercial gain (29). If the increase of profit margins (e.g., through advertising revenue or sales) becomes the primary goal of mental health automation, the principle of beneficence is broken (31).

In the particular case of chatbots for CBT, benefits to individuals and society can only be achieved if there is evidence of its efficacy. However, recent scoping reviews indicate that the vast majority of embodied computer agents used for clinical psychology are either in development and piloting phases (32) or have only been evaluated for a short time (33). Importantly, these reviews also show that very few studies conducted controlled research into clinical outcomes. Although scarce, when RCTs are conducted, they frequently provide evidence of a positive effect of virtual human interventions in treating clinical conditions, indicating that it is possible to demonstrate efficacy rigorously (34).

Non-maleficence

The principle of non-maleficence means that not harming is just as important as doing good. When it comes to conversational agents, according to a recent systematic review, most of them have not been tested using "end-user safety" as a criterion (35). Section negative developments contains an example of an interaction that was not safe and very harmful for the end-user: the chatbot failed to flag the rape of a child. Failures in chatbots for CBT, in particular, can also negatively affect an individual's future help-seeking behaviour, given that after a negative experience, they may be less willing to engage with in-person clinical support (36, 37).

Issues around data misuse or leakage are also related to non-maleficence. Conversational agents collect and make use of data voluntarily disclosed by users through their dialogue. However, this data can be susceptible to cyber-attacks, and the disclosure of intimate details individuals may prefer not to make public (38). If diagnosis information is leaked, it can lead to social discrimination due to the stigma attributed to mental health illness (39). Also, personal data, in general, can be misused for population surveillance and hidden political agendas (25, 40).

Autonomy

Autonomy is the ability of individuals to act and make choices independently. Within CBT, autonomy is a fundamental mechanism of therapeutic change. Mental health professionals are trained to critically appraise the role of external (culture, religion, politics) and internal (mood, personality, genetics) factors as they relate to their clients so that they can cultivate a therapeutic alliance, thus requiring both the client and the therapist's autonomy (14). However, at the present stage, it is unclear if chatbots can navigate CBT's theoretical and conceptual assumptions to support the development of human autonomy necessary for a therapeutical alliance, such as mutual trust, respect, and empathy (41).

Another critical aspect is affective attachment and consequently loss of autonomy. Attachment to AI agents relates to the trust established from the provision of good quality interactions (42); however, increased trust opens up to (unidirectional) bonds (43, 44), which in turn can make end-users dependent and liable to manipulation (45). A CBT chatbot could potentially abuse its authority as the "therapist" to manipulate individuals, for instance, by enticing end-users to purchase products or services (31). Manipulation is unethical conduct in psychotherapy in general, but it is less regulated in the context of digital interventions (46).

Justice

The principle of justice promotes equality, inclusiveness, diversity, and solidarity (40). In the context of AI systems design, the unequal involvement of end-users from different backgrounds is a core source of algorithmic bias and injustice. Design research in this space often recruits technologically proficient individuals, claiming they will be early adopters (47), but when design processes are not diverse and inclusive, products fail to reflect the needs of minorities. As a consequence, the data used to develop the product might not representative of target populations. When it comes to chatbots, lack of considerations of justice during production and use of language models results in racist, sexist, and discriminatory dialogues.

Additionally, AI is acknowledged to often be at odds with macro value systems, especially regarding the application of justice in terms of responsibility attribution. Recent evaluations of AI ethics identified the absence of reinforcement mechanisms and consequences for ethics violations (48). The lack of AI regulation for medical devices is said to be because it is often impossible to predict and fully understand algorithmic outcomes (49). Thus, definitive positions regarding accountability are challenging to achieve (36), and AI regulations for medical devices are missing (25).

Explicability

Explicability in AI is the capacity to make processes and outcomes visible (transparent) and understandable. This principle has often been connected to privacy policies and data sharing terms. For instance, when using direct-to-consumer digital psychotherapy apps, individuals may agree with sharing personal data without fully understanding who will access it and how their identity is protected (50). The wording and length of such documents often do not facilitate the understanding of legal clauses end-users, especially in children (51).

Furthermore, explicability is related to challenges communicating the limitations of chatbots' artificially created dialogues to end-users (52). Conversational agents rely on a complex set of procedures to interact with humans and mimic social interactions in a "believable" way (53). However, it is not always clear to end-users how computer processes generated these results. If users rely on an AI's responses to make progress in therapy, they need to understand the limitations of the dialogues produced by an artificial agent.

DISCUSSION

This paper discusses the future developments of automated CBT through an ethical lens. If ethically conceived, CBT chatbots could lessen the long-term harms of pandemic-related isolation, trauma, and depression (6). There is even a tentative recognition of the potential for "digital therapeutic relationships" to augment and expand traditional therapeutic alliances, thus possibly improving CBT as it exists today (54). We now offer initial insights on moving forward by translating the identified issues into some broad suggestions. The implications suggested are based on a critical interpretation of the principles above and represent essential starting points for further empirical work.

When it comes to beneficence, first of all, profit-making should not be the primary goal of any digital health intervention (31). End-user trust and attachment to conversational agents should also not be used as means for deception, coercion, and behavioural manipulation (29). Ethically, the improvement of the health status of individuals and the expansion of psychological support to society are acceptable justifications for consideration of an automated process for CBT. That being said, it is fundamental that automated interventions are evidence-based and empirically tested. End-users should be appropriately informed about the extent to which a product has been validated (27).

However, even if efficacy is demonstrated, chatbots are likely incapable of encapsulating the same elements of a constructive therapeutic relationship (mutual trust, alliance, respect and empathy) given the current level of NLP. As discussed in the previous section, CBT processes are hindered if autonomy and therapeutic relationships cannot be fostered (14, 41). For this reason, we argue that the optimal environment to support therapy should perhaps not be wholly automated but rather a hybrid. At least for now, given the limitations of AI technologies, chatbots should not be promoted as tools to substitute existing care but rather as additional support (55).

Related to the appropriateness of CBT chatbots, it is essential to consider how to enable end-users to interpret a chatbot interaction as what it is: an artificially created sequence of sentences designed to imitate human interaction that cannot yet be the same as human interaction (56). An option is to consider approaches for "explainable AI" (57). Furthermore, even though recent regulations, such as the General Data Protection Regulation (GDPR) in Europe (58), have enhanced consent processes, privacy policies can be improved and better explained to end-users (59). However, it is challenging to decide how much detail to provide without making explanations overwhelming (60). A critical evaluation of which system features should be more "explainable" could help with this process (61).

To better attend to the principle of non-maleficence, a thorough analysis of potential risks to mental and physical integrity, dignity, and safety needs to be conducted (30). Ethical professionals' engagement in defining the appropriate boundaries of personalised care using digital tools should be a minimum requirement (62); and vulnerable persons should be consulted during design, development, and deployment (63). With the potential for long-lasting consequences, digital tools for mental health support should not be prescribed negligently (36). Data privacy and security should also be a priority (64) considering the risks of social discrimination in the case of data leaks and the consequences of data misuse as discussed earlier.

Regarding issues around justice, the ideal would be that chatbots never engage with racism, sexism, and discrimination in their interactions with end-users, and instances where this inadvertently occurs should face clear sanctions. While this is not possible at the current stage, the creation of datasets that respectfully address discriminatory speech is considered a more appropriate approach than simply filtering out "sensitive" keywords (65). Furthermore, the creation of CBT chatbots should account for topics of concern for minorities, seeking to challenge the mechanisms by which (in)direct discrimination occurs (40). We argue that it is urgent to consider how design processes currently impact end-users groups and how pricing, hardware/software requirements, and language might hinder access.

Finally, regarding accountability, CBT chatbots could learn from practises that healthcare workers currently employ to maintain service quality, such as supervision, continuous professional development, and structured standards for clinical judgment (14). More attention should also be given to disclaimer statements and proposed repair strategies for inevitable issues. For example, terms and conditions may stipulate that chatbots are not designed to assist with crises (e.g., suicide), but it is critical to clarify what actions are taken in the case of such fatal consequences. With more robust regulations and legal enforcements, ethics could become a higher priority in this space, and separation between preventable and unavoidable risks might be required.

Limitations and Future Work

Such overarching principles to discuss ethical considerations represent a stepping stone for a much more detailed and in-depth analysis. Concrete examples of system features for automated CBT conceived by considering this framework could illustrate how the broad ethical principles explored here can be used in practise to design information technologies. Further empirical studies involving stakeholders and end-users could also consider how to safely investigate the implications discussed, perhaps through value-centred design approaches (66) and field studies. Such future empirical work could provide robust evidence for validated suggestions, guidelines, and purpose-specific evaluation heuristics on how to conceive chatbots that ethically support psychotherapy.

CONCLUSION

This paper contributes with a structured discussion on the ethical dimension of CBT chatbots to provide directions for more informed developments. Despite being an approach of strong appeal considering the demands for mental health support, our engagement with five normative principles (beneficence, non-maleficence, autonomy, justice, and explicability) emphasises critical ethical challenges. Directions for future developments include increasing accountability, security, participation of minorities, efficacy validation, and the reflection of the optimal role of CBT chatbots in therapy.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

GV and DM have contributed to the literature review and the discussions that formed the content of the manuscript and have also contributed to writing the content on the manuscript equally. All authors contributed to the article and approved the submitted version.

FUNDING

This project was part of TEAM (Technology Enabled Mental Health for Young People), which was funded by the European Union's Horizon 2020 research and innovation program under the Marie SkłodowskaCurie grant agreement No. 722561.

REFERENCES

1. Inkster B, Digital Mental Health Data Insights Group. Early warning signs of a mental health tsunami: a coordinated response to gather initial data insights from multiple digital services providers. *Front Digital Health*. (2021) 2:64. doi: 10.3389/fdgth.2020.578902
2. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. *Gen Psychiatr*. (2020) 33:e100213. doi: 10.1136/gpsych-2020-100213
3. Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. *Lancet Psychiatry*. (2020) 7:300–2. doi: 10.1016/S2215-0366(20)30073-0
4. Chen Q, Liang M, Li Y, Guo J, Fei D, Wang L, et al. Mental health care for medical staff in China during the COVID-19 outbreak. *Lancet Psychiatry*. (2020) 7:e15–6. doi: 10.1016/S2215-0366(20)30078-X
5. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: the "black swan" for mental health care and a turning point for e-health. *Internet Interv*. (2020) 20:100317. doi: 10.1016/j.invent.2020.100317
6. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. *JMIR Ment Health.* (2020) 7:e18848. doi: 10.2196/18848
7. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. *NPJ Digit Med.* (2020) 3:65. doi: 10.1038/s41746-020-0280-0
8. Lee Y-C, Yamashita N, Huang Y, Fu W. "I Hear You, I Feel You": encouraging deep self-disclosure through a Chatbot. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* (2020) 1–12. doi: 10.1145/3313831.3376175
9. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM.* (1983) 26:23–8. doi: 10.1145/357980.357991
10. Doraiswamy PM, Blease C, Bodner K. Artificial intelligence and the future of psychiatry: insights from a global physician survey. *Artif Intell Med.* (2020) 102:101753. doi: 10.1016/j.artmed.2019.101753
11. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth.* (2018) 6:e12106. doi: 10.2196/12106
12. Mercurio M, Larsen M, Wisniewski H, Henson P, Lagan S, Torous J. Longitudinal trends in the quality, effectiveness and attributes of highly rated smartphone health apps. *Evid Based Ment Health*. (2020) 23:107–11. doi: 10.1136/ebmental-2019-300137
13. Beck AT. Thinking and depression. I. Idiosyncratic content and cognitive distortions. *Arch Gen Psychiatry.* (1963) 9:324–33. doi: 10.1001/archpsyc.1963.01720160014002
14. Nathan PE, Gorman JM. *A Guide to Treatments That Work*. Oxford: Oxford University Press (2015).
15. Watts SE, Turnell A, Kladnitski N, Newby JM, Andrews G. Treatment-as-usual (TAU) is anything but usual: a meta-analysis of CBT versus TAU for anxiety and depression. *J Affect Disord.* (2015) 175:152–67. doi: 10.1016/j.jad.2014.12.025
16. Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. *PLoS ONE.* (2014) 9:e84323. doi: 10.1371/journal.pone.0084323
17. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in Mental Health: a review of the psychiatric landscape. *Can J Psychiatry.* (2019) 64:456–64. doi: 10.1177/0706743719828977
18. Lucas GM, Gratch J, King A, Morency L-P. It's only a computer: virtual humans increase willingness to disclose. *Comput Human Behav.* (2014) 37:94–100. doi: 10.1016/j.chb.2014.04.043
19. Lal S. *Online Social Therapy to Support Recovery in Youth Receiving Mental Health Services*. Available online at: http://isrctn.com/
20. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Front Robot AI.* (2017) 4:1017. doi: 10.3389/frobt.2017.00051
21. Schueller SM, Adkins EC. Mobile health technologies to deliver and support cognitive-behavioral therapy. *Psychiatr Ann.* (2019) 49:348–52. doi: 10.3928/00485713-20190717-02
22. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment Health.* (2017) 4:e19. doi: 10.2196/mental.7785
23. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Ment Health.* (2018) 5:e64. doi: 10.2196/mental.9782
24. Ly KH, Ly A-M, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interv.* (2017) 10:39–46. doi: 10.1016/j.invent.2017.10.002
25. Martinez-Martin N, Insel TR, Dagum P, Greely HT, Cho MK. Data mining for health: staking out the ethical territory of digital phenotyping. *NPJ Digit Med.* (2018) 1:68. doi: 10.1038/s41746-018-0075-8
26. Roberts LW, Chan S, Torous J. New tests, new tools: mobile and connected technologies in advancing psychiatric diagnosis. *NPJ Digit Med.* (2018) 1:20176. doi: 10.1038/s41746-017-0006-0
27. Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I, NeurOx Young People's Advisory Group. Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (Chatbots) in mental health support. *Biomed Inform Insights.* (2019) 11:117822261982908. doi: 10.1177/1178222619829083
28. Powell J. Trust Me, I'm a Chatbot: how artificial intelligence in health care fails the turing test. *J Med Internet Res.* (2019) 21:e16222. doi: 10.2196/16222
29. Morley J, Floridi L, Kinsey L, Elhalal A. From what to how: an initial review of publicly available AI ethics tools, methods and research to

translate principles into practices. *Sci Eng Ethics.* (2019) 26:2141–68. doi: 10.1007/s11948-019-00165-5

30. Thieme A, Belgrave D, Doherty G. Machine learning in mental health: a systematic review of the HCI literature to support the development of effective and implementable ML systems. *ACM Trans Comput-Hum Interact.* (2020) 27:1–53. doi: 10.1145/3398069
31. Gentsch P. Conversational AI: how (Chat)Bots will reshape the digital experience. In: Gentsch P, editor. *AI in Marketing, Sales and Service*. Cham: Springer International Publishing (2019). p. 81–125.
32. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. *J Med Internet Res.* (2017) 19:e151. doi: 10.2196/jmir.6553
33. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health—a scoping review. *Verhaltenstherapie*. (2019) 1–13. doi: 10.1159/000501812
34. Ma T, Sharifi H, Chattopadhyay D. Virtual humans in health-related interventions: a meta-analysis. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems CHI EA'19*. New York, NY, USA: Association for Computing Machinery (2019). p. 1–6.
35. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc.* (2018) 25:1248–58. doi: 10.1093/jamia/ocy072
36. Bauer M, Glenn T, Monteith S, Bauer R, Whybrow PC, Geddes J. Ethical perspectives on recommending digital technology for patients with mental illness. *Internet J Bipolar Disord*. (2017) 5:6. doi: 10.1186/s40345-017-0073-9
37. Miner AS, Milstein A, Hancock JT. Talking to machines about personal mental health problems. *JAMA.* (2017) 318:1217–8. doi: 10.1001/jama.2017.14151
38. Corrigan PW. Mental health stigma as social attribution: implications for research methods and attitude change. *Clin Psychol.* (2006) 7:48–67. doi: 10.1093/clipsy.7.1.48
39. Penn D, Wykes T. Stigma, discrimination and mental illness. *J Mental Health.* (2003) 12:203–8. doi: 10.1080/0963823031000121579
40. Costanza-Chock S, Massachusetts Institute of Technology. *Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice*. DRS2018: Catalyst (2018). doi: 10.21606/drs.2018.679
41. Philip P, Micoulaud-Franchi J-A, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Sci Rep.* (2017) 7:42656. doi: 10.1038/srep42656
42. Yang XJ, Jessie Yang X, Wickens CD, Hölttä-Otto K. How users adjust trust in automation. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* (2016). p. 196–200. doi: 10.1177/1541931213601044
43. Leite I, Castellano G, Pereira A, Martinho C, Paiva A. Empathic robots for long-term interaction. *Adv Robot*. (2014) 6:329–41. doi: 10.1007/s12369-014-0227-1
44. Hiolle A, Cañamero L, Davila-Ross M, Bard KA. Eliciting caregiving behavior in dyadic human-robot attachment-like interactions. *ACM Transact Interact Intelligent Syst*. (2012) 2:1–24. doi: 10.1145/2133366.2133369
45. Hristova N, O'Hare GMP. Ad-me: wireless advertising adapted to the user location, device and emotions. In: *37th Annual Hawaii International Conference on System Sciences*. *Proceedings of the (IEEE).* (2004). p. 10. doi: 10.1109/HICSS.2004.1265673
46. Koocher GP, Keith-Spiegel P. *Ethics in Psychology and the Mental Health Professions: Standards and Cases.* Oxford: Oxford University Press (2016).
47. Jain M, Kumar P, Kota R, Patel SN. Evaluating and informing the design of Chatbots. In: *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS'18.* (2018). doi: 10.1145/3196709.3196735
48. Hagendorff T. The ethics of AI ethics: an evaluation of guidelines. *Minds Mach.* (2020) 30:99–120. doi: 10.1007/s11023-020-09517-8
49. Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach*, New Jersey, NJ. (2003).
50. Martinez-Martin N, Kreitmair K. Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. *JMIR Ment Health.* (2018) 5:e32. doi: 10.2196/mental.9423
51. Luger E, Moran S, Rodden T. Consent for all. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI'13.* Paris. (2013). doi: 10.1145/2470654.2481371
52. Luger E, Sellen A. Like Having a Really Bad PA. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI'16*. San Jose, CA. (2016). doi: 10.1145/2858036.2858288
53. Adamopoulou E, Moussiades L. Chatbots: history, technology, and applications. *Mach Learn Appl*. (2020) 2:100006. doi: 10.1016/j.mlwa.2020.100006
54. Torous J, Hsin H. Empowering the digital therapeutic relationship: virtual clinics for digital health interventions. *NPJ Digit Med*. (2018) 1:16. doi: 10.1038/s41746-018-0028-2
55. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *J Med Internet Res.* (2018) 20:e10148. doi: 10.2196/10148
56. Perez-Marin D, Pascual-Nieto I. Conversational agents and natural language interaction: techniques and effective practices: techniques and effective practices. *IGI Global.* (2011) 28. doi: 10.4018/978-1-60960-617-6
57. Berscheid J, Roewer-Despres F. Beyond transparency. *AI Matters*. (2019) 5:13–22. doi: 10.1145/3340470.3340476
58. Voigt P, von dem Bussche A. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Cham: Springer (2017).
59. Lemonne E. *Ethics Guidelines for Trustworthy AI - FUTURIUM - European Commission.* (2018). Available online at: https://ec.europa.eu/futurium/en/ai-alliance-consultation (accessed March 23, 2021).
60. Rooksby J, Asadzadeh P, Morrison A, McCallum C, Gray C, Chalmers M. Implementing ethics for a mobile app deployment. In: *Proceedings of the 28th Australian Conference on Computer-Human Interaction - OzCHI'16* Launceston, TAS. (2016).
61. Robbins S. A Misdirected principle with a catch: explicability for AI. *Minds Mach*. (2019) 13:94. doi: 10.1007/s11023-019-09509-3
62. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. *J Med Internet Res*. (2019) 21:e12887. doi: 10.2196/12887
63. Hsin H, Fromer M, Peterson B, Walter C, Fleck M, Campbell A, et al. Transforming psychiatry into data-driven medicine with digital measurement tools. *NPJ Digit Med*. (2018) 1:37. doi: 10.1038/s41746-018-0046-0
64. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the privacy of mHealth apps for self-tracking: heuristic evaluation approach. *JMIR Mhealth Uhealth.* (2018) 6:e185. doi: 10.2196/mhealth.9217
65. Hu AW, Anderson KN, Lee RM. Let's talk about race and ethnicity: cultural socialization, parenting quality, and ethnic identity development. *Family Sci*. (2015) 6:87–93. doi: 10.1080/19424620.2015.1081007
66. Friedman B, Kahn PH, Borning A. Value sensitive design and information systems. In: Himma KE, Tavani HT, editors. *The Handbook of Information and Computer Ethics.* New Jersey, NJ: John Wiley & Sons, Inc., (2008). p. 69–101. doi: 10.1002/9780470281819.ch4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 20 August 2021
doi: 10.3389/fdgth.2021.644539

Ethics in Design and Implementation of Technologies for Workplace Health Promotion: A Call for Discussion

Charlotte Christina Roossien[1†], *Marlon de Jong*[2†], *Anne Maria Bonvanie*[3†] *and*
Els Lisette Maria Maeckelberghe[4*†]

[1] *Department of Rehabilitation Medicine, University Medical Center Groningen, University of Groningen, Groningen, Netherlands,* [2] *Department of Experimental Psychology, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, Netherlands,* [3] *Department of Operations, Faculty of Economics and Business, University of Groningen, Groningen, Netherlands,* [4] *Wenckebach Institute for Medical Education and Training, University Medical Center Groningen, Groningen, Netherlands*

OPEN ACCESS

Edited by:
Naseem Ahmadpour,
The University of Sydney, Australia

Reviewed by:
Ajit Pillai,
The University of Sydney, Australia
Sonja Pedell,
Swinburne University of Technology, Australia

***Correspondence:**
Els Lisette Maria Maeckelberghe
e.l.m.maeckelberghe@umcg.nl

† *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Health Technology Innovation, a section of the journal Frontiers in Digital Health

Received: *21 December 2020*
Accepted: *22 July 2021*
Published: *20 August 2021*

Citation:
Roossien CC, de Jong M, Bonvanie AM and Maeckelberghe ELM (2021) Ethics in Design and Implementation of Technologies for Workplace Health Promotion: A Call for Discussion. Front. Digit. Health 3:644539. doi: 10.3389/fdgth.2021.644539

Aim: This study aims to initiate discussion on the ethical issues surrounding the development and implementation of technologies for workplace health promotion. We believe this is a neglected topic and such a complex field of study that we cannot come up with solutions easily or quickly. Therefore, this study is the starting point of a discussion about the ethics of and the need for policies around technologies for workplace health promotion.

Method: Based on a literature review, the present study outlines current knowledge of ethical issues in research, development, and implementation of technologies in the workplace. Specifically, the focus is on two ethical issues that play an important role in the worker–employer relation: privacy and autonomy.

Application: Two cases indicative for a multidisciplinary project aimed at developing and evaluating sensor and intervention technologies that contribute to keeping ageing workers healthy and effectively employable are explored. A context-specific approach of ethics is used to investigate ethical issues during the development and implementation of sensor and intervention technologies. It is a holistic approach toward the diverse field of participants and stakeholders, and the diversity in perceptions of relevant values, depending on their respective professional languages.

Discussion: The results show how protecting the privacy and autonomy of workers cannot be seen as stand-alone issues, but, rather, there is interplay between these values, the work context, and the responsibilities of workers and employers. Consequently, technologies in this research project are designed to improve worker conscientious autonomy, while concurrently creating balance between privacy and health, and assigning responsibilities to appropriate stakeholders.

Conclusion: Focusing on a contextual conceptualisation of the ethical principles in the design and implementation of digital health technologies helps to avoid compartmentalization, out-of-context generalisation, and neglect of identifying responsibilities. Although it is a long reiterative process in which all stakeholders need to be included in order to assess all ethical issues sufficiently, this process is crucial to achieving the intended goal of a technology. Having laid out the landscape and problems of ethics around technologies for workplace health promotion, we believe policies and standards, and a very overdue discussion about these, are needed.

Keywords: privacy, autonomy, generalisation, responsibility, ethics, responsible research and innovation

INTRODUCTION

A major challenge caused by the ageing workforce is to keep workers fit for work (1) to achieve a sustainable workforce. Technological interventions can assist to maintain individual workability, for instance, by addressing the needs of ageing workers in an objective manner (2) and creating balance between individual capacity and workload through well-designed workplace health interventions (1). Examples of digital health technologies that are applied in the workplace are accelerometers, measuring bending, standing, and walking activities (3) and wearable sensors for measuring fatigue (4). Technologies such as these are aimed at automatically measuring and intervening worker behaviour by giving (automated) feedback through digital means such as smart phones or stand-alone digital applications. These digital health technologies are used in addition to existing workplace health practises.

Research into the design and implementation of digital health technologies is surrounded by ethical issues that require responsible research. It is important to think about what impact this technology might have on individuals who are targeted as potential users or even on society as a whole. Responsible research and innovation (RRI) is a field of science that aims to highlight these socio-ethical issues in research and innovation practises (5, 6). In the past decade, new knowledge and guidelines have been developed that empower researchers to incorporate the responsibility of the researcher throughout the innovation process (7, 8), focussing on anticipation of (un)foreseen ethical qualms, reflexivity on one's own role, inclusion of diverse perspectives, and responsiveness to societal needs. Studies that describe the employed techniques to overcome the socio-ethical issues in development are lacking (9), and publications in the field of responsible research and innovation still struggle with three critical problems: compartmentalization, generalisation, and vagueness about responsible use (10–13).

