Challenges and alternatives to evaluation methods and regulation approaches of medical apps as mobile medical devices: An international and multidisciplinary focus group discussion

Laura Maaß, Robert Hrynyschyn, Martin Lange, Alexandra Löwe, Kathrin Burdenski, Kaley Butten, Sebastian Vorberg, Mariam Hachem, Aldo Gorga, Vittorio Grieco, Vincenzo Restivo, Giuseppe Vella, Marlien Varnfield, Felix Holl

Submitted to: Journal of Medical Internet Research
on: November 28, 2023

Table of Contents

Challenges and alternatives to evaluation methods and regulation approaches of medical apps as mobile medical devices: An international and multidisciplinary focus group discussion

Laura Maaß[1, 2, 3] BA, MA; Robert Hrynyschyn[2, 4] BSc, MSc; Martin Lange[5] Prof Dr; Alexandra Löwe[5] Dipl, MA; Kathrin Burdenski[6]; Kaley Butten[7] PhD; Sebastian Vorberg[8, 9] LLM; Mariam Hachem[10, 11] BA, BSc, MPH; Aldo Gorga[12] Dr med; Vittorio Grieco[13] MD; Vincenzo Restivo[14] Prof Dr; Giuseppe Vella[15] MD; Marlien Varnfield[7] BSc, MSc, PhD; Felix Holl[2, 16] BSc, MSc, FAMIA, PhD

[1]University of Bremen SOCIUM - Research Center on Inequality and Social Policy Departement Health, Long Term Care and Pensions Bremen DE
[2]Leibniz ScienceCampus Digital Public Health Bremen Bremen DE
[3]European Public Health Association (EUPHA) Digital Health Section Utrecht NL
[4]Charite? – Universita?tsmedizin Berlin, corporate member of Freie Universita?t Berlin and Humboldt-Universita?t zu Berlin Institute of Health and Nursing Science Berlin DE
[5]IST University of Applied Sciences Department of Fitness & Health Düsseldorf DE
[6]Loughborough University School of Sport, Exercise and Health Sciences Loughborough GB
[7]Australian eHealth Research Centre (CSIRO) Brisbane AU
[8]Vorberg.law / QuR.digital Hamburg DE
[9]University of Melbourne Faculty of Dentistry, Medicine and Health Sciences Department of Medicine, Austin Health Melbourne AU
[10]University of Melbourne Faculty of Dentistry, Medicine and Health Sciences Australian Centre for Accelerating Diabetes Innovations Melbourne AU
[11]University of Turin Department of Sciences of Public Health and Pediatrics Turin IT
[12]University of Catania Department of Medical, Surgical Sciences and Advanced Technologies “GF Ingrassia” Catania IT
[13]University Kore of Enna Department of Medicine Enna IT
[14]University of Palermo Department of Health Promotion, Maternal and Infant Care, Internal Medicine and Medical Specialties (PROMISE) “G. D’Alessandro” Palermo IT
[15]Neu-Ulm University of Applied Sciences DigiHealth Institute Neu-Ulm DE

Corresponding Author:
Laura Maaß BA, MA
University of Bremen
SOCIUM - Research Center on Inequality and Social Policy
Departement Health, Long Term Care and Pensions
Mary-Somerville-Straße 3
Bremen
DE

Abstract

Background: The rapid proliferation of medical apps has transformed the healthcare landscape by giving patients and providers unprecedented access to personalized health information and services. However, concerns regarding the effectiveness and safety of apps have raised questions regarding the efficacy of randomized controlled trials (RCTs) for their evaluation and as a requirement for their regulation as a mobile medical device.

Objective: This study addresses this issue by investigating alternative methods besides RCTs for evaluating and regulating medical apps.

Methods: Using a qualitative approach, a focus group study with 46 international and multidisciplinary public health experts, split into three groups, was conducted at the 17th World Congress on Public Health in May 2023 in Rome (Italy) to gather in-depth insights into alternative approaches to evaluation and regulation. We conducted a policy analysis of the current regulation of medical apps as mobile medical devices for the four most represented countries in the workshop: Italy, Germany, Canada, and Australia. We developed a logic model that combines the evaluation and regulation domains based on these findings.

Results: The focus group discussions explored the strengths and limitations of current evaluation and regulation methods and identified potential alternatives that could enhance the quality and safety of medical apps. Although RCTs were only explicitly mentioned in the German regulative system as one of many options, an analysis of chosen evaluation methods for the German apps on prescription pointed towards a “scientific reflex” where RCTs are always the chosen evaluation method. However, this method poses substantial limitations regarding digital interventions such as medical apps. Comparable results were observed during the focus group discussions, where participants expressed similar experiences for their own evaluation approaches. Additionally, the participants highlighted numerous alternatives to RCTs that can be used at different points during the life cycle of a digital intervention to assess its efficacy and potential harm to users.

Conclusions: It is crucial to recognize that digital interventions constantly evolve, unlike analog tools, posing challenges to inflexible evaluation methods, such as RCTs. Potential risks include high dropout rates, decreased adherence, and non-significant results. However, existing regulations do not explicitly advocate for other evaluation methodologies. Our research highlighted the necessity of overcoming the gap between regulatory demands to demonstrate safety and efficacy in medical apps and evolving scientific practices, ensuring that digital health innovation is evaluated and regulated in a way that considers the unique characteristics of mobile medical devices.

(JMIR Preprints 28/11/2023:54814)
DOI: https://doi.org/10.2196/preprints.54814

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Challenges and alternatives to evaluation methods and regulation approaches of medical apps as mobile medical devices: An international and multidisciplinary focus group discussion

Laura Maaß[1,2], Robert Hrynyschyn[1,3], Martin Lange[4], Alexandra Löwe[4], Kathrin Burdenski[5], Kaley Butten[6], Sebastian Vorberg[7,8], Marlien Varnfield[6], Mariam Hachem[9,10], Aldo Gorga[11], Vittorio Grieco[12], Vincenzo Restivo[13], Giuseppe Vella[14], and Felix Holl[1,15]

1 Leibniz ScienceCampus Digital Public Health Bremen, Bremen, Germany
2 University of Bremen, SOCIUM Research Center on Inequality and Social Policy, Bremen, Germany
3 Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt- Universität zu Berlin, Institute of Health and Nursing Science, Berlin, Germany
4 IST University of Applied Sciences, Department of Fitness & Health, Düsseldorf, Germany
5 Loughborough University, School of Sport, Exercise, and Health Sciences, Loughborough, United Kingdom
6 Australian eHealth Research Centre, CSIRO, Brisbane, Australia
7 Vorberg.law / QuR.digital, Hamburg, Germany
8 Bundesverband Internetmedizin e.V., Hamburg, Germany
9 University of Melbourne, Department of Medicine, Austin Health, Faculty of Dentistry, Medicine and Health Sciences, Parkville, Melbourne, Victoria, Australia.
10 University of Melbourne, Australian Centre for Accelerating Diabetes Innovations, Faculty of Dentistry, Medicine and Health Sciences, Parkville, Melbourne, Victoria, Australia
11 University of Turin, Department of Sciences of Public Health and Pediatrics, Turin, Italy
12 University of Catania, Department of Medical, Surgical Sciences and Advanced Technologies “GF Ingrassia”, Catania, Italy
13 University of Enna,School of Medicine, Enna, Italy
14 University of Palermo, Department of Health Promotion, Maternal and Infant Care, Internal Medicine and Medical Specialties (PROMISE) “G. D’Alessandro”, Palermo, Italy
15 Neu-Ulm University of Applied Sciences, DigiHealth Institute, Neu-Ulm, Germany

Corresponding author:
Laura Maaß,
University of Bremen, SOCIUM
E-Mail laura.maass@uni-bremen.de | Tel. +49 421 218-58610

Abstract

Background: The rapid proliferation of medical apps has transformed the healthcare landscape by giving patients and providers unprecedented access to personalized health information and services. However, concerns regarding the effectiveness and safety of apps have raised questions regarding the efficacy of randomized controlled trials (RCTs) for their evaluation and as a requirement for their regulation as a mobile medical device.

Objective: This study addresses this issue by investigating alternative methods besides RCTs for evaluating and regulating medical apps.

Methods: Using a qualitative approach, a focus group study with 46 international and multidisciplinary public health experts, split into three groups, was conducted at the 17th World Congress on Public Health in May 2023 in Rome (Italy) to gather in-depth insights into alternative

approaches to evaluation and regulation. We conducted a policy analysis of the current regulation of medical apps as mobile medical devices for the four most represented countries in the workshop: Italy, Germany, Canada, and Australia. We developed a logic model that combines the evaluation and regulation domains based on these findings.
Results: The focus group discussions explored the strengths and limitations of current evaluation and regulation methods and identified potential alternatives that could enhance the quality and safety of medical apps. Although RCTs were only explicitly mentioned in the German regulative system as one of many options, an analysis of chosen evaluation methods for the German apps on prescription pointed towards a “scientific reflex” where RCTs are always the chosen evaluation method. However, this method poses substantial limitations regarding digital interventions such as medical apps. Comparable results were observed during the focus group discussions, where participants expressed similar experiences for their own evaluation approaches. Additionally, the participants highlighted numerous alternatives to RCTs that can be used at different points during the life cycle of a digital intervention to assess its efficacy and potential harm to users.
Conclusions: It is crucial to recognize that digital interventions constantly evolve, unlike analog tools, posing challenges to inflexible evaluation methods, such as RCTs. Potential risks include high dropout rates, decreased adherence, and non-significant results. However, existing regulations do not explicitly advocate for other evaluation methodologies. Our research highlighted the necessity of overcoming the gap between regulatory demands to demonstrate safety and efficacy in medical apps and evolving scientific practices, ensuring that digital health innovation is evaluated and regulated in a way that considers the unique characteristics of mobile medical devices.

Keywords

medical apps; mobile medical devices; evaluation methods; mobile medical device regulation; focus group study; alternative approaches; logic model

Introduction

Background

Medical applications (medical apps) have emerged as a subgroup of health apps as powerful tools with the potential to revolutionize healthcare by supporting, managing, and enhancing individual and population health [1]. These software programs can run on mobile devices and offer new possibilities for processing health-related data [2, 3]. According to the General Data Protection Regulation (GDPR), data processing includes any operation performed on personal (health) data, including collecting, organizing, storing, adapting, visualizing, retrieving, disseminating, restricting, or erasing the data [4, 5]. Unlike most health app types, medical apps do not primarily target healthy or health-conscious individuals but patients, health professionals, or family caregivers. This stems from their primary use as treatment supplements or medical devices in clinical contexts such as diagnostics, treatment, or rehabilitation. Figure 1 summarizes the distinction between medical apps as a subgroup and other areas of health apps (such as wellness) from a legal point of view [6].

