Recommendations for developing a lifecycle, multidimensional assessment framework for mobile medical apps

Received: 6 August 2021 | Revised: 18 February 2022 | Accepted: 2 March 2022

DOI: 10.1002/hec.4505

Health Economics WILEY

SUPPLEMENT PAPER

Recommendations for developing a lifecycle, multidimensional assessment framework for mobile medical apps

Rosanna Tarricone[1,2] | **Francesco Petracca**[1] | **Maria Cucciniello**[2,3] | **Oriana Ciani**[1,4]

[1]Centre for Research in Health and Social Care Management (CERGAS), Government, Health and Non Profit Division, SDA Bocconi, Milan, Italy

[2]Department of Social and Political Sciences, Bocconi University, Milan, Italy

[3]University of Edinburgh Business School, Scotland, UK

[4]Institute of Health Research, University of Exeter Medical School, Exeter, UK

Correspondence
Rosanna Tarricone, SDA Bocconi School of Management, Via Sarfatti 10, 20136 Milan, Italy; Department of Social and Political Analysis, Bocconi University, Via Roentgen 1, 20136 Milan, Italy.
Email: rosanna.tarricone@unibocconi.it

Funding information
European Commission; European Union's Horizon 2020 research and innovation program, Grant/Award Number: 779306
Open Access Funding provided by Universita Bocconi within the CRUI-CARE Agreement.

Abstract

Digital health and mobile medical apps (MMAs) have shown great promise in transforming health care, but their adoption in clinical care has been unsatisfactory, and regulatory guidance and coverage decisions have been lacking or incomplete. A multidimensional assessment framework for regulatory, policymaking, health technology assessment, and coverage purposes based on the MMA lifecycle is needed. A targeted review of relevant policy documents from international sources was conducted to map current MMA assessment frameworks, to formulate 10 recommendations, subsequently shared amongst an expert panel of key stakeholders. Recommendations go beyond economic dimensions such as cost and economic evaluation and also include MMA development and update, classification and evidentiary requirements, performance and maintenance monitoring, usability testing, clinical evidence requirements, safety and security, equity considerations, organizational assessment, and additional outcome domains (patient empowerment and environmental impact). The COVID-19 pandemic greatly expanded the use of MMAs, but temporary policies governing their use and oversight need consolidation through well-developed frameworks to support decision-makers, producers and introduction into clinical care processes, especially in light of the strong international, cross-border character of MMAs, the new EU medical device and health technology assessment regulations, and the Next Generation EU funding earmarked for health digitalization.

KEYWORDS
assessment, digital health, eHealth, HTA, lifecycle, mHealth, mobile medical apps, regulatory

1 | INTRODUCTION

The COVID-19 pandemic has unquestionably accelerated the transition to digital health (DH) (Pandey & Pal, 2020; Petracca et al., 2020). More specifically, mobile apps have given health care systems the opportunity to support public health surveillance during the most hard-hit periods, at once emphasizing the need to picture and build cutting-edge care models for the upcoming future.

The promise of mobile apps to manage individual health created growing expectations over time, leveraging on the value of personalized data captured outside of traditional hospital and clinical settings (Ku & Sim, 2021). However, their actual adoption in clinical care was still unsatisfactory before the pandemic. While the uptake of DH solutions during the pandemic was widespread and shaped by the organizational and institutional systems in which it took place (Petracca et al., 2020), it was largely favored by temporary, flexible policies that predominantly affected low-risk technologies with high potential for patients (Kadakia et al., 2020). For example, the United States (US) Food and Drug Administration (FDA) issued temporary guidance to expand access to digital therapeutics for psychiatric disorders without the submission of a pre-market notification (FDA, 2020e).

International agencies have discussed the need for regulatory guidelines to classify mobile health apps and detail their pre-market route. However, since many mobile apps at present are not considered medical devices, they have largely escaped oversight, resulting in calls for expanded regulation (Kasperbauer & Wright, 2020). Similarly, for complex digital technologies, such as those based on artificial intelligence and machine learning, little regulatory guidance is available (Muehlematter et al., 2021).

In addition, where regulatory frameworks exist, they have rarely been followed by formal reimbursement or coverage arrangements. Likewise, there is no common framework for assessment purposes in the scholarly debate, with those available often proposing generic assessment models for all digital services, which fail to properly weigh the specifics of each single digital technology. Most of the contributions are merely rating tools aimed at supporting patients and clinicians in choosing the best option, often a mobile health app (Hensher et al., 2021; Lagan et al., 2020; Lagan et al., 2021). In fact, a systematic review of the scientific literature showed that available evaluation frameworks for mobile apps are not suitable for use in health technology assessment (HTA), as none cover all of the HTA core domains (Moshi et al., 2018). As a consequence, it is still unclear how to disseminate appropriate app interventions to patients and providers (Leigh & Ashall-Payne, 2019), despite having amply identified strategies and practicalities for their thorough diffusion and integration into routine clinical practice (Gordon et al., 2020). Without full integration into the care continuum alongside traditional technologies and interventions, DH fails to reach its highest possible value (Gordon et al., 2020). For mobile apps to make the leap to routine use for clinical purposes, regulators and agencies need to ensure a robust pathway for their adoption, embracing the universal paradigm of value to enable an unbiased comparison with other health care technologies competing for the same resources.

Herein, we focus on mHealth apps within the digital medicine (DM) area, namely mobile medical apps (MMAs), defined by the FDA as mobile apps that incorporate device software functionalities (FDA, 2019b). This allows for simultaneously providing a technological specification—that of software accessed through specific platforms such as smartphones and tablets—and a narrowing of attention from DH to DM, as such excluding wellness and support products not requiring the same degree of regulatory oversight. MMAs also include digital therapeutics (DTx) accessible through mobile platforms such as smartphone and tablets, as well as artificial intelligence and/or machine learning (AI/ML)-based medical apps (see Box 1 for definitions of DH solutions).

Like integrated approaches to decision-making already proposed for medical devices (Tarricone et al., 2020), a dynamic approach that looks at the entire lifecycle of digital technologies and takes into account their distinctive features is needed (Tarricone et al., 2021) (i.e., one that could foster a participatory approach and surpass the sharp distinction between the pre-market and post-market stages). As for medical devices (MDs), also the MMAs lifecycle consists of four main stages: the (i) pre-clinical, pre-market; (ii) clinical, pre-market; (iii) diffusion, post-market; and (iv) obsolescence & replacement. Similarly to MDs, the lifecycle approach for MMAs, to be effective, requires anticipation of what evidence will be needed at each stage (Tarricone et al., 2020). What matters in the MMAs case are the type of evidence and the length of the stages that may differ from conventional MDs thus making their assessment even more challenging. Experimental studies are – for instance - even more widely inapt for MMAs, that can evolve very rapidly, incorporate continuous measurement of the intervention effects, and need to be constantly readjusted just to keep up with their intended use. This consideration also affects the obsolescence stage that, for MMAs, is easier to be replaced by software adaption for instance. Moreover, retention is one of the most significant issues for any MMAs, as only small portions of users show consistent access to mHealth apps over time (Tarricone et al., 2021).

Given this lack of comprehensive guidance on the evaluation of MMAs and the concurrent need to provide thorough evidence to sustain their deployment, we aim at providing principles and recommendations for developing a multidimensional (e.g., clinical, economic, usability, equity, etc. dimensions) assessment framework to: (i) address all relevant domains that should be considered in the assessment of MMAs, (ii) leverage on the specific features of MMAs, (iii) be compatible with regulatory and HTA purposes, and (iv) inform policy decisions on MMAs.

2 | METHODS

In developing a multidimensional framework for the assessment of MMAs, with a view to support policy decisions on their sustained uptake and diffusion within healthcare systems, we followed a multistep approach.