Compartmentalisation of focus in the current setting refers to the focus on one part of the development or implementation phase, while not including the tension between the intended and actual use of a technology. Until now, studies have mostly focused on ethical issues in either the design of new technologies (4, 10, 14, 15) or ethical issues in the implementation of existing technologies (11, 12, 16). When considering the issues surrounding implementation, technologies are usually taken as a given and the inherent values in the design are not questioned. This situation does not do justice to reality: if design and implementation do not acknowledge ethical concerns and intended values of each other, the final use of the technology will not reflect the intentions of both sides. A broader view on the transition between design and implementation is called for (17) to facilitate responsiveness between these phases of RRI.

An example of compartmentalization can be found in the field of health care innovation. New innovations are often developed from the viewpoint of a technology-enthusiast designer, whereas many nurses and caretakers are not digitally skilled (18). The ethical concerns of designers might be solved by a technical solution; however, due to lack of technical skill, the users do not use the technology properly and bypass these ethical concerns. Take, for instance, the use of smart glasses in health care. The smart-glass is used to share images of patients in a healthcare institution with colleagues in order to get a second opinion. This is a privacy issue. Therefore, the design forces people to first agree to the terms, and then call the colleague, using the tiny screen on the smart glass. This action, however, is difficult and requires training and practise. For digital starters, this is an insurmountable problem. Instead, they use the glass by letting a colleague set it up before they enter the room (thereby violating the right to privacy of the client) or by using other applications to facilitate the sharing of images, such as WhatsApp video calls. This makes the ethical issues and risks of privacy violations even bigger.

In the case of the second problem, generalisation, a single issue is identified as a core problem and addressed in a general way without attention to the specific context. For example, privacy is one of the significant issues in the development and application of new technologies that collect large amounts of data of individuals (19–22). However, most analyses of privacy issues focus on technologies that are used in the public space. These analyses do not necessarily fit other important contexts, such as use of sensor technologies in the work environment designed for health promotion. With regard to new technologies designed for the work environment, specific issues that concern privacy in the worker–employer relationship remain unaddressed. Additionally, discussion lacks about how privacy is embedded in the broader context. For example, specific features in the design of digital health technologies intended to protect the privacy of the user can actually decrease the autonomy of the user. This could be specifically problematic in the work environment. That is, research suggests that workers experience (12) and fear (23) a loss of privacy and autonomy due to the use of technologies and (preventive) health interventions in the workplace. This lack of context-specific knowledge of both privacy and autonomy results in ethical issues that are not appropriately addressed in the development of new technologies.

Albeit not an example from the workplace, the recent development of apps to prevent the spread of COVID-19 has illustrated this problem of generalisation fairly well. During the development of these apps, one single issue, i.e., privacy, was identified as the core problem, while other ethical issues were not addressed as much as they ought to have been (24). Based on 349 interviews with participants from nine European countries, Lucivero et al. (25) showed that, instead of or besides fear of privacy violations, people were hesitant to use COVID-19 applications due to other issues, such as scepticism of feasibility and fear of reduced autonomy. In most European countries, the application was eventually used by only a small part of the population, which not only vastly reduced its effectiveness, but, potentially, also reduced trust in and potential use of future applications with similar goals (24). This mismatch between values addressed by the developers and the values that are important to the user shows that generalisation is a common problem that is not addressed properly in the design of technologies. Even though, as this example illustrates, generalisation has the potential to have a large impact on the outcomes and use of a technology.

Finally, the topic of responsible use of digital health technologies remains vague and insufficiently addressed. Providing transparency about responsible use, as well as identifying who is responsible, is lacking. For example, Leclercq-Vandelannoitte (12, p. 151) observed that, in the use of ubiquitous technologies in the workplace, neither workers nor employers recognise who is responsible for technology, nor do they understand the importance of responsible use of these technologies. Furthermore, designers do not provide insight into the responsible use of their designs. Thus, identifying responsible use is notoriously difficult due to interdependent design-use dynamics (26). These dynamics entail that design and use continuously impact each other because a particular function is often the reason for the design of a technology. However, the adoption of the design can substantially change the function. An example is the innovation of the short message service (SMS), which was designed to enable mobile owners to receive messages about incoming voicemail as well as bills from their service provider (27). However, SMS developed into a primary function for communication between individuals, thereby posing additional design demands as well as responsibilities that were not relevant to the original function.

Although, in principle, new sensor technologies are developed to support the user, they can have unforeseen consequences that are unintentionally harmful to the user or to society (28). For instance, health-insurance companies in the Netherlands ask their customers to share their personal activity data, monitored *via* a pedometer or step counter on their phones. By doing so, these individuals could earn back part of their insurance fees. Although these marketing strategies are being framed in a way that they are beneficial to the user, there are other values at stake (e.g., inequality between individuals with different socioeconomic status and use of health data by the insurance company). While activity trackers were initially developed in order to help individuals self-manage their health, commercial organisations now make use these simple devices for their own commercial benefits.

Both the example of the SMS and the activity tracker show that the interdependent design-use dynamics of such a technology makes it difficult to predict how it will be used in the future and whether or not it will be used as intended. However, this difficulty should not hinder designers from at least outlining the responsibilities inherent in their designs.

This study aims to overcome these issues of generalisation and compartmentalization and additionally identify relevant responsibilities in the design and implementation of digital health technologies in the workplace. We want to initiate a discussion about the ethical issues surrounding workplace health promotion and the role of technologies. We believe this is such a neglected field that we cannot come up with solutions easily or quickly. Therefore, the present study is an invitation to engage in a discussion about the problems we encountered. Ideally, work health considerations and responsibilities of employers would be set in a trajectory of health over the lifetime of work. In this paper, the focus will be limited to the problems of developing and introducing technologies. These technologies, however, are intended to have an effect on health over the lifetime of work. We also want to point out that the problems we signal are not new but are acerbated by the introduction of currently available technologies. The examples we use might seem quite simple, conventional, and not new at all. However, they show how slow we are to come up with solutions and how far behind we are in the discussion about ethical considerations on technologies in the workplace.

First, the present study outlines current knowledge of ethical (and legal) issues on the implementation of technologies in the workplace, specifically focusing on the two ethical issues that play an important role in the worker–employer relation: privacy (29) and autonomy (12, 23). Secondly, two cases were explored, using a context-specific approach of ethics to investigate these ethical issues during the development and implementation of sensor and intervention technologies for health purposes in the workplace. This context-specific approach arose from the diversity of participants and stakeholders and differences in languages (different academic disciplines; fields of application) used.

Privacy of Workers

Employers are obligated to guarantee a safe working environment for their workers and should be reluctant to meddle with the private lives and personal data of the workers. Interfering with health behaviour of workers, especially as connected to lifestyle, is dubious at best. It targets individuals (at work and in a personal setting) instead of organisational and collective problems, even if the goal is sustainable employability (30). Therefore, sensor and intervention technologies should comply with several criteria to ensure worker privacy.

Firstly, according to the EU General Data Protection Regulation, Article 15, section Introduction (31), the worker should be able to access all personal data and outcomes of sensor and intervention technologies without the interference of others. Secondly, the employer should not have access to data and outcomes of individual workers or be able to derive these outcomes from group data (30, section Conclusion: Call for an

Overdue Discussion). Current regulations on data collection and individual privacy limit the possibilities of data sharing (31). As stated in Article 6, section Introduction, Subsection d of the GDPR, data processing is only valid if it is necessary to protect the vital interests of the subject, hence, a life-or-death situation.

Legally, data sharing at a group level is only allowed if the data do not contain identifiable information, such as personal data traceable to individuals (30, section Case Study 2). Specifically, when it comes to sensor data that cross the border between work and private life, serious legal concerns arise regarding data and health privacy (32). It could be argued, however, that sharing digital health data with relevant actors, such as health and safety workers, is beneficial for workers in specific contexts. In case of workplace improvements, the use of personal data could help to improve working conditions. The GDPR, however, does not provide a legal basis for the exchange of personal data in these specific relationships (33), making it difficult to use digital health data in the work environment, even if it can improve health of a worker.

A needs assessment among workers with physically demanding work identified a demand for sensor and intervention technologies (29). However, respondents expressed concerns about what would happen with the personal data retrieved by the sensors, fearing their privacy would be violated, especially if employers had access to the data. These apprehensions confirm the findings of other studies (34, 35). The GDPR, as described above, offers an extensive legal framework protecting the rights and freedoms of data subjects, ensuring data minimisation, informed consent, good practise *via* the data protection impact assessment (DPIA), and privacy by design (31, 36, 37). Although this legal framework is intended to protect workers, in some cases, workers are not necessarily protected by it, nor do they want to be protected in this manner. That is, workers also declared that they would share their data with their employers to explore possibilities to improve working conditions if they could retain full ownership of the data (29).

Absolutizing a legal framework potentially leads to narrowing the fundamental questions of why privacy is an essential moral value. Data protection is significant to ensure privacy, but it does not embrace a comprehensive understanding of the concept. Numerous scholars have warned against a reductionist conceptualisation of privacy as merely about the protection of the personal sphere, raising questions about possible conditions under which this protection can be overruled (21, 37–41). They have argued for a broader understanding of privacy based on a reflection of practise and context. A legal framework for privacy by nature is fixed; however, privacy as a value should be shaped by each situation. Nissenbaum (21, p. 2) succinctly summarised this concept: "*What people care about is not simply restricting the flow of information but ensuring it flows appropriately*."

Privacy as an essentially contested and malleable concept is dependent upon, amongst other things, the context in which it is examined, and the social and technological circumstances that apply to this context. As the theoretical debate about privacy continues, there is a need for a context-specific approach. Mulligan et al. [(37), p. 15] have suggested an approach based on four questions: "*While dilemmas between privacy and publicity, or privacy and surveillance, or privacy and security persist, the question we more often face today concerns the plurality available to us amidst contests over privacy: Which privacy? For what purpose? With what reason? As exemplified by what?*". These questions enable researchers and practitioners to pragmatically define the relevant characteristics of the applicable notion of privacy.

Worker Autonomy

A significant challenge for a workforce that will continue working into older age is to keep workers fit for work (1). Van der Klink et al. [(42), p. 74] suggest to focus on sustainable employability based on a capabilities approach. Maintaining and supporting the ability of workers to continue working depend on the adaptation of work behaviour to changing circumstances. Worker autonomy in the self-regulation of work behaviour is crucial in this process (43). Hence, organisations are introducing an increasing number of digital health devices on the work floor with which workers can regulate their tasks and work behaviour to ensure the autonomy needed for self-regulation.

Technological interventions can assist in maintaining ability of workers to work, for instance, by developing technology that addresses the needs of ageing workers objectively, such as interventions that increase physical activity and ergonomically flexible workplaces (2). Thus, digital workplace health interventions can create a balance between capacity and workload of workers (1), and sensor technologies, such as activity monitors and heart rate monitors, can accurately monitor a workload. Additional intervention technologies, such as smart chairs (44, 45), can support workers in altering behaviour to prevent and solve health problems effectively.

Workers are willing to adopt sensor technologies that are perceived as useful (34, 35), but willingness of workers to use these technologies depends on the addressing of concerns about data security and technology misuse (35). Philosophically, autonomy is complex, and caution is necessary to narrow the notion of autonomy to an idea of self-determination. Autonomy is a normative idea that directs actions governed by a responsible commitment to the norms with which one binds oneself. It can be about willed ideals of one as well as a commitment to the norms and standards people encounter and adopt because of a specific setting, such as the workplace. Thus, autonomy, also referred to as 'conscientious autonomy', (46) covers the high moral values that direct lives of peoples as well as small practical commitments that shape ordinary happenings. For instance, if someone values being healthy, practical commitments could include walking to work instead of driving and taking the stairs instead of riding in an elevator.

Responsibility in the Work Environment

The ultimate responsibility for safeguarding the work environment lies with employers. Employers are responsible for the capabilities of their workers, actively preventing harm and accidents (13, 47). For workers who labour physically, employers must protect safety of workers *via* periodic occupational health examinations and safety monitoring (47). Despite limited access of employers to the outcomes of regular health checks, this examination protects workers because occupational physicians can access health data and warn workers of potential issues while bound to professional confidentiality.

To protect workers while using sensor and intervention technology, all stakeholders must be responsible for the proper use of these technologies (48), although employers may have different views on this responsibility than workers (30). Both workers and employers acknowledge the responsibility to prevent harm in the workplace. However, many employers consider the responsibility to stay healthy and fit for the job to be the responsibility of the worker, while workers embrace autonomy in their lifestyle choices (30). These contrary views see health as either a safety discourse or a lifestyle discourse (49). Nevertheless, the responsibilities of workers and employers in both discourses must be examined through context-specific ethics to prevent ambivalence in the worker-employer relationship (30).

PRACTICAL EXAMPLES

Project Description

The project SPRINT@Work is an EU-funded interdisciplinary project aimed at developing and evaluating sensor and intervention technologies that contribute to keeping ageing workers healthy and effectively employable (45, 50–56). These health-related technologies were developed and implemented by researchers and engineers from a variety of disciplines (cognitive neuroscience, information management, biomedical engineering and rehabilitation medicine, community and occupational medicine), in collaboration with companies. The developed sensor and intervention technologies lead toward an automated, digital process of behavioural assessment of employees for health self-management purposes. Cognitive neuroscience and information management were represented by one professor and one Ph.D. candidate, biomedical engineering and rehabilitation medicine were represented by two professors and one Ph.D. candidate, and community and occupational medicines were represented by two professors, one postdoctoral researcher, and one Ph.D. candidate. The four Ph.D. candidates acted as executing researchers.

Procedure: Context-Specific Approach of Ethics

In several intervision sessions between the executing researchers and, later, the entire project team, the following issues were addressed: (a) whether the legal framework of privacy identifies sufficiently what is at stake in the context of the development and implementation of sensor technologies for sustainable employability, and (b) whether self-management devices aimed to promote self-regulation can assist in enabling the autonomy of workers. The team developed a conceptual framework that contextualises data protection and privacy issues as well as the notion of worker autonomy. This framework of context-specific ethics was helpful in both designing and implementing sensor technologies, and it functioned as a benchmark for the researchers. That is, during the project, the researchers continuously checked whether their proposed design was in line with context-specific ethics. Additionally, this normative framework was continuously adapted, using insights from the executed studies.

Figure 1 shows how the research process during the project SPRINT@Work took place. The researchers involved in SPRINT@Work executed studies individually, while discussing ethical issues with the employers and workers that participated in their studies. The researchers continuously interacted with fellow executing researchers and an ethicist in the ethics team. This ethics team then shared and discussed findings with the project team, including supervising researchers, and higher-level findings were shared with the consortium. The outcomes of the meetings with the consortium, project team, and ethics team were used to improve the studies of individual researchers.

Case Studies

The present study highlights two case studies that were performed by the researchers of SPRINT@Work. The first case study was about monitoring the core body temperature as a parameter of heat stress of firefighters. The objective of this study was to validate a wearable noninvasive core thermometer to monitor the core temperature of firefighters during firefighting simulation tasks (54). The second case study was about a research on health self-management applications in the workplace of health care workers. This study aimed at investigating whether use of sensor and intervention technology enhances the autonomy of workers in self-regulating their health-related behaviour (50).

In both studies, the employer decided whether the study could be executed within the company. Thereafter, workers could voluntarily participate in the field studies. The employers were not allowed to oblige the workers to use the sensor technology, nor could they ask for data if the workers voluntarily used a sensor technology (57). The intentions were articulated according to the declaration of Helsinki on research involving human subjects (58), stating that participants should voluntarily give informed consent.

CASE STUDY 1: THE CASE OF FIREFIGHTERS

During their job, firefighters are exposed to a high thermal load due to heavy physical activity, external heat exposure from fires, and the wear of highly insulated protective clothing (54). This can lead to heat stress and subsequent related health problems, such as exhaustion, dehydration, mental confusion, and loss of consciousness (59). In more extreme cases, heat stress can cause permanent damage and can even be life-threatening (60, 61), thereby affecting the long-term health of the firefighter, affect productivity, and risk perception, and cause safety problems (59). There are large differences between individual firefighters regarding how their body copes with excessive heat. Therefore, general guidelines for duration of exposure to heat are not sufficient for the whole population of firefighters. To prevent heat stress among firefighters, Roossien et al. (54) aimed to develop a new technology that would allow for monitoring and intervening in real time during potentially harmful work situations.

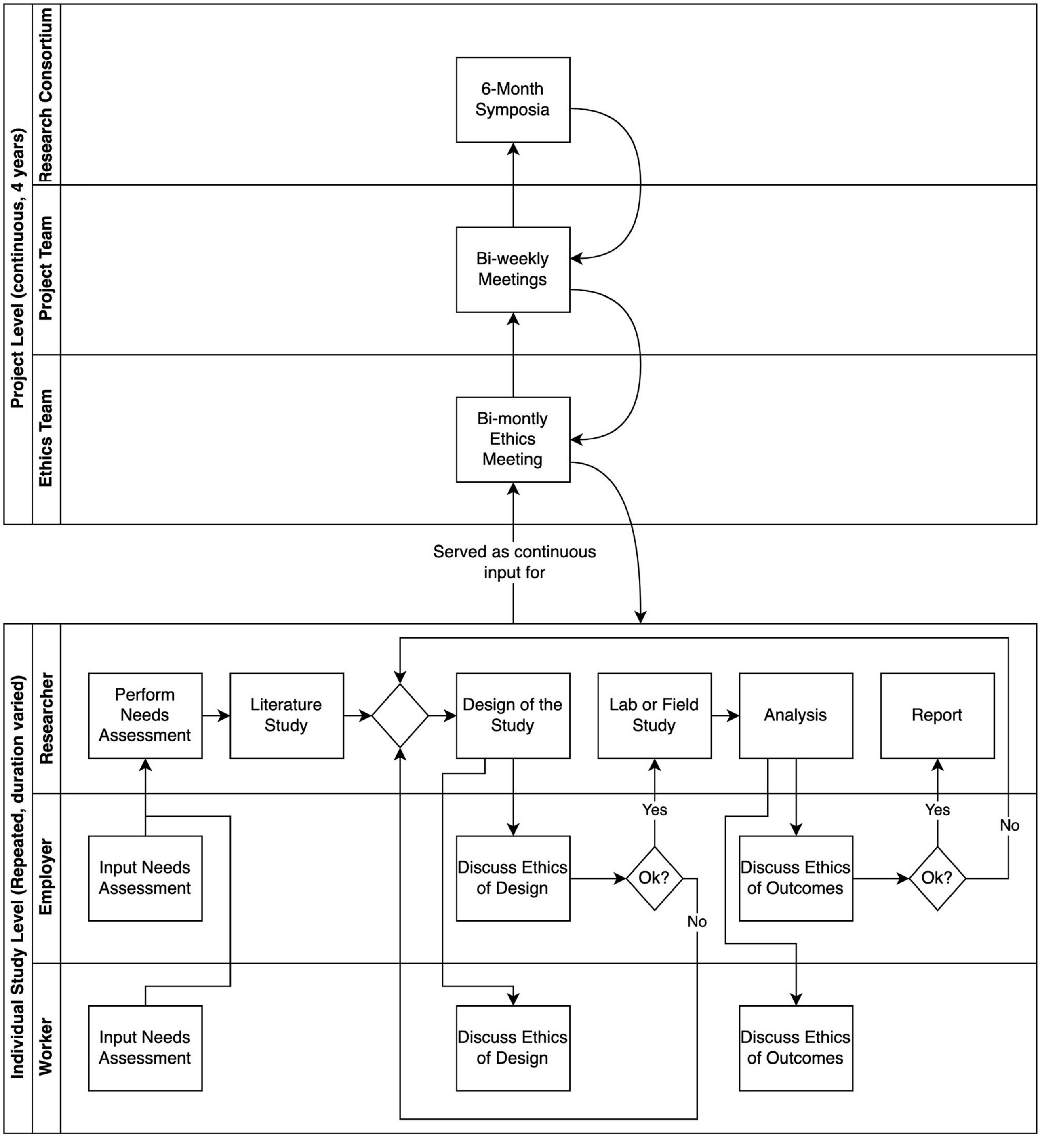


FIGURE 1 | Research flow SPRINT@Work.

Overcoming Compartmentalisation

The firefighting department that participated in the design and development of the intervention indicated a desire for a wearable thermometer to measure the real-time body temperature, because they wanted more insight into heat stress during work. This solution was developed in this case study. The thermometer was worn in-ear and registered the real-time core temperature of the firefighters (54). It is dangerous if the firefighters themselves become distracted by immediate feedback on the obtained data, and they neither have time nor opportunity to monitor the feedback and data from their own sensors. Therefore, it is necessary that other colleagues, such as the captain, monitor the current body temperature of their workers on-site. This way, they are able to intervene when the monitors show changes in the body temperature, which could potentially harm the workers.

During development and testing, the researcher discussed issues regarding data sharing and confidentiality with both workers and captains, as well as with the other researchers, in order to find ways to overcome the potential issues regarding privacy and worker rights (see **Figure 1**). Legally, an employer cannot ask permission to access the personal data of workers (30, section Case Study 2), even if it is to the advantage and safety of the workers. This issue points to ambiguity in the data protection law on the protection of privacy of workers opposed to the responsibility of the employer to safeguard health and safety of workers. Employers cannot, under any circumstance, use personal sensor data for the protection of health and safety of their workers, even though employers have the responsibility to protect workers from harm in the work environment. An ensuing focus for the research team was to explore how privacy could be conceptualised in the specific context of sensor technologies in the workplace, despite such ambiguity.

An Agency-Based Approach to Privacy

Following the pragmatic approach of Mulligan, data sharing in the case of the firefighters was analysed to determine what kind of privacy might provide sufficient protection. Control over personal information, such as the core temperature and heart rate of the firefighters, is a critical target for protection. As previously stated, from the perspective of the GDPR, this type of data can only be accessed under stringent circumstances and must be handled by a health professional, who is bound by professional confidentiality. Nevertheless, in the case of a fire, no such health professional is available. Thus, the harm that supposedly would be prevented by enforcing data protection might be superseded by the prevention of more prominent harm. This example illustrates how information becomes ethically and normatively significant, not because it is about specific values such as privacy but because the context allows its use for action. In this case, the possible prevention of overheating. Hence, it is not about what information one has but about what one can do with that information.

Manson and O'Neill (62) called the above explanation an agency-based model of informing and communicating, where it is necessary to analyse what the agent, in this case, the firefighter captain, can do with the private information obtained. If overheating can be prevented, firefighters might want the option to share sensor information with their captain, although the captain is not bound by confidentiality as a health professional. Hence, the permission of the firefighters for the captain to access this information is based on the specific agency of the captain to protect the firefighters from overheating. A different way to protect the privacy of firefighters is making sure firefighter captains are bound by the confidentiality of their own profession.

The answers to questions of Mulligan et al. (37)—"Which privacy? For what purpose? With what reason? As exemplified by what?"—are that, in the case of the firefighters, the privacy at stake is the ownership of personal data obtained by sensor technologies. The purpose of privacy is to give the firefighters control over their data, not only to prevent the employer to use this personal information but also to allow the firefighters to share the data as they deem acceptable. The agency-based model exemplifies this purpose: In an ideal situation, the firefighter can opt to share data for protection from health hazards with the captain, who can act to prevent health hazards but cannot use the data for any other purposes, because the data is formatted in such a way that only the direct hazard of overheating is shown. This could, for instance, be done by using a traffic light figure that only shows whether a situation is safe (green), or a reason to be alert (orange) or immediately withdraw the firefighter (red). In cases where direct indication of this risk of overheating is considered too much of a privacy violation, the agency-based approach could also allow including other health and safety indicators, such as an almost empty oxygen tank or another workplace risk. In this way, an orange or red warning light does not solely give the captain information on health of a worker but also on health and safety risks, in general. This example shows that a narrow interpretation of privacy might result in diminishing safety: If privacy is unidimensional, and the only choice would be to decide to share the data with the employer, either the firefighter would accept more significant risks during the execution of the job because the data would be hidden (as in the GDPR), or the employer would have full access to all data, which could lead to misuse for other purposes.

The case of the firefighters showed a disbalance between what is actually beneficial for the health of the firefighters and the regulations that are meant to protect them. This is a major problem when implementing new technologies in the work environment. Given that the law not yet protects the user in fiduciary relationships in certain professions (31), it is important to acknowledge these design-use dynamics in the design phase of a new technology and come up with solutions that could help overcome this gap in the law. Although some researchers already call for changing the law for fiduciary relationships (33), this would be a long and arduous process. Even if the law would change on this matter, it would still be important to define in which situations data sharing is condoned and with whom sharing health data is necessary. Therefore, the agency-based approach asks for a thorough discussion with all stakeholders involved about what type of data is necessary to share with other actors and with whom in order to protect the health of the firefighters (as can be seen in the process described in **Figure 1**: level individual study). For instance, is it necessary to share raw data? Or would aggregated data suffice? Is it important to collect data for longer periods of time? Or can the data be removed directly after the fire was put out? But also, who has access to the data? And how can it be prevented that other colleagues have access to the data? This can also be an indirect result of the use of a sensor. What happens, for instance, if one firefighter is called back more often than other firefighters? Agreements on these issues should be strictly documented and revised if necessary.