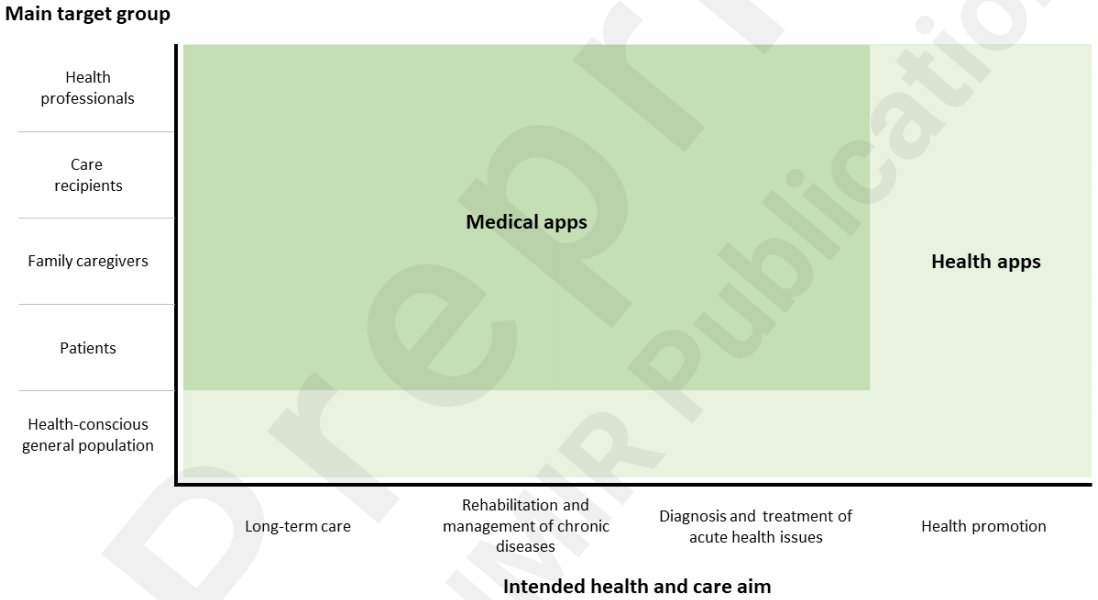


Figure 1. Medical apps as a sub-group of health apps. Adapted from Maaß et al. (2022) [6].

Medical apps promise personalized healthcare, improving access to medical resources and fostering active patient participation [7]. However, this thematic orientation of medical apps introduces a range of potential risks. Health and patient information may be inaccurate or flawed [8], sensitive data can be disclosed to the wrong recipients or lost [9], and applications frequently target broad user adoption via the internet, increasing the risk of rapid dissemination of any potential dangers. Due to the sensitive nature of health data, medical apps must adhere to rigorous data security standards. As many apps are used across borders, different jurisdictions and regulatory standards may need to be met [10, 11].

When considering covering the costs of these products through entities such as health insurance providers, demonstrating their clinical and medical benefits becomes essential to be regulated as mobile medical devices [12-15]. This advantage must be substantiated and evaluated to align costs adequately within the insurance community [16]. In addition, high data protection requirements, such as the GDPR, impact the possibilities of data collection,

analysis, and, therefore, the evaluation of medical apps [17-19]. Software is an increasingly critical area of healthcare product development, and there is a need to establish a standard and converged understanding of clinical evaluation globally [20].

Challenges in the evaluation of medical apps

With the increase in medical apps' usage in healthcare, it is crucial to validate their effectiveness. Randomized controlled trials (RCTs) are commonly considered the gold standard for establishing cause-and-effect relationships owing to their rigorous methodology [21-23]. However, there is an apparent disparity between the design of RCTs and the need to evaluate medical apps: RCTs are structured and rigid, whereas medical apps are versatile, updated frequently and depend on specific contexts [24]. A critical challenge arises from the discrepancy between the thoroughness of evaluations and the pace of generating evidence, as depicted in Figure 2 [25]. This misalignment underscores the need for developing more adaptable evaluation strategies in the rapidly evolving field of medical technology.

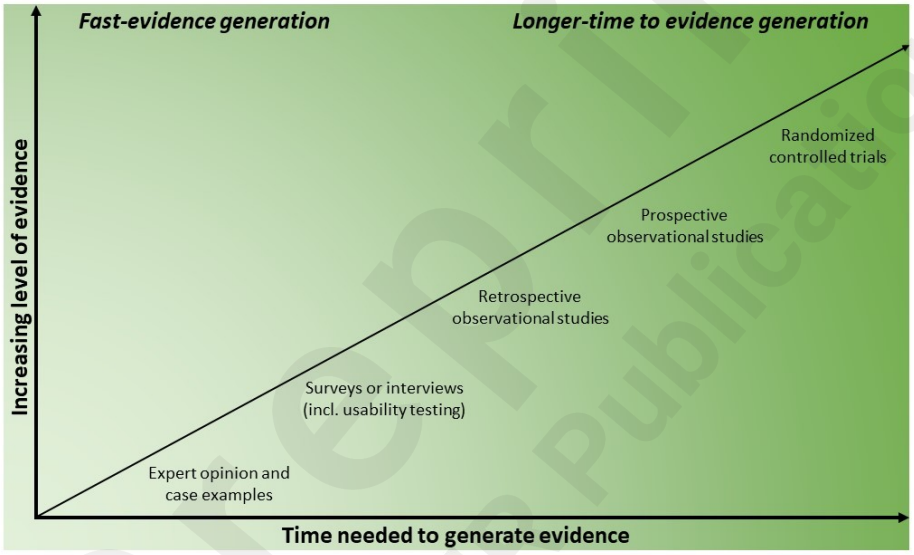


Figure 2. The balancing act between increasing evidence levels and the time needed to generate them. Adapted from Guo et al. (2020) [26]

Furthermore, traditional trials like RCTs do not necessarily align with the unique characteristics of medical apps [27, 28]. The evaluations of medical apps tend to be based on short-term studies [29-31], limited patient samples [32], feasibility trials, and user preference surveys [33]. While these methods can swiftly provide valuable insights during the development phase, they only offer a narrow view of the apps' true efficacy [34].

Another challenge is that medical apps are often customized to individual needs, in contrast to the between-group design of RCTs. Continuous updates and tweaks to medical apps to improve their performance can sometimes shift their fundamental functionality. It is essential to recognize that these updates might influence patient outcomes and call for continuous reassessment [24]. Given users' diverse health needs, preferences, and technological expertise, a universal evaluation method seems impractical. Moreover, blinding participants and researchers, a cornerstone of RCTs to reduce bias, can be difficult with medical apps since users are often aware they are using an app, and developers may struggle to create "placebo" apps with identical functionalities.

Additionally, app evaluation is multifaceted, demanding attention to factors beyond the mere health metrics typically assessed in RCTs [34]. Aspects such as user engagement and sustained use are vital for long-lasting behavioral shifts [35-37]. During the developmental

phase of a medical app, alternative study designs, including feasibility studies and user preference surveys, might be more suitable [33], even if they offer limited efficacy data [32]. They can promptly provide essential evidence regarding developmental and application areas. The multifaceted demands of medical app evaluation call for innovative evaluation methods to form the foundation of medical app regulations.

The complexity and struggles of overregulation

Regulatory and quality standards that are legally mandated and government-reviewed ensure medical apps' safety, efficacy, and data privacy [26]. Historically, pharmaceuticals and medical devices have faced increased regulation only after significant scandals and incidents have emerged. “Fueled” by these scandals, medical app regulation seeks to prevent scandals and incidents through rigorous guidelines [38-40]. However, there is a risk that anticipatory regulation may overemphasize securing against hypothetical risks [41]. It is crucial to weigh hypothetical risks against the effort required for regulatory compliance and the expected benefits of maintaining a balanced regulatory approach [42]. Relevant criteria for regulating medical apps are displayed in Textbox 1. Regulatory solutions should adopt the mildest measures possible to minimize resistance to technological advancement. Consequently, it is essential to incorporate the potential severity of harm into regulatory assessments and, if necessary, accept manageable errors as learning opportunities. Post-market surveillance is a critical component of quality control and assurance in the regulation of innovative medical developments.

Textbox 1. Relevant criteria for regulating mobile apps as mobile medical devices

Effectiveness and Safety: Regulatory frameworks should be adapted to the specific characteristics of digital health applications, necessitating clinical studies and testing procedures tailored to these applications.

Data Privacy and Security: Regulation must ensure applications implement appropriate data protection and security measures.

Interoperability: Digital health applications should be integrable into existing healthcare systems to ensure seamless care delivery.

Provider Qualifications: Regulations should mandate that qualified professionals develop and operate applications to ensure patient safety and quality.

Monitoring and Traceability: Regulatory mechanisms should enable digital health applications' ongoing monitoring and traceability to identify and address adverse events promptly.

Patient Involvement: Patients should be actively engaged in the development and evaluation of digital health applications to ensure that their needs and expectations are met

Adapted from Torous et al. (2022) [43]; Meskó & deBronkart (2022) [44]

Need for alternative evaluation approaches and regulation

Regulatory requirements homogenize research conditions, while the results are prone to reduction and distortion. At the same time, medical apps are, by nature, complex interventions due to their structure, a large number of interactions, and various intervention components. The efficacy of medical apps depends on context and leads to distinct effects on individuals. This requires a broader evaluation framework that does not focus solely on

specific intervention outcomes but combines different perspectives from feasibility testing to intervention, implementation, and impact evaluation [26, 45, 46]. Current regulatory approaches to evaluating medical apps do not appropriately reflect the complex nature of medical apps.

Regulation of medical apps in selected countries

To illustrate similarities and differences, four countries have been selected for case studies about the existing regulation methods for medical apps as mobile medical devices. The analysis for Germany, Italy, Australia, and Canada followed an adapted version of the policy benchmarking framework by Essén et al. 2022 [38] (see Table 1).