2.1 | Policy document review

First, with the aim of mapping the current state of the art in terms of MMA assessment from a regulatory and coverage or reimbursement point of view, we conducted a targeted or a state-of-the-art review (Grant & Booth, 2009) of the most relevant policy documents published by international organizations (i.e., the World Health Organization [WHO] and the International Medical Device Regulators Forum [IMDRF], a voluntary network of 10 jurisdictions that includes the US and the European Union [EU] and seeks to harmonize medical device assessment), supranational institutions (such as the EU), single member states, and the US. In this mapping, the FDA in the US was explicitly included as a leading regulatory organization, responsible for ensuring the safety, efficacy, and security of drugs, biological products, and medical devices, and as a pioneer institution in embracing the lifecycle logic in the evaluation of health care technologies. Public sources, newsletters, and websites were regularly searched for regulatory or HTA guidelines specifically dedicated to MMAs throughout the COMED project, from March 2018 to May 2021, complemented by snowballing and inclusion of documents already known to the research team and published in English.

2.2 | Expert focus group

The results of the review, along with the identification of the main features that distinguish MMAs from other health care technologies (Tarricone et al., 2021), provided the basis for preliminary development of an initial set of recommendations, which were subsequently discussed within an international focus group comprised of key relevant stakeholders to solicit their input and feedback. Eight participants represented all key stakeholder groups (i.e., industry, patient associations, HTA agencies, policy makers, and academia) (Table A1).

The focus group methodology was selected as a suitable means to collect feedback and solicit input due to: i) its flexibility, ii) the ability to directly interact with domain experts and probe them on key design ideas, iii) the ability to collect both qualitative and quantitative data, iv) the fact that valuable information is created in the interaction between participants (Krueger & Casey, 2014).

Initially proposed as a 1-day, in-person workshop, the March 30, 2020, meeting was shortened to a half-day online workshop due to COVID-19 restrictions. Participants were presented with slides, summarizing the distinguishing features of MMAs and the proposed recommendations for their assessment, followed by discussion sessions, divided into two main parts: first, addressing the relevant domains that should be considered in assessing MMAs, in line with their specific features, and second, discussing each of the 10 recommendations identified by the research team.

Participants shared their views and suggested modifications to refine and improve the assessment framework. Based upon the panel's input, the set of recommendations was finalized and eventually shared within the panel.

3 | RESULTS

3.1 | Assessment of MMAs from a policy point of view: State of the art

In this section, we comparatively analyze the assessment methods and procedures for MMAs adopted by key international organizations such as the IMDRF and the WHO, jurisdictions such as the US and the EU, and other countries such as the United Kingdom (UK), France, and Germany that—during the COMED project—have issued frameworks or guidance documents relevant for our purposes (Tables 1, 2 and 3) in English.

3.1.1 | Assessment frameworks: Aims and targets

First, it is important to note that the assessment frameworks developed by these jurisdictions (i.e., IMDRF, FDA, European Commission [EC], and MHRA) have different purposes (i.e., in some cases, the goal is to regulate market access of DH, whilst in others, it is to help decide upon coverage and/or reimbursement/procurement) (Table 1). As to the target object, the guidance documents cover broad definitions of DH, ranging from the WHO definition, which refers to the use of digital, mobile, and wireless technologies to support the achievement of health objectives, to the more specific definition adopted by Federal Institute for Drugs and Medical Devices in Germany (Bundesinstitut für Arzneimittel und Medizinprodukte, BfArM), which focuses on the EU Medical Device Regulation (MDR) Class I or IIa medical devices achieving their medical purpose through core digital functions, used either by patients alone or patients with health care professionals. However, in all cases, the critical aspect is the link between the software and its impact on health, without which the digital solution ceases to be the target of the assessment framework.

3.1.2 | Classification criteria

The classification of DH tends to be risk-based, inspired by the work done by the IMDRF. In 2014, the IMDRF defined the concept of Software as Medical Device (SaMD) and grouped them according to the level of risk into four classes (IMDRF, 2014) (Table 2). A different approach was adopted by the National Institute for Care and Health Excellence (NICE) in the UK that lists 10 classes of apps, software, and programs grouped in three evidence tiers (NICE, 2018, 2021) (Table 2). The only exception to the risk-based classification is that of the WHO, which, instead, identifies a taxonomy based on the different ways in which digital and mobile technologies are being used to support health system needs (WHO, 2018) (Table 2).

3.1.3 | Assessment approaches

Major differences can, however, be found in assessment approaches and domains (Table 3). As to the approach, the majority of frameworks advise some form of iterative process but differ in how they are formulated, ranging from a linear, step-wize method to a clockwise total product lifecycle from the US FDA. The FDA arguably represents the most innovative regulation system and focuses on, differently from other jurisdictions, both the product and the manufacturer, witnessed by the launch of the Digital Health Center of Excellence (DHCoE) (FDA, 2020b) within the Center for Devices and Radiological Health (CDRH), aimed at providing regulatory advice and support to the FDA's regulatory review of DH technology. The establishment of the DHCoE is consistent with the Digital Health Innovation Plan (FDA, 2020c) that aims—among others—to foster FDA's strengths and expertize in CDRH's DH unit and re-imagine the assessment process of device software functions. For this purpose, the Software Pre-Certification Program (or Pre-Cert program) (FDA, 2019a) was conceived with the aim of providing either pre-market review exemptions to lower risk technologies or a streamlined pathway for higher-risk products to foster system efficiency (Table 3). The first step of the Pre-Cert program is the excellence appraisal, which represents the essence of this "firm-based" approach, whereby developers are audited and pre-certified as long as they "have demonstrated a robust culture of quality and organizational excellence (CQOE) and are committed to monitoring real-world performance" (FDA, 2019a). The CQOE assessment is based on demonstration of five excellence principles (Table 1), and each domain is measured against a set of key performance indicators (KPIs). If deemed eligible, organizations receive either Pre-Cert levels L1 or L2. (Level 1 Pre-Cert is granted in case of objective demonstration of excellence across all five principles but with limited track record in the development, delivery, and maintenance of products, whereas Level 2 Pre-Cert is granted to those with more extensive experience). In the second step, the results of the excellence appraisal (i.e., recognition of L1 or L2 status) are coupled with an IMDRF-inspired risk categorization framework for the sake of review pathway determination. In general, lower-risk devices will be directly marketed, whereas higher-risk devices will have to undergo the third step of a streamlined pre-market review. Finally, both the product and the developer are monitored throughout the fourth step, real-world performance, which aims at verifying ongoing excellence and generation of clinical evidence through post-market data. In 2019, a pilot test was launched with 9 companies of varying sizes, age, and organizational structure,[1] the findings of which will help inform large-scale beta testing.[2] As of 2020, the pilot has allowed to iteratively test different methods and approaches through mock excellence appraisals, aggregating results in a library of activities, processes, and KPIs used by high-performing organizations. Preliminary results also helped understand how to operationalize the real-world performance domain, particularly regarding user experience and product performance

TABLE 1 Target focus of regulatory and assessment approaches across jurisdictions