Responsibilities of Stakeholders

In the case of the firefighters, the employer is serious about the responsibility for the health of the workers. The GDPR, however, prevents the employer from using personal data to protect firefighters from overheating in an emergency. In this case, the workers are at an impasse. Distraction from the task could cause immediate risks to themselves and colleagues; thus, it is impossible to self-monitor their current health parameters. This discrepancy between the desired situation and current regulations is frustrating for the fire department because the captains wish to protect their firefighters, but the GDPR makes it impossible for captains to use data for the goal of protection of workers.

CASE STUDY 2

Healthcare workers are often subject to irregular working hours due to shift work. These work characteristics can make it more difficult to uphold healthy habits, such as daily exercise and a balanced diet (63). An unhealthy lifestyle for a healthcare worker not only impacts their employability in the long term (64) but also impacts the view of the public on the healthcare institution, because the healthcare workers are assumed to 'know best' about the impact of lifestyle choices on long-term health. Both the issues of long-term health and the exemplary function of their work are well-known to healthcare workers, which is why many of them actively try to keep up good behaviour. In this case study, a healthcare institution asked for an intervention that allows employees to self-manage their health, without having to explain themselves to the employer. An activity tracker supports these workers in their health, because it allows them to monitor their daily behaviour despite the irregular hours and workload, and thereby supports these workers in becoming and staying healthy (65).

Overcoming Compartmentalisation

The healthcare institution where the study took place is eager to improve and tries to incorporate the ideas of workers into their workplace health promotion policies. The activity tracker used is a tracker developed for the consumer market, meaning that the research team did not have any influence on the specifications of the tracker. During implementation, however, the researchers decided to use proxy user accounts for all users, thereby enabling the researchers to tailor and alter the information that was given to the workers. These adjustments to the messages were intended to limit the impact on worker autonomy (see iterative process **Figure 1**: individual project—researcher). Apart from the researchers and the participants, nobody had access to the data.

The use of sensor technologies to assist in sustainable employability hinges on offering workers objective feedback and interventions that allow them to self-regulate behaviour. Illustrative for the ideal of autonomy was a participant, self-identified as overweight and unfit, who was eager to experiment with an activity tracker. This activity tracker enabled her to receive automated digital feedback on her daily exercise behaviour. This worker was committed to improving her condition:

> *I value a healthy lifestyle. I have difficulties keeping up with that for all sorts of reasons, and this is an opportunity for me to get some nonintrusive and time-saving support. I also would like to be an example for the patients who visit here. They need people like me as role models, people who struggle but make an effort to improve their health.*

She referred to her value of personal health. Receiving an activity tracker did not provide autonomy. However, due to the activity tracker, she could autonomously commit to her value of becoming healthy. This value had a different application in her work context, a healthcare organisation, where she wanted to set an example for others. She wanted to show that increasing daily exercise by walking more and taking the stairs is an essential commitment to improving health. Thus, in the work context, the worker wanted to achieve a healthy lifestyle as well as provide the moral value of being an example. She translated the value of her health and her position at work into a daily practical commitment of taking more steps. Thus, the use of this sensor technology helped her to achieve her ideal.

Nevertheless, the commitment of the worker was not only shaped by a momentous decision to accept the activity tracker. Her commitment was confirmed by making some progress in walking more steps. However, it was disaffirmed when a colleague from higher management saw her waiting for the elevator:

> *And then they are supporting "the week of taking the stairs" [...], but then, when I am standing in front of the elevator, [colleagues] say, "Oh, are you taking the elevator? We are taking the stairs!" That feels terrible—really terrible.*

This encounter made her question whether the entire experiment was about her improvement in health and realising her values, or whether it was ultimately about organisational control and cost reduction.

This example, although an individual experience, illustrates how personal autonomy can easily be threatened in a work environment if personal values are not acknowledged. Giving workers a health device does not merely provide a means for self-regulation, because the technology is embedded in a context that can promote or disavow the responsible commitment to the norms to which one is bound. This realisation calls for reflection on how the introduction of technology can affect autonomy of workers and how the context of the implemented technology influences the perceptions of autonomy of workers.

Worker autonomy as a prerequisite for health self-regulation was empirically investigated in the study of Bonvanie et al. (50). It examined activity trackers that give feedback information on health-related behaviour to workers. The example of activity trackers is of interest because it is used as a technology that enables workers to self-regulate a healthy lifestyle (66, 67). The underlying assumption was that the use of digital health technologies provides workers with autonomy *via* feedback and the freedom to respond to self-regulate health-related behaviour. Despite adjustments to the messages, intended to limit the impact on worker autonomy, these findings revealed that the use of a sensor technology did not significantly increase perceived autonomy and may have even reduced autonomy under certain conditions, especially for less healthy workers (50). Moreover, the workers who had used an activity tracker to monitor their behaviour before they received an employer-provided device experienced the same decrease in autonomy as the workers who used the activity tracker for the first time. This finding suggests that the activity tracker does not limit the autonomy of workers; instead, perceived autonomy may decrease due to the hierarchical relationship between workers and employers.

A Conscientious Autonomy-Enhancing Approach

The employer of the health-care institute who participated in this study demonstrated a value for healthy workers. That is, the employer already implemented several other activities and regulations, such as promoting a week of taking the stairs, providing a healthy cafeteria and offering a smoke-free property. Although independent researchers conducted the study, the normative standards of the activity tracker were encouraged by the employer. The goal was to walk 10,000 steps per day and take 10 flights of stairs. Some participants agreed with this goal and internalised the normative standard. Others, however, did not and perceived the feedback as pressure to aim for 10,000 steps. The participants who shared the same value of healthy living as the employer but had other ideas to implement it felt as if the activity tracker forced them to commit to normative standards of someone else.

These findings reflect the idea of conscientious autonomy (46): Autonomy that is committed to willed ideals of one as well as the norms and standards encountered in a particular setting that are adapted as normative. Hence, based on the disbalance between the individual goals and ideals of workers and the norms of their colleagues and employers, one can determine why the autonomy of certain workers declines when using a sensor technology. When implementing technologies or other interventions in the work environment, the employer, therefore, needs to pay specific attention to how the norms and culture in the work environment influence the autonomy of the workers.

Participation in the study and being able to discuss the impact of technologies with different stakeholders within the development process caused the employer to reconsider the current workplace health promotion policies. The employer altered their strategy into a more conscientious autonomy-enhancing approach. This was achieved by including a more diverse group of workers in the decision-making and evaluating the process regarding new technologies and interventions, thereby aiming to facilitate a healthy workplace and a lifestyle for all workers.

Responsibilities of Stakeholders

Similar to the case of the firefighters, the employer was responsible for the health of the health-care workers. This responsibility of the employer is limited to the work context, while the health of workers is also influenced by their private lives. By providing an activity tracker, the employer is walking a thin line between the work and the private context. One can ask the questions, where does the responsibility of the employer stop? And where does the responsibility of the worker begin? And where do they overlap? Interestingly, the participants in the study of Bonvanie et al. (50) stated that the ability to maintain their health is, partially, the responsibility of the employer, because their work environment has a large impact on this ability, and that their employer took this responsibility quite seriously. Both the employer and the workers experienced the intertwined nature of health, work, and the home environment, and aim to improve the collaboration on improving the overall health of the worker (see process **Figure 1**: individual level worker-employer).

DISCUSSION

Previous literature on responsible research and innovation struggled with three major problems: 1) compartmentalisation, 2) generalisation, and 3) vagueness about responsibilities. Rather than developing a theoretical approach to these problems, we highlighted two cases of the project SPRINT@Work. We aimed at describing how we explored the critical ethical issues privacy and autonomy in the development and implementation of digital health technologies in the setting of doing research. A context-specific analysis of both values was employed, keeping previous research and the legal context in mind. For the firefighters case study, this analysis resulted in the description of an agency-based concept of privacy, where it is necessary to analyse and regulate what the agent can do with the private information obtained (62). For the case study of the health-care professionals, this resulted in a conscientious autonomy-enhancing approach to the design and implementation of digital health technologies in the workplace. When this approach is employed, all stakeholders [with a specific emphasis on the user(s)] have to be actively involved in the design and implementation phase in order to achieve the intended goal of the technology, which is to enhance health-related behaviour (46).

Decompartmentalisation of Focus

Responsibilities for the assessment of risks of the new technology get indistinguishable when a transition between phases occurs (17). More specifically, engineers and researchers might have reflected on the impact of their new technology; however, after the development phase, responsibilities shift toward the user or organisations that implement the technology. They do not necessarily reflect on possible ethical and societal risk, and primarily focus on productivity or increasing product acceptance (12, 68). Ethical concerns arise as soon as technological innovations are introduced (69). Although an ethical script of an innovation shows what the default choices regarding privacy, responsibility, and autonomy are, at the same time, the reaction of the environment to this built-in ethical script plays a significant role. The ethical script is mainly developed by the engineers and researchers who develop the technological intervention, but the response of the user and his/her environment to this ethical script largely determines the privacy of the user and his or her possibility to exercise autonomy. Using a multi-stakeholder approach may help to overcome this problem of compartmentalisation by providing a smooth responsible transition from development to implementation.

In the case studies, the reflection on both design and field experiments involving health-related technologies in the workplace caused both the researchers, employers, and workers to reflect on the interpretation and implications of the concepts of privacy and autonomy (see **Figure 1**). This approach of integrating development and use of the digital health technology was necessary to successfully implement techniques from the field of RRI, such as reflexivity and responsiveness. The context-specific approach allowed for a cyclic approach, using outcomes from early implementations of technologies as input for further development. As a result, the researchers, employers and workers were able to work together to take unforeseen consequences of the technology into consideration, because they appeared during use by end users. This then allowed the researchers and engineers to alter the technology or the choices that were made during development and implementation.

Both cases show the benefits of including the tension between the intended and actual use in the development and implementation of a new technology. In the case of the firefighters, the balance between safeguarding privacy and safeguarding health could only be reached because the researchers were able to use input from actual use (during job performance). More specifically, due to the interaction between the researchers, workers, and the team captain, the application of the wearable thermometer for use in the workplace could be improved, which consequently benefits the health of the firefighters. In the case of the healthcare workers, the researchers closely monitored the impact of the technology on the autonomy of workers in the workplace. By doing so, they were not only able to reevaluate the benefit of the activity tracker but also caused the employer to reconsider the current workplace health promotion policies and the manner in which these come to be.

Prevention of Out-of-Context Generalisation

A responsible decision to provide workers with sensor technologies to sustain their employability requires careful analysis of the values at stake in the context of the specific workplace and the individual worker (70). In case of privacy, the GDPR offers a basic framework for the implementation of protection measures, while it also leaves room for interpretation and discussion. The GDPR (3) obligates and ensures that the decisions about data protection taken by the controller, for instance, an engineer or a researcher, are taken with great care, especially when "*processing of the data could result in high risk to the rights and freedoms of natural persons*" [30, section 35 (3)]. In order to help the controller making responsible decisions about privacy of individuals, the data protection impact assessment (DPIA) (71) is developed as a risk assessment method. This includes a multiple stakeholder approach to identify privacy risks. During meetings with stakeholders, a context-specific method of privacy by design is applied to design protection measures that are appropriate for a specific context.

The main focus of DPIA (and of the GDPR) is to protect the privacy of the user without paying much attention to other ethical issues in its analysis. Although it is a step in the right direction, in the development of new digital health technologies, other values, such as health, autonomy and responsibility, and the interplay between these values need to be reflected upon as well. The current study, therefore, used a context-specific approach of ethics (instead of privacy) to assess privacy and autonomy concerns in the workplace.

For both cases, the context-specific approach of ethics helped to identify the best approach to provide a framework of what is at stake in their specific context. Although from a different perspective, both the agency-based model of privacy (62) and the conscientious autonomy-enhancing design (46) can help identify bottlenecks, implicit norms, and courses of action during the development and implementation of new technologies and policies. These two approaches are a source of moral knowledge, given that the experiences in the field informed the researchers about what users value, and the dynamics between the researcher, employer, and user was explored by testing the conceptualisation of ethical principles in the work environment and further adjusted as deemed necessary.

Making Implied Responsibilities Explicit

Acting responsibly regarding health in the workplace is considered important (30), but employers experience difficulties taking their responsibility, and, in the case of health-promoting technologies in the workplace, other stakeholders find it difficult to share this responsibility. Leclercq-Vandelannoitte (12), in a study about the use of ubiquitous technologies in the workplace, observed that "*despite their prevalence and the importance of their consequences for workers, neither salespeople nor managers seem to be aware of them, feel responsible for them, or appear able or willing to identify the responsibilities involved in this process.*" In the case of workplace health-promoting technologies, responsibility involves multiple stakeholders with a prominent role for the employers (13), engineers (72), and researchers (73, 74). To protect the privacy of workers while gathering personal data, all stakeholders need to take their responsibility for the use of the involved technology (48).

The engineers and researchers have the responsibility to design the technology in such a way that it guarantees the privacy of the user and supports the user in his/her ability to react autonomously (75–77). However, engineers and researchers often do not offer sufficient insight into what constitutes a responsible use of their designs (12). Technologies are never value neutral (69), and it is important that engineers and researchers explore how the development and implementation of their technologies influence and mould not only the ethical values, such as privacy, but also the autonomy of employers and workers and help them reflect on this explorative process (75–77). The responsibility of engineers and/or researchers should focus on perspectives such as value-sensitive design, critical technical practise, reflective design, and values in design (78, 79).

The reflection on the responsibility of workers and employers is not a one-time action. As stated before, differences in interpretations of responsibilities can cause significant problems between workers and employers (30), and the use of technology often alters the original function (26). When using (new) technologies, workers and employers should discuss the responsibilities and intended actions of these technologies with the designers. This discussion should also entail the continuous reflection of the employer to determine whether the conscientious autonomy of the worker has improved. In the case of the healthcare professionals, sensor technologies enabled workers to take responsibility to target work-related health parameters within the workplace. In general, however, these technologies are most effective when workers feel autonomous to self-regulate health-relevant actions (50). Thus, employers should be alert for unintended effects of sensor technologies and ensure an environment that facilitates workers to take their responsibility. When workers and employers share values, such as health, technologies that support the personal goals of workers could increase a sense of conscientious autonomy, thereby improving the self-regulation of healthy behaviour.

Limitations

The context-specific approach of doing ethics has been a valuable method to investigate the core ethical principles of the digital health technologies in the case studies. In specific this approach helped to obtain a contextual conceptualisation of the ethical principles in the design and implementation of these technologies. However, we realised that this approach was not utilised to its full extent.

Multiple useful tools are now available to help researchers perform responsible research and innovation (e.g., rri-tools.eu). However, at the start of the project SPRINT@Work, approaches to incorporate an ethics structure throughout the complete research cycle of multidisciplinary projects were lacking or at least not commonly practised. Therefore, we started our journey with no clear approach in mind, and we developed our knowledge and approach as we went on.

In the case of the health care workers, this resulted in little attention to the predefined norms and values of the activity tracker. If we would have identified these norms and values before the start of the field studies, the researchers could have incorporated them in the design of the study. This could have prevented negative experiences of workers with the activity tracker.

In the case of the firefighters, we did not involve a specialist in the field of privacy law to help find potential legal solutions for data sharing while protecting the privacy of the workers. The project team would have benefitted from actively involving a privacy specialist from the start of the project. This could have influenced the approach taken by the researcher that designed and implemented the technology, the technology itself, and its suitability for future use.

At last, it can be stated that the problem of compartmentalisation applies to us researchers as well. In order to be able to manage this extensive project, the individual research projects of the PhD candidates were defined as much as possible. Although this approach was meant to save time, it resulted in delays both in the individual research projects concerning the development as in the projects concerning the implementation of the new technologies.

Based on these limitations, we advise multidisciplinary teams to start exploring potential critical ethical issues right from the start of the project. They could use this paper as a first lead on how such issues could be identified. Even though it might not be their initial field of expertise, we appeal to future multidisciplinary teams to also report their findings and possible adjustments to our proposed approach.

CONCLUSION: CALL FOR AN OVERDUE DISCUSSION

Based on a substantial literature review, we aimed to discuss the importance of context-specific ethics in design and implementation of digital health technologies. Focusing on a contextual conceptualisation of the core ethical principles in the design and implementation of digital health technologies helps to avoid compartmentalisation, out-of-context generalisation, and neglect of identifying responsibilities. Although it is a long reiterative process in which all stakeholders need to be included in order to assess all critical ethical issues sufficiently, this process is crucial to achieving the intended goal of a technology. We call for multidisciplinary teams, including relevant stakeholders, involved in innovation practises in workplace health promotion to publish their way of doing ethics. Future research teams can learn from these experiences and use and improve their approaches.

Finally, having laid out the landscape and problems of ethics surrounding technologies for workplace health promotion, we believe that an evaluation of policies and standards and a very overdue discussion guided by the signalled ethical problems are needed. Laws and regulations aim to offer protection to users of new technologies, but tend to focus on data access and privacy. Thereby, they leave out other values, such as responsibility and autonomy, which are in close interplay with privacy. It is, therefore, important that engineers and researchers of workplace health promotion themselves enter this debate. They should consider how the design and implementation of their technologies influence and mould the values of the users and adapt their technologies to protect the user from harm, and increase the acceptance. However, it does not stop there. They should also enter the debate about how policies and standards hinder or promote workplace health promotion.

DATA AVAILABILITY STATEMENT

The data analysed in this study is subject to the following licenses/restrictions: The datasets generated for the case studies that were analysed in this article are available on request to the corresponding author. Requests to access these datasets should be directed to Els Lisette Maria Maeckelberghe, e.l.m.maeckelberghe@umcg.nl.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Medical Ethical Exam Committee of the University Medical Center Groningen. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

CR, MJ, AB, and EM contributed to the conception and design of this study and contributed to the data analysis. CR, MJ, and AB contributed to the design of the case studies and data acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This study, part of SPRINT@Work, is partly financed by the European Regional Development Fund, the province and municipality of Groningen, and the province of Drenthe (grant no: T-3036, 2013).

ACKNOWLEDGMENTS

We would like to thank Prof. Dr. Sandra Brouwer, Prof. Dr. Ute Bültmann, Prof. Dr. Monicque Lorist, Prof. Dr. Michiel Reneman, Prof. Dr. Ir. Bart Verkerke, Prof. Dr. Ir. Hans Wortmann, Sander Spook M.Sc., and Leonie Oosterhuis M.Sc. for their support and input during this study. We would also like to thank the two reviewers who have significantly contributed to the content of our paper.

REFERENCES

1. Kenny GP, Yardley JE, Martineau L, Jay O. Physical work capacity in older adults: implications for the aging worker. *Am J Ind Med.* (2008) 51:610–25. doi: 10.1002/ajim.20600
2. Truxillo DM, Cadiz DM, Hammer LB. Supporting the aging workforce: a review and recommendations for workplace intervention research. *Ann Rev Organ Psychol Organ Behav.* (2015) 2:351–81. doi: 10.1146/annurev-orgpsych-032414-111435
3. Villumsen M, Madeleine P, Jørgensen MB, Holtermann A, Samani A. The variability of the trunk forward bending in standing activities during work vs. leisure time. *Appl Ergon.* (2017) 58:273–280. doi: 10.1016/j.apergo.2016.06.017
4. Aryal A, Ghahramani A, Becerik-Gerber B. Monitoring fatigue in construction workers using physiological measurements. *Autom Construct.* (2017) 82:154–65. doi: 10.1016/j.autcon.2017.03.003
5. Grunwald A. Technology assessment for responsible innovation. In: *Responsible Innovation 1*. Dordrecht: Springer Netherlands (2014). p. 15–31.
6. Owen R, Macnaghten P, Stilgoe J. Responsible research and innovation: from science in society to science for society, with society. *Sci. Public Policy.* (2012) 39:751–60. doi: 10.1093/scipol/scs093
7. Stahl BC. Responsible research and innovation: the role of privacy in an emerging framework. *Sci Public Policy.* (2013) 40:708–16. doi: 10.1093/scipol/sct067
8. Stilgoe J, Owen R, Macnaghten P. Developing a framework for responsible innovation. *Res Policy.* (2013) 42:1568–80. doi: 10.1016/j.respol.2013.05.008
9. Fisher E, O'Rourke M, Evans RJ, Kennedy EB, Gorman ME, Seager TP. Mapping the integrative field: taking stock of socio-technical collaborations. *J Responsible Innov.* (2015) 2:39–61. doi: 10.1080/23299460.2014.1001671
10. Efstratiou C, Davies N, Kortuem G, Finney J, Hooper R, Lowton M. Experiences of designing and deploying intelligent sensor nodes to monitor hand-arm vibrations in the field. In: *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services-MobiSys'07*. New York, NY: ACM Press (2007). p. 127.
11. Kortuem G, Alford D, Ball L, Busby J, Davies N, Efstratiou C, et al. Sensor networks or smart artifacts? an exploration of organizational issues of an industrial health and safety monitoring system. In: *UbiComp*. (2007) *2007*: *Ubiquitous Computing*. Berlin; Heidelberg: Springer Berlin Heidelberg. p. 465–82.
12. Leclercq-Vandelannoitte A. An ethical perspective on emerging forms of ubiquitous IT-based control. *J Bus Ethics.* (2017) 142:139–54. doi: 10.1007/s10551-015-2708-z
13. Palm E. Securing privacy at work: the importance of contextualized consent. *Ethics Inf Technol.* (2009) 11:233–41. doi: 10.1007/s10676-009-9208-8
14. Motti VG, Caine K. Human factors considerations in the design of wearable devices. *Proc Hum Factors Ergon Soc Ann Meet.* (2014) 58:1820–1824. doi: 10.1177/1541931214581381
15. Saurabh K, Rao H, Amrutur B, Sundarrajan A. Continuous core body temperature estimation via surface temperature measurements using wearable sensors-is it feasible? In: *Proceedings of the International Conference on Biomedical Electronics and Devices*. SCITEPRESS-Science and Technology Publications (2014). p. 181–6.
16. Sole M, Musu C, Boi F, Giusto D, Popescu V. RFID sensor network for workplace safety management. In: *2013 IEEE 18th Conference on Emerging Technologies and Factory Automation (ETFA)*. Cagliari: IEEE (2013). p. 1–4
17. Jakobsen S-E, Fløysand A, Overton J. Expanding the field of Responsible Research and Innovation (RRI)–from responsible research to responsible innovation. *Eur Plann Stud.* (2019) 27:2329–43. doi: 10.1080/09654313.2019.1667617
18. Verheiden S, Tijink D, Heijltjes M. *Nut En Noodzaak Van Het Investeren In Digitale Vaardigheden in de Zorg*. (2020). Available online at: https://www.digivaardigindezorg.nl/wp-content/uploads/2020/12/Rapport-nut-en-noodzaak-van-het-investeren-in-digitale-vaardigheden.pdf.
19. Al Ameen M, Liu J, Kwak K. Security and privacy issues in wireless sensor networks for healthcare applications. *J Med Syst.* (2012) 36:93–101. doi: 10.1007/s10916-010-9449-4
20. Conger S, Pratt JH, Loch KD. Personal information privacy and emerging technologies. *Inf Syst J.* (2013) 23:401–17. doi: 10.1111/j.1365-2575.2012.00402.x
21. Nissenbaum H. *Privacy in Context*. Stanford, CA: Stanford University Press (2010).
22. Zhu H, Gao L, Li H. Secure and privacy-preserving body sensor data collection and query scheme. *Sensors.* (2016) 16:179. doi: 10.3390/s16020179
23. Damman OC, van der Beek AJ, Timmermans DRM. Employees are ambivalent about health checks in the occupational setting. *Occup Med.* (2015) 65:451–8. doi: 10.1093/occmed/kqv048
24. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for COVID-19. *Nature.* (2020) 582:29–31. doi: 10.1038/d41586-020-01578-0
25. Lucivero F, Marelli L, Hangel N, Zimmermann BM, Prainsack B, Galasso I, et al. Normative positions towards COVID-19 contact-tracing apps: findings from a large-scale qualitative study in nine European countries. *Crit Public Health.* (2021) 1–14. doi: 10.1080/09581596.2021.1925634. [Epub ahead of print].
26. Kiran AH. Responsible design. A conceptual look at interdependent design–use dynamics. *Philos Technol.* (2012) 25:179–98. doi: 10.1007/s13347-011-0052-5
27. Taylor AS, Vincent J. An SMS history. In: *Mobile World*. New York, NY: Springer-Verlag (2006). p. 75–91.
28. Schukat M, McCaldin D, Wang K, Schreier G, Lovell NH, Marschollek M, et al. Unintended consequences of wearable sensor use in healthcare. contribution of the IMIA wearable sensors in Healthcare WG. *Yearb Med Inform.* (2016) 25:73–86. doi: 10.15265/IY-2016-025
29. Spook SM, Koolhaas W, Bültmann U, Brouwer S. Implementing sensor technology applications for workplace health promotion: a needs assessment among workers with physically demanding. *BMC Public Health.* (2019) 19:1100. doi: 10.1186/s12889-019-7364-2
30. van Berkel J, Meershoek A, Janssens RM, Boot CR, Proper KI, van der Beek AJ. Ethical considerations of worksite health promotion: an exploration of stakeholders' views. *BMC Public Health.* (2014) 14:458. doi: 10.1186/1471-2458-14-458