Table 1. Overview of medical app regulation for selected countries

Country	Germany	Italy	Australia	Canada
Actors developing national policy regulation	Health Innovation Hub, an external, interdisciplinary expert think tank to the Federal Ministry of Health and the Federal Institute for Drugs and Medical Devices	Ministry of Health overseeing conformity following the European Medical Device Regulation and classification of Software as a Medical Device (SaMD)	Therapeutic Goods Administration (TGA)	Health Canada (Federal Government Agency)
Intended use of the framework	Determine eligibility of an app for reimbursement through the public insurance scheme. Either permanently or on a preliminary basis.	No	National guidance and a helpful reference tool for app developers working on mobile health apps for release in Australia. It provides a solid basis for further research and analysis	To determine if the app meets the legal definition of a SaMD
Key regulations underpinning the policy framework	Digital care applications (DiGAs) under the Digital Care Act	No	The Privacy Act 1988, overseen by the Office of the Australian Information Commissioner, Competition and Consumer Act 2010 administered by the Australian Competition and Consumer Commission	Risk classification as a SaMD
Risk	European Medical	European	TGA (Medical	Health Canada

classification framework for medical apps	Device Regulation (Risk classes I, IIa, IIb, III)	Medical Device Regulation (Risk classes I, IIa, IIb, III)	Devices) Regulation 2002 (Risk classes I, IIa, IIb, III)	Medical Device Classification (Risk classes I, II, III, IV)
Reimbursement approval policy regulation developed?	Applications get approved through the DiGA fast-track procedure. Approved DiGA are reimbursed through health insurance	No	No	No
End-user interface to clinical practice and patients, which lists approved medical apps	DiGA directory	No	No	No, but Health Canada does engage with the Canadian Agency for Drugs and Technologies to provide evidence and publicly available Health Technology Assessments
Are RCTs explicitly mentioned in the regulation?	Mentioned as an example of a methodology designed to demonstrate a positive care effect	No	No	No

Germany

Since December 2019, digital care applications (DiGAs) have become eligible for reimbursement through statutory health insurance in Germany. This mechanism is included in the Digital Care Act (DVG) [47, 48]. Reimbursable DiGAs are listed in a directory administered by the Federal Institute for Drugs and Medical Devices (BfArM), DiGA directory. Applications can be added to the directory through a fast-track procedure [49]. Applications can either be permanently included in the list if evidence of a positive care effect is provided at the time of application or a provisional reimbursability is granted if apps meet a set of essential requirements but evidence of the positive care effect is still missing. Evidence of the positive care effect must be provided within 12 months for permanent inclusion. Otherwise, the app will be removed from the list. The term “positive care effect” introduced by the DVG can either be a “medical benefit” or "patient-relevant improvement of structure and processes". For instance, the application can help improve health literacy and improve coordination of treatment processes or can reduce therapy costs. Currently, most DiGAs provide evidence for the positive care effect using data from RCTs [50]. As of October 13, 2023, the DiGA directory included 55 applications, of which 23 were provisionally included, 26 were permanently included, and six were removed. All were assessed through an RCT, although the DVG does not explicitly call for an RCT [51]. A limitation of the DVG is the restriction on low-risk devices (risk classes I and IIa) under the Medical Devices Act. All applications outside the Medical Devices Act, such as preventive applications, and high-risk medical devices do not fall under the DVG. The fast-track procedure only applies to applications classified and certified as medical devices in risk classes I and IIa; therefore many potentially helpful applications do not fall within the scope

of the DVG [52].

Italy

In Italy, medical apps can be promptly marketed, provided they meet legal requirements, are classified as Software as a Medical Device (SaMD), and obtain a CE marking [53, 54]. On May 25, 2021, the Italian Ministry of Health issued a newsletter to offer interested stakeholders with recommendations for clinical investigations of medical devices in line with current regulations [55]. The Ministry is currently responsible for the approval process. However, there is no national framework for market access or reimbursement approval, and a law for the health technology assessment of medical apps is also lacking. While an application can be sold, it cannot be prescribed or reimbursed, hindering equitable healthcare access. Discussions are ongoing in legislative bodies to address this issue, and a Parliamentary Intergroup for Digital Health and Digital Therapies was established in May 2023 [56]. This initiative aims to develop regulations that align with those of other European countries, ideally during the current legislative term.

Australia

In Australia, medical devices, including software-based medical devices, are classified according to the medical device classification rules of the Therapeutic Goods (Medical Devices) Regulations 2002 [57]. This act oversees the quality, safety, efficacy, and availability of therapeutic goods in Australia. Medical devices are classified and regulated according to the level of harm they may pose to the users or patients. This includes medical device software and medical devices that incorporate software. A four-tiered classification system is used to determine the minimum conformity assessment procedures, or comparable overseas regulator evidence requirements, to be determined by the manufacturer before an application is made for the device's inclusion in the Australian Register of Therapeutic Goods (ARTG) [58]. The classification rules are applied according to the manufacturer’s intended purpose and the device’s functionality. Where more than one rule applies, the device must be classified at the highest applicable level [57, 59]. However, the Australian regulatory system lacks refinement regarding emerging technologies, such as AI in medical decision-making and wearable technology. This generates problems in deciding which tools and devices must be registered and to what classification standard [60].

Canada

Within Canada, mobile health apps are classified as SaMD and regulated according to a 4-tier risk-based classification system utilized for Medical Devices similar to Australia. The manufacturer’s intended use determines the classification of potential SaMD and corresponding regulation and whether the federal department Health Canada agrees with the use-case [61]. The clinical and evaluation evidence required to support the use-case varies depending on the tiered classification and intended use. However, the guidance on SaMD released by Health Canada in 2019 suggests broad exclusion criteria, eliminating many health apps from requiring SaMD status and thus leaving many apps unregulated and without evaluation [62].

Study aim

Due to the described challenges in regulating and evaluating medical apps, this study pursues two central objectives. Building on the previously described country examples for the current regulation and evaluation practices of medical apps, this study aims first to determine whether there are international challenges in regulation and evaluation and second to gather

the opinions and perspectives of digital (public) health researchers on current practices and future developments through a focus group approach. The results from both objectives are then utilized to develop a logic model proposing a differentiated approach to regulating and evaluating medical apps. Therefore, this study attempts to initiate a discussion about current regulation and evaluation practices of medical apps and accelerate the adoption of innovations in healthcare systems.

Methods

The reporting quality of this study is assessed through the consolidated criteria for reporting qualitative research checklist (COREQ) (see Multimedia Appendix 1).

Focus group discussion workshop at the 17th World Congress on Public Health

Participants for the focus groups were recruited through a convenient sampling method, where interested attendees of the 17th World Congress on Public Health (WCPH) attended the workshop in person. The study was advertised to participants through the conference program. Due to the nature of the context, information on non-participation was not recorded. 46 WCPH delegates from 14 countries joined the workshop. Besides the moderators and workshop participants, two conference technical-team members were in the room to oversee the technical equipment. They did not participate in the workshop. At the beginning of the session, participants were introduced to the general concepts of health and medical apps and the problems that RCTs pose for evaluating these types of digital health interventions.

We used a QR-Code leading to a LimeSurvey questionnaire (LimeSurvey GmbH, Hamburg, Germany) hosted by the University of Bremen (Germany) to collect demographic data and information regarding the participants' knowledge. All questions were optional. The 46 participants were equally distributed among three focus groups, each led and protocolled by one of the article's authors. One female and two male authors introduced themselves as researchers interested in evaluating digital apps and explained the scope of the workshop. They discussed the two questions presented in the study's aim section. The two questions were not pilot-tested. The three authors have a background in public health and medical informatics with previous research on the topic: One author conducts his PhD on alternative evaluation approaches [63], one author published a review on definitions for health and medical apps [6], and the third author conducted an assessment of evaluation methods for mobile health interventions [35]. The authors had the underlying assumption that the majority of participants will come from Italy and that the discussions will focus more on the first question due to the conference's target group of public health researchers and practitioners.

Focus groups are used in qualitative research to gain in-depth knowledge of specific issues; in our case, the evaluation and regulation of medical apps. Unlike in surveys or individual interviews, a strength of focus groups is that participants can interact with other participants and the moderator rather than merely sharing their views [64]. In our case, we followed the methodology proposed by Traynor [64] with a minor adaptation to avoid audio-recording the discussions due to the noise in the room. Instead, we asked one participant per group to assist the moderator in taking notes on a flip chart visible to the respective focus group. The discussion lasted forty minutes, with half the time focusing on question one and the second half on question two. The moderators asked the two guiding and follow-up collective questions, encouraged quieter participants to express their opinions, and quietened more dominant speakers if necessary. After the workshop, all participants were invited to contribute to the project as co-authors, which nine participants did. The workshop transcripts were only provided to the nine co-authoring participants (see Multimedia Appendix 2).

Ethical considerations

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Joint Ethics Committee of the Universities of Applied Sciences of Bavaria (GEHBa-202304-V-105, dated April 28, 2023). After the introductory presentations described above, we handed out written participant information sheets to educate them on the study and which data would be collected, stored, and analyzed. All data were anonymized, as we did not collect information that would allow the identification of participants. All participants were informed that their study participation was voluntary and that they could withdraw from the study at any time without providing reasons. Participants were informed before data collection that they would not receive any financial compensation for participating in this study but were invited as co-authors if they expressed written interest. In this case, the participant's e-mail addresses were collected on a paper sheet and were not connected to the online survey where we collected their sociodemographic data. Lastly, all participants provided their written informed consent before data collection.

Qualitative analysis and data synthesis

Data analysis and synthesis followed the conceptual and design thinking for the thematic analysis (TA) methodology of Braun and Clarke [65, 66]. TA aims to indicate and analyze patterns (so-called themes) in minimally organized datasets. TA was deemed the most suitable framework as it can be applied without pre-existing theoretical frameworks, which makes it an ideal methodology for multi- and interdisciplinary research questions, such as the current ones/topic [65, 66]. In our case, two authors individually clustered the workshop results inductively, without predefined themes, in two separate Microsoft Excel sheets. Both authors were provided with transcripts of all the flip charts for the focus groups. They then merged individual statements based on the same typology and meaning, for example, when two groups mentioned the same evaluation designs or barriers. The results were then grouped according to a shared umbrella theme. This was performed separately for the evaluation and regulation of medical apps. The final themes were approved during a discourse among three authors with one of these having the ultimate decision-making power.

Based on the workshop findings and policy analysis, we developed a logic model to capture the complexity of medical app evaluation and regulation. Logic models can be helpful to display the life cycle of a medical app based on overarching themes, for instance problem, target, intervention and content, moderating and mediating factors, outcomes, and impact [67] and have been used to evaluate digital health interventions [68, 69]. The model can be adapted to the specific context. It enables the display of interaction and feedback loops and distinguishes between outcomes and impacts in different domains [67]. Furthermore, adjusted logic models can provide a framework to display how complex interventions work across multiple domains in a single setting, with interlinking actions producing a range of outputs and outcomes. In our logic model, we used a generic medical app example to display significant findings and assign them to specific phases of the medical app lifecycle, such as development, implementation, and impact.

Results

46 experts participated in the one-hour international workshop at the 17th WCPH [70]. The participants were divided into three focus groups of 15-16 people each. The participant demographics are displayed in Table 2.

Table 2. Workshop participant demographics

Criteria	Number of participants

	(n=46; 100%)
Highest qualification	
Bachelor	2 (4.3%)
Master	13 (28.3%)
Diploma	2 (4.3%)
Medical Doctor	18 (39.1%)
PhD	11 (23.9%)
Primary discipline	
Computer Science	5 (10.9%)
Medicine	3 (6.5%)
Psychology	3 (6.5%)
Public Health	34 (73.9%)
Sociology	1 (2.2%)
Gender	
Women	26 (56.5%)
Men	20 (43.5%)
Country of residence	
Australia	5 (10.9%)
Austria	1 (2.2%)
Canada	4 (8.7%)
Finland	1 (2.2%)
Germany	8 (17.4%)
Hungary	1 (2.2%)
Indonesia	1 (2.2%)
Italy	17 (37.0%)
Netherlands	1 (2.2%)
New Zealand	2 (4.3%)
Philippines	1 (2.2%)
Portugal	1 (2.2%)
Switzerland	1 (2.2%)
Taiwan	1 (2.2%)
United States of America	1 (2.2%)
Age	
Average (Range, SD)	35.1 (25-73, 9.6)
Median	31.5
Years of general work experience in the primary field	
Average (Range, SD)	8.3 (1-40, 8.5)
Median	5
Years of experience in developing, implementing, regulating, or evaluating health or medical apps	
Average (Range, SD)	1.8 (0-15, 2.7)
Median	1

We asked the participants about their experiences and opinions regarding evaluating and regulating medical mobile apps. The thematic analysis of the contributions in the three groups resulted in four clusters for the evaluation and three for the regulation of medical mobile apps, which are presented in more detail later (Textboxes 2 and 3).