	IMDRF	WHO	EU	FRANCE	GERMANY	USA	UK	
Organization	IMDRF	WHO	EC	HAS	BfArM	FDA	MHRA	NICE, NHSX
Purpose	Regulatory approval	Assessment for Decision-Making	Regulatory approval	Assessment for Coverage and Reimbursement Decisions	Assessment for Coverage and Reimbursement Decisions	Regulatory approval	Regulatory approval	Assessment for Coverage and Procurement Decisions
Target Object	Software as Medical Device (SaMD)	Digital Health (DH)	Medical Device Software (MDS)	Health Apps and Smart Devices (mHealth)	Digital Health Application (DIGA)	Device Software Functions (DSF), including Mobile Medical app (MMA)	Medical Device Standalone Software	Digital Health Technologies (DHT)
Definition	Software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device	The use of digital, mobile, and wireless technologies to support the achievement of health objectives. Digital health is inclusive of both mHealth and eHealth	Software that is intended to be used, alone or in combination, for a purpose as specified in the definition of a "medical device" in the Medical Device Regulation (MDR) or In Vitro Diagnostics Regulation (IVDR), regardless of whether the software is independent or driving or influencing the use of a device	Medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices Smart devices are devices connected to the internet that can collect, store, process and send data or that can take specific actions based on information received	EU MDR Class I or IIa medical devices achieving their medical purpose through core digital functions and used either by patients alone or patients with healthcare professionals.	All SAMD and SIMD that are not used for administrative support of a healthcare facility, for maintaining or encouraging a healthy lifestyle, to serve as electronic patient records, or for transferring, storing, converting formats, or displaying data	Software not incorporated in a device possessing its own medical purpose	Apps, programmes and software used in the health and care system that may be standalone or combined with other products such as medical devices or diagnostic tests

TABLE 1 (Continued)

	IMDRF	WHO	EU	FRANCE	GERMANY	USA	UK	
Main Reference Documents	Software as a Medical Device (SaMD): Key Definitions (IMDRF, 2014)	Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment (WHO, 2016)	Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR (European Union, 2017a, 2017b)	Good Practice Guidelines on Health Apps and Smart Devices (Mobile Health or mHealth) (Haute Authorité de Santé, 2016)	The Fast-Track Process for Digital Health Applications (DiGA) according to Section 139e SGB V: A Guide for Manufacturers, Service Providers and Users (BfArM, 2020b)	Policy for Device Software Functions and Mobile Medical Applications (FDA, 2019b)	Guidance: Medical device stand-alone software including apps (including IVDMDs) (Medicine and Healthcare Products Regulatory Agency, 2021)	Evidence Standards Framework for Digital Health Technologies (NICE, 2018)

TABLE 2 Risk-categorization models developed across jurisdictions

Jurisdictions	IMDRF	WHO	EU	FRANCE	GERMANY	USA	UK
Main reference documents	"Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations (IMDRF, 2014)	Classification of digital health interventions v1.0. A shared language to describe the uses of digital technology for health (WHO, 2018)	Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR (European Union, 2017a, 2017b)	Good Practice Guidelines on Health Apps and Smart Devices (Mobile Health or mHealth) (Haute Authorité de Santé, 2016)	The Fast-Track Process for Digital Health Applications (DiGA) according to Section 139e SGB V. A Guide for Manufacturers, Service Providers and Users (BfArM, 2020b)	Developing a Software Pre-Certification Program: A Working Model (FDA, 2019b)	Evidence Standards Framework for Digital Health Technologies (NICE, 2018, 2021)
Risk categorization	Two-dimensional: 1. significance of the information provided by the SaMD to the healthcare decision 2. state of the healthcare situation or condition	DH is classified according to the different ways in which DH interventions are being used to support health system needs and not on risk level	Two-dimensional: 1. significance of the information provided by the SaMD to the healthcare decision 2. state of the healthcare situation or condition	Two-dimensional: 1. main target user: - general public - patient/carers/family/ patient associations - healthcare professionals directly with their patients - healthcare professionals directly with their peers (e.g., teamwork) 2. main intended use: - information, general advice - primary prevention, health promotion, manual data entry and acquisition without analysis - secondary and tertiary prevention, therapeutic patient education - analysis of data, medical evaluation contributing to assessment, diagnosis, monitoring throughout the care pathway, impact on treatment	Two-dimensional: 1. significance of the information provided by the SaMD to the healthcare decision 2. state of the healthcare situation or condition	Two-dimensional: 1. significance of the information provided by the SaMD to the healthcare decision 2. state of the healthcare situation or condition	One-dimensional: (1) Detailed functional use

TABLE 2 (Continued)

Jurisdictions	IMDRF	WHO	EU	FRANCE	GERMANY	USA	UK
Risk classes	Categories I, II, III, IV	DH are classified around the target audience: (1) Clients (2) Healthcare providers (3) Health systems managers (4) Data services	Classes I, IIa, IIb, III (and A, B, C, and D for IVD). Rule 11 has been introduced to address the risks related to the information provided by Medical Device Software (MDSW) and is divided into three sub-rules (an analogous reasoning is applied to IVDR MDSW with classes A, C, D): - 11a: intended to provide information which is used to take decisions with diagnostic or therapeutic purposes (in this case MDSW is classified as class IIa, except if such decisions have an impact that may cause death or an irreversible deterioration of a person's state of health, in which case it is in class III; or a serious deterioration of a person's state of health or a surgical intervention, in which case it is classified as class IIb); - 11b: intended to monitor physiological processes or parameters (in this case MDSW is classified as class IIa, except if it is intended for monitoring of vital physiological parameters, where the nature of variations of those parameters is such that it could result in immediate danger to the patient, in which case it is classified as class IIb); - 11c: all other uses (in this case it is classified as class I).	Low Criticality, Medium Criticality, High Criticality	Classes I and IIa	Categories I, II, III, IV	10 Classes grouped in Evidence Tiers A, B, C. • **Tier A**: system impact ◦ system services: DHTs with no measurable patient outcomes but which provide services to the health and social care system • **Tier B**: understanding and communicating ◦ inform: provides information, resources or activities to the public, patients or clinicians; includes information about a condition or general health and lifestyle ◦ health diaries: includes general health monitoring using fitness wearables and simple symptom diaries ◦ communicate: allows 2-way communication between citizens, patients or healthcare professionals • **Tier C**: interventions ◦ preventative behavior change: address public health issues like smoking, eating, alcohol, sexual health, sleeping and exercise ◦ self-manage: allows people to self-manage a specified condition; may include behavior change techniques ◦ treat: provides treatment; guides treatment ◦ active monitoring: tracking patient location, using wearables to measure, record or transmit data (or both) about a specified condition; uses data to guide care ◦ calculate: a calculator that impacts on treatment, diagnosis or care diagnose: diagnoses a specified condition; guides diagnoses.

TABLE 3 Assessment approaches across jurisdictions

Jurisdictions	IMDRF	WHO	EU	FRANCE	GERMANY	USA	UK
Organization	IMDRF	WHO	EC	HAS	BfArM	FDA	NICE NHSX
Main reference documents	Software as a Medical Device (SaMD): Clinical Evaluation (IMDRF, 2017) Software as a Medical Device (SaMD): Application of Quality Management System (IMDRF, 2015)	Classification of digital health interventions v1.0. A shared language to describe the uses of digital technology for health (WHO, 2018)	Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR (Medical Device Coordination Group, 2019) Guidance on Clinical Evaluation (MDR)/ Performance Evaluation (IVDR) of Medical Device Software (Medical Device Coordination Group, 2020)	Good Practice Guidelines on Health Apps and Smart Devices (Mobile Health or mHealth) (Haute Authorité de Santé, 2016)	The Fast-Track Process for Digital Health Applications (DiGA) according to Section 139e SGB V: A Guide for Manufacturers, Service Providers and Users (BfArM, 2020b)	Developing a Software Pre-Certification Program: A Working Model (FDA, 2019b)	Evidence Standards Framework for Digital Health Technologies (ESFDHT) (NICE, 2018, 2021) Digital Technology Assessment Criteria for Health and Social Care (DTAC) (NHSX, 2021)
Target focus	Product	Product	Product	Product	Product	Product & Firm	Product
Assessment approach	Clinical Evaluation Model embedded in SaMD Realization and Use Processes	Lifecycle Approach	Continuous and iterative clinical (MDR)/ performance (IVDR) evaluation embedded in the QMS (IMDRF inspired)	Step-Wise Evaluation model based on:	Fast-track process for the assessment of DiGA, focusing on product quality and evidence underlining positive healthcare effects	Total Product Lifecycle (TPLC) Approach for continued evaluation from pre-market development to post-market performance	Risk-based model for evaluation, with cumulative evidence requirements for demonstrating effectiveness and economic impact