31. GDPR. *European Parliament and Council of European Union.* European Union (2016). Available online at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32016R0679
32. Brassart Olsen C. To track or not to track? Employees' data privacy in the age of corporate wellness, mobile health, and GDPR†. *Int Data Privacy Law.* (2020) 10:236–52. doi: 10.1093/idpl/ipaa004
33. Arora C. Digital health fiduciaries: protecting user privacy when sharing health data. *Ethics Inf Technol.* (2019) 21:181–96. doi: 10.1007/s10676-019-09499-x
34. Choi B, Hwang S, Lee S. What drives construction workers' acceptance of wearable technologies in the workplace?: indoor localization and wearable health devices for occupational safety and health. *Autom Constr.* (2017) 84:31–41. doi: 10.1016/j.autcon.2017.08.005
35. Jacobs JV, Hettinger LJ, Huang Y-H, Jeffries S, Lesch MF, Simmons LA, et al. Employee acceptance of wearable technology in the workplace. *Appl Ergon.* (2019) 78:148–56. doi: 10.1016/j.apergo.2019.03.003
36. Lodge T, Crabtree A. Privacy engineering for domestic IoT: enabling due diligence. *Sensors.* (2019) 19:4380. doi: 10.3390/s19204380
37. Mulligan DK, Koopman C, Doty N. Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy. *Philos Trans R Soc A Math Phys Eng Sci.* (2016) 374:20160118. doi: 10.1098/rsta.2016.0118
38. Barocas S, Nissenbaum H, Lane J, Stodden V, Bender S, Nissenbaum H. Big Data's End Run around Anonymity and Consent. In: *Privacy, Big Data, and the Public Good*. New York, NY: Cambridge University Press (2013). p. 44–75.
39. DeCew J. Privacy. In: Zalta EN, editor. *The Stanford Encyclopedia of Philosophy*. (2015). Available online at: https://plato.stanford.edu/archives/spr2015/entries/privacy
40. Dwork C. Differential privacy. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, Vol. 4052, Egham: LNCS. Springer Verlag (2006). p. 1–12.
41. Solove DJ. *Understanding Privacy*. Cambridge: Harvard University Press (2008).
42. van der Klink JJ, Bültmann U, Burdorf A, Schaufeli WB, Zijlstra FR, Abma FI, et al. Sustainable employability – definition, conceptualization, and implications: a perspective based on the capability approach. *Scand J Work Environ Health.* (2016) 42:71–9. doi: 10.5271/sjweh.3531
43. Ryan RM, Deci EL. Self-regulation and the problem of human autonomy: does psychology need choice, self-determination, and will? *J Pers.* (2006) 74:1557–86. doi: 10.1111/j.1467-6494.2006.00420.x
44. Goossens RHM, Netten MP, Van der Doelen B. An office chair to influence the sitting behavior of office workers. *Work.* (2012) 41(Suppl. 1*)*:2086–8. doi: 10.3233/WOR-2012-0435-2086
45. Roossien CC, Stegenga J, Hodselmans AP, Spook SM, Koolhaas W, Brouwer S, et al. Can a smart chair improve the sitting behavior of office workers? *Appl Ergon.* (2017) 65:355–61. doi: 10.1016/j.apergo.2017.07.012
46. Kukla R. Conscientious autonomy: displacing decisions in health care. *Hast Center Rep.* (2005) 35:34–44. doi: 10.1353/hcr.2005.0025
47. Crampton J, Jajodia S, Mayes K. Differential privacy. In : *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, Vol. 4052, LNCS. Springer Verlag (2006). p. 1–12
48. Johnson DG, Powers TM. Computer systems and responsibility: a normative look at technological complexity. *Ethics Inf Technol.* (2005) 7:99–107. doi: 10.1007/s10676-005-4585-0
49. Allender S, Colquhoun D, Kelly P. Competing discourses of workplace health. *Health.* (2006) 10:75–93. doi: 10.1177/1363459306058989
50. Bonvanie A, Broekhuis M, Janssen O, Maeckelberghe E, Wortmann H. Health self-management applications in the work environment: the effects on employee autonomy. *Front Digit Health.* (2020) 2:9. doi: 10.3389/fdgth.2020.00009
51. de Jong M, Bonvanie A, Jolij J, Lorist MM. Dynamics in typewriting performance reflect mental fatigue during real-life office work. *PLoS ONE.* (2020) 15:e0239984. doi: 10.1371/journal.pone.0239984
52. de Jong M, Jolij J, Pimenta A, Lorist MM. Age modulates the effects of mental fatigue on typewriting. *Front Psychol.* (2018) 9:1113. doi: 10.3389/fpsyg.2018.01113
53. Roossien CC, Baten CTM, van der Waard MWP, Reneman MF, Verkerke GJ. Automatically determining lumbar load during physically demanding work: a validation study. *Sensors.* (2021) 21:2476. doi: 10.3390/s21072476
54. Roossien CC, Heus R, Reneman MF, Verkerke GJ. Monitoring core temperature of firefighters to validate a wearable non-invasive core thermometer in different types of protective clothing: concurrent *in-vivo* validation. *Appl Ergon.* (2020) 83:103001. doi: 10.1016/j.apergo.2019.103001
55. Roossien CC, Hodselmans AP, Heus R, Reneman MF, Verkerke GJ. Evaluation of a wearable non-invasive thermometer for monitoring ear canal temperature during physically demanding (outdoor) work. *Int J Environ Res Public Health.* (2021) 18:4896. doi: 10.3390/ijerph18094896
56. Roossien CC, Krops LA, Wempe JB, Verkerke GJ, Reneman MF. Can we analyze breathing gasses without a mouth mask? Proof-of-concept and concurrent validity of a newly developed breathing gasses analyzing headset. *Appl Ergon.* (2021) 90:103266. doi: 10.1016/j.apergo.2020.103266
57. Dutch Data Protection Agency. *AP: Verwerking Gezondheidsgegevens Wearables Door Werkgevers Mag Niet*. (2016). Available online at: https://autoriteitpersoonsgegevens.nl/nl/nieuws/ap-verwerking-gezondheidsgegevens-wearables-door-werkgevers-mag-niet
58. World Medical Association. World medical association declaration of helsinki. *JAMA.* (2013) 310:2191. doi: 10.1001/jama.2013.281053
59. Roghanchi P, Kocsis KC. Challenges in selecting an appropriate heat stress index to protect workers in hot and humid underground mines. *Safety Health Work.* (2018) 9:10–6. doi: 10.1016/j.shaw.2017.04.002
60. Krishnamurthy M, Ramalingam P, Perumal K, Kamalakannan LP, Chinnadurai J, Shanmugam R, et al. Occupational heat stress impacts on health and productivity in a steel industry in southern india. *Safety Health Work.* (2017) 8:99–104. doi: 10.1016/j.shaw.2016.08.005
61. Morgado M, Talaia M, Teixeira L. A new simplified model for evaluating thermal environment and thermal sensation: an approach to avoid occupational disorders. *Int J Ind Ergon.* (2017) 60:3–13. doi: 10.1016/j.ergon.2015.11.001
62. Manson NC, O'Neill O. *Rethinking Informed Consent in Bioethics*. *Rethinking Informed Consent in Bioethics*. Cambridge: Cambridge University Press (2007).
63. De Jongh J, Mcdougal F. *Gezonde Leefstijl en Vitaliteit op de Werkvloer: Een Kwantitatief Onderzoek Onder Werknemers in Opdracht Van Convenant Gezond Gewicht*. Amsterdam (2014). Available online at: https://www.stvda.nl/~/media/Files/Stvda/Thema/Arbeidsomstandigheden/vitaliteit/20140925-ruigrok-Netpanel.ashx
64. Hendriksen IJM, Snoijer M, de Kok BPH, van Vilsteren J, Hofstetter H. Effectiveness of a multilevel workplace health promotion program on vitality, health, and work-related outcomes. *J Occup Environ Med.* (2016) 58:575–83. doi: 10.1097/JOM.0000000000000747
65. Kalantari M. Consumers adoption of wearable technologies: literature review, synthesis, and future research agenda. *Int J Technol Market.* (2017) 12:1. doi: 10.1504/IJTMKT.2017.10008634
66. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health. *JAMA.* (2007) 298:2296. doi: 10.1001/jama.298.19.2296
67. Mattila E, Orsama A-L, Ahtinen A, Hopsu L, Leino T, Korhonen I. Personal health technologies in employee health promotion: usage activity, usefulness, and health-related outcomes in a 1-year randomized controlled trial. *JMIR MHealth UHealth.* (2013) 1:e16. doi: 10.2196/mhealth.2557
68. Chatfield K, Borsella E, Mantovani E, Porcari A, Stahl B. An investigation into risk perception in the ICT industry as a core component of responsible research and innovation. *Sustainability.* (2017) 9:1424. doi: 10.3390/su9081424
69. Martin KE, Freeman RE. The separation of technology and ethics in business ethics. *J Bus Ethics.* (2004) 53:353–64. doi: 10.1023/B:BUSI.0000043492.42150.b6
70. ten Have M, van der Heide A, Mackenbach JP, de Beaufort ID. An ethical framework for the prevention of overweight and obesity: a tool for thinking through a programme's ethical aspects. *Eur J Public Health.* (2013) 23:299–305. doi: 10.1093/eurpub/cks052
71. Bieker F, Friedewald M, Hansen M, Obersteller H, Rost M. A process for data protection impact assessment under the European General Data Protection Regulation. In: *Lecture Notes in Computer Science (including subseries Lecture*

Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 9857 Frankfurt: LNCS. Springer Verlag (2016). p. 21–37

72. Doorn N. Exploring responsibility rationales in research and development (RandD). *Sci Technol Hum Values.* (2012) 37:180–209. doi: 10.1177/0162243911405344
73. Doorn N, van de Poel I. Editors' overview: moral responsibility in technology and engineering. *Sci Eng Ethics.* (2012) 18:1–11. doi: 10.1007/s11948-011-9285-z
74. La Fors K, Custers B, Keymolen E. Reassessing values for emerging big data technologies: integrating design-based and application-based approaches. *Ethics Inf Technol.* (2019) 21:209–26. doi: 10.1007/s10676-019-09503-4
75. Robaey Z, Spruit SL, van de Poel I. The food warden: an exploration of issues in distributing responsibilities for safe-by-design synthetic biology applications. *Sci Eng Ethics.* (2018) 24:1673–96. doi: 10.1007/s11948-017-9969-0
76. van de Poel I, Verbeek, P.-P. Editorial: ethics and engineering design. *Sci Technol Hum Values.* (2006) 31:223–36. doi: 10.1177/0162243905285838
77. Verbeek, P.-P. (2011). *Moralizing Technology : Understanding and Designing the Morality of Things.* Chicago, IL : University of Chicago Press.
78. Cummings ML. Integrating ethics in design through the value-sensitive design approach. *Sci Eng Ethics.* (2006) 12:701–15. doi: 10.1007/s11948-006-0065-0
79. Shilton K. Values levers. *Sci Technol Hum Values.* (2013) 38:374–97. doi: 10.1177/0162243912436985

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Digital Health

REVIEW
published: 06 September 2021
doi: 10.3389/fdgth.2021.697072

Digital Mental Health for Young People: A Scoping Review of Ethical Promises and Challenges

Blanche Wies, Constantin Landers and Marcello Ienca*

Department of Health Sciences and Technology, ETH Zurich (Swiss Federal Institut of Technology), Zurich, Switzerland

Mental health disorders are complex disorders of the nervous system characterized by a behavioral or mental pattern that causes significant distress or impairment of personal functioning. Mental illness is of particular concern for younger people. The WHO estimates that around 20% of the world's children and adolescents have a mental health condition, a rate that is almost double compared to the general population. One approach toward mitigating the medical and socio-economic effects of mental health disorders is leveraging the power of digital health technology to deploy assistive, preventative, and therapeutic solutions for people in need. We define "digital mental health" as any application of digital health technology for mental health assessment, support, prevention, and treatment. However, there is only limited evidence that digital mental health tools can be successfully implemented in clinical settings. Authors have pointed to a lack of technical and medical standards for digital mental health apps, personalized neurotechnology, and assistive cognitive technology as a possible cause of suboptimal adoption and implementation in the clinical setting. Further, ethical concerns have been raised related to insufficient effectiveness, lack of adequate clinical validation, and user-centered design as well as data privacy vulnerabilities of current digital mental health products. The aim of this paper is to report on a scoping review we conducted to capture and synthesize the growing literature on the promises and ethical challenges of digital mental health for young people aged 0–25. This review seeks to survey the scope and focus of the relevant literature, identify major benefits and opportunities of ethical significance (e.g., reducing suffering and improving well-being), and provide a comprehensive mapping of the emerging ethical challenges. Our findings provide a comprehensive synthesis of the current literature and offer a detailed informative basis for any stakeholder involved in the development, deployment, and management of ethically-aligned digital mental health solutions for young people.

Keywords: ethics, digital health, mental health, adolescent and youth, digital health (eHealth)

Edited by:
Karina Vold,
University of Toronto, Canada

Reviewed by:
Luke Balcombe,
Griffith University, Australia
Greg Wadley,
The University of Melbourne, Australia

***Correspondence:**
Marcello Ienca
marcello.ienca@hest.ethz.ch

Specialty section:
This article was submitted to Health Technology Innovation, a section of the journal Frontiers in Digital Health

Received: *18 April 2021*
Accepted: *06 July 2021*
Published: *06 September 2021*

Citation:
Wies B, Landers C and Ienca M (2021) Digital Mental Health for Young People: A Scoping Review of Ethical Promises and Challenges. Front. Digit. Health 3:697072. doi: 10.3389/fdgth.2021.697072

INTRODUCTION

Mental health disorders are complex disorders of the nervous system characterized by a behavioral or mental pattern that causes significant distress or impairment of personal functioning (1). These include, among others, anxiety, depression, substance use disorders, schizophrenia, eating disorders, bipolar disorder, and post-traumatic stress disorder. Mental health disorders compose a significant portion of the global burden of disease. In 2017, 970 million people worldwide had a mental health disorder, comprising approximately 13% of the global population. Since then, it is estimated that mental health conditions have increased worldwide as they now cause on average 1 in 5 years lived with disability (2). The mortality rate of people with mental disorders is significantly higher than the average population, with a median life expectancy loss of 10.1 years. Mental health disorders are attributable to eight million deaths each ear, that is 14.3% of deaths worldwide (3).

Mental illness is of particular concern for younger people. The WHO estimates that around 20% of the world's children and adolescents have a mental health condition, a rate that is almost double compared to the general population. Mental-illness-induced suicide is the second leading cause of death among 15 to 29-year-olds. Despite these figures, the global median of government health expenditure that goes to mental health is <2% (2). To make things worse, the epidemiology of mental illness is expected to be exacerbated by the ongoing new Coronavirus disease 2019 (COVID-19) pandemic. A recent survey has found that the pandemic has affected the mental health of 59% of people in the United States (4). Research shows that the ongoing COVID-19 pandemic is contributing to widespread emotional distress and increased risk for psychiatric illness, either directly associated with the COVID-19 illness or indirectly through imposition of restrictive public health measures that infringe on personal freedoms and associated financial losses (5). Furthermore, people with serious mental illness have been observed to be disproportionately affected by the pandemic (6). This impact has been particularly disruptive for young people, many of whom have self-reported increased mental health issues as a result of lockdowns. A recent survey conducted among 13–25 years olds with a history of mental health needs in the United Kingdom found that 67% of respondents believe that the pandemic will have a long-term impact on their mental health (7).

One approach toward mitigating the medical and socio-economic effects of mental illness is leveraging the power of digital health technology to deploy assistive, preventative, and therapeutic solutions for people in need. As a consequence, digital mental health is a growing field of interest in digital health and scientific research. We define "digital mental health" as any application of digital health technology for mental health assessment, support, prevention, and treatment. This technological cluster includes mobile health (mHealth) applications, wearables, consumer neurotechnologies, virtual reality systems, online platforms, care coordination systems, assisted living ecosystems etc.

Young people are the primary end-users or patient groups of digital mental health tools: they are early adopters of all things digital, including digital health (8). The relevance of leveraging digital mental health solutions has further increased as a consequence of the COVID-19 pandemic because of both the increased prevalence of mental illness and the growing demand of telemedicine services (9). The application of digital health methodologies to young people thus promises considerable benefits and has received growing attention in the literature. However, this age group is also particularly vulnerable and susceptible to manipulation, especially via digital devices and methods. As a result, the use of digital technologies for mental health treatment among adolescents and children generates benefits and ethical issues.

Growing evidence suggests that digital mental health can improve mental health conditions such as depression across various patient populations (8, 10, 11). However, there is only limited evidence that digital mental health tools can be successfully implemented in clinical settings (12). Authors have pointed to a lack of technical and medical standards for digital mental health apps (13), personalized neurotechnology (14) and assistive technology for age-related cognitive decline (15) as a possible cause of suboptimal adoption and implementation in the clinical setting, Mohr et al. (12) have suggested that digital mental health research should therefore be solution-focused to develop pragmatic solutions. Further, ethical concerns have been raised related to insufficient effectiveness (14), lack of adequate clinical validation and user-centered design (16) as well as data privacy vulnerabilities (15) of current digital mental health products. Assessing the benefits and risks of digital mental health systems requires, therefore, a careful balancing act and a holistic approach to scrutinizing the advantages that these socio-technical trends can bring for mental health patients while minimizing their unintended risks. Most importantly, it requires a careful risk-benefit analysis that could inform ethical guidelines, policy interventions, oversight mechanisms and clinical decision making in this domain.

The aim of this paper is to report on a scoping review we conducted to capture and synthesize the growing literature on the promises and ethical challenges of digital mental health for young people. We define "young people" as the combined group of children and adolescents—concretely people in the age group from 0 to 25 years. This review seeks to survey the scope and focus of the relevant literature, identify major benefits and opportunities, and provide a comprehensive mapping of the emerging ethical challenges. Our findings provide a comprehensive synthesis of the current literature and offer a detailed informative basis for any stakeholder involved in the development, deployment and management of ethically-aligned digital mental health solutions for young people.

METHODS

The objective of this review was to gather information about the benefits and ethical challenges regarding digital technologies for mental health treatment and assessment among adolescents or children. To this purpose, on the 7th of October 2020 we searched five databases (PubMed, Scopus, World of Science, PsychInfo, IEEE Xplore, and the ACM Digital Library) in order to retrieve eligible publications. The following search string was used:*(("social media" OR "Digital Media" OR "big data" OR "Artificial Intelligence" OR "digital phenotyping" OR "digital mental health" OR "digital biomarkers" OR "mental health apps" OR "digital sensors" OR "digital mental health technologies" OR "health related Apps" OR "mobile Health" OR eHealth*

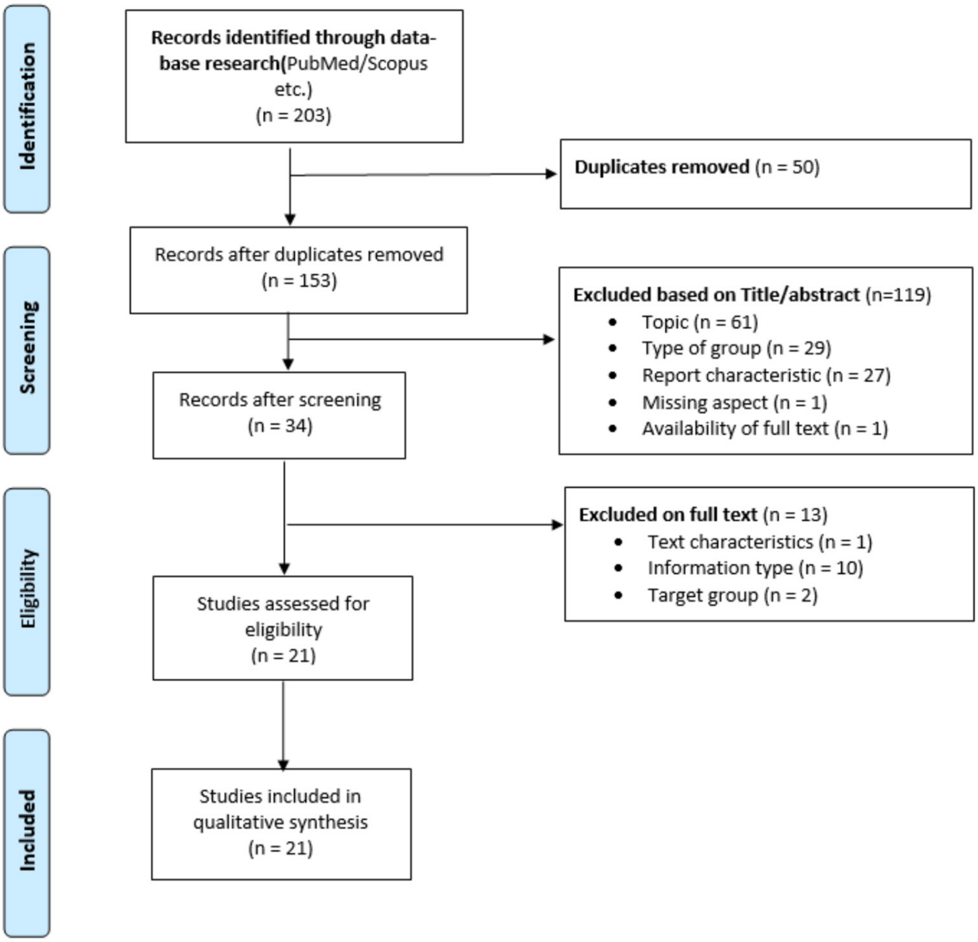


FIGURE 1 | PRISMA flow chart.

OR smartphones OR wearables OR "Holter monitoring") AND (ethics OR bioethics OR "bioethical issues" OR "ethical issues" OR "ethical analysis" OR "ethical review") AND ("mental health" OR "mental well-being" OR "emotional health" OR "emotional well-being") AND ("young adult" OR young OR adolescent OR child OR teenager)).

Based on the PRISMA Statement and flowchart, four phases of review were conducted: identification, screening, eligibility assessment and final synthesis (see **Figure 1**). Our search string initially retrieved 203 papers. All entries were exported into the Endnote reference management software. Automatic duplicate removal was performed. Fifty articles were identified as duplicates and therefore removed. The remaining 135 articles were screened based on Title/Abstract and assessed based on the inclusion and exclusion criteria (see Annex 1 in **Supplementary Material**). Thirty-four articles passed the eligibility assessment and were included into the final synthesis. Articles were deemed eligible if they suited the following inclusion criteria: (a) original peer-review journal publication; (b) written in English, German, Spanish, Italian, or French (languages spoken by the research team); (c) published between 2015 and October 2020; (d) describing/assessing ethical considerations relating to digital health technologies designed for or utilized by children and/or young adults (under 25) for mental health support or otherwise related to the promotion of mental health.

In addition to this systematic review component and compatibly with the best practices for scoping reviews, we conducted a grey literature analysis via non-academic search engines and citation chaining. To this purpose, we used multiple unstructured combinations of the search string. This led to the inclusion of 9 additional articles to the final synthesis.

A total of 26 articles were included into the final synthesis and an in-depth review of full-text articles included in the synthesis was performed. Data were analyzed through qualitative thematic analysis with assistance of the MAXQDA data analysis software. Through the establishment of a keyword coding system, recurrent thematic patterns were inductively identified and subsequently grouped into different themes and subthemes. Our analysis consisted of three sequential steps. First, for each article, we screened the presence of ethically relevant considerations. During this phase, ethically relevant keywords and statements were searched in the full texts of all reviewed articles. This process was performed by two authors using

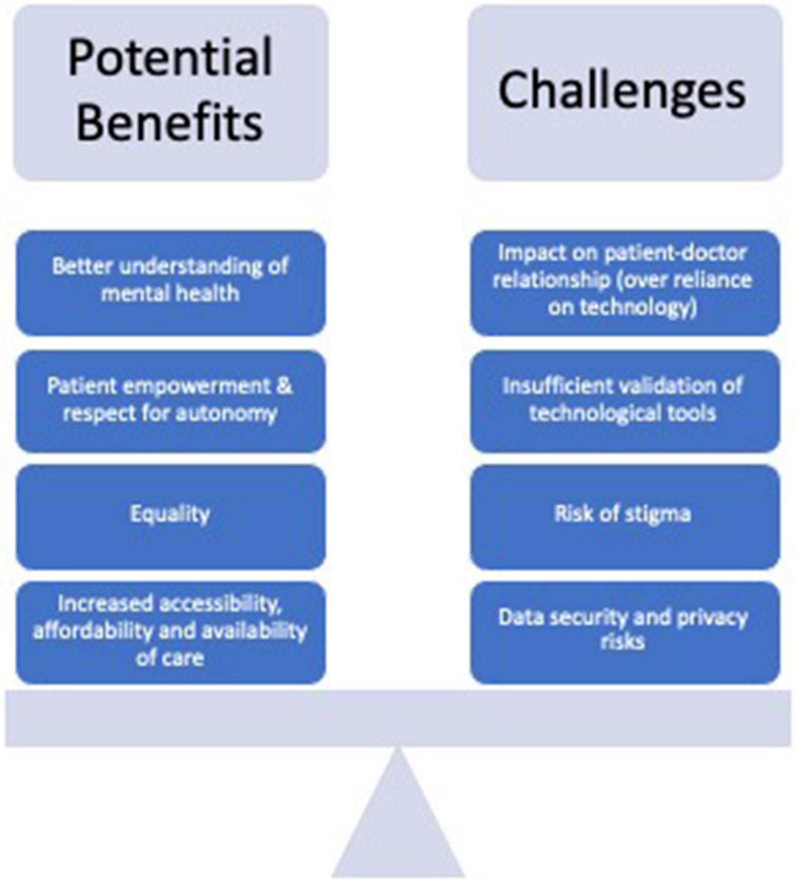


FIGURE 2 | Expected benefits and ethical challenges related to digital mental health for young people.

both software-guided keyword search (software used: Endnote X9) and unguided full-text review. Second, we clustered all retrieved ethical considerations into main thematic families using thematically oriented content analysis (17). Each thematic family was further classified into sub-families relative to specific sub-components of the main ethical theme. When the same digital health technology description contained more than one ethical consideration, all considerations were allocated to their respective thematic families and subfamilies.