Alternative evaluation methods

The following general question was presented to the participants at the beginning of the focus group discussions: “What could be alternative evaluation methods for assessing the effectiveness of medical applications?” Four overarching themes crystallized as pivotal

according to the participants (see Textbox 2):

1. **Traditional evaluation practice:** Participants emphasized the advantages of RCTs in evaluating medical apps, acknowledging their ability to reduce bias and establish causal relationships. However, concerns were raised about participant dissatisfaction and high attrition rates due to random assignment, questioning sustainable implementation, and user centricity. Incongruence between RCT temporal requirements and the rapid evolution of medical apps necessitated frequent updates, prompting the consideration of alternative evaluation methods.
2. **Universal considerations for all evaluation types:** The importance of engagement and user data in app evaluations was highlighted, and participants advocated qualitative assessment in pre-study designs to reduce dropouts. Additionally, the participants stated that standard outcome parameters were essential for effectiveness evaluation and could be supplemented by output data to demonstrate efficacy. The participants further suggested segmental app evaluations and various comparative interventions. The importance of enhancing power and data quality through a priori sample calculations and suitable study designs was highlighted, regardless of the study design.
3. **Alternative approaches for participant recruitment & grouping in an RCT:** Alternative recruitment and grouping methods were explored to address dissatisfaction and attrition in RCTs. These included preference clinical trials, pragmatic RCTs, partially randomized patient-preference trials, and best-choice experiments. By increasing flexibility and allowing participants to choose their interventions, these approaches might reduce withdrawal, shorten study duration, reduce bias, and facilitate participant recruitment.
4. **Alternative research designs to RCTs:** Other research designs, such as qualitative and participatory approaches, can help to understand engagement barriers and enhance user experience. Prototype checklists were proposed for researchers unfamiliar with alternative evaluation designs. Alternative data sources, including real-world evidence, smartphone sensors, and commercially available user analytics, can complement traditional evaluation methods.

Textbox 2. Results of thematic data analysis about alternative evaluation methods.

Traditional RCTs:

- Strength of RCTs (e.g., internal validity, the strength of causal relationships)
- Weakness of RCTs to represent reality, resource consumption, and low generalizability
- Randomization might cause dissatisfaction and increase dropout
- Effectiveness trials tend to produce too little engagement
- The speed of app development does not align with RCT duration

General evaluation aspects:

- Different types of outcomes need to be acknowledged by evaluation goals and methods.
- Power (a priori sample size, bias) and data validity (level of causality, purity of data)
- Differentiation of traditional (e.g., effectiveness, satisfaction) and implementation outcomes (e.g., acceptability, costs, feasibility)
- Qualitative pre-study generates fast evidence
- Level of engagement and interaction (user frequency, duration of use, etc.)

- Level of engagement and interaction (user frequency, duration of use, etc.)
- Importance of multiple and standardized outcomes
- Find a suitable comparison group (new, existing, or non-treatment intervention)
- Consider intervention effects (e.g., Hawthorne)
- Outcomes for stakeholder and target groups differ and need to be considered
- Relevance of component's effectiveness
- Relevance of output evaluation

Alternatives for recruiting and grouping in RCTs:

- Preference-based controlled trials
- Evaluation of specific implementation stages rather than implementation as one
- Pragmatic RCTs / Real-world RCTs
- Ensure data quality through volunteer participation
- Quasi-controlled trial
- Best choice experiment after randomization
- Stepped wedge trials
- Waitlist-control-group design
- Evaluation as treated

Alternatives to RCTs:

- Qualitative empirical data (e.g., user interviews) to understand engagement
- Improving user experience to promote engagement
- Create a prototype evaluation checklist for researchers with less experience in app development.
- Use of real-world evidence (e.g., hospital data, cohort data set)
- Shortcut evaluation if a known aspect works (McCarthy Evaluation)
- Evaluate the practice/roll-out phase
- Model app usage through existing, similar datasets
- Apply user analytics (e.g., Google Analytics)
- Use of additional data through smartphone sensors

A glossary explaining the alternative evaluation methods in more detail is available in Multimedia Appendix 3.

Alternative regulation approaches

Participants worked on the regulative part during the second half of the focus group discussions. The guiding question was: “What could be alternative approaches to regulating medical applications as mobile medical devices?”. Three themes emerged from these discussions (see Textbox 3):

1. **Minimal standards for medical apps:** Participants emphasized essential minimum standards that medical apps should adhere to ensure safety and harmlessness, including data privacy, user accountability, informed consent, and prevention of biases or discrimination. Experts acknowledged that regulatory requirements should be proportional to the app's function and data processing. Central to the regulation process is prioritizing the efficacy of the medical app, established through rigorous evaluation procedures.
2. **Tools for regulation:** Labels are commonly used in public health and can offer a concise way to communicate information and evaluate app content and user-friendliness. Participants stressed the need for standardized guidelines across geographic regions,

exemplified by the Device Regulation in Europe, to establish a foundational standard for further country-specific development.

3. **Barriers to regulation:** The primary concern regarding the law was potential overregulation. Complex approval processes could stifle app innovations. Overregulation might lead to an abundance of specialized "silo apps," where patients managing multiple health conditions require numerous apps, hindering efficiency and user friendliness. Additionally, participants recognized that the app's usability is contingent on the device, a factor beyond the scope of app regulation.

Textbox 3. Results of thematic data analysis about alternative regulation approaches.

Functionality of medical apps:
- Regulation should be proportional to app functionality
- No discrimination and not doing any harm as the minimum requirements
- Privacy of data needs to be ensured
- Accountability and informed consent as a developer's responsibility
- Efficacy should be measured

Tools for regulation:
- Using labels that rate various aspects of the app (e.g., accessibility, health benefits)
- Standardize regulation across geographical areas, e.g., in Europe: create minimal requirements

Barriers to regulation:
- There should be no overregulation
- Silo Apps: if a chronically ill person needs 20 apps to manage their symptoms, this is unfeasible and expensive
- It is equally important to consider the device as the app

Logic model for the evaluation and regulation of medical apps

Based on the country assessment and the focus-group discussions, we developed a logic model highlighting the complexity of medical app evaluation and regulation. A logic model depicts the relationship between a program’s activities and its intended effects [71]. In our case, the logic model illustrates different stages of app development (read from left to right) along with the various types of evidence and regulation required at these stages [72, 73]. The logic model consists of three independent horizontal blocks (see Figure 3). The first horizontal block provides an example of a generic medical app. The second horizontal block focuses on evaluation designs for the effectiveness and usability of such apps, and the third contains regulatory aspects. Each block is divided into ‘columns’ defined as the general phases of logic models (Problem, Target, Intervention and Content, Moderating and Mediating Factors, Outcomes, Impact). The boxes on evaluation and regulation are additionally influenced by the conceptual, implementation, and evaluation quality. It is essential for evaluation and regulation approaches for medical apps to also consider these additionally to the mere analysis of outcomes (e.g., as measured through RCTs).

Feasibility studies can be applied to evaluate the medical apps’ target and its content. They can include multidisciplinary and participatory designs or theoretical models based on evidence-based frameworks. Machine learning algorithms and app usage modeling from similar datasets can provide additional feasibility analyses [74]. Clinical simulations can be run to test digital interventions in a safe, cost-effective, and efficient manner [26]. Alternative trial and study designs that integrate qualitative studies, and user analytics (to evaluate the

outcomes of the medical app) allow for integrating multiple perspectives and offer efficient approaches to evidence generation. Regulation of medical apps should incorporate the app's life cycle similarly. An a priori regulatory concept should be applied already during the target and content phase of the app development based on the purpose (e.g., disease-related) and functionality (e.g., documenting, storing, monitoring, data analyzing, or transferring data) of the app. Its potential harm concerning data protection and safety of use (e.g., informed consent, appropriateness of data processing) should be best regulated even earlier during the problem-identification phase. Relevant regulation is needed for adverse events, patient benefits, and clinical evidence during the outcome-phase. The quality and risk management as well as the technical documentation should be regulated continuously. Together with the market surveillance after the implementation of the medical app, these procedures will ensure that the application adheres to regulations and fulfills its intended benefits.

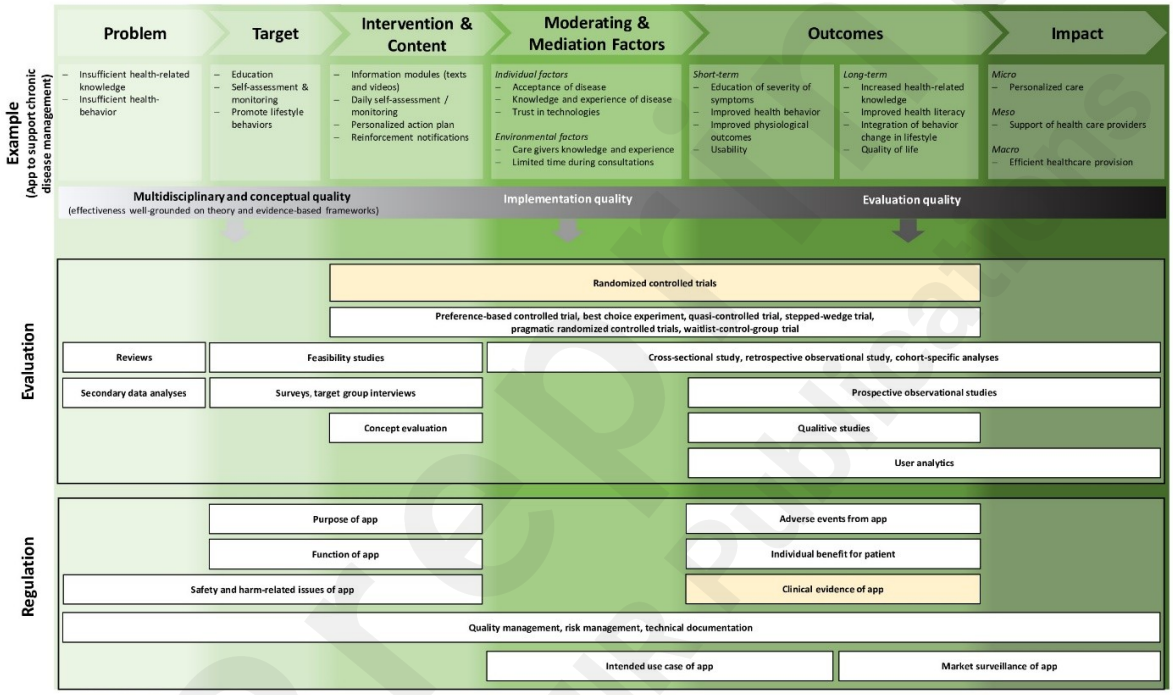


Figure 3. Logic model for evaluating and regulating medical apps through the intervention's life-cycle phases.