TABLE 3 (Continued)

Jurisdictions	IMDRF	WHO	EU	FRANCE	GERMANY	USA	UK
Assessment domains/ steps	(1) Clinical Validity (2) Analytical Validation (3) Clinical Validation	**Monitoring**: - Functionality, i.e., the ability to support the desired intervention; - Stability, i.e., the ability to remain functional both under normal and anticipated peak conditions for data loads. - Fidelity, i.e., whether the intervention is delivered as intended from both a technical and user perspective. Monitoring fidelity consists of three different categories: a) monitoring the technical fidelity in the implementation process; b) monitoring external barriers that might cause the system not to function as expected; c) monitoring the compliance of the digital health system users; - Quality, i.e., the overall quality of the intervention in terms of excellence, values, conformance, fitness for purpose and ability to meet or exceed expectation. Two main aspects should be addressed when monitoring quality. The first one pertains to user capabilities, to guarantee users do enter information accurately and use the system correctly overall. The second dimension relates to the overall quality of the intervention as a prerequisite for the effectiveness of the intervention **Evaluation**: - **Feasibility**: Assess whether the digital health system works as intended in a given context; - **Usability**: Assess whether the digital health system is used as intended; - **Efficacy**: Assess whether the digital health intervention achieves the intended results in a research (controlled) setting; - **Effectiveness**: Assess whether the digital health intervention achieves the intended results in a non-research (uncontrolled) setting; - **Implementation research**: Assess the uptake, institutionalization, and sustainability of evidence-based digital health interventions in a given context, including policies and practices	(1) Planning Evidence Generation (2) Demonstrating Clinical association to the outcomes (3) Updating Clinical Evaluation through Post-Marketing Clinical Follow-Up	(1) Informing User: - Description - Consent (2) Health Content: - Design of initial content - Standardization - Generated content - Interpreted content (3) Technical Content: - Technical design - Data flow (4) Security/ Reliability: - Cybersecurity - Reliability - Confidentiality (5) Usability/Use: - Usability/ design - Acceptability - Integration/ import	(1) Mandated Requirements: - Security - Functionality - Quality - Data Protection - Data Security - Interoperability (2) Positive Care Effects: - Medical benefits - patient-relevant structural and procedural improvements If the assessment process is positive the product is included in the DiGA Directory, a repository of apps available for prescription Type of Recommendations: (1) Permanent acceptance (2) Preliminary Acceptance	(1) Excellence appraisal (2) Review Pathway Determination (3) Streamlined Pre-Market Review (4) Real World Performance The Excellence Appraisal introduces a firm-based assessment approach based upon: (1) Product Quality (2) Patient Safety (3) Clinical responsibility (4) Proactive Culture	ESF: (1) Clinical Effectiveness (2) Economic Impact DTAC: (1) Clinical safety (2) Data protection (3) Technical assurance (4) Interoperability (5) Usability and accessibility

(Continues)

TABLE 3 (Continued)

Jurisdictions	IMDRF	WHO	EU	FRANCE	GERMANY	USA	UK
Economic domains	N/A	N/A	N/A	N/A	N/A	N/A	Economic impact relative to the financial risk: (1) key economic information (2) appropriate economic analysis (3) economic analysis reporting standards
Study Designs	N/A	Hybrid approaches	Prospective study may be required for higher risk MDSW, whereas retrospective analyses may be sufficient for lower risk ones	N/A	Minimum requirement of one retrospective comparative study (although prospective ones are more appreciated) and must be conducted, at least partially, in Germany Recognized relevance of alternative study designs (e.g. PCT, SMART, or MOST)	N/A	Tier C: High quality observational or quasi-experimental studies (minimum), High quality (quasi-) experimental intervention studies (best practice), High quality RCT or meta-analysis of RCTs (best practice)
Real World Data	Clinical evaluation should be intended as a dynamic summary that changes through continuous RWP learning	Clinical evaluation must be generated as a blend effectiveness and implementation trial elements	Manufacturers are asked to continuously monitor the device safety, effectiveness and performance	N/A	Declared preference for RWD based on healthcare practice	RWP Analytics Framework (1) Real World Health analytics (2) User Experience Analytics (3) Product Performance Analytics	Request for ongoing data collection to show DHT usage and value

(Continues)

TABLE 3 (Continued)

Jurisdictions	IMDRF	WHO	EU	FRANCE	GERMANY	USA	UK
Reference to AI/ML technologies	IMDRF Artificial Intelligence Medical Devices (AIMD) Work Item (IMDRF, 2021) Work in progress	N/A	The White paper on Artificial Intelligence is a non-industry specific document that discusses the adoption of AI across sectors and activities (EC, 2020). The most relevant aspect for SaMD is that of suggesting a conformity assessment for high-risk AI applications	N/A	N/A	Artificial Intelligence and Machine Learning in Software as a Medical Device (FDA, 2021) Two principles: - SaMD Pre-Specifications - Algorithm Change Protocol Two *ex-novo* regulatory pathways: - Document Approach - Focused FDA Review Pathway	An additional Evidence Framework is expected as from the collaboration between the NICE and MHRA

analytics, but further real-world health analytics investigation is still needed. Measurement of health benefits has also proven to be challenging, and additional work to identify these measures is needed (FDA, 2020a).

A similar commitment toward establishing ad hoc approaches for DH can be found in the EU, where a number of policy cooperation platforms have also been established, such as the eHealth Network, focusing on interoperability and standardization; the eHealth Action (eHAction), providing technical and scientific advice to the network; the eHealth Stakeholder Group (eHSG), contributing to the development of eHealth policy; and the Joint Action for the European Health Data Space (TEHDAS), supporting the EC plan for the European Health Data Space (EHDS). However, the assessment approach of mHealth apps does not truly differentiate from the one introduced for conventional medical devices by the new EU MDR (EU 2017/745), implemented in May 2021, and In Vitro Diagnostics (IVDR) (EU 2017/746) to be implemented in May 2022 (Table 1), which suggest drafting an iterative clinical development plan across the pre- and post-market phases (Table 3).

3.1.4 | Domains of assessment

As to the domains' target of assessment, the jurisdictions vastly differ. The clinical dimension is present in all frameworks and mainly pertains to assessment of the product's impact on users' health. However, besides the general indication of proportionality with the level of risk, few recommendations can be found as to how to assess the clinical impact. The WHO vaguely refers to hybrid approaches, by blending effectiveness and implementation trial elements (Table 3). The German Digital Health Application (DiGA) Guide (BfArM, 2020b) is the most precise and sets a preference for (possibly domestically generated) real-world data based on health care practice rather than trial settings, as demanded by the interactivity of MMAs with their surrounding environment (Table 3). In proving positive care effects, although prospective comparative studies are more appreciated as they provide higher-level evidence, the minimum requirement is set at the submission of at least one retrospective comparative study (e.g., case-control studies, retrospective cohort studies, or intra-individual comparisons). The guide also recognizes the relevance of alternative study designs and methods such as pragmatic clinical trials (PCT), sequential multiple assignment randomized trial (SMART), or multiphase optimization strategy (MOST). Nonetheless, a converged and established approach for assessing the quality of the study design, especially with the respect to the selected comparator, has yet to be fully defined.