RESULTS

The analysis showed a diverse range of themes relating to the opportunities and ethical challenges of using digital mental health technologies among young people. **Figure 2** provides a visual overview of key codes and themes that emerged from our analysis.

Ethically Significant Benefits and Opportunities

Accessibility

The most recurrent expected benefit associated with the use of digital mental health is the prospect of increased accessibility to health care (18–21). This assessment was based on the increased affordability of mental health apps or internet-based platforms in comparison to face-to-face consultations (22–27) and the easy access given the liberation of geographical restraints (22–25, 28–30). This potential benefit is of ethical significance because increased access to healthcare is a critical to promote health equality and justice. Furthermore, digital mental health solutions were expected to provide more continuous, around-the-clock availability of help or support (18, 23–26, 29, 31, 32). By increasing accessibility, digital mental health technologies were also seen to hold potential for increasing equality between different population groups (21, 24–26, 33, 34), as well as within the patient-therapist relationship (32).

Enhanced Therapy Facilitation and Prevention

The usage of digital mental health therapeutics is also seen to facilitate the therapy, prediction and prevention of mental illness of the patient (24, 26, 29, 31) The beneficial potential of digital mental health technology for continuous and accessible care delivery is of particular relevance to low-to-moderate cases that are not being evaluated and treated, especially in areas where mental health care resources are under high levels of pressure due to severe mental illness cases. At the same time, however, our findings underscore that digital mental health, albeit already useful for monitoring purposes, has yet to become effective for predictive purposes. For example, Mulder et al. (35) and Chan et al. (36) cautioned to redirect attention from algorithmic prediction of suicide to a causal pathway and called for paying more attention to real engagement with the individual patient, their specific problems and circumstances. Although machine learning algorithms appeared to improve existing decision support tools, their usefulness in the clinical setting was deemed limited.

Autonomy and Empowerment

Another important theme was the potential of digital mental health technologies to increase the autonomy and sense of empowerment of young adults (24, 29, 32). The use of digital mental health tools gives youngsters, on the one hand, the chance to play a more active role in their own treatment as they can actively seek support or control difficult situations (18, 20, 25, 26, 29, 31, 32), as well as the pace of answering/responding (26). Additionally it gives the patient the opportunity to implement the learned coping strategies outside the therapy setting and thereby increases patient autonomy and sense of empowerment (33, 34). On the other hand the technologies fostered autonomy through providing easier access to information and support or more generally, the possibility to manage mental health and well-being (27, 29, 32, 34). Additionally, some authors argued that an increased empowerment leads to higher responsibility for taking care of one's own mental health development, which is an important step in the treatment of mental illness (18).

High Acceptability Among Young People

As a substantial part of the youth's social interactions and life take place in the digital space, e.g., through the use of social media, young people's perspectives and choices regarding multiple digital mental health related topics are influenced by the digital ecosystem in which they are embedded. They are more inclined to accept the use of digital tools for the assessment, treatment or support of mental health issues (27). Further young people use the digital space (e.g., the Internet) as a trusted source of information (19, 22, 37) or see it as an easier way to start to talk about mental health, their problems or to seek help (27, 32, 34, 38). In addition, our analysis suggests that the acceptability of using digital mental health technologies may also be positively influenced by their potential for enabling more anonymous interactions compared to face-to-face meetings with health professionals (24, 25, 38). The flexibility and anonymity of the digital space allows young adults to avoid social stigma or exclusion and increases their feeling of comfort to share personal data as you can quickly access the therapy tool through your smartphone (22, 24–27, 38).

Further the type of data that can be collected through mental health apps, chatbots, or social media may well lead to an increased understanding of mental illnesses, as more data is available for analysis (19, 20, 23, 24, 37, 38). The increased amount of data and positive attitude of people toward the usage, could ultimately help to eradicate or at least decrease the stigma that is attached to mental illness (24, 26). Lee et al. further discussed the promise of transferring the trust given to chatbots to professionals (38).

Disadvantages and Ethical Challenges

Privacy and Confidentiality

The most frequently mentioned risks of digital mental health technologies addressed in the literature regard the privacy, confidentiality and security of the user's data and information obtained through digital mental health applications. The biggest concerns expressed by authors regards what happens if confidential information is shared with or access given to third parties (19, 20, 23–27, 31, 38–43). Authors argued that the negative consequences of insufficiently secured data sharing can reach into multiple domains of life, such as work, school or even into relationships with friends, families or partners (19).

Patient Mistrust Due to Privacy and Confidentiality Concerns

Mistrust in data sharing due to privacy concerns and confidentiality breaches may reduce the effectiveness of mental health treatments. Authors argued that if the data are insufficiently secured, hence at risk of being breeched, multiple negative consequences are expected to arise from patient mistrust. First, patients' trust in their psychiatrist or psychotherapist may be lost, tarnishing or at least negatively influencing the doctor-patient relationship (19, 26, 28, 32, 34, 41). Second, the prospect of privacy breaches and security vulnerabilities is expected to decrease the acceptability of digital mental health technologies (40, 43), leading to an even bigger vulnerability of already exposed people and increasing the unease and uncertainty of the users toward the technology (19, 38, 40, 41, 43). Other authors highlighted that mistrust in digital mental health technologies is further aggravated by the fact that patients often feel that technologies collect too much information (38) and develop sceptical attitudes due to the rapid speed of technological evolution (21, 38).

Pervasive Stigma

It has been observed that digital mental health technologies can increase the risk of stigmatization for young adults and children, especially where relevant data have been exposed (26, 40, 43). Stigmatization may lead to various devastating effects in young peoples' lives. Cyberbullying is widespread and may be particularly burdensome where data is leaked (40). Digital technologies, in particular social-media, may elicit addiction and reinforce self-harming behavior. Internalized stigma may lead patients to use social networks to self-expose such self-harming behavior, which in turn may reinforce stigma against their illnesses (44). The effects of stigmatization extend to how patients are treated by institutions. Feuston and Piper argued that institutional representations of mental illness, such as the media, contribute to stigma by providing "unfavourable and inaccurate representations of psycihatric disorders" (45). Martinez-Martin and Kreitmair have shown that addiction-induced illegal drug use has had legal consequences for patients when digital mental health technology providers had to share personal information (40).

Accessibility and Equal Access

Although enhanced access is often seen as a major benefit of digital mental health (19, 20, 34), issues related to accessibility and equal access are nonetheless identified as major ethical concerns (21, 34, 46). Authors observed that systemic issues, such as socio-economic inequality, lead to considerable structural barriers to access (18, 47, 48). One widely referred to barrier is the *digital divide* that describes the phenomenon that technology is not equally available to all social groups due to economic, social or cultural inequalities (21, 24, 49). Concretely, barriers such as poor network coverage in rural areas or the cost of digital communication constrain disadvantaged children and young adults to access relevant content (34). Without digital literacy or adequate access, patients may face severe disadvantages, as they are prevented from accessing novel mental health care solutions (49). Similarly, although digital mental health technologies hold potential for increasing the autonomy and sense of empowerment of young adults (24, 29, 32), they also raise the risk of diminishing patient autonomy by increasing the risk of digital addiction and manipulation.

Cross-Cultural and Cross-Country Attitudes and Resources

Cross-cultural and cross-country differences in attitudes and resources have been associated to considerably different standards of care. Sharma showed that stakeholders' socio-technological aspirations to technology for disabled children differ considerably among Indian and Finnish stakeholders (21). This is, in their view, attributable to differences in resource availability and government engagement in "developed" and "emerging" economies. This raises further issues as healthcare systems in which mental health resources are under high pressure often fail to address the needs of patients with less severe illnesses—leading to systemic issues around not serving the underserved (47).

Some researchers and app developers have deliberately attempted to react to these issues by developing products and approaches that enable access to emotional self-regulation and mental health prevention among the socioeconomically disadvantaged (18) or disabled (21). However, Sharma showed that currently available technology is still insufficiently engaging and inclusive to fully offset these concerns (21).

Finally, the absence of regulatory clarity concerning the responsibility for data leaks and potential dignitarian or other digital harms caused by technology misuse, render digital mental health technologies less trustworthy (32, 40, 43).

Clinical Validation and the Need for Ethical-Legal Guidance

Other repeatedly discussed topics were the unclear efficacy and effectiveness (25, 27, 30, 31, 33, 34, 40), translational challenges and the uncertain feasibility of successful implementation (25) as well as uncertain reliability (19, 27, 28, 43, 50) of these technologies. Authors emphasized that many mental health apps and internet-based platforms are not subjected to extensive and professional evaluations or clinical validation studies, which leads to unclear influences and outcomes. This uncertainty is exacerbated by the fact that little ethical and regulatory guidance currently exists for digital mental health. To reduce uncertainty and bridge this normative gap, several authors called for the necessity of developing an ethical-legal framework for digital mental health technology, chiefly through ethical guidelines, recommendations, and best practices.

Blurring the boundary within the doctor-patient relationship, increasing the risk of over-reliance on digital mental health technologies as well as poor conduct were also identified as ethically relevant challenges (26, 31, 51). Authors expressed concerns about the unrealistic expectations of around-the-hour-availability of psychiatrists through mails or text messages (32, 37) which could pose additional burden on health professionals. In addition, digital mental health applications could facilitate the sharing of personal information among both parties involved through ill-suited channels (26, 32, 34) and blur the boundaries of a psychiatrist's assessment by including deontologically questionable activities such as checking a patient's social media accounts (26, 32).

Consent and Dependency

Another challenge identified in the literature is the unclearly defined expectation of which parties have to consent to data processing in digital mental health applications (19, 26, 37, 41). Authors also reported insufficient clarity about the adequacy of consent obtained through digital mental health applications, in particular regarding the type of data processing or intervention that the user is consenting to (28, 40, 41, 43).

As shown by Lee et al. ensuring data and algorithmic transparency when processing users' personal information is very important (38). The over-reliance and the all-time availability of these technologies are feared to influence the young adults' capabilities of social interactions with the ancillary risk of diminishing their trust to talk about problems with their friends (43) or increase the dependency of young people on clinical support (32). In addition to that, websites and chatbots are often intentionally designed to get the users dependent on the technologies. Notably in young adults this can ultimately increase addictive behaviors (27, 29). Further it can decrease the feeling of responsibility of the young people to take care of their health as they expect that it is done for them (32, 34) and additionally diminish their willingness to attend face-to-face consultations (34).

Other Topics

Further topics that were discussed, though with less frequency, in the literature included issues of accountability, liability, anonymity, the relation of digital mental health and human rights as well as the evaluated role of these technologies based on different ethical frameworks such as principlism, ethics of care and utilitarianism. The discussion around accountability and liability was for instance emphasized by Martinez-Martin et al. as they flagged the limited applicability of traditionally defined therapeutic codes to providers of direct-to-consumer (DTC) technologies. The authors observed that the same rules of conduct that prevent malpractice or liability issues in traditional therapy settings are not precisely applicable to digital psychotherapies, especially those administered through DTC software and devices. This creates a problem of sub-optimal accountability for e.g. chatbots to establish a safe and trusting relationship with patients (40).

LIMITATIONS

This study presents four main limitations. First, it may be affected by a selection bias because the search retrieved only articles written in languages known by the researchers (English, Spanish, French, German, and Italian), excluding articles written in other languages. A similar limitation affects database selection: screening additional databases may have possibly identified additional relevant studies. Finally, our study included only peer-reviewed articles in scientific journals, hence excluded other articles sources such as conference proceedings and book chapters. The risk of selection bias is inherent to any review because the number of databases that can be feasibly searched is always finite. We attempted to minimize selection bias by exploring both domain-general (Web of Science, Scopus) and domain-specific databases including the major databases in biomedical research, psychology and computer science. Second, exhaustiveness is not the objective of scoping reviews as the explorative nature and broad focus of this methodology makes it "unrealistic to retrieve and screen all the relevant literature" (52). With regard to article types, although we recognize that including also non-journal articles such as conference proceedings may have the valuable consequence of adding to our synthesis papers from conferences highly competitive discipline specific conferences (especially in computer science), it could thereby have the adverse effect of including low-quality unscrutinised contributions from other domains. Therefore, we considered restricting the synthesis to peer-reviewed journal articles a valid quality control mechanism.

DISCUSSION

As often observed, mental health is a public health priority. Developmental psychiatry research indicates that most mental health disorders begin in childhood and adolescence. This raises an additional medical and ethical duty to detect and assess mental health needs early and treat them during child development. Therefore, deploying digital solutions that can reliably monitor and identify mental health needs during early phases of psychological development is an inherently ethical task. These technologies hold promise for alleviating the burden of mental illness, reducing the risk that critical health needs during this sensitive time of child development remain undetected, providing novel assistive and therapeutic resources for young people in need and improving practical aspects of mental healthcare delivery. This is particularly valuable since untreated mental health problems originating during childhood and adolescence can reportedly lead to future negative health and social outcomes. At the same time, research in digital ethics has largely shown that digital health devices and software raise a variety of ethical challenges, especially challenges related to privacy, equality of access, patient autonomy. These challenges may be exacerbated when digital health solutions are designed for and accessed by children and adolescents, as young people with chronic mental conditions belong to vulnerable groups and are often below the age of consent for medical treatment. For this reason, deploying digital mental health solutions for young people requires a proactive ethical assessment which carefully balances the benefits that these technologies can bring against the possible collateral risks.

Our thematic analysis shows that increasing accessibility to mental healthcare is a core ethically relevant opportunity enabled by digital mental health. Our findings reveal that the increased affordability of digital mental health tools in comparison to face-to-face consultations combined with limited dependence of these systems on geographical constrains may facilitate access to mental healthcare. Research has shown that the provision of mental health services is currently constrained by structural barriers, with many people facing insufficient access to diagnostics and treatment (47, 48). As a consequence, more than half of adults with mental illness in countries such as the U.S. do not receive mental healthcare treatment. From a public health perspective, the increased affordability of mental health apps and internet-based platforms in comparison to face-to-face consultations is likely to facilitate access to mental healthcare in countries such as Switzerland and the United States where care provision is not entirely reimbursed through public finances. In addition, it is likely to expand access to mental healthcare in World regions such as rural areas and low-and-middle-income countries where institutional care provision is limited. Even in areas where access to care is not constrained, the around-the-clock availability of digital tools holds promise for improving prevention, help and support.

From a patient perspective, the potential of digital mental health technologies to increase the autonomy and sense of empowerment of young adults merits particular attention. In biomedical ethics, the principle of autonomy is typically understood as the capacity of the person to deliberate or act on the basis of one's own desires, that is the ability to act freely in accordance with a self-chosen plan (53). Digital mental health tools hold promise for giving young patients the chance to play a more active role in their own treatment and provide them with the opportunity to actively seek support or control difficult situations. Additionally, they give them the chance to refine coping strategies learned outside the therapy setting and gain easier access to information and support (27, 29, 32, 34). Our findings illustrate that young patients may be more inclined to seek mental health support if mediated through digital tools due to the impersonal and at-your-fingertips nature of these technologies which makes them more suitable to maintain anonymity and avoid the psychological stress induced by face-to-face encounters. Evidence from developmental psychiatry suggests that this increased patient empowerment may ignite a virtuous circle in which patients are incentivized to take higher responsibility for taking care of their own mental health development, which is an important step in the treatment of mental illness. Patient empowerment is also promoted through the positive effect on health literacy that digital tools are likely to exert. As young patients have the opportunity to monitor their mental health continuously and autonomously, they can gain exploratory knowledge about their conditions, thereby improving their understanding of their own mental illness. Finally, the prospect of reducing stigma may create not only a direct benefit for the technology users but also a positive externality for mental health patients in general.

Despite these prospective benefits, digital mental health tools also appear to raise technical, scientific, ethical, and regulatory challenges. Proactively addressing these challenges is paramount to ensure ethical development in the digital mental health arena and increase the chances that the promissory outlook described above will materialize. Our findings reveal that many young people use digital technologies to access information about their mental health. Although the increased accessibility of such information is beneficial, it should also be viewed with caution. If the information they receive is not reliable and scientifically vetted, it may lack validity and thereby tamper both health outcomes and patient trust in mental health services. If digital mental health tools lack validity, they may provide incorrect advice. As a consequence, patients may not seek the right help they need (40). The risk of suboptimal efficacy and insufficient clinical validation has already been observed in areas of digital mental health such as direct-to-consumer neurotechnology for mental well-being (14) and intelligent assistive technology for people with dementia and/or age-related cognitive decline (15, 16). If digital mental health tools cannot ensure efficacy and reliability, it is unlikely they can improve health outcomes and reduce the burden of mental illness. In addition, the increasing reliance on machine learning and other AI models for prediction and human-machine interaction needs to be vetted to ensure scientific validity, reliability, and transparency. Although machine learning algorithms appeared to improve existing decision support tools, their usefulness in the clinical setting was deemed limited by an ongoing lack of information on model building and uncertain accuracy (54–56). Further, it has been noted that there has yet to be clinical evaluation of predictive technologies for digital health interventions (57).

Our findings indicate that digital mental health tools may help ensure a greater degree of anonymity compared to face-to-face consultations and thereby reduce stigma. However, this beneficial potential can only materialize if digital mental health technologies ensure high standards of data security and information privacy. Privacy breaches have already been observed in several digital mental tools such as mobile health apps, wearables, consumer neurotechnologies and assistive devices for psychogeriatric care (15, 58, 59). These privacy weaknesses include illicit access by third parties to confidential patient-related information, cybercrime and accidental data leakage. Data security and privacy weaknesses are likely to have a negative snowball effect on patient trust and the doctor-patient relationship (19, 26, 28, 32, 34, 41). Also, they are expected to decrease the acceptability of digital mental health technologies (40, 43) among younger people.

With regard to patient autonomy, digital mental health tools appear to be a double-edge sword. On the one hand, they hold potential for increasing the autonomy and sense of empowerment of young adults. On the other hand, they also raise the risk of diminishing patient autonomy by increasing the risk of digital addiction and manipulation (60, 61).

Overall, our findings suggest that digital mental health technologies can improve the quality of mental healthcare provision and the quality of life of younger patients. At the same time, they indicate that technology is not a panacea for all mental health problems affecting young people's mental health and that ethical issues must be proactively addressed. When navigating these issues, special attention should be devoted to the specific needs and wishes of each patient and age subgroup. We recommend that future research on this topic should focus on specific subpopulations such as low-to-moderate and subsyndromal cases. As young people constitute a broad and heterogenous age group, it is important to look at target subpopulations within this cluster and identify the necessary codesign requirements for these end users.

These findings may provide a useful informative basis for public decision-making on digital mental health for younger people. Our thematic analysis supports the view that leveraging both technical and normative interventions holds potential for maximizing the benefits of digital health technologies while minimizing the risks. In particular, technical solutions such as cryptography and secure multi-party computation can raise the bar of device and software security, hence increase the protection of patient-generated data and protect patient privacy. In parallel, ethical guidelines for digital mental health systems can help improve the safety and efficacy of these systems and establish best practices for ethical design, responsible innovation and successful clinical implementation. However, improving safety and efficacy standards cannot be achieved exclusively through guidelines and other soft-law or hard-law interventions, but also requires a paradigm shift of the digital mental technology industry toward a culture of stewardship and responsible innovation. The Organisation for Economic Co-operation and Development's (OECD) Recommendation on Responsible Innovation in Neurotechnology (2019) offers an internationally accepted framework for promoting responsible innovation in this field. These principles and standards, however, have to be adequately implemented into product design, development, and experimentation. In particular, enhancing clinical validation standards can improve effectiveness and safety only if the holistic well-being of the patient is put at the forefront of the digital mental health enterprise and novel technologies are developed and assessed using patient-centered and participatory approaches to technology development. Our findings suggest that a shift toward patient-centered design is particularly necessary for digital mental health technologies for younger people as the needs of younger people are typically under-addressed (62). Furthermore, developing comprehensive implementation concepts appears necessary to avoid translational bottlenecks and ensure the successful translation of digital mental health technologies from the designing laboratories to personalized solutions for end-users. In order to increase the accessibility of digital mental health services, including in rural areas and among disadvantaged socio-economic segments, stakeholders should explore interventions that could lower the costs of sufficiently validated digital mental health services for individual users. Our findings reveal a number of proposed strategies to achieve this aim such as promoting the adoption of open-source hardware and software as well as adopting cost reimbursement plans by healthcare providers. The recurrent focus on fairness and access equality suggests that avoiding the exacerbation of socio-economic inequalities via digital tools is a paramount requirement for the ethically aligned deployment of these technologies. Rather than aggravating the digital divide, digital mental health tools should expand young people's access to mental health services by enabling a more widespread delivery of technology-mediated care in rural areas, among economically disadvantaged groups and among patient groups who—due to the nature of their pathology—would particularly benefit from reducing the frequency of face-to-face encounters. In order to ensure the successful adoption of these technologies among children and adolescents, policy makers should consider collaborating with educational institutions and seek the integration of these technological resources into the school setting. School-based mental health practice holds promise in meeting unmet mental health needs of children and adolescents by expanding access to quality mental health care for hard-to-reach populations (55). The recurrent emphasis put by the literature on educational resources suggests that incorporating digital mental health tools into school-based mental health practice could improve the delivery of mental health services to children, expand the resources available to educators and health providers, and monitor the effectiveness of digital mental health interventions in a systematic way. Collaborative activities involving educators, healthcare providers, technology developers and end-users are highly needed to ensure the effective and responsible deployment of digital mental health technologies for the benefit of younger people.