'Randomized controlled trials' and 'clinical evidence of apps' are highlighted in yellow, as most medical apps focus on these methods and outcomes [51], as confirmed in our country analysis. This is despite the fact that the relevant regulatory framework requires a sufficient evaluation, which does not necessarily have to be an RCT [12]. In addition, our logic model underlines that RCTs start too late in medical app development and are unsuitable for all development stages.

Discussion

Discussion of results

This study explored alternative methods of evaluating and regulating medical apps in an international focus group of public health professionals. The focus group debate highlighted that medical app development could benefit from flexible study designs that can evaluate multiple components simultaneously and independently. While acknowledging the value of RCTs in establishing causal relationships, participants were dissatisfied with RCTs as the

"go-to" evaluation method. Instead, they perceived them as a barrier to medical app development. Although participants were aware of various alternative evaluation methods, they found that their practice application is lacking because the default method remains RCTs.

According to participants, medical app regulation should align with the rapid technological development of medical apps through an agile and iterative evaluation process, which could more appropriately capture the breadth of efficacy and safety outcomes associated with medical apps and potentially reduce the time-to-evidence and time-to-market.

The focus group enabled a difficult discussion to challenge the medical dogma of RCTs. Following the global thalidomide tragedy in 1962, RCTs became established as best practice and are a requirement to demonstrate safety and efficacy for drugs entering the drug market [75]. The scientific community successfully lobbied for stricter evaluation and regulation at the time. In the context of medical apps, regulation is in its infancy. The launch of a diagnostic radiology app at Apple's World Wide Developers Conference in 2008 caused a regulatory headache and is arguably the catalyst for medical app regulation worldwide [76]. As scientists, we know the benefits of evaluation. Still, we now see that RCTs are not apt to the agile nature of digital interventions and that viewing them as the status quo stifles innovation in the digital health sector.

The focus group provided an opportunity to share our collective desire for change and collate alternative methods of evaluation and regulation, which are presented in the logic model. This model offers a more dynamic approach to medical app evaluation and regulation. Its structure considers medical apps' decomposed evaluation and regulation process and encourages researchers to combine alternative study designs to meet their research goals. Alternative study designs mentioned in the focus group include pragmatic RCTs [77, 78], such as cohort multiple control randomized studies [79], regression discontinuity designs [80], and registry-based randomized trials [81], which were incorporated into our logic model. To enable sustainable evidence generation, a paradigm shift in the approach to evaluation is needed: from RCTs as the gold standard to a multi-method, flexible approach that reflects the agility of the medical app market [28]. Selected methods should aim to reduce time-to-market access, provide robust evidence on the desired (health) outcomes, and be compatible with the rapid technological development of medical apps.

Finally, it is essential to recognize that digital products constantly evolve, necessitating ongoing efforts to optimize them retrospectively. Undertaking resource-intensive RCTs for updates to, for instance, the German DiGA, may pose significant problems in adherence, increase dropout rates, and may lead to statistically non-significant results when analyses are conducted for the whole study population. The high costs and risks of re-evaluating apps using RCTs increase development costs and, thus, risks for app developers. This, in turn, increases the probability that apps will not be launched onto the market as medical apps at all but as health and well-being apps, circumventing regulatory approaches. Overregulation then becomes dysregulation.

Strengths and limitations

Focus groups have become a frequently used method in healthcare research due to their usefulness in identifying and analyzing collective opinions and experiences [82, 83]. Our study included multinational and multidisciplinary experts, allowing us to collect global perspectives from practitioners and researchers. The participants provided valuable insights on the challenges of using RCTs for medical app evaluation and how alternatives could be applied and integrated into current regulations. Real-world examples were shared as the group included participants who had used RCTs in medical app evaluation or refrained from developing medical apps due to the challenges discussed in this paper. Lastly, our analysis of

the four most prominently represented countries (Germany, Italy, Canada, and Australia) followed a predefined framework to guarantee comparability between countries [38]. This analysis highlights the lack of a legal obligation to use RCTs to evaluate and regulate medical apps. It suggests that there may be a reflex to use RCTs for evaluation.

Our study comes with certain limitations. Conducting the workshop at an international conference meant we had no control over recruitment and participant demographics. Despite the global attendance at the conference, this workshop was primarily attended by people living in European and Western countries, with an underrepresentation of participants from the global South and low- and middle-income countries. Given the international nature of digital health interventions and in particular its potential in low- and middle-income countries [84, 85], the focus group likely did not adequately capture the full spectrum of experiences and viewpoints. This limits the generalizability of findings, particularly given the general underrepresentation of researchers from low-and middle-income countries in the discussion around evaluation criteria in medical apps [86]. With a more diverse participant pool, future research could provide a more comprehensive understanding of challenges and opportunities in medical app regulation. Furthermore, participants differed in their age, academic background, and previous knowledge of this topic. Selection bias, defined as participants choosing to participate based on personal interest, limits the study's external validity.

Large focus groups, such as our study with 15 participants per group, tend to be challenging to moderate and may not achieve in-depth discussions [82, 83]. This was addressed by having experienced focus group moderators who were able to facilitate an open discussion atmosphere. Although approximately eight participants per focus group are recommended [82, 83], larger groups allow for capturing more views in a shorter period. As the goal of the focus group was to gain a broad overview of current views on app evaluation and regulation, large focus groups seemed appropriate.

Additionally, the feedback quality was mixed, potentially originating from the self-reported low average experiences with medical app development, evaluation, or regulation (average 1.8 years, range 0-15 years). Another reason might be a differing understanding of medical apps compared to health apps due to varying definitions [6] and regulation practices globally. This might have led to an inaccurate participant understanding of medical apps. A 20-minute introductory workshop on RCTs, medical apps, and their regulation as mobile devices hopefully cleared up some of the misconceptions.

Further research and a first look at the desired future

Medical apps do not necessarily have borders, making this a topic of global interest, with the same challenges occurring in various countries. A focus group discussion with researchers can only be the first step in assessing this topic's relevance and diverse aspects. Further research needs to develop an evaluation framework that allows different methodological approaches for specific stages during the evaluation of the medical app life span from development over implementation to broader impact among all stakeholders, like developers, patients, or app prescribers. Simultaneously, this evaluation framework should include regulatory aspects beyond the clinical evidence generated by RCTs.

Establishing the use of evaluation frameworks like the proposed logic model can provide developers and regulators with a scientifically sound set of evaluation criteria that can address the agile and time-efficient development of medical apps while gradually generating evidence. In turn, this will help overcome the misconception of using RCTs as the cultural gold standard for evaluating digital health applications. Furthermore, future research should also focus on other regulatory aspects, such as patient safety and effectiveness, which are vital for the market approval of medical apps but are not necessarily covered by RCTs. Therefore, future evaluation frameworks must enhance the acceptability of alternative

evaluation methods for stakeholders despite the traditional research evidence provided by RCTs.
Medical apps developed by professionals without medical backgrounds can lead to several challenges, including ineffective, unsafe, and difficult-to-use interventions. Therefore, the development of medical applications should be conducted in multi-professional teams to balance the priority of health outcomes over profitable motives.
Establishing clear regulatory criteria centered on safety, efficacy, and data protection is essential to minimize errors and harm while fully harnessing the potential of medical applications to improve healthcare delivery. By carefully evaluating hypothetical risks, avoiding unnecessary regulatory burdens (overregulation), and incorporating the potential for learning from manageable errors, innovation, and safety can be balanced in regulating digital health applications. This approach ensures that digital medical apps continue to evolve and improve the quality of healthcare services while safeguarding the well-being of users.

Conclusion

While regulations such as the German DVG or the regulations in Canada and Australia emphasize the need for comprehensive clinical data, their lack of strict RCT mandates allows for flexibility in evaluating digital health interventions, such as medical apps. This flexibility, reinforced by consensus in focus groups, highlights the need for integrating adaptable evaluation techniques into regulatory frameworks. Policy recommendations should include specific guidance regarding the unique considerations of mobile medical devices to bridge the gap between regulatory requirements and evolving scientific methods. Focusing on ongoing assessment, adaptation of evaluation strategies, and balancing patient safety and innovation will create a robust evaluation system for medical apps. This will ultimately empower patients and healthcare providers while ensuring digital health technology's safe and effective advancement. Regulatory frameworks should be updated to include guidelines for incorporating real-world data analysis, user engagement studies, and other adaptable evaluation techniques alongside traditional clinical trials.

Acknowledgments

We thank the workshop participants for joining our focus group discussions and for their valuable input. Without their contribution, this study would not have been possible. The authors gratefully acknowledge the support of the Leibniz ScienceCampus Bremen Digital Public health (lsc-diph.de), which is jointly funded by the Leibniz Association (W4/2018), the Federal State of Bremen, and the Leibniz Institute for Prevention Research and Epidemiology – BIPS.

Competing interests

The authors declare no competing interests.

Funding

The authors declare that they did not receive funding for this project. However, the State and University Library Bremen covered the open-access publication fee.

Author contributions

LM: Initiated the project, led the workshop with RH and FH, structured the manuscript, conducted the qualitative content analysis of workshop results, modified the logic

model, wrote sections in the manuscript, and edited the manuscript.
RH: Led the workshop with LM and FH, structured the manuscript, conducted the qualitative content analysis of workshop results, wrote sections in the manuscript, and edited the manuscript.
ML: Participated in the workshop, wrote sections in the manuscript, developed and modified the logic model, and edited the manuscript.
AL: Participated in the workshop, conducted the qualitative content analysis of workshop results, modified the logic model, and wrote sections in the manuscript.
KBur: Participated in the workshop, modified the logic model, and wrote sections in the manuscript.
KBut: Participated in the workshop, wrote sections in the manuscript, and edited the manuscript.
SV: Wrote sections in the manuscript, modified the logic model, and conducted the qualitative content analysis of workshop results.
MH: Participated in the workshop, wrote sections in the manuscript, and edited the manuscript.
AG: Participated in the workshop, wrote sections in the manuscript, and edited the manuscript.
VG: Participated in the workshop, conducted the qualitative content analysis of workshop results, and wrote sections in the manuscript.
VR: Participated in the workshop and wrote sections in the manuscript.
GV: Participated in the workshop and wrote sections in the manuscript.
MV: Wrote sections in the manuscript.
FH: Led the workshop with LM and RH, obtained IRB approval, structured the manuscript, modified the logic model, wrote sections in the manuscript, and edited the manuscript.

Multimedia Appendix

Multimedia Appendix 1: Coreq-Checklist
Multimedia Appendix 2: Transcribed flipcharts of the focus group discussions
Multimedia Appendix 3: Glossary of technical terms given during the focus group discussions

Abbreviations

Medical apps	Medical applications
ARTG	Australian Register of Therapeutic Goods
BfArM	Federal Institute for Drugs and Medical Devices
COREQ	Consolidated criteria for reporting qualitative research checklist
DiGA	Digital health applications
DVG	Digital Care Act
EU MDR	European Medical Device Regulation
GDPR	General Data Protection Regulation
RCT	Randomized controlled trial
SaMD	Software as a Medical Device
TA	Thematic analysis
WCPH	17th World Congress on Public Health

References

1. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. npj Digital Medicine. 2023 2023/03/10;6(1):38. doi: https://doi.org/10.1038/s41746-023-00777-z.

2. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. npj Digital Medicine. 2019 2019/06/03;2(1):45. doi: https://doi.org/10.1038/s41746-019-0121-1.

3. Van Ameringen M, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depression and Anxiety. 2017;34(6):526-39. doi: https://doi.org/10.1002/da.22657.

4. 4 GDPR. Sect. 2 Regulation (EU). 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Brussels: European Union; 2016 [cited 2023 21 August]; Available from: https://op.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1.

5. Mulder T. Health Apps, their Privacy Policies and the GDPR. Research Paper No. 15/2020: European Journal of Law and Technology, University of Groningen Faculty of Law Research; 2019.

6. Maaß L, Freye M, Pan C-C, Dassow H-H, Niess J, Jahnel T. The Definitions of Health Apps and Medical Apps From the Perspective of Public Health and Law: Qualitative Analysis of an Interdisciplinary Literature Overview. JMIR Mhealth Uhealth. 2022;10(10):e37980. PMID: 36315221. doi: https://doi.org/10.2196/37980.

7. Timmers T, Janssen L, Pronk Y, van der Zwaard BC, Koëter S, van Oostveen D, et al. Assessing the Efficacy of an Educational Smartphone or Tablet App With Subdivided and Interactive Content to Increase Patients' Medical Knowledge: Randomized Controlled Trial. JMIR Mhealth Uhealth. 2018;6(12):e10742. PMID: 30578185. doi: https://doi.org/10.2196/10742.

8. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res. 2014 Sep 15;16(9):e210. PMID: 25223398. doi: https://doi.org/10.2196/jmir.3133.

9. LaMonica HM, Roberts AE, Lee GY, Davenport TA, Hickie IB. Privacy Practices of Health Information Technologies: Privacy Policy Risk Assessment Study and Proposed Guidelines. J Med Internet Res. 2021;23(9):e26317. PMID: 34528895. doi: https://doi.org/10.2196/26317.

10. Rowland SP, Fitzgerald JE, Lungren M, Lee E, Harned Z, McGregor AH. Digital health technology-specific risks for medical malpractice liability. npj Digital Medicine. 2022 2022/10/20;5(1):157. doi: https://doi.org/10.1038/s41746-022-00698-3.

11. van der Storm SL, Jansen M, Meijer HAW, Barsom EZ, Schijven MP. Apps in healthcare and medical research; European legislation and practical tips every healthcare

provider should know. International Journal of Medical Informatics. 2023 2023/09/01/;177:105141. doi: https://doi.org/10.1016/j.ijmedinf.2023.105141.

12. 53 Medical Device Regulation (MDR). Regulation (EU) 2017/745 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. Brussels: European Commission; 2017 [cited 2023 21 August]; Available from: https://op.europa.eu/en/publication-detail/-/publication/83bdc18f-315d-11e7-9412-01aa75ed71a1/language-en/format-PDF/source-58036705.

13. Federal Food D, and Cosmetic Act (F&C Act),. 52 Stat. 1040 (Pub. Law 75-717): An act to prohibit the movement in interstate commerce of adulterated and misbranded food, drugs, devices, and cosmetics, and for other purposes. Rule No. 201(h). 1976 [cited 2023 21 August]; Available from: https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title21-section321&num=0&edition=prelim.

14. Federal Institute for Drugs and Medical Devices (BfArM). Medical devices: Differentiation and classification. 2023 [cited 2023 27th Sept]; Available from: https://www.bfarm.de/EN/Medical-devices/Tasks/Differentiation-and-classification/_node.html%3Bjsessionid=41CB9196DC8D464C7991BE2CB09567D3.intranet382.

15. U.S. Food & Drug Administration (FDA). Policy for Device Software Functions and Mobile Medical Applications. Guidance for Industry and Food and Drug Administration Staff. Rockville: FDA; 2022.

16. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Medicine. 2018 2018/12/03;16(1):225. doi: https://doi.org/10.1186/s12916-018-1211-7.

17. Chassang G. The impact of the EU general data protection regulation on scientific research. ecancermedicalscience. 2017;11:709. doi: https://doi.org/10.3332/ecancer.2017.709.

18. Chico V. The impact of the General Data Protection Regulation on health research. British Medical Bulletin. 2018;128(1):109-18. doi: https://doi.org/10.1093/bmb/ldy038.

19. Rumbold JMM, Pierscionek B. The Effect of the General Data Protection Regulation on Medical Research. J Med Internet Res. 2017;19(2):e47. PMID: 28235748. doi: https://doi.org/10.2196/jmir.7108.

20. Stern AD, Brönneke J, Debatin JF, Hagen J, Matthies H, Patel S, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. Lancet Digit Health. 2022 Mar;4(3):e200-e6. PMID: 35216754. doi: https://doi.org/10.1016/s2589-7500(21)00292-2.

21. Gamble T, Haley D, Buck R, Sista N. Designing Randomized Controlled Trials (RCTs). In: Guest G, Namey EE, editors. Public Health Research Methods. London: SAGE Publications, Inc.; 2015. p. 223-50.

22. Hariton E, Locascio JJ. Randomised controlled trials – the gold standard for effectiveness research. BJOG: An International Journal of Obstetrics & Gynaecology. 2018;125(13):1716-. doi: https://doi.org/10.1111/1471-0528.15199.

23. White BK, Burns SK, Giglia RC, Scott JA. Designing evaluation plans for health promotion mHealth interventions: a case study of the Milk Man mobile app. Health Promotion Journal of Australia. 2016;27(3):198-203. doi: https://doi.org/10.1071/HE16041.

24. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of Intervention Principles: Evaluation Methods for Evolving Behavioral Intervention Technologies. J Med Internet Res. 2015;17(7):e166. PMID: 26155878. doi: https://doi.org/10.2196/jmir.4391.

25. Ahmed I, Ahmad NS, Ali S, Ali S, George A, Saleem Danish H, et al. Medication Adherence Apps: Review and Content Analysis. JMIR Mhealth Uhealth. 2018;6(3):e62. PMID: 29549075. doi: 10.2196/mhealth.6432.

26. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions—A call for innovative evidence generation approaches. npj Digital Medicine. 2020 2020/08/27;3(1):110. doi: https://doi.org/10.1038/s41746-020-00314-2.

27. Bonten TN, Rauwerdink A, Wyatt JC, Kasteleyn MJ, Witkamp L, Riper H, et al. Online Guide for Electronic Health Evaluation Approaches: Systematic Scoping Review and Concept Mapping Study. J Med Internet Res. 2020;22(8):e17774. PMID: 32784173. doi: https://doi.org/10.2196/17774.

28. Subbiah V. The next generation of evidence-based medicine. Nature Medicine. 2023 2023/01/01;29(1):49-58. doi: 10.1038/s41591-022-02160-z.

29. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. npj Digital Medicine. 2019 2019/06/17;2(1):54. doi: https://doi.org/10.1038/s41746-019-0129-6.

30. Larsen ME, Nicholas J, Christensen H. Quantifying App Store Dynamics: Longitudinal Tracking of Mental Health Apps. JMIR Mhealth Uhealth. 2016;4(3):e96. PMID: 27507641. doi: https://doi.org/10.2196/mhealth.6020.

31. Singh K, Drouin K, Newmark LP, Filkins M, Silvers E, Bain PA, et al. Patient-Facing Mobile Apps to Treat High-Need, High-Cost Populations: A Scoping Review. JMIR Mhealth Uhealth. 2016;4(4):e136. PMID: 27993761. doi: https://doi.org/10.2196/mhealth.6445.

32. Eis S, Solà-Morales O, Duarte-Díaz A, Vidal-Alaball J, Perestelo-Pérez L, Robles N, et al. Mobile Applications in Mood Disorders and Mental Health: Systematic Search in Apple App Store and Google Play Store and Review of the Literature. International Journal of Environmental Research and Public Health. 2022;19(4):2186. PMID: doi:10.3390/ijerph19042186.

33. Agarwal P, Gordon D, Griffith J, Kithulegoda N, Witteman HO, Sacha Bhatia R, et al. Assessing the quality of mobile applications in chronic disease management: a scoping review. npj Digital Medicine. 2021 2021/03/10;4(1):46. doi: https://doi.org/10.1038/s41746-021-00410-x.

34. Pham Q, Wiljer D, Cafazzo JA. Beyond the Randomized Controlled Trial: A Review of Alternatives in mHealth Clinical Trial Methods. JMIR Mhealth Uhealth. 2016;4(3):e107. PMID: 27613084. doi: https://doi.org/10.2196/mhealth.5720.

35. Holl F, Kircher J, Swoboda WJ, Schobel J. Methods Used to Evaluate mHealth Applications for Cardiovascular Disease: A Quasi-Systematic Scoping Review. International Journal of Environmental Research and Public Health. 2021;18(23):12315. PMID: doi:10.3390/ijerph182312315.

36. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps. JMIR Mhealth Uhealth. 2017;5(12):e190. PMID: 29254914. doi: https://doi.org/10.2196/mhealth.8758.

37. Liew MS, Zhang J, See J, Ong YL. Usability Challenges for Health and Wellness Mobile Apps: Mixed-Methods Study Among mHealth Experts and Consumers. JMIR Mhealth Uhealth. 2019;7(1):e12160. PMID: 30698528. doi: https://doi.org/10.2196/12160.

38. Essén A, Stern AD, Haase CB, Car J, Greaves F, Paparova D, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med. 2022 Mar 18;5(1):31. PMID: 35304561. doi: https://doi.org/10.1038/s41746-022-00573-1.

39. Shuren J, Patel B, Gottlieb S. FDA Regulation of Mobile Medical Apps. JAMA. 2018;320(4):337-8. doi: https://doi.org/10.1001/jama.2018.8832.

40. U.S. Food & Drug Administration (FDA). Digital Health Center of Excellence. Empowering digital health stakeholders to advance health care. 2023 [cited 2023 27th Sept]; Available from: https://www.fda.gov/medical-devices/digital-health-center-excellence.

41. Kao C-K, Liebovitz DM. Consumer Mobile Health Apps: Current State, Barriers, and Future Directions. PM&R. 2017;9(5S):S106-S15. doi: https://doi.org/10.1016/j.pmrj.2017.02.018.

42. Bates DW, Landman A, Levine DM. Health Apps and Health Policy: What Is Needed? JAMA. 2018;320(19):1975-6. doi: https://doi.org/10.1001/jama.2018.14378.

43. Torous J, Stern AD, Bourgeois FT. Regulatory considerations to keep pace with innovation in digital health products. NPJ Digit Med. 2022 Aug 19;5(1):121. PMID: 35986056. doi: https://doi.org/10.1038/s41746-022-00668-9.

44. Meskó B, deBronkart D. Patient Design: The Importance of Including Patients in Designing Health Care. J Med Internet Res. 2022;24(8):e39178. PMID: 36044250. doi: https://doi.org/10.2196/39178.