Other domains pertain to the technical contents, security and reliability, and usability, and are found, although under different definitions and with various levels of details, in most frameworks. The economic dimension is present only in the UK, where NICE published, in March 2019, its "Evidence Standards Framework (ESF) For Digital Health Technologies (DHTs)"; it was further updated in April 2021 (NICE, 2018, 2021). The ESF describes standards for the evidence that should be available, or developed, for DHTs to demonstrate their value in the UK health and social care system. This includes evidence of economic impact relative to financial risk. The NICE approach tries to provide a comprehensive guide for developers on how to build a strong economic case, as was done for innovative medical devices and diagnostics (Campbell et al., 2018). As a way of informing decision-making and investment decisions, NICE has identified three building blocks: (1) key economic information (i.e., user population size, care pathways—existing and proposed, and parameters for the economic model such as intervention, cost, resource use, and utilities), (2) appropriate economic analysis (i.e., budget impact analysis, cost-consequence analysis, and cost-utility analysis), and (3) economic analysis reporting seven standards (i.e., economic perspective, time horizon, discounting, sensitivity analyses, equity analyses, additional analytical methods, and critique of the economic analysis) (Table 3). A separate framework is also provided by the Digital Technology Assessment Criteria for Health and Social Care (DTAC), published by the NHSX—a joint unit of National Health Service (NHS) England and the Department of Health and Social Care—to provide clearer direction on both the development and procurement of sound DHT (NHSX, 2021). This is to be understood as a broader and more practical counterpart to the more theoretical ESF by NICE and provides a measurement of the overall success of the assessment of the product or service and adherence to NHS service standards (Table 3).

3.1.5 | Coverage and reimbursement

As to coverage and reimbursement decisions, the French National Committee for the evaluation of medical devices and health technologies (CNEDIMTS) assesses the added value of the app similarly to that of any other health technologies. For example, the MOOVCARE POUMON application (intended for medical telemonitoring of relapse and complications in patients with non-progressive lung cancer), upon positive evaluation, was included in 2020 in the French List of Products and Healthcare Services Qualifying for Reimbursement (LPPR) with Added Clinical Value (ASA) level III (moderate improvement)

compared to conventional care (i.e., monitoring by imaging and medical face-to-face consultations) (MedTech Reimbursement Consulting, 2020).

In a similar fashion, DiGA may be granted permanent or preliminary acceptance (i.e., when all requirements have been fulfilled but the positive care effect is not definitely proven). In this case, the company is granted a 12-month (or exceptionally 24-month) period to deliver the needed supplementary data and be inserted in the DiGA Directory, an online platform, similar to the NHS app library and to the French LPPR, intended for physicians and psychotherapists but also patients and their caregivers. Exhaustive information, such as the current status of evidence of the approved technologies, is provided through different levels according to the targeted audience, fostering transparency and allowing for more informed decisions. The DiGA Directory, hence, substantially provides an updated pool of apps that are available for prescription and reimbursable at the price set through negotiation between manufacturers and the Central Federal Association of the Health Insurance Funds.

3.1.6 | Artificial intelligence- and machine learning-based DH technologies

One last aspect concerns the consideration of AI/ML-based DH products. Although several jurisdictions recognize the importance of developing an ad hoc framework, only the FDA currently offers guidance on the AI/ML in software as a medical device (SaMD)(FDA, 2021) (Table 3). The framework proposes the creation of a "predetermined change control plan" that includes "*the types of anticipated modifications—"Software as a Medical Device Pre- Specifications"—based on the retraining and model update strategy, and the associated methodology—Algorithm Change Protocol—being used to implement those changes in a controlled manner that manages risks to patients.*" Two important principles are hence recognized. SaMD pre-specifications (SPS) are the pre-envisioned types of changes the developer plans to achieve through real-world learning in terms of SaMD performance, inputs, or intended use of AI/ML-based SaMD. The algorithm change protocol (ACP) delineates the steps that should be followed to ensure that the modification does not hinder device safety and effectiveness, accounting for four components: data management, re-training, performance evaluation, and procedures update. Conditional on the type of modification, three regulatory outcomes are possible. First, when the risk to patients of the modification is low, developers would simply have to document the modification and the analysis in the risk management and 510(k) files. Second, a focused FDA review when the technologies can be refined based on real-world learning and training but still within the same intended use. Third, a new 510(k) for pre-market review when none of the previous exemptions apply.

3.2 | Recommendations for developing a lifecycle, multidimensional assessment framework for MMAs

The review of the policy documents highlighted how several jurisdictions and international organizations are keen to govern the growing introduction of DH technologies and MMAs through issuing and repeatedly updating regulatory and/or guidance documents aimed at assessing digital technologies before or after they enter routine practice. However, as noted by the international focus group participants, a systematic, comprehensive framework that could serve diverse purposes ranging from regulation of market authorization to purchasing and appropriate and safe use of MMAs is lacking. Analysis of available guidance and the focus group discussion were used to identify a list of recommendations that should be considered in the development of an assessment framework for MMAs adopting a lifecycle approach to evidence generation and assessment (Table 4).

3.2.1 | Recommendation 1. MMA development: Shared decision-making approaches to the development of apps can enhance their replicability and ultimately lead to improved outcomes

The pre-deployment phase of MMAs starts from their development in terms of content and features embedded in the app. Development factors are multiple and typically include the adoption of health behavioral theories in the app design process, as well as user and health care professional involvement (Adu et al., 2018). Engaging with end-users and following robust theories could enable a more appropriate app design, as well as reduce the ineffective use of the app (Farao et al., 2020; McCurdie et al., 2012; Schnall et al., 2016). An increasing number of design methods for MMAs now exists, with different approaches to user needs and the establishment of user models (personas) as archetypes of groupings based on behaviors, attitudes, and goals to guide the functional design (Duan et al., 2020). Particularly for lower-risk apps that may never undergo a formal clinical and economic evaluation, the adoption of these development factors in the design stages could certify a first tier of quality to at least

TABLE 4 Multi-dimensional assessment framework of mobile medical apps (MMAs). The DECALOGUE

Domain	Recommendation
MMA development and update	Shared decision-making approaches to the development of apps can enhance the replicability of apps and ultimately lead to improved outcomes
Classification and evidentiary requirements	An unambiguous classification of MMAs should be associated with corresponding evidence generation requirements and the possibility to flexibly review the associated level of risk of every single MMA
Performance and maintenance monitoring	MMA manufacturers must ensure that technical system implementation does not threaten the overall effectiveness. Analytical validity must be determined by the manufacturers and a core set of indicators should be developed that are coherent with the MMA classification
Usability testing	Usability should be continuously monitored, both in the development phase of the solution and after its implementation in the field
Additional outcome domains	Besides conventional outcomes similar to all healthcare technologies, patient empowerment associated with use and potential environmental impact are distinguishing outcome domains of MMAs that, where relevant, should be appropriately measured and valued
Clinical evidence requirements	Flexible study designs that account for the specific characteristics of MMAs and can generate fast and efficient results should be adopted and coupled with flexible policy arrangements, based on the level of risk and the position in the product lifecycle of the app
Safety and security	The assessment of the risks related to data privacy, cyber security and misinformation and their potential impact are emphasized issues for MMAs
Equity considerations	Given the context of escalating inequalities and persisting technological divide, it is paramount to evaluate the net effect brought by MMAs on equity
Organizational assessment	The assessment of direct and indirect implications of the adoption of MMAs on the organizational level should focus on process, people, structure, culture or management impacts
Cost and economic evaluation	The evaluation of cost and economic impacts of MMAs should require no innovative forms and methodologies (reporting should follow established guidelines for economic evaluations), but rather new metrics for outcomes and different structures for cost

guide individuals in their choices. This approach has already been fostered by health care authorities in their initial attempts to support individual app choices, as in the case of the NHS App Library and DiGA Directory. However, trials engaging MMAs for chronic disease self-management were found to seldom adopt developmental factors or report on them (Adu et al., 2018; Cucciniello et al., 2021).