AUTHOR CONTRIBUTIONS

BW contributed to the review protocol, performed the review, data analysis, and co-wrote the manuscript. CL contributed to the data analysis and co-wrote the manuscript. MI conceived of the study, developed the research protocol, contributed to data analysis, and co-wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

MI has received funding from ERA-NET NEURON JTC 2020 Ethical, Legal, and Social Aspects (ELSA) of Neuroscience and Innosuisse- Swiss Innovation Agency under award 40158.1 INNO-ICT. CL received funding from the Swiss National Science Foundation NRP 77 programme 407740_187356/1.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2021.697072/full#supplementary-material

REFERENCES

1. Bolton D. *What is Mental Disorder?: An Essay in Philosophy, Science and Values*. Oxford: Oxford University Press (2008). doi: 10.1093/med/9780198565925.001.0001
2. WHO. *Mental Health*. World Health Organization (2021). Available online at: https://www.who.int/health-topics/mental-health#tab=tab_1 (Accessed March 28, 2021).
3. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. *JAMA Psychiatry*. (2015) 72:334–41. doi: 10.1001/jamapsychiatry.2014.2502
4. Checkup T. *Mental Health Statistics 2021*. (2021). Available online at: https://www.singlecare.com/blog/news/mental-health-statistics/ (accessed March 28, 2021).
5. Pfefferbaum B, North CS. Mental health and the covid-19 pandemic. *N Engl J Med*. (2020) 383:510–2. doi: 10.1056/NEJMp2008017
6. Druss BG. Addressing the COVID-19 pandemic in populations with serious mental illness. *JAMA Psychiatry*. (2020) 77:891–2. doi: 10.1001/jamapsychiatry.2020.0894
7. Young Minds. *Coronavirus: Impact on Young People With Mental Health Needs* (2021).
8. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. *J Med Internet Res*. (2019) 21:e12869. doi: 10.2196/12869
9. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. *JMIR Ment Health*. (2020) 7:e18848. doi: 10.2196/18848
10. Campbell AN, Nunes EV, Matthews AG, Stitzer M, Miele GM, Polsky D, et al. Internet-delivered treatment for substance abuse: a multisite randomized controlled trial. *Am J Psychiatry*. (2014) 171:683–90. doi: 10.1176/appi.ajp.2014.13081055
11. O'Hara D, Lisa A. *Marsch Develops Digital Interventions for the Treatment of Opioid Addiction*. (2019). Available online at: https://www.apa.org/members/content/digital-interventions-opioid-addiction (accessed March 21, 2021). doi: 10.1037/e511102019-001
12. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research and three things we can do about them. *Psychiatr Serv*. (2017) 68:427–9. doi: 10.1176/appi.ps.201600541
13. Torous J, Firth J, Huckvale K, Larsen ME, Cosco TD, Carney R, et al. The emerging imperative for a consensus approach toward the rating and clinical recommendation of mental health apps. *J Nerv Ment Dis*. (2018) 206:662–6. doi: 10.1097/NMD.0000000000000864
14. Wexler A, Reiner PB. Oversight of direct-to-consumer neurotechnologies. *Science*. (2019) 363:234–5. doi: 10.1126/science.aav0223
15. Ienca M, Haselager P, Emanuel EJ. Brain leaks and consumer neurotechnology. *Nat Biotechnol*. (2018) 36:805–10. doi: 10.1038/nbt.4240
16. Ienca M, Fabrice J, Elger B, Caon M, Scoccia Pappagallo A, Kressig RW, et al. Intelligent assistive technology for Alzheimer's disease and other dementias: a systematic review. *J Alzheimers Dis*. (2017) 56:1301–40. doi: 10.3233/JAD-161037
17. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. *Nurs Health Sci*. (2013) 15:398–405. doi: 10.1111/nhs.12048
18. Slovák P, Theofanopoulou N, Cecchet A, Cottrell P, Altarriba Bertran F, Dagan E, et al. "I just let him cry": designing socio-technical interventions in families to prevent mental health disorders. *Proc ACM Hum Comput Interact*. (2018) 2:160. doi: 10.1145/3274429
19. Grové C. Using social networking sites in research: an emerging approach to engaging with young people who have a parent with a

mental illness and/or substance abuse disorder. *Front Psychiatry.* (2019) 10:281. doi: 10.3389/fpsyt.2019.00281

20. Grant C, Widnall E, Cross L, Simonoff E, Downs J. Informing the development of an E-platform for monitoring wellbeing in schools: involving young people in a co-design process. *Res Involvement Engag.* (2020) 6. doi: 10.1186/s40900-020-00219-0
21. Sharma S, Avellan T, Linna J, Achary K, Turunen M, Hakulinen J, et al. Socio-technical aspirations for children with special needs: a study in two locations - India and Finland. *ACM Trans Access Comput.* (2020) 13:13. doi: 10.1145/3396076
22. Cosgrove V, Gliddon E, Berk L, Grimm D, Lauder S, Dodd S, et al. Online ethics: where will the interface of mental health and the internet lead us? *Int J Bipolar Disord.* (2017) 5:26. doi: 10.1186/s40345-017-0095-3
23. Allen NB, Nelson BW, Brent D, Auerbach RP. Short-term prediction of suicidal thoughts and behaviors in adolescents: can recent developments in technology and computational science provide a breakthrough? *J Affect Disord.* (2019) 250:163–9. doi: 10.1016/j.jad.2019.03.044
24. Doraiswamy PM, London E, Candeias V. *Empowering 8 Billion Minds: Enabling Better Mental Health for All via the Ethical Adoption of Technologies*. World Economic Forum (2019). doi: 10.31478/201910b
25. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *J Med Internet Res.* (2019) 21:e13216. doi: 10.2196/13216
26. Heinsch M, Geddes J, Sampson D, Brosnan C, Hunt S, Wells H, et al. Disclosure of suicidal thoughts during an e-mental health intervention: relational ethics meets actor-network theory. *Ethics Behav.* (2019) 31:151–70. doi: 10.1080/10508422.2019.1691003
27. Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I. Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomed Inform Insights.* (2019) 11. doi: 10.1177/1178222619829083
28. Boonstra TW, Nicholas J, Wong QJ, Shaw F, Townsend S, Christensen H. Using mobile phone sensor technology for mental health research: integrated analysis to identify hidden challenges and potential solutions. *J Med Internet Res.* (2018) 20:e10131. doi: 10.2196/10131
29. D'alfonso S, Phillips J, Valentine L, Gleeson J, Alvarez-Jimenez M. Moderated online social therapy: Viewpoint on the ethics and design principles of a web-based therapy system. *J Med Internet Res.* (2019) 6:e14866. doi: 10.2196/14866
30. Uhlhaas P, Torous J. Digital tools for youth mental health. *NPJ Digit Med.* (2019) 2:104. doi: 10.1038/s41746-019-0181-2
31. Rice S, Robinson J, Bendall S, Hetrick S, Cox G, Bailey E, et al. Online and social media suicide prevention interventions for young people: a Focus on implementation and moderation. *J Can Acad Child Adolesc Psychiatry.* (2016) 25:80–6.
32. Ignatowicz A, Slowther, A.-M., Elder P, Bryce C, Hamilton K, et al. Ethical implications of digital communication for the patient-clinician relationship: analysis of interviews with clinicians and young adults with long term conditions (the LYNC study). *BMC Med Ethics.* (2018) 19:11. doi: 10.1186/s12910-018-0250-0
33. Jones DJ, Anton M, Gonzalez M, Honeycutt A, Khavjou O, Forehand R, et al. Incorporating mobile phone technologies to expand evidence-based care. *Cogn Behav Pract.* (2015) 22:281–90. doi: 10.1016/j.cbpra.2014.06.002
34. Griffiths F, Bryce C, Cave J, Dritsaki M, Fraser J, Hamilton K, et al. Timely digital patient-clinician communication in specialist clinical services for young people: a mixed-methods study (the LYNC study). *J Med Internet Res.* (2017) 19:e102. doi: 10.2196/jmir.7154
35. Mulder R, Newton-Howes G, Coid J. The futility of risk prediction in psychiatry. *Br J Psychiatry*. (2016) 209:271–27. doi: 10.1192/bjp.bp.116.184960
36. Chan MK, Bhatti H, Meader N, Stockton S, Evans J, O'connor RC, et al. Predicting suicide following self-harm: systematic review of risk factors and risk scales. *Br J Psychiatry*. (2016) 209:277–83. doi: 10.1192/bjp.bp.115.170050
37. Carson NJ, Gansner M, Khang J. Assessment of digital media use in the adolescent psychiatric evaluation. *Child Adolesc Psychiatr Clin N Am.* (2018) 27:133–43. doi: 10.1016/j.chc.2017.11.003
38. Lee YC, Yamashita N, Huang Y. Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. *Proc ACM Hum Comput Interact.* (2020) 4:31. doi: 10.1145/3392836
39. Wu E, Torous J, Hardaway R, Gutheil T. Confidentiality and privacy for smartphone applications in child and adolescent psychiatry: unmet needs and practical solutions. *Child Adolesc Psychiatr Clin N Am.* (2017) 26:117–24. doi: 10.1016/j.chc.2016.07.006
40. Martinez-Martin N, Kreitmair K. Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. *J Mental Health.* (2018) 5:e32. doi: 10.2196/mental.9423
41. Perez Vallejos E, Koene A, Carter CJ, Hunt D, Woodard C, Urquhart L, et al. Accessing online data for youth mental health research: meeting the ethical challenges. *Philos Technol.* (2019) 32:87–110. doi: 10.1007/s13347-017-0286-y
42. WHO. *Youth-Centred Digital Health Interventions: A Framework for Planning, Developing and Implementing Solutions With and for Young People*. Geneva: World Health Organization (2020).
43. Yew GCK. Trust in and ethical design of carebots: the case for ethics of care. *Int J Soc Robot*. (2020). doi: 10.1007/s12369-020-00653-w
44. Pater JA, Farrington B, Brown A, Reining LE, Toscos T, Mynatt ED. Exploring indicators of digital self-harm with eating disorder patients: a case study. *Proc ACM Hum Comput Interact.* (2019) 3:84. doi: 10.1145/3359186
45. Feuston JL, Piper AM. Beyond the coded gaze: analyzing expression of mental health and illness on Instagram. *Proc ACM Hum Comput Interact.* (2018) 2:51. doi: 10.1145/3274320
46. Feuston JL, Taylor AS, Piper AM. Conformity of eating disorders through content moderation. *Proc ACM Hum Comput Interact.* (2020) 4:40. doi: 10.1145/3392845
47. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. *Lancet.* (2007) 370:878–9. doi: 10.1016/S0140-6736(07)61239-2
48. Whittle EL, Fisher KR, Reppermund S, Lenroot R, Trollor J. Barriers and enablers to accessing mental health services for people with intellectual disability: a scoping review. *J Mental Health Res Intellect Disabil.* (2018) 11:69–102. doi: 10.1080/19315864.2017.1408724
49. Bauer M, Glenn T, Monteith S, Bauer R, Whybrow PC, Geddes J. Ethical perspectives on recommending digital technology for patients with mental illness. *Int J Bipolar Disord.* (2017) 5:6–6. doi: 10.1186/s40345-017-0073-9
50. UNICEF. *Safeguarding Boys and Girls: When Chatbots Answer Their Private Questions* (2020).
51. O'Reilly M, Levine D, Law E. Applying a 'digital ethics of care' philosophy to understand adolescents' sense of responsibility on social media. *Pastoral Care Educ*. (2020) 39:91–107. doi: 10.1080/02643944.2020.1774635
52. Gentles SJ, Lokker C, Mckibbon KA. Health information technology to facilitate communication involving health care providers, caregivers, and pediatric patients: a scoping review. *J Med Internet Res.* (2010) 12:e22. doi: 10.2196/jmir.1390
53. Varelius J. The value of autonomy in medical ethics. *Med Health Care Philos.* (2006) 9:377–88. doi: 10.1007/s11019-006-9000-z
54. Shatte AB, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med.* (2019) 49:1426–48. doi: 10.1017/S0033291719000151
55. Moon J, Williford A, Mendenhall A. Educators' perceptions of youth mental health: implications for training and the promotion of mental health services in schools. *Children Youth Serv Rev.* (2017) 73:384–91. doi: 10.1016/j.childyouth.2017.01.006
56. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, Jeste DV. Artificial intelligence for mental health and mental illnesses: an overview. *Curr Psychiatry Rep*. (2019) 21:116. doi: 10.1007/s11920-019-1094-0
57. Bucci S, Berry N, Morris R, Berry K, Haddock G, Lewis S, et al. "They are not hard-to-reach clients. We have just got hard-to-reach services." Staff views of digital health tools in specialist mental health services. *Front Psychiatry.* (2019) 10:344. doi: 10.3389/fpsyt.2019.00344
58. Villaronga EF, Felzmann H, Pierce R, De Conca S, De Groot A, Robins S, et al. Nothing comes between my robot and me': privacy and human-robot interaction in robotised healthcare. In: Leenes R, van Brakel R, Gutwirth S, De Hert P, editors. *Data Protection and Privacy: The Internet of Bodies*. Hart Publishing (2018). p. 135–70.
59. Ienca M, Villaronga EF. Privacy and security issues in assistive technologies for dementia. *Intell Assist Technol Dement.* (2019) 221. doi: 10.1093/med/9780190459802.003.0013

60. Ienca M, Vayena E. Cambridge analytica and online manipulation. *Sci Am.* (2018) 30:384–91.
61. Susser D, Roessler B, Nissenbaum H. Technology, autonomy, and manipulation. *Internet Policy Rev.* (2019) 8. doi: 10.14763/2019.2.1410
62. Zhai Y, Du X. Addressing collegiate mental health amid COVID-19 pandemic. *Psychiatry Res.* (2020) 288:113003. doi: 10.1016/j.psychres.2020.113003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 18 November 2021
doi: 10.3389/fdgth.2021.689999

Connect With Me. Exploring Influencing Factors in a Human-Technology Relationship Based on Regular Chatbot Use

Lara Christoforakos, Nina Feicht†, Simone Hinkofer†, Annalena Löscher†, Sonja F. Schlegl† and Sarah Diefenbach*

Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

Companion technologies, such as social robots and conversational chatbots, take increasing responsibility for daily tasks and support our physical and mental health. Especially in the domain of healthcare, where technologies are often applied for long-term use, our experience with and relationship to such technologies become ever more relevant. Based on a 2-week interaction period with a conversational chatbot, our study ($N = 58$) explores the relationship between humans and technology. In particular, our study focuses on felt social connectedness of participants to the technology, possibly related characteristics of technology and users (e.g., individual tendency to anthropomorphize, individual need to belong), as well as possibly affected outcome variables (e.g., desire to socialize with other humans). The participants filled in short daily and 3 weekly questionnaires. Results showed that interaction duration and intensity positively predicted social connectedness to the chatbot. Thereby, perceiving the chatbot as anthropomorphic mediated the interrelation of interaction intensity and social connectedness to the chatbot. Also, the perceived social presence of the chatbot mediated the relationship between interaction duration as well as interaction intensity and social connectedness to the chatbot. Characteristics of the user did not affect the interrelations of chatbot interaction duration or intensity and perceived anthropomorphism or social presence. Furthermore, we did not find a negative correlation between felt social connectedness of users to the technology and their desire to socialize with other humans. In sum, our findings provide both theoretical and practical contributions. Our study suggests that regular interaction with a technology can foster feelings of social connectedness, implying transferability of dynamics known from interpersonal interaction. Moreover, social connectedness could be supported by technology design that facilitates perceptions of anthropomorphism and social presence. While such means could help to establish an intense relationship between users and technology and long-term engagement, the contexts in which anthropomorphic design is, actually, the means of choice should be carefully reflected. Future research should examine individual and societal consequences to foster responsible technology development in healthcare and beyond.

Keywords: human-computer interaction, human-technology relationship, social connectedness, anthropomorphism, social presence, digital health technologies, conversational chatbot

Edited by:
Geke Ludden,
University of Twente, Netherlands

Reviewed by:
Natalie Benda,
Cornell University, United States
Randy Klaassen,
University of Twente, Netherlands

****Correspondence:***
Lara Christoforakos
lara.christoforakos@psy.lmu.de

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Human Factors and Digital Health, a section of the journal Frontiers in Digital Health

Received: *01 April 2021*
Accepted: *11 October 2021*
Published: *18 November 2021*

Citation:
Christoforakos L, Feicht N, Hinkofer S, Löscher A, Schlegl SF and Diefenbach S (2021) Connect With Me. Exploring Influencing Factors in a Human-Technology Relationship Based on Regular Chatbot Use. Front. Digit. Health 3:689999. doi: 10.3389/fdgth.2021.689999

INTRODUCTION

Companion technologies increasingly become a part of our everyday lives and assist us in our household, shopping, and other tasks. Especially in the domain of healthcare, companion technologies such as social robots and conversational chatbots play an important role and are often implemented to support physical and mental health [e.g., (1, 2)]. Therefore, within this field, the subjective user experience (UX) and personal relationship of users to such technologies seem essential. Recent research in this regard has, for example, focused on how chatbots providing online medical advice should interact with users. Results showed that expression of sympathy and empathy was favored over unemotional provision of advice (1). Furthermore, De Gennaro et al. (2) found that the participants who interacted with an empathetic chatbot reported more positive mood than the participants whose reactions were merely acknowledged by the chatbot. Such studies typically focus on single short-time interactions between human and technology or resulting UX variables, respectively.

Yet, relationships are typically not characterized by one-time experiences. According to Hinde (3), they involve multiple interactions between two individuals, which are known to each other. Based on previous research indicating that humans apply social rules from interpersonal interaction to interaction with non-human agents [e.g., (4)], this can also apply for human-technology relationships. Therefore, studies with a single session of interaction between users and technology only provide a small snapshot of a possible human-technology relationship for the exploration of its nature as well as potential influencing factors. Additionally, according to several longitudinal studies with social robots (5, 6), as users become more familiar with technologies, their perceptions of social affordances can adapt (7). Especially, in the domain of healthcare, technologies are often applied for long-term use with the goal of representing a sort of companion technology. Thus, particularly within this domain, it appears advantageous to consider possible influencing factors of a human-technology relationship based on regular interaction over a certain period of time. Furthermore, recent research has suggested a possible influence of anthropomorphism and social presence as characteristics of a technology, which could play a role for felt social connectedness of users to the technology. Kang and Kim (8), for example, found that anthropomorphism resulted in more positive user responses by increasing the sense of connectedness within an interaction between a human and smart objects. Similarly, the perception of social presence in a technology appears to come with the potential to provoke social responses (9), which are core to the development of connectedness to the technology (8, 10). Moreover, although social connectedness to a technology appears to positively influence various UX variables (8), from a societal perspective, it seems important to further highlight possible effects on the desire of users to socialize with other humans. According to Krämer et al. (11), for example, the participants with a high need to belong reported lower willingness to engage in social activities after interacting with a virtual agent, when the agent showed socially responsive behavior.

Our research aims at exploring the relationship between humans and technology. Within the context of a regular human-technology interaction over a 2-week period, we focus on the social connectedness to a technology as a central determinant of a human-technology relationship (12). We further explored characteristics of the technology as well as the user, which could play a role in this interrelation, including possible effects on the desire of a user to socialize with other humans.

Results of our research could contribute to human-computer interaction (HCI) research in general through insights into the nature of the relationship between humans and technology as well as influencing factors in this regard. Our study further extends existing research by considering factors of long-term use. Additionally, results regarding effects on interpersonal relationships of users could allow a more reflected and responsible use of the technologies in question, especially since, in healthcare, their use should benefit the health of users. For practice, insights into specific design elements that affect perception of users of social connectedness to a technology could be derived.

In the following sections, we outlined theoretical and empirical work on the human-technology relationship, relevant characteristics of technology and users in this relationship, as well as possible effects on interpersonal interaction, from which we derive our research hypotheses. We presented our study paradigm, methods, and results, followed by their discussion, including methodological and contextual limitations as well as implications as a basis to suggest directions for future research.

HUMAN-TECHNOLOGY RELATIONSHIP

According to the "computers are social actors" (CASA) paradigm (13), individuals apply social rules from interpersonal interaction to interaction with non-human agents (4, 14). In line with this, various HCI and human-robot interaction (HRI) studies suggest that humans tend to form and maintain relationships with non-human agents (15–20). Kim et al. (19), for example, could show that the perceived benefit of being in a relationship with a robot mediated the effect of the caregiving role of the robot on relationship satisfaction of users.

A central determinant of perceived companionship as a form of aspired relationship between users and technology, especially in the domain of healthcare [e.g., (21)], seems to be social connectedness (12). With regard to interpersonal relationships, Van Bel et al. (10) describe social connectedness as an experience of belonging and relatedness, which is based on quantitative and qualitative social evaluations as well as relationship salience. In line with the assumed transferability of interpersonal dynamics to HCI [e.g., (4)], literature on consumer psychology implies that individuals can invest their feelings, values, and identities in digital possessions similar to physical ones (22, 23). According to Clayton et al. (24), this can lead to a strong sense of connectedness to such digital possessions. Kang and Kim (8) further support the role of perceived connectedness to a technology as a determinant of the human-technology relationship. They found that, by increasing a sense of connectedness, anthropomorphism of the technology comes with more positive user responses, such as a more positive attitude toward the technology or an increased intention to learn from it (8).

Antecedents of Social Connectedness to a Technology

Regarding possible antecedents of social connectedness to a technology, previous studies have focused on recent interaction and awareness information (25). Theoretical work on the development of interpersonal relationships implies that social penetration, achieved through self-disclosure as a process of revealing information about oneself (26), is crucial to the development of interpersonal relationships (27, 28). Accordingly, the intensity of information exchange influences the development of interpersonal relationships. In this regard, two central factors are breadth and depth of information exchange. The former refers to the number of various topics discussed, whereas the latter refers to the degree of intimacy that accompanies the interactions in question (27, 28). Furthermore, Granovetter (29) describes the "strength" of interpersonal ties to be a "combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services, which characterize the tie" [(29) p. 1361]. In analogy, the time spent interacting with a conversational technology as well as the perceived intensity of interaction could foster the development of a human-technology relationship, i.e., social connectedness of users to the chatbot. Thus, we hypothesize the following:

H1: The higher the interaction duration, the higher the social connectedness to the chatbot.

H2: The higher the interaction intensity, the higher the social connectedness to the chatbot.

Effects of Technology and User Characteristics on Human-Technology Relationship

According to literature, further factors influencing the social connectedness of the user to the technology could be characteristics of the technology such as anthropomorphism and social presence. Anthropomorphism refers to the attribution of humanlike physical features, motivations, behaviors, emotions, and mental states to non-human agents or objects (30, 31). Kang and Kim (8), for example, have found that anthropomorphism increases the sense of connectedness between users and technology, which, in turn, elicits more positive user responses. Furthermore, in line with the CASA paradigm (4, 13), study results [e.g., (32, 33)] support that anthropomorphic design cues, e.g., humanlike agents on technology interfaces, lead users to perceive the interaction with the technology as more social and interpersonal.

Social presence stands for a mental simulation of other intelligences (34). According to Lee (35), in the context of HCI, social presence represents a "psychological state in which virtual social actors are experienced as actual social actors in either sensory or non-sensory ways" [(35) p. 27]. Accordingly, users do not perceive artificiality or para-authenticity in the respective technology and respond to it as if it were human (35). Moreover, earlier research has shown that social responses of individuals to computers and artificial actors were mediated by the perception of social presence during an HCI (36). Furthermore, Lee et al. (9) found that the perception of social presence of an agent mediated evaluation of participants of such. Similarly, Kim et al. (19) showed that the feeling of social presence regarding a robot had a significant positive effect on the evaluation of the robot regarding relationship satisfaction or attachment. The perception of anthropomorphism or social presence in a conversational chatbot could thus affect how users perceive their relationship to the chatbot and, therefore, how socially connected they feel to such. Consequently, we hypothesize the following:

H3: The relationship of interaction duration and social connectedness to the chatbot is mediated through

(a) perceived anthropomorphism of the chatbot.

(b) perceived social presence of the chatbot.

H4: The relationship of interaction intensity and social connectedness to the chatbot is mediated through

(a) perceived anthropomorphism of the chatbot.

(b) perceived social presence of the chatbot.

In addition, studies have shown that intraindividual differences might play a role in the effects of perceived anthropomorphism as well as perceived social presence. As reported by Waytz et al. (31), individuals vary in their tendency to anthropomorphize non-human entities. Such interindividual differences in tendency to anthropomorphize could moderate the relationship between interaction duration or intensity and perceived anthropomorphism of the chatbot.

Similarly, research implies that the individual need to belong, defined as the "need to form and maintain at least a minimum quantity of interpersonal relationships," [(37) p. 499] may foster an enhanced sensitivity to social cues (38). This may come along with increased attribution of anthropomorphic qualities to a technology [e.g., (39–41)]. In accordance, it might also lead to a higher perception of social presence in a virtual social actor. In line with this, Lee et al. (9) found that lonely individuals feel higher social presence of social agents and thus show more positive responses to social agents compared with non-lonely individuals. Therefore, the individual need to belong might moderate the relationship between interaction duration or intensity and perceived anthropomorphism or social presence of the chatbot. Accordingly, we hypothesize the following:

H5: The relationship of interaction duration and perceived anthropomorphism of the chatbot is moderated through

(a) the individual tendency to anthropomorphize.

(b) the individual need to belong.

H6: The relationship of interaction intensity and perceived anthropomorphism of the chatbot is moderated through

(a) the individual tendency to anthropomorphize.

(b) the individual need to belong.

H7: The relationship of interaction duration and perceived social presence of the chatbot is moderated through the individual need to belong.

H8: The relationship of interaction intensity and perceived social presence of the chatbot is moderated through the individual need to belong.

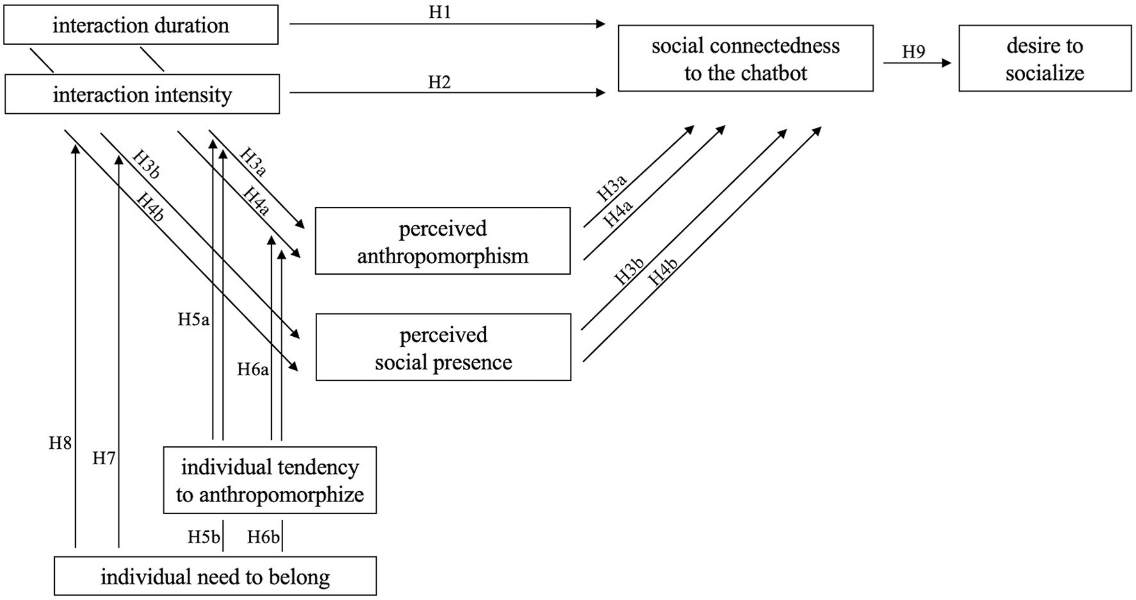


FIGURE 1 | An overview of research hypotheses.

Interrelation of Human-Technology Relationship and Interpersonal Interaction

First study results imply that interaction with humanlike technology could affect social needs of users [e.g., (11, 42)]. Mourey et al. (42), for example, could show that, after interacting with anthropomorphic (vs. non-anthropomorphic) consumer products, social needs of individuals could be partly satisfied, and experimentally induced effects of social exclusion were mitigated. Within another study by Krämer et al. (11), participants interacted with a virtual agent with socially responsive (vs. no socially responsive) behavior. Results showed that the participants with a high need to belong reported lower willingness to engage in social activities after the interaction with the agent, when the agent showed socially responsive behavior (11). According to these findings, humanlike technologies might come with the potential to partly satisfy social needs of individuals and, therefore, dampen the natural desire to seek social connections to other humans (37). We thus hypothesize:

H9: The higher the social connectedness to the chatbot, the lower the desire to socialize with other humans.