45. Silberman J, Wicks P, Patel S, Sarlati S, Park S, Korolev IO, et al. Rigorous and rapid evidence assessment in digital health with the evidence DEFINED framework. npj Digital Medicine. 2023 2023/05/31;6(1):101. doi: https://doi.org/10.1038/s41746-023-00836-5.

46. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:n2061. doi: https://doi.org/10.1136/bmj.n2061.

47. Federal Ministry of Health Germany (BMG). Driving the digital transformation of Germany’s healthcare system for the good of patients. The Act to Improve Healthcare Provision through Digitalisation and Innovation (Digital Healthcare Act – DVG). 2019 [27th Sept]; Available from: https://www.bundesgesundheitsministerium.de/en/digital-healthcare-

act.html.

48. Federal Ministry of Health Germany (BMG). Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz – DVG). Vom 9. Dezember 2019. 2019 [20th October]; Available from: https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&start=//*[@attr_id=%27bgbl119s0400.pdf%27]#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl119s2562.pdf%27%5D__1697794668989.

49. Federal Institute for Drugs and Medical Devices (BfArM). The BfArM's Guide on the Fast-Track Process: The Fast-Track Process for Digital Health Applications (DiGA) according to Section 139e SGB V. A Guide for Manufacturers, Service Providers and Users. Bonn: BfArM; 2020.

50. Lauer W, Löbker W, Höfgen B. Digitale Gesundheitsanwendungen (DiGA): Bewertung der Erstattungsfähigkeit mittels DiGA-Fast-Track-Verfahrens im Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. 2021 2021/10/01;64(10):1232-40. doi: https://doi.org/10.1007/s00103-021-03409-7.

51. Federal Institute for Drugs and Medical Devices (BfArM). DiGA-Verzeichnis. 2023 [cited 2023 27th Sept]; Available from: https://diga.bfarm.de/de/verzeichnis.

52. Gerke S, Stern AD, Minssen T. Germany’s digital health reforms in the COVID-19 era: lessons and opportunities for other countries. npj Digital Medicine. 2020 2020/07/10;3(1):94. doi: https://doi.org/10.1038/s41746-020-0306-7.

53. European Commission. Guidelines on the ualification and classification of stand alone software used in healthcare within the regulatory framework of medical devices. 2016 [cited 2023 20th October]; Available from: https://ec.europa.eu/docsroom/documents/17921.

54. Medical Device Coordination Group (MDCG). MDCG 2019-11. Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR. 2019 [cited 2023 20th October]; Available from: https://health.ec.europa.eu/system/files/2020-09/md_mdcg_2019_11_guidance_qualification_classification_software_en_0.pdf.

55. Ministero della Salute. Applicazione del Regolamento UE 2017/745 del Parlamento europeo e del Consiglio, del 5 aprile 2017, nel settore delle indagini cliniche relative ai dispositivi medici. 2021 [cited 2023 5th November]; Available from: https://www.trovanorme.salute.gov.it/norme/renderNormsanPdf?anno=2021&codLeg=80614&parte=1%20&serie=null.

56. Intergruppo Parlamentare Sanità Digitale e Terapie Digitali. Un nuovo paradigma di sanità pubblica. Patto di Legislatura: Sanità digitale e terapie digitali. 2023 [cited 2023 5th November]; Available from: https://www.panoramasanita.it/wp-content/uploads/2023/05/Patto-di-legislatura.pdf.

57. Australian Government. Federal Register of Legislation. Therapeutic Goods (Medical Devices) Regulations 2022. Statutory Rules 236, 2002. Compilation No. 32 (compilation data 1st uly 2017). 2002 [cited 2023 20th October]; Available from:

https://www.legislation.gov.au/Details/F2017C00534.

58. Australian Governmnet. Department of Health and Aged Care. Therapeutig Goods Administration. Australian Register of Therapeutic Goods (ARTG). 2023 [cited 2023 20th October]; Available from: https://www.tga.gov.au/resources/artg.

59. Australian Government. Australian Digital Health Agency. Assessment framework for mHealth apps. Sydney: Australian Digital Health Agency; 2022.

60. O'Shea B, Sridhar R. ICLG - Digital Health Laws and Regulations. Digital Health Laws and Regulations Australia. ICLG; 2023 [cited 2023 21 August]; Available from: https://iclg.com/practice-areas/digital-health-laws-and-regulations/australia.

61. Jogova M, Shaw J, Trevor J. The Regulatory Challenge of Mobile Health: Lessons for Canada. Healthcare Policy. 2019;14(3):19-28. doi: 10.12927/hcpol.2019.25795.

62. Health Canada. Guidance Document Software as a Medical Device (SaMD): Definition and Classification. Ottawa: Health Canada; 2019.

63. Hrynyschyn R, Prediger C, Stock C, Helmer SM. Evaluation Methods Applied to Digital Health Interventions: What Is Being Used beyond Randomised Controlled Trials? —A Scoping Review. International Journal of Environmental Research and Public Health. 2022;19(9):5221. PMID: doi:10.3390/ijerph19095221.

64. Traynor M. Focus group research. Nursing Standard. 2015;29(37):44-8. doi: https://doi.org/10.7748/ns.29.37.44.e8822.

65. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006 2006/01/01;3(2):77-101. doi: https://doi.org/10.1191/1478088706qp063oa.

66. Braun V, Clarke V. Conceptual and design thinking for thematic analysis. Qualitative Psychology. 2022;9(1):3-26. doi: https://doi.org/10.1037/qup0000196.

67. Mills T, Lawton R, Sheard L. Advancing complexity science in healthcare research: the logic of logic models. BMC Medical Research Methodology. 2019 2019/03/12;19(1):55. doi: https://doi.org/10.1186/s12874-019-0701-4.

68. Berry A, McClellan C, Wanless B, Walsh N. A Tailored App for the Self-management of Musculoskeletal Conditions: Evidencing a Logic Model of Behavior Change. JMIR Form Res. 2022;6(3):e32669. PMID: 35258462. doi: https://doi.org/10.2196/32669.

69. Davis JA, Ohan JL, Gibson LY, Prescott SL, Finlay-Jones AL. Understanding Engagement in Digital Mental Health and Well-being Programs for Women in the Perinatal Period: Systematic Review Without Meta-analysis. J Med Internet Res. 2022;24(8):e36620. PMID: 35943773. doi: https://doi.org/10.2196/36620.

70. World Federation of Public Health Associations (WFPHA). About the Congress. WFPHA; 2023 [cited 2023 21 August]; Available from: https://wcph.org/about-the-congress/.

71. Centers for Disease Control and Prevention. Logic Models. CDC Approach to Evaluation. 2018 [cited 2024 6th May]; Available from: https://www.cdc.gov/evaluation/logicmodels/index.htm.

72. Fernandez ME, Ruiter RAC, Markham CM, Kok G. Intervention Mapping: Theory- and Evidence-Based Health Promotion Program Planning: Perspective and Examples. Frontiers in Public Health. 2019 2019-August-14;7. doi: 10.3389/fpubh.2019.00209.

73. O'Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, et al. Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. Pilot and Feasibility Studies. 2019 2019/03/12;5(1):41. doi: 10.1186/s40814-019-0425-6.

74. Wu X, Guo X, Zhang Z. The Efficacy of Mobile Phone Apps for Lifestyle Modification in Diabetes: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth. 2019;7(1):e12297. PMID: 30664494. doi: https://doi.org/10.2196/12297.

75. Bhatt A. Evolution of clinical research: a history before and beyond James Lind. Perspect Clin Res. 2010 Jan;1(1):6-10. PMID: 21829774.

76. Lievevrouw E, Marelli L, Van Hoyweghen I. The FDA's standard-making process for medical digital health technologies: co-producing technological and organizational innovation. BioSocieties. 2022 2022/09/01;17(3):549-76. doi: 10.1057/s41292-021-00232-w.

77. Gamerman V, Cai T, Elsäßer A. Pragmatic randomized clinical trials: best practices and statistical guidance. Health Services and Outcomes Research Methodology. 2019 2019/03/01;19(1):23-35. doi: https://doi.org/10.1007/s10742-018-0192-5.

78. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence — What Is It and What Can It Tell Us? New England Journal of Medicine. 2016;375(23):2293-7. PMID: 27959688. doi: https://doi.org/10.1056/NEJMsb1609216.

79. Relton C, Torgerson D, O'Cathain A, Nicholl J. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. BMJ. 2010;340:c1066. doi: https://doi.org/10.1136/bmj.c1066.

80. Berk RA, Rauma D. Capitalizing on Nonrandom Assignment to Treatments: A Regression-Discontinuity Evaluation of a Crime-Control Program. Journal of the American Statistical Association. 1983 1983/03/01;78(381):21-7. doi: https://doi.org/10.1080/01621459.1983.10477917.

81. Colli A, Pagliaro L, Duca P. The ethical problem of randomization. Internal and Emergency Medicine. 2014 2014/10/01;9(7):799-804. doi: https://doi.org/10.1007/s11739-014-1118-z.

82. Gill P, Baillie J. Interviews and focus groups in qualitative research: an update for the digital age. British Dental Journal. 2018 2018/10/01;225(7):668-72. doi: https://doi.org/10.1038/sj.bdj.2018.815.

83. Goodman C, Evans C. Focus Groups. In: Gerrish K, Lathlean J, editors. The Research Process in Nursing. Oxford: John Wiley & Sons, Ltd.; 2015. p. 401-12.

84. Global Observatory for eHealth. Global diffusion of eHealth: Making universal health coverage achievable. Report of the third global survey on eHealth. Geneva: World Health Organization; 2016.