Contrary to this, we recommend that considerations related to the design and development of MMAs should not only be acknowledged by developers but also be part of the multidimensional assessment upon market entry. Shared decision-making approaches to the development of apps can enhance the replicability of apps and ultimately lead to improved outcomes (Adu et al., 2018; Edbrooke-Childs et al., 2019; Joiner & Lusch, 2016). While the adoption of these factors should characterize the lifecycle assessment of MMAs, decision-makers and scholars need to ensure that the additional value generated thanks to the continuous involvement of users and professionals in the development stage is confirmed empirically and adequately promoted (Table 4).

3.2.2 | Recommendation 2. Classification and evidentiary requirements: An unambiguous classification of MMAs should be associated with corresponding evidence generation requirements and the possibility to flexibly review the associated level of risk of every single MMA

The development stage should be followed by the identification of the MMA class of risk and the associated requirements to be considered along the product lifecycle. Among these, clinical validation constitutes the core of thorough assessment processes of any health care technologies, including MMAs. A large debate has focused on whether exceptionalism should be applied to the clinical and economic evaluation of MMAs. Is DH different and even at odds with methodological rigor, or should the same rules in place for other healthcare technologies apply? (Greaves et al., 2018). The assessment modules for MMAs should include additional technology-specific items is undisputed (Moshi et al., 2020). What is arguably more controversial are the evidence requirements that MMAs should adhere to and, above all, when this should happen along their product lifecycle. This issue is under considerable debate, given the variety of available apps that differ, among others, according to the main func-

tionalities adopted, behavior change techniques implemented, involvement of health care professionals, and level of technology automation in decision-making, all of which contribute to define the expected level of risk. As for the evidence requirements, MMAs should comply with the same evidential requirements in place for the equivalent medical device class, depending on their risk level.

In this sense, the most mature example of an all-encompassing classification is the NICE ESF, which aims to help innovators and commissioners by delineating the corresponding levels of evidence (NICE, 2018). However, it was demonstrated that the specificities currently reported in the standards may not be sufficiently suitable to unambiguously group mobile apps according to their risk tier (Nwe et al., 2020), and at least one review noted that peer-reviewed publications assessing MMAs found that few meet the required evidence level set out by the NICE framework (Forsyth et al., 2021), reflecting the fact that major barriers still need to be overcome for the solid translation of the NICE, or any other, methodology into practice.

3.2.3 | Recommendation 3. Performance and maintenance monitoring: MMAs must assure that technical system implementation does not threaten overall effectiveness. Analytical validity must be determined by the manufacturers and a core set of indicators should be developed coherent with the MMA classification

Technical soundness is highly significant for mobile technologies for multifold reasons: (i) limited constraints in terms of manufacturing and distribution, together with the high scalability of MMAs, make full-scale adoption of apps possible and easier than it has ever been for previous technologies; (ii) most of the app solutions are developed to be managed by patients directly and require additional control mechanisms compared to traditional ones; and (iii) apps operate in complex environments, in which frequent changes and modifications can be implemented, often based on the output generated through the device. Process monitoring and performance testing thus refer to the "continuous process of collecting and analyzing data to compare how an intervention is being implemented compared to its expected results" (WHO, 2016). The technical robustness of any MMA is fundamental to guarantee the delivery of intended results in an accurate and reliable way. Although this concept has been stressed by all reviewed documents, at the empirical level, experiments have shown that MMAs regularly fail to perform as intended, often resulting in failure or crash (Hussain et al., 2018). More recently, an analysis performed on Android contact tracing apps deployed by European countries demonstrated that these apps were not free of weaknesses, vulnerabilities, and technical misconfigurations (Kouliaridis et al., 2021). As of today, definitive standards to guide developers and regulators in their respective duties on this specific domain are lacking (Llorens-Vernet & Miró, 2020), as generic standards available for medical devices are thought to inhibit rather than foster innovation (Van Velthoven et al., 2018).

Given this gap, our recommendation for MMAs is to assure that technical system implementation does not threaten overall effectiveness. This means that analytical validity should be determined by the manufacturer in the verification phase of the software and would need to cover the four major components of monitoring identified by the WHO (WHO, 2016) (Table 3), with suitable metrics that regulatory bodies should check and validate. The collection of robust metrics should progressively help agencies identify a minimum set of standards that all apps should comply with, based on their classification. While performance testing must be a preliminary requirement ahead of market launch, the distinguishing features of MMAs mandate that it be longitudinally monitored, given that app features may evolve, even significantly, over time.

3.2.4 | Recommendation 4. Usability should be continuously monitored, both in the development phase of the solution and after its implementation in the field

Usability refers to the quality of the interaction between the user and the technology. MMAs are interactive technologies, whose value is inherently embedded in the direct relationship with end-users, which can be highly diversified across individuals. Therefore, the actual longitudinal use of the app is itself a prerequisite for the intervention to deliver its promised effects. In fact, continuity in use is extremely challenging for eHealth technologies, to the point that the term "law of attrition" was coined to denote the substantial proportion of users that drop out before study completion (Eysenbach, 2005). MMAs are no exception, as only small portions of users show consistent interaction over time. A meta-analysis on drop-out rates in app-based interventions for chronic diseases showed a pooled drop-out rate of 43% (95% CI 29-57), emphasizing that this issue is inadequately investigated, with a likely under-estimation of the true extent of the phenomenon, higher observed dropout rates in real world settings compared to RCTs, and limited examination of the reasons behind attrition (Meyerowitz-Katz et al., 2020). Similarly, in the context of mental health, apps showed even poorer user retention, with a median 15-day retention of 3.9% and 30-day

retention of 3.3% (Baumel et al., 2019). A further challenge regards the optimal level of interaction, which is necessarily subjective (Michie et al., 2017), as the same levels of interaction may generate different outcomes across categories of individuals.

To sustain the ability of MMAs to produce intended benefits, we recommend that usability should be monitored both during the development stage and after the field implementation of the technology. Usability assessment should hence epitomize the advocated lifecycle approach. Before placing the product on the market, MMA usability scales should be collected to evaluate self-reported usability by the intended end-users and professionals. Numerous scales have been proposed in the literature (Azad-Khaneghah et al., 2020), some of which have been designed and validated to specifically assess the usability of MMAs (Zhou et al., 2019). The monitoring of usability should, however, continue for the entire lifecycle of the MMA, with scales collected over time to monitor effects due to modifications in the technology and its content, coupled with appropriate usage metrics (e.g., access, number of logins, attrition levels), as indirect measures of real-world usability.

3.2.5 | Recommendation 5. Additional outcome domains: Besides conventional outcomes common for all health care technologies, patient empowerment associated with use and the potential environmental impact are distinguishing outcome domains of MMAs that, where relevant, should be appropriately measured and valued

In terms of outcome domains, relevant outcomes are comparable to those in use for other health care technologies, and standard clinical quality measures should be appraised (Mathews et al., 2019). In accordance, published peer-reviewed studies mostly target primary outcomes pertaining to the clinical domains and to life impact (Cucciniello et al., 2021). However, the promise of MMAs to go beyond the traditional patient-physician relationship, in which the latter makes decisions in the patient's best interest, creates the opportunity to gauge additional outcomes during the assessment. First, apps that are directly used by patients and their caregivers promise to significantly impact their empowerment, defined as the process of gaining better knowledge about one's health and being able and motivated to influence it (Tomes, 2007). Patient autonomy (i.e., patients' autonomous health behavior) is recognized by the DiGA Guide (BfArM, 2020b) as a relevant patients' health outcome such that DH can effectively support their involvement in the decision-making processes. Although several international calls to action have fostered the adoption of DH to enhance patient empowerment, there is a lack of clarity and consensus surrounding measurement of the concept, with no comprehensive metrics yet identified and validated (Morley & Floridi, 2019). Contingent valuation (CV) can represent a viable solution in this case. Although the literature is still rather scarce (Callan & O'Shea, 2015; Fawsitt et al., 2017; Shariful Islam et al., 2016), a recent paper (Somers et al., 2019) has demonstrated that CV is effective in providing decision-makers with a broad base of evidence concerning what people value in MMAs.