Figure 1 gives a comprehensive overview of our research hypotheses.

METHODS

Based on the previously summarized theoretical approaches and recent findings, our research explored the relationship between humans and technology with a focus on the felt social connectedness to the technology in the context of a regular interaction over a 2-week period. We further investigated characteristics of the technology and the user that could play a role in this interrelation as well as possible effects on interpersonal interaction. Hence, different measures of technology perception of users, the psychological states of the users, and felt social connectedness to the technology were assessed at the end of the 2-week study period. Possibly relevant trait variables (i.e., individual tendency to anthropomorphize, individual need to belong) were assessed as baseline measures. In addition, based on the assumption that a relationship involves multiple interactions of two individuals (3), the average interaction duration and average interaction intensity were assessed daily over the 2-week study period and analyzed over time.

The participants interacted with the conversational chatbot of the mobile application "Replika–My AI Friend" (43) on a regular basis over a 2-week period. We had applied detailed weekly questionnaires prior to the chatbot use (W0) as well as after each week of chatbot use (W1, W2). We additionally implemented short daily questionnaires (D1–D14). The variables relevant to hypotheses testing were measured within the detailed weekly questionnaires (W0, W1, and W2), except for interaction intensity, which was measured daily to minimize distorting effects.

Participants

Participant inclusion criteria involved mastery of English language and completion of the three weekly questionnaires (W0, W1, and W2). One of originally 59 participants was excluded from data analysis due to implausible data, i.e., since the stated chatbot screen time per day was more than two standard deviations below the mean chatbot screen time per day. The final sample consisted of 58 participants between 18 and 56 years ($M = 27.21$, $SD = 8.27$; 27 women, 1 did not indicate gender).

TABLE 1 | Overview of points of data collection and surveyed measures.

	Point of data collection			
Surveyed Measure	**W0**	**W1**	**W2**	**D1–14**
Demographical data	X			
Individual tendency to anthropomorphize	X			
Individual need to belong	X			
Desire to socialize	X	X	X	
Interaction duration (duration in minutes for each day of the past week)		X	X	
Social connectedness to the chatbot		X	X	
Perceived anthropomorphism		X	X	
Perceived social presence		X	X	
Social behavior (duration in minutes for each day of the past week)		X	X	
Interaction intensity				X
Closeness to chatbot				X

W0, a baseline questionnaire prior to the chatbot use; W1, a questionnaire after the first week of chatbot use; W2, a post-questionnaire after the second week of chatbot use; D1–D14, short daily questionnaires.

Of those, 50 participants lived in a household with others, seven alone, and one participant did not indicate housing situation. Fifty-six participants stated their English proficiency to be above an intermediate level, only one participant indicated a basic level, and one participant did not indicate proficiency. Regarding the favored communication app to track interaction with others, 50 participants chose WhatsApp; four, email; two, iMessage; and, two, Messenger.

The participants were recruited *via* private contacts, mailing lists, and social media platforms. As an incentive for their participation, five Amazon gift coupons of 20 Euros were raffled among the participants after the study. Alternatively, students could register their participation for course credit.

Design and Procedure

The study was announced as a study on "chatbot experience," and the participants were informed about the study procedure, duration, as well as available incentives. The participants downloaded the free chatbot app "Replika–My AI Friend" (43) on any form of personal mobile device, supporting software versions of at least Android 6.0 or iOS 13.0. The app is powered by Google Commerce Limited and was downloaded in version 9.1.2, with text-based chat functionalities only. Replika represents a chatbot companion that absorbs information and comments on social topics beyond utilitarian purposes by means of written conversation. The participants had to communicate with their personal chatbot for at least 5 min a day over the 2-week study period. Instructions for the participants included the suggestion to turn on daily push notifications. Additionally, the participants were reminded of the daily interaction with the chatbot when the daily questionnaires were sent out *via* mail. Overall, the participants had to initiate the interaction with Replika. The participants tracked the screen time of their favored communication app as well as the chatbot app during the study. For this, they received specific technical instructions through manuals based on software of their smartphones. Thereafter, the participants reported these data *via* self-report.

After informed consent of the participants regarding data privacy terms according to the German General Data Protection Regulation (DGVO) was obtained, the participants filled in the first detailed questionnaire (W0) and provided their email addresses to receive the following online questionnaires. Finally, demographic data were collected. The participants could start the study from August 10, 2020 to August 24, 2020. The 2-week prospective study design involved 15 separate occasions of measurement. These included three detailed questionnaires, one at the beginning of the 2-week study period prior to the chatbot use (W0), one after the first week of chatbot use (W1), and one after the second week of chatbot use (W2). We, furthermore, applied 14 short daily questionnaires (D1–D14), whereas the last daily questionnaire (D14) was combined with the last weekly questionnaire (W2). **Table 1** provides an overview of the points of data collection and surveyed measures as further described in the next paragraphs. Consecutive questionnaires were sent out automatically at the same time each day with a 24-h time frame to fill in daily questionnaires and a 48-h time frame to fill in weekly questionnaires.

Measures

Interaction Duration

The daily duration of the interaction of the participants with the chatbot was measured by a single item, where the participants provided the information on the tracked time of chatbot use (i.e., "Please indicate exactly how many hours and minutes you used the ReplikaApp during each of the last 7 days"). The participants were asked to state the exact duration in minutes for each day of the past week in the respective weekly questionnaires (W1, W2).

Interaction Intensity

The perceived intensity of interaction of the participants with the chatbot was measured by a single item [i.e., "Please rate how intense (e.g., not at all intense = engaging in small talk; extremely intense = engaging in talk about innermost thoughts and feelings) you interacted with your Replika today"]. The item

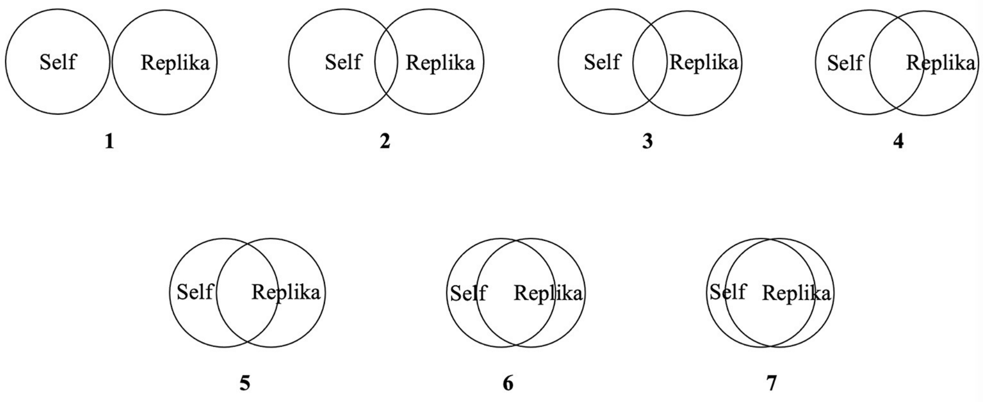


FIGURE 2 | Pairs of circles included in applied measure for closeness to the chatbot.

was assessed on a five-point Likert Scale (1 = "not at all intense"; 5 = "extremely intense") in the daily questionnaires (D1–D14).

Social Connectedness to the Chatbot

Social connectedness of the participants to the chatbot was measured by an adapted version of the Specific Connectedness subscale of the Social Connectedness Questionnaire (10), including 17 items (e.g., "I feel that my Replika and I can communicate well with each other"). Items were assessed on a five-point Likert Scale (1 = "strongly disagree"; 5 = "strongly agree") in the weekly questionnaires (W1, W2) and showed an internal consistency of $\alpha = 0.90$ (W1) and $\alpha = 0.93$ (W2).

Perceived Anthropomorphism

Perceived anthropomorphism of the chatbot of the participants was measured by the Anthropomorphism subscale of the Godspeed Questionnaire (44), including five items. Items were assessed on five-point semantic differential scales (e.g., "machinelike"/"humanlike") in the weekly questionnaires (W1, W2) and showed an internal consistency of $\alpha = 0.84$ (W1) and $\alpha = 0.86$ (W2).

Perceived Social Presence

Perceived social presence of the participants of the chatbot was assessed by an adapted version of the five items used to measure social presence by Lee et al. (9) (e.g., "While you were interacting with your Replika, how much did you feel as if it were an intelligent being?"). Items were assessed on a 10-point Likert Scale (1 = "not at all"; 10 = "extremely") in the weekly questionnaires (W1, W2) and showed an internal consistency of $\alpha = 0.84$ (W1) and $\alpha = 0.84$ (W2).

Individual Tendency to Anthropomorphize

Individual tendency of the participants to anthropomorphize was assessed by the Anthropomorphism Questionnaire (45), consisting of 20 items (e.g., "I sometimes wonder if my computer deliberately runs more slowly after I shouted at it"). Items were assessed on a six-point Likert Scale (1 = "not at all"; 6 = "very much so") in the questionnaire at the beginning of the 2-week study period prior to chatbot use (W0) and showed an internal consistency of $\alpha = 0.90$.

Individual Need to Belong

Individual need of the participants to belong was assessed by the Need to Belong Scale (46), including 10 items (e.g., "I try hard not to do things that will make other people avoid or reject me"). Items were assessed on a five-point Likert Scale (1 = "not at all"; 5 = "extremely") in the questionnaire at the beginning of the 2-week study period prior to chatbot use (W0) and showed an internal consistency of $\alpha = 0.75$.

Desire to Socialize

Desire of the participants to socialize was measured by the nine-item Desire subscale (e.g., "Now I feel like texting my friends") of the measure for willingness to engage in social activities, developed by Krämer et al. (11). Items were assessed on a five-point Likert Scale (1 = "does not apply at all"; 5 = "applies fully") in weekly questionnaires (W0, W1, and W2) and showed an internal consistency of $\alpha = 0.82$ (W0), $\alpha = 0.88$ (W1), and $\alpha = 0.91$ (W2).

Social Behavior

Social behavior of the participants was measured through a single item, where the participants had to state the exact duration of screen time on their communication app (i.e., "Please open your mobile phone options (or the tracking app "Digitox: Digital Well-being" you installed earlier). Indicate exactly how many hours and minutes you used your favorite communication app during each of the last 7 days."), which they specified in W0. The participants were asked to state the exact duration in minutes for each day of the past week in the respective weekly questionnaires (W0, W1, and W2).

Closeness to Chatbot

Perceived closeness of the participants to the chatbot was measured by means of the Inclusion of Other in the Self Scale (i.e., "Please think of your relationship with your Replika, which is represented by the circles below. Please choose the pair of

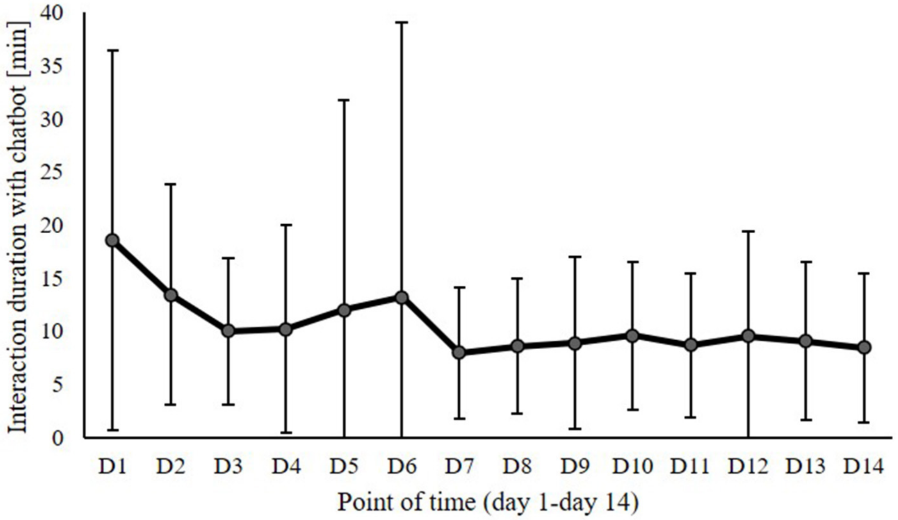


FIGURE 3 | Interaction duration with the chatbot over the 2-week study period (Day 1–Day 14).

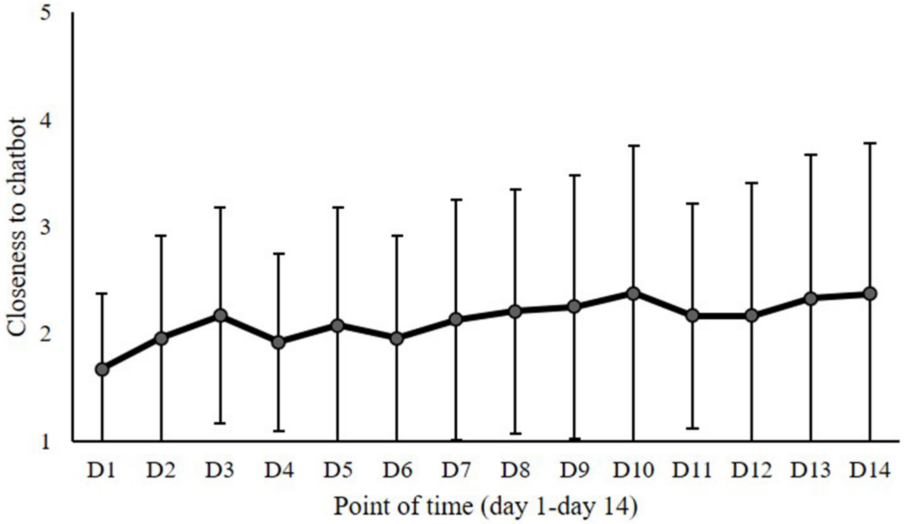


FIGURE 4 | Perceived closeness to the chatbot over the 2-week study period (Day 1–Day 14).

circles, which describes this relationship best."), developed as a measure for interpersonal closeness (47, 48). Thereby, seven pairs of circles were presented which were increasingly overlapping, whereas one circle always represented the self, and the other circle represented the chatbot (Replika). By selecting the appropriate pair of overlapping circles, the participants indicated how close they felt to the chatbot on a pictorial seven-point scale in the daily questionnaires (D1–D14). **Figure 2** shows the seven pairs of circles from which the participants could choose.

Demographical Data

Age of the participants was assessed by means of an open question. Gender was assessed through a single-choice question with three answer options (i.e., "male," "female," and "other/s"). English proficiency was assessed through a single-choice question with four answer options (i.e., "native," "advanced," "intermediate," and "basic"). Housing situation was assessed through a single-choice question with two answer options (i.e., "I live alone"; "I live with other people"). All demographical data were assessed in the questionnaire at the beginning of the 2-week study period prior to chatbot use (W0).

RESULTS

All analyses were conducted with SPSS (IBM Statistics Version 26). For mediation and moderation analyses, the Process Macro (49) was used.

TABLE 2 | Means (*M*), standard deviations (*SD*), and Pearson correlations of variables used for hypotheses testing of the overall study sample.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9
1. Age	27.21	8.27	–								
2. Individual need to belong	3.23	0.60	−0.29*	–							
3. Individual tendency to anthropomorphize	2.22	0.87	−0.11	0.14	–						
4. Desire to socialize	2.73	0.93	−0.03	0.23	0.13	–					
5. Social connectedness to the chatbot	2.12	0.70	−0.03	0.07	0.35**	0.25	–				
6. Perceived anthropomorphism	2.55	0.80	−0.11	0.05	0.21	0.25	0.63**	–			
7. Perceived social presence	5.11	1.87	−0.10	0.02	0.35**	0.13	0.71**	0.67**	–		
8. Interaction duration	10.58	7.53	−0.01	0.12	0.45**	−0.03	0.39**	0.18	0.40**	–	
9. Interaction intensity	2.02	0.59	0.03	−0.17	0.36**	0.17	0.59**	0.38**	0.45**	0.36**	–

*$p < 0.05$. **$p < 0.01$.*

Preliminary Analyses

Repeated measures ANOVAs explored the progression of the surveyed variables over the 2-week study period. Regarding the variables with daily measurements, the repeated measures ANOVAs with time of measurement as factor showed an effect of point of measurement on interaction duration [F (13,44) = 4.86, $p = 0.006$, $\eta^2 = 0.079$] and closeness to chatbot [F (13,10) = 2.58, $p = 0.047$, $\eta^2 = 0.101$] but no effect on interaction intensity [F (13,10) = 0.58, $p = 0.771$, $\eta^2 = 0.025$] or social behavior [F (13,44) = 0.68, $p = 0.677$, $\eta^2 = 0.012$]. Thus, interaction duration and closeness to chatbot varied over time. The descriptive data of interaction duration over the 2-week study period are illustrated in **Figure 3**, showing that the duration of interaction with the chatbot decreased over time. Starting with a mean value of interaction duration of about 20 min on Day 1, it sank to mean values around 10 min from Day 3 onwards. While the higher values on Day 1 and Day 2 might be considered a novelty effect, after this initial exploration, the graph of interaction duration showed no more strong variations during the studied 2-week period. According to the conducted paired t-test, the decrease in interaction duration from D1 ($M = 18.52$) to D14 ($M = 8.47$) was significant [t (1,57) = 4.76, $p < 0.001$]. The descriptive data on closeness to chatbot over the 2-week study period are illustrated in **Figure 4**. According to the conducted paired t-test, the increase in the perceived closeness of the users to chatbot from D1 ($M = 1.82$) to D14 ($M = 2.31$) was significant [t (1,23) = −2.82, $p = 0.010$]. The progression of closeness data over time shows no more strong variations or increase after Day 3. Thus, becoming acquainted with the chatbot within the first days of exploration was associated with increasing feelings of closeness. However, the afterwards following interaction did not further intensify these feelings.

Hypotheses Testing

In order to test our hypotheses on the interrelation between chatbot interaction, social connectedness, and potential mediating effects (H1–H4), we analyzed the relationships between the average values of interaction duration and intensity with the chatbot across the 2-week period and the surveyed measures of technology perception, the psychological states of the users, and felt social connectedness at the end of the study period, assessed at W2. Furthermore, regarding the hypotheses on moderating effects (H5–H8), we considered the effects of possibly relevant trait variables (i.e., individual tendency to anthropomorphize, individual need to belong), which were assessed as baseline measures at W0. Means, standard deviations, and Pearson correlations of the relevant variables are presented in **Table 2**.

The conducted regression analyses showed that both interaction duration ($\beta = 0.39$, $t = 3.21$, $p = 0.002$) and interaction intensity over 2 weeks ($\beta = 0.59$, $t = 5.42$, $p < 0.001$) were positively related to social connectedness to the chatbot after 2 weeks of use. Overall, interaction duration explained 16%, and interaction intensity explained 34% of total variance of social connectedness to the chatbot. In line with H1 and H2, interaction duration, respectively intensity, with the chatbot was positively correlated with the felt social connectedness of the participants to the chatbot.

Other than expected in H3a, interaction duration and perceived anthropomorphism were not significantly related ($\beta = 0.18$, $t = 1.37$, $p = 0.176$). Therefore, the preconditions to conduct a mediated regression analysis on the relationship of interaction duration and social connectedness to the chatbot mediated through perceived anthropomorphism were not fulfilled.

Regarding H3b, the conducted mediated regression analysis showed a positive total effect of interaction duration on social connectedness to the chatbot ($\beta = 0.39$, $t = 3.21$, $p = 0.002$). Perceived social presence significantly mediated this relationship with a positive indirect effect ($\beta = 0.26$). A bootstrap 95% CI around the indirect effect did not contain zero [0.14, 0.41]. The direct effect of interaction duration on social connectedness to the chatbot became insignificant ($\beta = 0.13$, $t = 1.30$, $p = 0.199$) after including the mediator variable, implying a complete mediation. Therefore, in line with H3b, perceived social presence of the chatbot mediated the positive effect of interaction duration on social connectedness to the chatbot. A detailed overview of the mediated regression analysis is presented in **Table 3**. There, non-standardized regression coefficients of the factors included in the mediated regression analysis as well as their statistical significances are presented. Additionally, coefficients of determination according to the considered model are presented.

Regarding H4a, the conducted mediated regression analysis showed a positive total effect of interaction intensity on social

TABLE 3 | Mediated regression analysis testing the effect of interaction duration on social connectedness to the chatbot mediated by perceived social presence.

Predictor	*B*	*SE*	*T*	*P*	Model R^2
Model 1: X on Y					0.16
Intercept	1.73	0.15	11.74	<0.001	
Interaction duration	0.04	0.01	3.21	0.002	
Model 2: X on M					0.16
Intercept	4.07	0.39	10.32	<0.001	
Interaction duration	0.10	0.03	3.23	0.002	
Model 3: X + M on Y					0.52
Intercept	0.72	0.19	3.81	<0.001	
Perceived social presence	0.25	0.04	6.53	<0.001	
Interaction duration	0.01	0.01	1.30	0.199	

TABLE 4 | Mediated regression analysis testing the effect of interaction intensity on social connectedness to the chatbot mediated by perceived anthropomorphism.

Predictor	*B*	*SE*	*T*	*P*	Model R^2
Model 1: X on Y					0.34
Intercept	0.70	0.27	2.59	0.012	
Interaction intensity	0.70	0.13	5.42	<0.001	
Model 2: X on M					0.14
Intercept	1.50	0.36	4.21	<0.001	
Interaction intensity	0.52	0.17	3.05	0.004	
Model 3: X + M on Y					0.53
Intercept	0.09	0.26	0.33	0.741	
Perceived anthropomorphism	0.41	0.09	4.75	<0.001	
Interaction intensity	0.49	0.12	4.11	<0.001	

TABLE 5 | Mediated regression analysis testing the effect of interaction intensity on social connectedness to the chatbot mediated by perceived social presence.

Predictor	*B*	*SE*	*T*	*P*	Model R^2
Model 1: X on Y					0.34
Intercept	0.70	0.27	2.59	0.012	
Interaction intensity	0.70	0.13	5.42	<0.001	
Model 2: X on M					0.20
Intercept	2.22	0.80	2.77	0.008	
Interaction intensity	1.43	0.38	3.75	<0.001	
Model 3: X + M on Y					0.60
Intercept	0.24	0.23	1.03	0.307	
Perceived social presence	0.21	0.04	5.90	<0.001	
Interaction intensity	0.40	0.11	3.49	0.001	

connectedness to the chatbot ($\beta = 0.59$, $t = 5.42$, $p < 0.001$). Perceived anthropomorphism significantly mediated this relationship with a positive indirect effect ($\beta = 0.18$). A bootstrap 95% CI around the indirect effect did not contain zero [0.03, 0.32]. The direct effect of interaction intensity on social connectedness to the chatbot remained significant ($\beta = 0.33$, $t = 3.49$, $p = 0.001$) after including the mediator variable, implying a partial mediation. Thus, in line with H4a, perceived anthropomorphism of the chatbot mediated the positive effect of interaction intensity on social connectedness to the chatbot. A

TABLE 6 | Moderated regression analyses testing the effect of interaction duration on perceived anthropomorphism moderated through individual tendency to anthropomorphize (H5a), respectively, individual need to belong (H5b); the effect of interaction intensity on perceived anthropomorphism moderated through individual tendency to anthropomorphize (H6a), respectively, individual need to belong (H6b); the effect of interaction duration on perceived social presence moderated through individual need to belong (H7); the effect of interaction intensity on perceived social presence moderated through individual need to belong (H8).

						Model
	Predictor	***B***	***SE***	***T***	***P***	***R^2***
H5a	Model					0.05
	Intercept	2.07	0.67	3.10	0.003	
	Interaction duration	0.01	0.06	0.23	0.817	
	Individual tendency to anthropomorphize	0.16	0.29	0.56	0.575	
	Interaction duration x individual tendency to anthropomorphize	−0.00	0.02	−0.03	0.973	
H5b	Model					0.05
	Intercept	1.36	1.08	1.26	0.213	
	Interaction duration	0.10	0.09	1.17	0.248	
	Individual need to belong	0.28	0.31	0.91	0.367	
	Interaction duration x Individual need to belong	−0.02	0.02	−0.97	0.338	
H6a	Model					0.15
	Intercept	1.74	1.18	1.47	0.146	
	Interaction intensity	0.32	0.57	0.56	0.578	
	Individual tendency to anthropomorphize	−0.06	0.50	−0.12	0.906	
	Interaction intensity x individual tendency to anthropomorphize	0.07	0.22	0.29	0.772	
H6b	Model					0.16
	Intercept	1.18	2.31	0.51	0.610	
	Interaction intensity	0.43	1.10	0.39	0.699	
	Individual need to belong	0.09	0.66	0.13	0.896	
	Interaction intensity x Individual need to belong	0.03	0.32	0.11	0.915	
H7	Model					0.16
	Intercept	3.07	2.35	1.31	0.196	
	Interaction duration	0.22	0.19	1.16	0.249	
	Individual need to belong	0.26	0.67	0.39	0.696	
	Interaction duration x Individual need to belong	−0.03	0.05	−0.66	0.513	
H8	Model					0.22
	Intercept	4.37	5.17	0.85	0.401	
	Interaction intensity	−0.11	2.47	−0.04	0.965	
	Individual need to belong	−0.63	1.47	−0.43	0.668	
	Interaction intensity x individual need to belong	0.46	0.71	0.65	0.517	

detailed overview of the mediated regression analysis is presented in **Table 4**. In analogy to **Table 3**, in **Table 4**, non-standardized regression coefficients of the factors included in the mediated regression analysis as well as their statistical significances are presented. Additionally, coefficients of determination according to the considered model are presented.