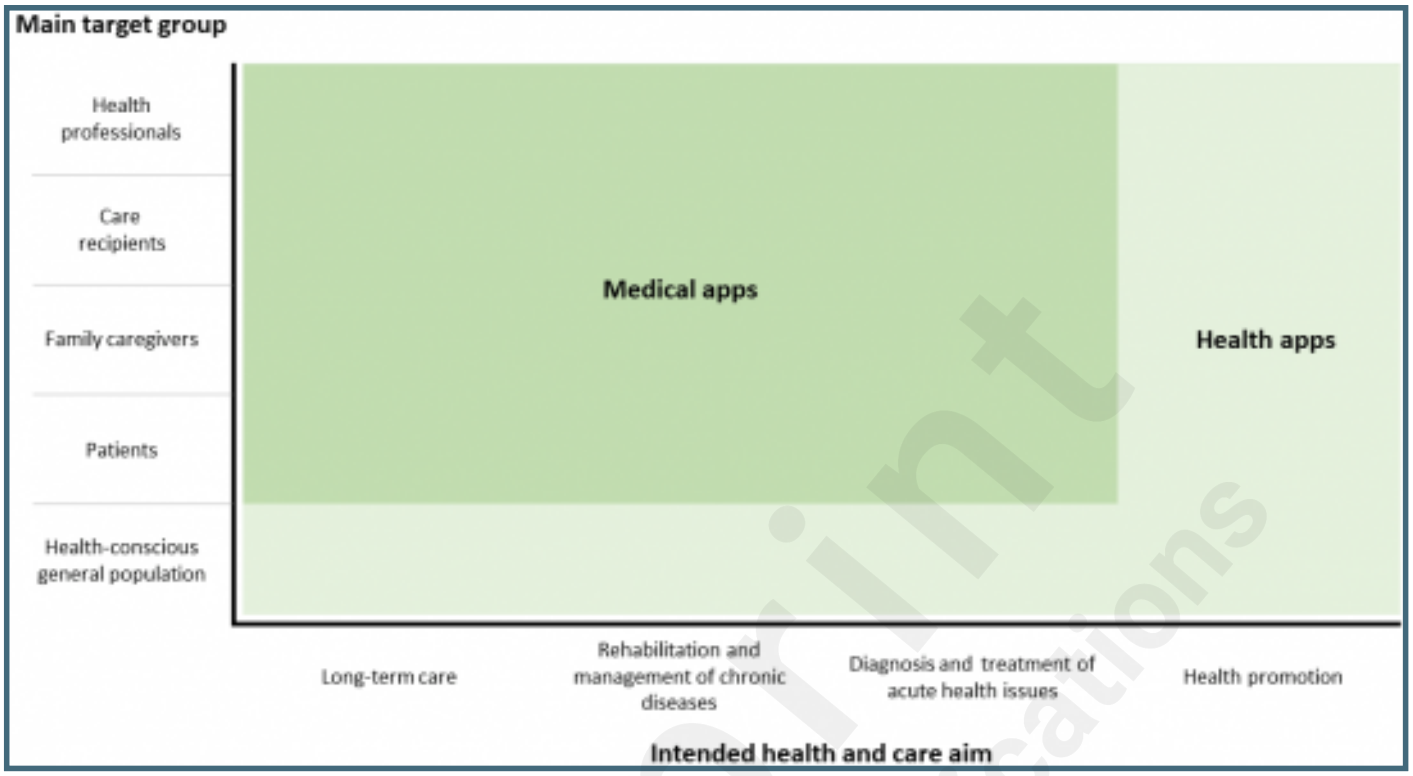
85. Osei E, Mashamba-Thompson TP. Mobile health applications for disease screening

and treatment support in low-and middle-income countries: A narrative review. Heliyon. 2021;7(3):e06639. doi: 10.1016/j.heliyon.2021.e06639.
86. Woulfe F, Fadahunsi KP, Smith S, Chirambo GB, Larsson E, Henn P, et al. Identification and Evaluation of Methodologies to Assess the Quality of Mobile Health Apps in High-, Low-, and Middle-Income Countries: Rapid Review. JMIR Mhealth Uhealth. 2021;9(10):e28384. PMID: 34636737. doi: 10.2196/28384.

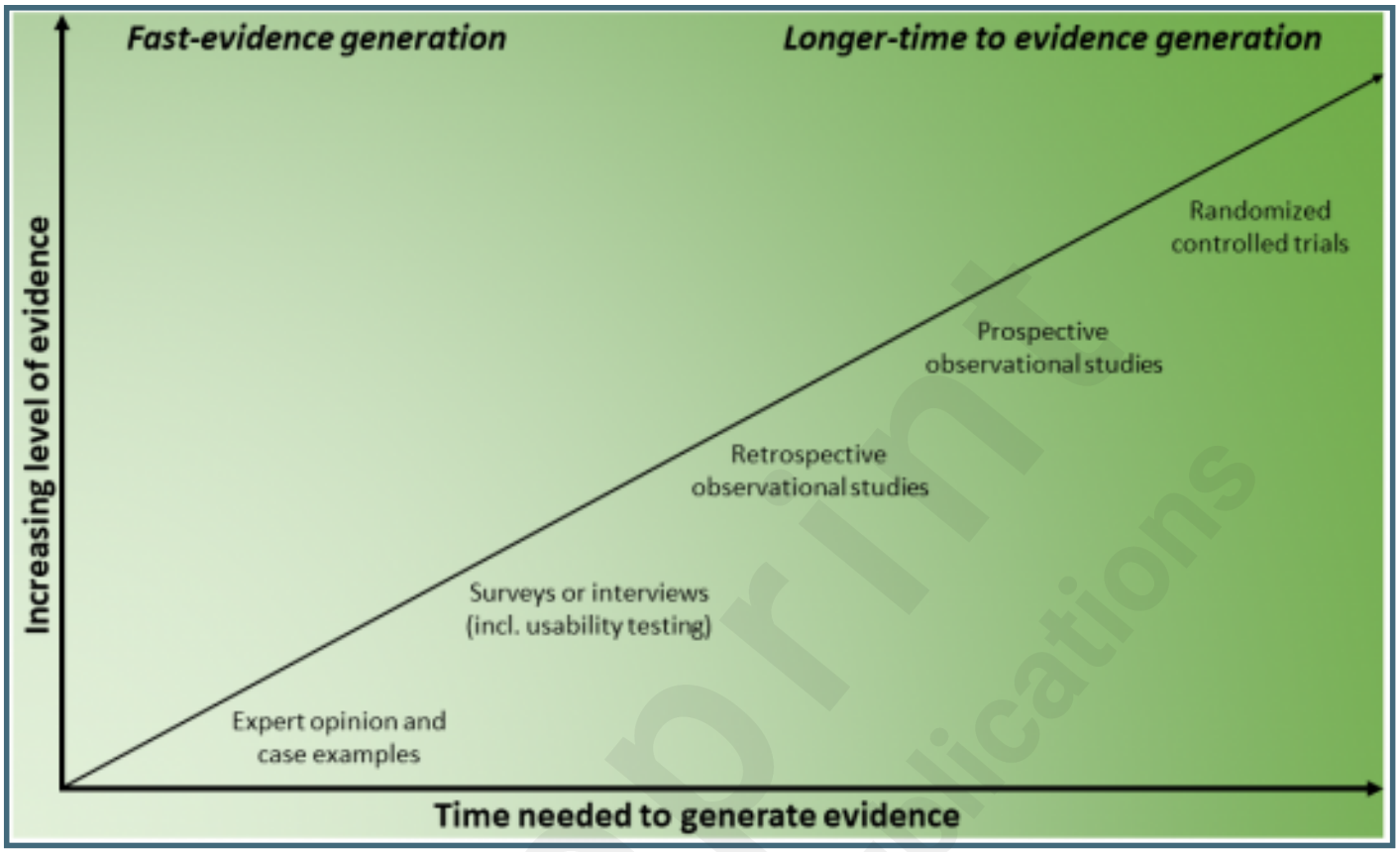
Supplementary Files

Figures

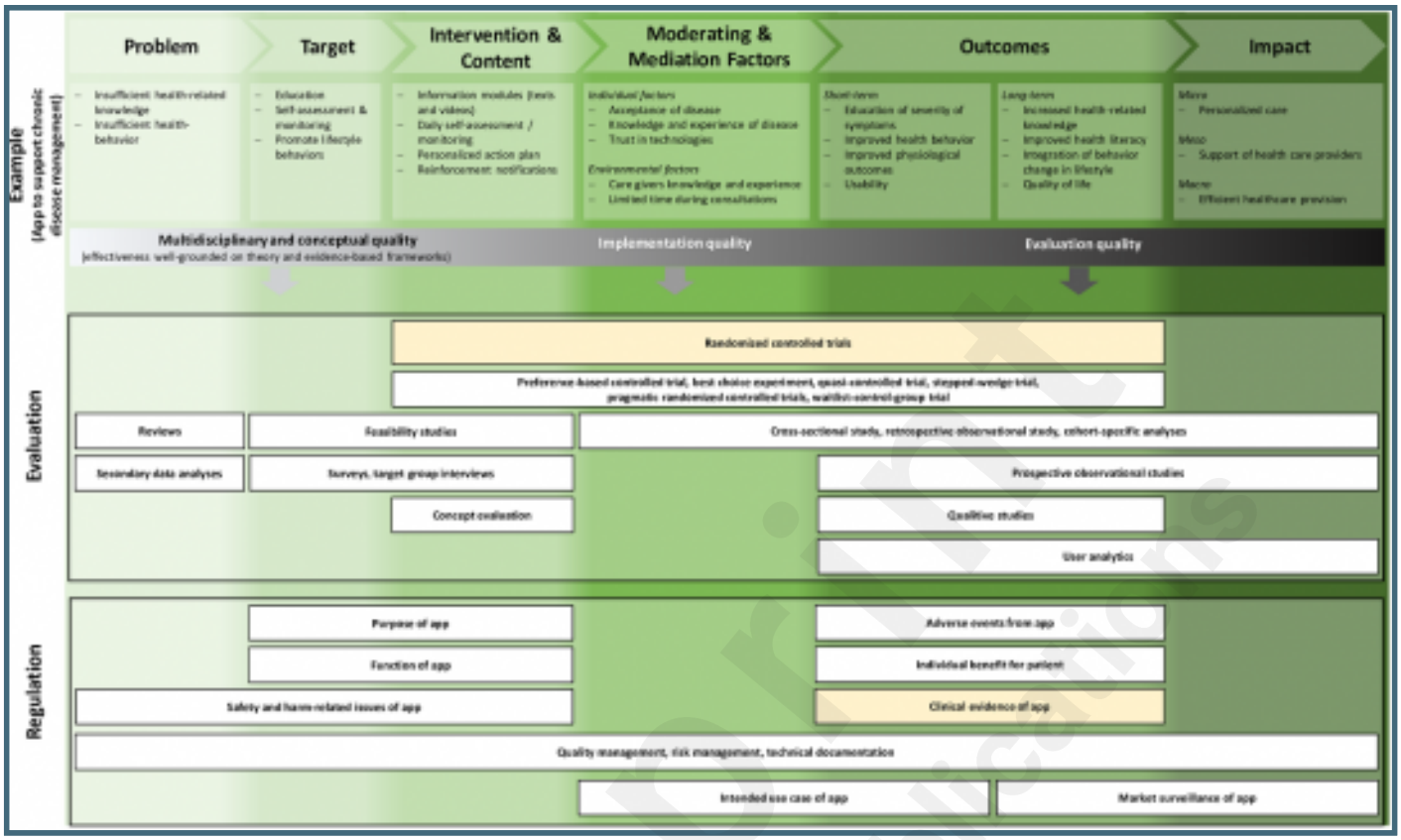
Medical apps as a sub-group of health apps. Adapted from Maaß et al. (2022).



The balancing act between increasing evidence levels and the time needed to generate them. Adapted from Guo et al. (2020).



Logic model for evaluating and regulating medical apps through the intervention's life-cycle phases.



Multimedia Appendixes

Coreq-Checklist.
URL: http://asset.jmir.pub/assets/46c77831a29164fb16dce2eb31fe9ee6.pdf

Transcribed flipcharts of the focus group discussions.
URL: http://asset.jmir.pub/assets/3db125bf2fa0ce2ad14e84f729e0a872.pdf

Glossary of technical terms given during the focus group discussions.
URL: http://asset.jmir.pub/assets/50d23cbd3a63448af796cb37a578d7ca.pdf

TOC/Feature image for homepages

Download Doctor, Medical, Medicine.