On a similar note, growing attention is being paid to the environmental sustainability impact of mobile apps, although this still needs to be formally incorporated in available guidelines even when they comprehensively map positive care effects (BfArM, 2020a). While the impact of apps may go as far as generating significant reductions in CO2 emissions as a result of less face-to-face visits and reduced traveling by both patients and healthcare professionals, increased use of digital technologies is associated with sharp surges in global energy consumption (Chevance et al., 2020). The net impact is nevertheless unknown, as no empirical studies on apps have tried to assess it to date. Digital technologies and MMAs can have a significant impact on health care systems' quest for a sustainable future (The Lancet Digital Health, 2021), but it is time to raise global awareness of their potential and start measuring them.

3.2.6 | Recommendation 6. Clinical evidence requirements: Flexible study designs that account for the specific characteristics of MMAs and can generate fast and efficient results should be adopted and coupled with flexible policy arrangements, based on the level of risk and the position in the product lifecycle of the MMA

MMAs that offer unvalidated diagnostic and therapeutic functions with no basis in evidence are in constant observance, as confirmed by the literature (Iribarren et al., 2021; Whitehead & Seaton, 2016). The reasons behind this problem are multiple and largely associated with challenges that make conventional methodologies unsuitable for evaluation (Tarricone et al., 2021).

Regarding RCTs, and prospective studies in general, apps are iterative solutions and hence generally incompatible with locked-down interventions (Tarricone et al., 2021). Costs and length of randomized studies are considered to be too high, given the perceived level of risk associated with most apps and available resources of developers. Retrospective studies may be easier but have rarely been conducted due to challenges in data access and limited use of MMAs in clinical practice (Guo et al., 2020).

Flexible prospective study methodologies that account for the specific characteristics of MMAs and that can generate faster and more efficient results should be adopted, by both tailoring randomized designs to the requirements of MMAs and leveraging on the opportunities provided by system simulations (Guo et al., 2020). These methodologies include MOST, SMART, micro randomized trials, and more in general trials developed with an adaptive design (Klasnja et al., 2015; Pham et al., 2016). At the same time, MMAs may constitute an ideal setting in support of increasing interest in the use of real-world data to support regulatory decision making. While the use of real-world data to support regulatory decision-making and assessment processes would not be new (long employed to support safety reports and risk management), there is lower acceptability of real-world data when the outcome of interest is effectiveness (Cave et al., 2019). Greater caution is typically exercised for the evaluation on patient outcomes, and this has typically precluded the adoption of observational data in regulatory processes. Practical and ethical reasons, however, are pushing to shift the regulatory landscape and decision-making toward greater real-world data use.

While innovative study methodologies may speed up the evidence generation process and contribute to close the current void, additional questions are related to when evidence should be delivered to maximize MMA uptake in everyday clinical practice.

We believe that flexible arrangements such as that set up by the German Digital Health Act (BfArM, 2020a) and the Pre-Cert Program (FDA, 2020d) are desirable and should be adopted by more jurisdictions. Early-stage, low-cost evidence generation approaches, albeit weaker, should be enough to secure market entry and receive temporary reimbursement. It is then necessary to address two additional topics: a specific deadline by which more robust evidence must be provided to confirm the initial clearance and a monitoring process to accompany the product over its entire lifecycle. Given that MMAs may be continuously revised and possibly change functional classification and risk level, it is necessary to dynamically determine whether additional evidence is needed as recommended for other types of medical devices (Tarricone et al., 2020). An ad hoc agency, directly managed by government bodies or subcontracted to independent third parties, could be charged with monitoring the evolution of commissioned MMAs, authorizing software changes and requiring additional evidence with predetermined timeframes when necessary. This complex governance can only be accomplished as part of collective effort of all parties involved.

3.2.7 | Recommendation 7. Safety and security: Risks related to data privacy, cyber security and misinformation and their potential impact must be thoroughly assessed for MMAs

As for other domains, safety acquires a slightly different meaning when applied to MMAs, as apps may generate unwanted effects through multiple channels. Other than the direct harm that could be caused by contact with the device and the undesirable effects that may be generated by diagnostic inaccuracy in medical software, an additional source of risks should be thoroughly assessed. A scoping review by Akbar et al. (2020) filled this gap by highlighting that available apps do indeed pose clinical risks to consumers with potential consequences on patient health. The analysis revealed a total of 80 safety concerns, the vast majority of which related to the quality of the content present and the appropriateness of information sources included within the apps. The dissemination of low-quality information, through either incorrect, incomplete, or inconsistent data, may be harmful to consumers by promoting inadequate or inappropriate health behaviors (BinDhim & Trevena, 2015). Since lay persons may not be able to distinguish whether the included information is correct or misleading, some of their decisions may be ill-managed and may induce them to seek help with considerable delay or inappropriately orient individual choices.

Pending further discussions and elaborations, methods to certify the accuracy of conveyed content and, thereby, the safety of apps have not yet been adopted by regulatory agencies and decision-makers. While MMA safety should be pre-certified before market entry, a vigilance framework that collects safety concerns for MMAs along the entire lifecycle should complement the monitoring.

Despite being strictly intertwined, the direct and indirect impacts of MMAs on patient safety through inadequate content delivery should be kept separate from security concerns, which are indeed significant for MMAs as for any digital technology. Although current assessment processes (e.g., BfArM, 2020b; NHSX, 2021) do prescribe strict adherence to data protection regulation systems (e.g., General Data Protection Regulation [GDPR]), several studies have reported that commercially available apps shared personal data with third parties, often for economic gain, and could compromise the privacy of users (Grundy et al., 2019; Huckvale et al., 2015).

3.2.8 | Recommendation 8. Equity considerations: given the context of escalating inequalities and persistent technological divide, it is paramount to evaluate the net effect of MMAs on equity

Great expectations have emerged for MMAs, and DH in general, for their potential to enhance access to care and reduce inequities. Improvements in health care technologies often tend to cause disparities, as was certified by the COVID-19 pandemic (Watts, 2020). Without empirical evidence, it is nonetheless hard to predict whether apps will support progress toward equity or instead exacerbate existing inequities (Makri, 2019). On April 17, 2019, the WHO (2019) released the report "WHO Guideline: recommendations on digital interventions for health system strengthening," to guide the evaluation process of evidence on emerging DH interventions that contribute to health system improvements, based on an assessment of benefits, harm, acceptability, feasibility, resource use, and equity considerations. The recommendations examine the extent to which DH interventions can expand universal health coverage and focus on a subset of prioritized DH interventions accessible via mobile devices that, however, do not pertain to MMAs (i.e., birth notification via mobile devices, death notification via mobile devices, client-to-provider telemedicine, provider-to-provider telemedicine, targeted client communication via mobile devices, digital tracking of patient/client health status and services via mobile devices, health worker decision support via mobile device, provision of training and educational content to health workers via mobile devices, and mobile learning–mLearning). More recently, the WHO (2021) has reiterated the relevance of equity of access and inclusion in the "Ethics and Governance of Artificial Intelligence for Health." The UK NICE (2018) refers to equity when it recommends including subgroup analyses to show the relevant economic impact provided the availability of good clinical data to show that the effects differ by demographic factors.