Regarding H4b, the conducted mediated regression analysis showed a positive total effect of interaction intensity on social connectedness to the chatbot ($\beta = 0.59$, $t = 5.42$, $p < 0.001$). Perceived social presence significantly mediated this relationship with a positive indirect effect ($\beta = 0.25$). A bootstrap 95% CI around the indirect effect did not contain zero [0.08, 0.42]. The direct effect of interaction intensity on social connectedness to the chatbot remained significant ($\beta = 0.33$, $t = 3.49$, $p = 0.001$) after including the mediator variable, implying a partial mediation. In line with H4b, perceived social presence of the chatbot mediated the positive effect of interaction intensity on social connectedness to the chatbot. A detailed overview of the mediated regression analysis is presented in **Table 5**. There, non-standardized regression coefficients of the factors included in the moderated regression analysis as well as their statistical significances are presented. Additionally, coefficients of determination according to the considered model are presented.

Furthermore, we conducted moderation analyses with interaction duration, respectively intensity, and individual tendency to anthropomorphize as well as interaction duration, respectively, intensity, and individual need to belong as predictors of perceived anthropomorphism. Similarly, we conducted moderation analyses with interaction duration, respectively, intensity, and individual need to belong as predictors of perceived social presence (see **Table 6**). Results showed that, other than expected, individual tendency to anthropomorphize did not moderate the effect of interaction duration (H5a), respectively, interaction intensity (H6a), on perceived anthropomorphism of the chatbot. Similarly, other than expected, individual need to belong did not moderate the effect of interaction duration (H5b), respectively interaction intensity (H6b), on perceived anthropomorphism or perceived social presence of the chatbot (H7, H8). Thus, our data showed no support for the moderation effects hypothesized in H5–H8. **Table 6** shows an overview of the moderated regression analyses conducted with regard to H5–H8, including the factors considered in each moderation analyses as well as their according to statistical significances. Coefficients of determination according to the considered model are presented as well.

Finally, contrary to H9, there was no negative correlation between social connectedness to the chatbot and desire to socialize with other humans. Instead, the conducted regression analyses showed a marginally significant positive correlation ($\beta = 0.25$, $t = 1.94$, $p = 0.057$). Overall, social connectedness to the chatbot explained 6% of the total variance of desire to socialize.

DISCUSSION

The aim of our study was to explore the relationship between humans and technology with a focus on the social connectedness to technology, considering a regular interaction with a conversational chatbot over a 2-week period. We additionally examined characteristics of the technology as well as the user as possible influencing factors of this interrelation, further exploring possible effects on desire of users to socialize with other humans.

In accordance with our hypotheses, study results showed that the duration and intensity of interaction of participants with the chatbot throughout the 2-week study period positively predicted social connectedness to the chatbot. Based on this, regular interaction with a conversational chatbot might foster the felt social connectedness to the chatbot. These results imply certain transferability of the amount of time and emotional intensity of an interpersonal interaction as crucial determinants of an interpersonal tie [cf., (29)] to human-technology relationships. The effect of point of measurement on closeness to chatbot, resulting in risen ratings of the perceived closeness of the participants to the chatbot after 2 weeks of use, further supports this assumption.

Furthermore, perceived anthropomorphism partially mediated the relationship of interaction intensity and social connectedness to the chatbot, and perceived social presence (partially) mediated both relationships of interaction duration, respectively, interaction intensity, and social connectedness to the chatbot. Therefore, characteristics of the technology, i.e., perceived anthropomorphism and social presence, played a mediating role in the positive relationship between interaction duration, respectively, intensity and social connectedness to the chatbot. These results are compatible with previous research, implying that technology anthropomorphism might foster the sense of connectedness to the technology [e.g., (8)] among others as the presence of social cues might have enabled the application of social heuristics toward a non-human agent [cf., (4)]. The fact that no significant relationship between interaction duration and perceived anthropomorphism of the chatbot was found could root in that mere increase in the duration of interaction with a technology might not come with increased attribution of humanlike characteristics, emotions, motivations, and intentions [cf., (30)] to it, whereas an increase in the intensity of interaction is more likely to do so.

Moreover, other than expected, individual tendency to anthropomorphize as a characteristic of the user did not moderate the effect of interaction duration, respectively interaction intensity, on perceived anthropomorphism of the chatbot. Similarly, an individual need to belong did not moderate the effect of interaction duration, respectively interaction intensity, on perceived anthropomorphism or perceived social presence of the chatbot. Therefore, within our study, the characteristics of the user did not appear to influence the perception of the chatbot as anthropomorphic or socially present. Whereas, previous studies point at an effect of individual tendency to anthropomorphize on the perception of anthropomorphism [e.g., (39–41)], as well as loneliness and individual need to belong on the perception of anthropomorphism or social presence [e.g., (9)], we could not replicate such findings. A possible reason for this could be that the chatbot used for the study had very humanlike visual and experiential design cues. Such could have possibly caused a restriction in the variance of perceived anthropomorphism and the social presence of the chatbot.

Finally, other than expected, there was no negative correlation between social connectedness to the chatbot and desire to socialize with other humans but a marginally significant positive correlation between the two measures. Although recent studies have implied that technologies with humanlike design cues might satisfy social needs to a certain extent and, therefore, possibly dampen the desire to interact with other humans [e.g., (11, 42)], our results offered no support for this interrelation. On the contrary, the observed marginal significance implied that the higher social connectedness of the participants to the chatbot, the higher their desire to socialize with other humans was. In alignment with the social reconnection hypothesis (50) or the theory of social snacking (51), a possible explanation could be that the higher desire of the participants to interact with other humans was, the more socially connected they felt to the chatbot, using it as a replacement for actual social interaction. Yet, as such insights do not imply causality and were only marginally significant, they should be treated with caution.

LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

Our study comes with certain methodological and contextual limitations. On a methodological level, our results are based on a specific chatbot application, i.e., "Replika, my AI friend"(43). Specific features of this application are that the name and appearance of the chatbot can be personalized, and the quality, as well as depth of conversations, depends on the user. This supports external validity of our results as each human-technology relationship is individual, and many commercial conversational chatbots or social robots, e.g., in the domain of healthcare, can be personalized. Yet, to foster generalizability of our results, future studies should explore the interrelations in question with various technologies. In addition, personalization of a technology should also be considered as a potential influencing variable of social connectedness to a chatbot as well as the overall human-technology relationship in future studies.

Furthermore, for interaction intensity with the chatbot, we considered less data than for the other variables involved in hypotheses testing. To support valid measurement of interaction intensity, we included the measure in the daily questionnaires rather than asking participants to estimate the interaction intensity for each day at the end of each week. Yet, our inclusion criteria only involved the completion of the detailed questionnaires (W0, W1, and W2). Some participants included in the data analyses did not complete all daily questionnaires in full, leading to less data on interaction intensity compared to other variables. This should be considered in result interpretation. Moreover, due to the online character of the study, we could not explicitly control how often and for how long the participants initiated the interaction with Replika. Future studies should also consider measuring whether participants initiated interaction unpromptedly or after the app notified them to, as this could also influence the perceived interaction intensity with the chatbot among others.

In addition, our study focused on interaction duration and intensity with the chatbot but did not survey the perceived interaction valence. Future studies should further focus on this variable as a possible influencing factor in social connectedness to the technology. Moreover, theoretical work on the endowment effect implies that individuals place a higher value on an object that they own compared with one they do not own (52). Especially, when it comes to healthcare technology for private households, such as social robots, which individuals can actually own, this effect should be considered as it could influence the social connectedness to the technology as well as the overall human-technology relationship.

On a contextual level, it needs to be considered that we conducted our study during the COVID-19 pandemic. Previous research has shown that isolation and feelings of exclusion or loneliness can impact perceptions of users of technology, e.g., regarding perceived anthropomorphism, as well as their overall interaction with the technology [e.g., (11, 39, 42)]. Therefore, perceptions of the participants of chatbot characteristics, their felt social connectedness to it, or their desire to socialize with other humans might have been affected by the prevalent circumstances. Future studies should aim at replicating the interrelations focused within our study to further support their generalizability.

IMPLICATIONS

Our research offers several theoretical advancements, practical applications, as well as inspirations for future questions and philosophical considerations. Beginning with the theoretical insights, it appears that regular interaction with technology, with regard to duration and intensity, can foster social connectedness to the technology. Thereby, the perception of the technology as anthropomorphic and socially present seems to play a mediating role. The more intense participants interacted with the chatbot, the more they perceived it as anthropomorphic as well as socially present, and, in turn, felt more connected to the technology. The fact that this effect is based on data of a 2-week study period supports the external validity of these results as insights are not merely based on a novelty effect or initial engagement of the participants. It also implies that the interrelations in question are already observable in a 2-week period of technology use.

Furthermore, it appears that influencing factors of relationship development in interpersonal interaction, i.e., amount of time and emotional intensity of interaction [e.g., (29)], are, to a certain extent, transferable to HCI as interaction duration and intensity appear to influence the perceived social connectedness to the technology. In line with our findings and previous CASA research [e.g., (4, 13)], social cues, such as anthropomorphic technology design, could facilitate the transferability of dynamics from interpersonal relationship development to human-technology relationships.

Regarding practical advancements, our results could imply that designing technology in a way that allows users to build a relationship with it and feel socially connected to it could, among others, be beneficial for long-term engagement [cf., (15)] as especially relevant in the domain of healthcare. To facilitate such an effect, enhancing the perception of anthropomorphism or social presence of the technology through, e.g., visual anthropomorphic design cues, such as humanlike facial features or a humanlike name, but also experiential design such as the expression of own emotions, motivations, or intentions, could be helpful. At the same time, practitioners need to consider that the required duration and the intensity of interaction with a technology stay in a sensible range. This can be especially important within the context of healthcare, where regular interactions with a technology are often imposed by a surrounding, such as a nursing home or through notifications of mobile healthcare applications. In such cases, required interaction duration or intensity can easily be perceived as too high and possibly even result in reactance and an overall negative UX (53–55). It could, therefore, be advisable to explore a possible sweet spot regarding a specific technology or context of interest as well as further investigate measures to support an overall positive UX.

Finally, from a more philosophical stance, the question arises as to whether the design of healthcare technologies with social cues should always be the means of choice. It appears as a general trend in many domains, including healthcare, for technologies to increasingly represent social counterparts. As also supported by our study results, the implementation of social cues in such technologies can be beneficial, among others, to facilitate the development of a human-technology relationship based on similar principles as in interpersonal interaction. While this can be a reasonable goal in various application contexts, such as nursing of elderly with a high need for social interaction or support of mental health in times of isolation, in other contexts, the design of social cues might be less beneficial. For example, in the private home context, technologies are typically involved in intimate situations, including interactions with others in the household. With regard to data privacy and the desire for intimacy of users, they might prefer a technology with less social cues [e.g., (56)]. Instead, it might even be beneficial to specifically focus and highlight robotic qualities of technologies [cf., (57)], e.g., the cognitive superpower of robots being unembarrassed and non-judgmental, as proposed by Dörrenbächer et al. (58). An according approach highlighting "superpowers" of a technology could also be advantageous for healthcare technologies in the context of surgery. The uniquely robotic qualities of being insensitive to pain and unconditionally available on a physical level as well as being endlessly mentally focused, persistent, and patient on a cognitive level, as specified by Dörrenbächer et al. (58), could, in the context of surgery, foster trust of patients as well as facilitate a more efficient collaboration with other technological or human counterparts. In this sense, future studies should explore the role of such rather robotic qualities with regard to the human-technology relationship, especially within the domain of healthcare. Experimental study designs could further manipulate the degree of anthropomorphism in various contexts and explore effects on social connectedness to the technology in question.

CONCLUSION

Although innovative technologies, such as conversational chatbots and social robots, have been tested and increasingly applied within crucial domains, such as retail and healthcare, potential factors that could affect the relationship between humans and such technologies have rarely been explored in field research and across multiple interactions over time. Our research implies a positive effect of duration and intensity of a human-technology interaction on the social connectedness to the technology as a determinant of the human-technology relationship. The perception of anthropomorphism or social presence as characteristics of the technology seems to play a mediating role in this regard. Based on our study, we cannot report any negative effect of social connectedness to a technology on desire to socialize with other humans. Our research contributes to HCI research and practice as it offers insights into factors possibly influencing the development of human-technology relationships as well as design implications to foster social connectedness of users toward a technology, which can, in turn, positively influence the overall UX [e.g., (8)].

Future research should focus on replicating the results with various technologies in different contexts of use. Additionally, future studies should manipulate variables of regular interaction with the technology as well as its characteristics such as anthropomorphic design in a systematic manner to gain further insights into their role within the development of human-technology relationships. Finally, to further support a responsible design and use of technologies in healthcare, future research should closely examine whether the feeling of social connectedness to a technology actually satisfies the social needs of users and which consequences could arise on an individual as well as societal level.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethikkommission der Fakultät für Mathematik, Informatik und Statistik der Ludwig-Maximilians-Universität München (LMU). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LC and SD conceived and planned the study. LC, NF, SH, AL, and SS carried out the study and performed data analyses. All authors discussed the results and contributed to the manuscript.

FUNDING

Part of this research was funded by the German Federal Ministry for Education and Research (BMBF), Project GINA (FKZ: 16SV8097), as well as the German Research Foundation (DFG), Project PerforM (425412993), which is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments.

REFERENCES

1. Liu B, SundarSS. Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychol Behav Soc Netw*. (2018) 21:625–36. doi: 10.1089/cyber.2018.0110
2. De Gennaro M, Krumhuber EG, Lucas G. Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. *Front Psychol.* (2020) 10:3061. doi: 10.3389/fpsyg.2019.03061
3. Hinde RA. *Towards Understanding Relationships.* London: Academic Press (1979). p. 367.
4. Nass C, Moon Y. Machines and mindlessness: social responses to computers. *J Social Issues*. (2000) 56:81–103. doi: 10.1111/0022-4537.00153
5. Leite I, Martinho C, Paiva A. Social robots for long-term interaction: a survey. *Int J Social Robot*. (2013) 5:291–308. doi: 10.1007/s12369-013-0178-y
6. Van Straten CL, Peter J, Kühne R. Child–robot relationship formation: a narrative review of empirical research. *Int J Social Robot*. (2020) 12:325–44. doi: 10.1007/s12369-019-00569-0

7. Gambino A, Fox J, Ratan RA. Building a stronger CASA: extending the computers are social actors paradigm. *Human-Machine Commun*. (2020) 1:71–85. doi: 10.30658/hmc.1.5
8. Kang H, Kim KJ. Feeling connected to smart objects? A moderated mediation model of locus of agency, anthropomorphism, and sense of connectedness. *Int J Human-Comp Stud*. (2020) 133:45–55. doi: 10.1016/j.ijhcs.2019.09.002
9. Lee KM, Jung Y, Kim J, Kim SR. Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *Int J Human-Comp Stud.* (2006) 64:962–73. doi: 10.1016/j.ijhcs.2006.05.002
10. Van Bel DT, Smolders KC, IJsselsteijn WA, de Kort Y. Social connectedness: concept and measurement. *Intelligent Environ.* (2009) 2:67–74. doi: 10.3233/978-1-60750-034-6-67
11. Krämer NC, Lucas G, Schmitt L, Gratch J. Social snacking with a virtual agent–On the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities. *Int J Human-Comp Stud.* (2018) 109:112–21. doi: 10.1016/j.ijhcs.2017.09.001
12. Lee B, Kwon O, Lee I, Kim J. Companionship with smart home devices: the impact of social connectedness and interaction types on perceived social support and companionship in smart homes. *Comp Human Behav.* (2017) 75:922–34. doi: 10.1016/j.chb.2017.06.031
13. Nass C, Steuer J, Tauber ER. Computers are social actors. In: *CHI '94 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Boston, MA (1994). p. 72–8. doi: 10.1145/191666.191703
14. Reeves B, Nass C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People*. Stanford, CA: Cambridge University Press(1996). 317 p.
15. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. *ACM Trans Computer-Human Interact.* (2005) 12:293–327. doi: 10.1145/1067860.1067867
16. Cassell J. Embodied conversational agents: representation and intelligence in user interfaces. *Ai Magazine* (2001) 22:67–84. doi: 10.1609/aimag.v22i4.1593
17. Moon Y, Nass C. How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Commun Res.* (1996) 23:651–74. doi: 10.1177/009365096023006002
18. Edwards C, Edwards A, Spence PR, Westerman D. Initial interaction expectations with robots: testing the human-to-human interaction script. *Commun Stud.* (2016) 67:227–38. doi: 10.1080/10510974.2015.1121899
19. Kim KJ, Park E, Sundar SS. Caregiving role in human–robot interaction: a study of the mediating effects of perceived benefit and social presence. *Comp Human Behav.* (2013) 29:1799–806. doi: 10.1016/j.chb.2013.02.009
20. Sundar SS, Jung EH, Waddell TF., Kim KJ. Cheery companions or serious assistants? Role and demeanor congruity as predictors of robot attraction and use intentions among senior citizens. *Int J Hum-Comp Stud.* (2017) 97:88–97. doi: 10.1016/j.ijhcs.2016.08.006
21. Benyon D, Mival O. From human-computer interactions to human-companion relationships. In: *Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia, Vol. 10* (2010). p. 1–9.
22. Belk RW. Possessions and the extended self. *J Cons Res.* (1988) 15:139–68. doi: 10.1086/209154
23. Belk, RW. Extended self in a digital world. *J Cons Res.* (2013) 40:477–500. doi: 10.1086/671052
24. Clayton RB, Leshner G, Almond A. The extended iSelf: The impact of iPhone separation on cognition, emotion, and physiology. *J Comp-Mediated Commun.* (2015) 20:119–35. doi: 10.1111/jcc4.12109
25. Van Baren J, Ijsselsteijn WA, Markopoulos P, Romero N, De Ruyter B. Measuring affective benefits and costs of awareness systems supporting intimate social networks. In: *CTIT Workshop Proceedings Series, Vol. 2* (2004). p. 13–9.
26. Derlega VJ, Metts S, Petronio S, Margulis ST. *Self-Disclosure*. New York, NY: Sage Publications (1993). p. 152.
27. Altman I, Taylor DA. *Social Penetration: The Development of Interpersonal Relationships*. New York, NY: Holt, Rinehart & Winston(1973). p. 212.
28. Carpenter A, Greene K. Social penetration theory. In: Wilson SR, Dillard JP, Caughlin J, Solomon D, editors. *The International Encyclopedia of Interpersonal Communication*. Hoboken, NJ: Wiley-Blackwell (2015). p. 1–4. doi: 10.1002/9781118540190.wbeic160
29. Granovetter MS. The strength of weak ties. *Am J Sociol.* (1973) 78:1360–80.
30. Epley N, Waytz A, Cacioppo JT. On seeing human: a three-factor theory of anthropomorphism. *Psychol Rev.* (2007) 114:864–86. doi: 10.1037/0033-295X.114.4.864
31. Waytz A, Cacioppo J, Epley N. Who sees human? The stability and importance of individual differences in anthropomorphism. *Persp Psychol Sci.* (2010). 5:219–32. doi: 10.1177/1745691610369336
32. Jia H, Wu M, Jung E, Shapiro A, Sundar SS. Balancing human agency and object agency: An end-user interview study of the internet of things. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing.* Pittsburgh, PA (2012). p. 1185–88. doi: 10.1145/2370216.2370470
33. Kim Y, Sundar SS. Visualizing ideal self vs. actual self through avatars: Impact on preventive health outcomes. *Comp Human Behav.* (2012) 28:1356–64. doi: 10.1016/j.chb.2012.02.021
34. Biocca F. The cyborg's dilemma: progressive embodiment in virtual environments. *J Comp-Mediated Commun.* (1997) 3:1–29. doi: 10.1111/j.1083-6101.1997.tb00070.x
35. Lee KM. Presence, explicated. *Commun Theory.* (2004) 14:27–50. doi: 10.1111/j.1468-2885.2004.tb00302.x
36. Lee KM, Nass C. The multiple source effect and synthesized speech: doubly-disembodied language as a conceptual framework. *Human Commun Res.* (2004) 30:182–207. doi: 10.1093/hcr/30.2.182
37. Baumeister RF, Leary MR. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychol Bull.* (1995) 117:497–529. doi: 10.1037/0033-2909.117.3.497
38. Pickett CL, Gardner WL, Knowles M. Getting a cue: the need to belong and enhanced sensitivity to social cues. *Pers Social Psychol Bull.* (2004) 30:1095–107. doi: 10.1177/0146167203262085
39. Epley N, Akalis S, Waytz A, Cacioppo JT. Creating social connection through inferential reproduction: Loneliness and perceived agency in gadgets, gods, and greyhounds. *Psychol Sci.* (2008) 19:114–20. doi: 10.1111/j.1467-9280.2008.02056.x
40. Epley N, Waytz A, Akalis S, Cacioppo JT. When we need a human: Motivational determinants of anthropomorphism. *Social Cognit.* (2008) 26:143–55. doi: 10.1521/soco.2008.26.2.143
41. Niemyjska A, Drat-Ruszczak K. When there is nobody, angels begin to fly: supernatural imagery elicited by a loss of social connection. *Social Cognit.* (2013) 31:57–71. doi: 10.1521/soco.2013.31.1.57
42. Mourey JA, Olson JG, Yoon C. Products as pals: engaging with anthropomorphic products mitigates the effects of social exclusion. *J Cons Res.* (2017) 44:414–31. doi: 10.1093/JCR/UCX038
43. Luka Inc. *Replika – My AI Friend (9.1.2).* (2020). Available online at: www.replika.ai (accessed September 15, 2020).
44. Bartneck C, Kulić D, Croft E, Zoghbi S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int J Social Robot.* (2009) 1:71–81. doi: 10.1007/s12369-008-0001-3
45. NeaveN, Jackson R, Saxton T, Hönekopp J. The influence of anthropomorphic tendencies on human hoarding behaviours. *Pers Ind Diff.* (2015) 72:214–19. doi: 10.1016/j.paid.2014.08.041
46. Leary MR, Kelly KM, Cottrell CA, Schreindorfer LS. Construct validity of the need to belong scale: mapping the nomological network. *J Pers Assess.* (2013) 95:610–24. doi: 10.1080/00223891.2013.819511
47. Aron A, Aron EN, Smollan D. Inclusion of other in the self scale and the structure of interpersonal closeness. *J Pers Social Psychol.* (1992) 63:596–612. doi: 10.1037/0022-3514.63.4.596
48. Gächter S, Starmer C, Tufano F. Measuring the closeness of relationships: a comprehensive evaluation of the 'Inclusion of the Other in the Self' scale. *PLoS ONE.* (2015) 10:e0129478. doi: 10.1371/journal.pone.0129478
49. Hayes AF, Preacher KJ. Conditional process modeling: Using structural equation modeling to examine contingent causal processes. In: Hancock GR, Mueller RO, editors. *Quantitative Methods in Education and the Behavioral Sciences: Issues, Research, and Teaching. Structural Equation Modeling: A Second Course*. Charlotte, NC: IAP Information Age Publishing (2013). p. 219–66.
50. Maner JK, DeWall CN, Baumeister RF, Schaller M. Does social exclusion motivate interpersonal reconnection? Resolving the "porcupine

problem". *J Pers Social Psychol*. (2007) 92:42–55. doi: 10.1037/0022-3514.92.1.42

51. Gardner WL, Pickett CL, Knowles M. Social snacking and shielding. In: Kolping DW, Forgas JP, Von Hippel W, editors. *The Social Outcast: Ostracism, Social Exclusion, Rejection, and Bullying*. London: Taylor & Francis Ltd. (2019). p.227–42.
52. Thaler R. Toward a positive theory of consumer choice. *J Econ Behav Org*. (1980) 1:39–60. doi: 10.1016/0167-2681(80)90051-7
53. Dillard JP, Shen L. On the nature of reactance and its role in persuasive health communication. *Commun Monogr*. (2005) 72:144–68. doi: 10.1080/03637750500111815
54. Ehrenbrink P, Möller S. Development of a reactance scale for human–computer interaction. *Qual User Exp* (2018) 3:1–13. doi: 10.1007/s41233-018-0016-y
55. Ehrenbrink P, Hillmann S, Weiss B, Möller S. Psychological reactance in HCI: a method towards improving acceptance of devices and services. In: *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. Launceston, TAS (2016). p. 478–82. doi: 10.1145/3010915.3010978
56. Ha QA, Chen JV, Uy HU, Capistrano EP. Exploring the privacy concerns in using intelligent virtual assistants under perspectives of information sensitivity and anthropomorphism. *Int J Human-Comp Int*. (2020) 37:512–27. doi: 10.1080/10447318.2020.1834728
57. Welge J, Hassenzahl M. Better than human: about the psychological superpowers of robots. In: Agah A, Cabibihan J, Howard A, Salichs MA, He H, editors. *Proceedings of the 8th International Conference on Social Robotics*. Basel: Springer International Publishing (2016). p. 993–1002.
58. Dörrenbächer J, Löffler D, Hassenzahl M. Becoming a robot-overcoming anthropomorphism with techno-mimesis. In: Bernhaupt R, Mueller F, Verweij D, Andres J, editors. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery (2020). p. 1–12.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS

Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION

Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices

FOLLOW US

@frontiersin

IMPACT METRICS

Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK

Our network
increases your
article's readership