The equity domain should hence examine the impact of the app in various clinically and socially relevant subpopulations. The impact may vary across different components in the assessment: the design component pertains to the possibility to personalize some of the technical features of the app (such as font, size, and colors), to facilitate the accessibility to disadvantaged groups; the technical component is related to both data consumption from app utilization and network characteristics, coverage, and data transfer speed in a particular area; the strategic component, which may allow for pre-assessing the potential equity impact of the app and adopting timely informational and educational strategies to mitigate its risks; the cultural component, which underlines the inherent disadvantages of specific social groups and aims at making the benefits of DH more accessible.

3.2.9 | Recommendation 9. Organizational assessment: The assessment of direct and indirect implications of the adoption of MMAs on the organizational level should focus on process, people, structure, culture or management impacts

With the exception of the UK (NICE, 2018) and Germany (BfArM, 2020b), the majority of the currently available frameworks for the assessment of DH neglect the organizational implications associated with their adoption. On the contrary, greater emphasis should be placed on this domain by both decision-makers and producers. Attention to organizational impact should cover the entirety of app-based interventions, including those providing telemedicine services, whose adoption has been significantly accelerated by large-scale accessibility of smartphones at the patient level (Allaert et al., 2020). The expected extent of organizational transformation induced by MMAs is extremely diversified, with MMAs variously impacting the way treatments are organized, the responsibilities of different health care professionals, and the amount and combinations of to employ. It is not possible to simply drop MMAs in established organizations without any tailored adjustments. Two interconnected points of interest pertain to the impact of apps on integrated care (IC) and their interoperability with existing systems. MMAs can be the cornerstone for allowing IC, thus reshaping chronic care and securing stronger data-based connections between levels of care. Initial acceptability studies of mHealth-enabled IC programs that involved all health care professionals from different levels and organizations have shown that, while patient-reported acceptability was extremely high, professionals criticized the lack of maturity and integration with legacy systems (de Batlle et al., 2020). Advances in mHealth are instead dependent on interoperability (Lehne et al., 2019), which should be a prerequisite for the reimbursement of any app and part of their assessment processes.

3.2.10 | Recommendation 10. Cost and economic evaluation: The evaluation of the cost and economic impacts of MMAs should require no innovative forms and methodologies but rather new metrics for outcomes and different structures for cost

Formal economic evaluations of the value for money associated with the introduction and uptake of health technologies are a cornerstone of the HTA paradigm (e.g., EUNeHTA HTA Core Model®) but were not explicitly cited by frameworks for mobile apps mapped by Moshi et al. (2018) and were not considered in the policy documents reviewed, with the exception of the UK ESF. In a more recent systematic review, Vis et al. (2020) found that several frameworks do include the economic dimension. However, the review focused on eHealth technologies and most of those frameworks covered telemedicine programs (Vis et al., 2020). No ad hoc metrics have been proposed for the economic evaluation of mHealth apps, nor should they be, in our opinion. Since MMAs compete for the same, restricted bulk of resources as all other health care technologies, the methodologies, and forms adopted should be similar and make comparisons possible across different technologies. What should change are the types of costs to be accrued, as well as how the suitable comparator is chosen. As for the former, the cost structure for MMAs is reportedly different from that of other health care technologies due to their potential to change scale at reduced incremental cost and the need to account for specific cost items, such as those associated with design and promotion for the active and continuous uptake of the app (Tarricone et al., 2021). Specific challenges are also associated with the selection of the comparator. No one-size-fits-all solutions are practicable. Standard-of-care technologies may sometimes be the best comparator for apps that have no major impact on the entire process of care, such as those meant to improve the level of awareness regarding a certain condition or activate individual self-management. For those that are significantly integrated with care processes, instead, the identification of the best comparator might be extremely intricate: these apps provide complex interventions that may alternatively aim at simply enhancing existing pathways with additional components, refashioning them or even entirely substituting the way they have traditionally been offered. Depending on which is the case, costs and outcomes to include in the cost-effectiveness analysis will have to change accordingly.

Preliminary analyses from the scientific literature confirm that the evidence to support cost-effectiveness is still insufficient and that estimating program costs and outcomes is extremely challenging (Hazel et al., 2021). Technological progress typically contributes significantly to the rise in health care spending: DH and apps could overhaul this trend, despite potential increases in demand and use associated with these technologies may complicate what appears to be an elusive quest for cost savings (Rahimi, 2019).

4 | DISCUSSION

Next Generation EU (European Commission, 2021), the €750 billion recovery instrument fostered at the European level in response to the COVID-19 pandemic, will support health digitization and further boost the development of MMAs. DH and namely MMAs have a strong international and cross-border element, to the point that they have been considered a "cross-jurisdictional" practice of medicine (Khirasaria et al., 2020). Their immateriality and compatibility with widely available consumer technologies make them prone to easy—global—scalability. The DH market is estimated to grow by 25% CAGR from 2019 to 2025, experiencing a four-fold increase from 175 to 657 bln USD (Global X [ETFs], 2021). Hence, despite the need for context-specific considerations and adjustments, MMAs demand converging international principles and the building of worldwide consensus on a set of recognized standards, ensuring consistency around their understanding, regulation, and assessment.

The targeted review of policy documents revealed that current assessment frameworks present some common elements, especially those that have a regulatory scope, but also significant differences.

As for the target focus of regulatory frameworks, software, and not MMAs (with the exception of Germany), is gaining traction, suggesting the relevance of the work conducted by the IMDRF in centering regulatory formation around SaMD-derived concepts. The US, for instance, is a clear example of this influence, as proven by its switch of focus from MMAs toward the more encompassing area of DSWF.

Moving to MMA assessment, identified approaches are widely different and IMDRF influence is far less noticeable. Searching for common themes, guidelines span both clinical and technical domains, and strive toward a continuous risk-monitoring approach through real-world data. A common challenge regards addressing iterative updates, where the US approach of complementing a firm-centered approach through quality and organizational assessment and structured real world performance requirements appears to be the only tentative solution. The UK, through its ESF, provides the most comprehensive and detailed guideline for effectiveness and economic assessment, including indications regarding the type of mandated study designs. Nevertheless, the study designs do not differ from those identified for other technologies (e.g., medical devices). We believe

instead that broadening the scope of outcome domains—mainly clinical—typically considered in evaluation and contingent valuation could represent a better option for evaluating the benefits of MMAs. The German approach may have possibly taken inspiration from the UK approach, as observed in the detailing of accepted study designs, with the setting of both minimum and preferred requirements. Interesting to note is that the DiGA Guide is the only example emphasizing the relevance of alternative trials, such as PCT, SMART, and MOST, for the evaluation of DH devices. However, while the types of studies are addressed in some guidelines (e.g., NICE, DiGA), study designs are currently less regulated and future assessment frameworks need to offer more guidance (e.g., what are the appropriate population, intervention, control, and outcomes) at this regard so to generate relevant high- quality clinical evidence. Economic evaluation is, however, missing from assessment purposes and possibly considered only throughout price negotiations.

At the broader European level, a remarkable innovation in the assessment process of MMAs may come from the new EU regulation on HTA, tentatively approved in 2021 (European Commission, 2018; European Council, 2021). The HTA regulation comes as a follow-up to over 10-year of voluntary collaboration among HTA agencies within the Joint Actions EUNETHA context, a project-based framework for the sharing and harmonization of processes, procedures, and methods. It is likely that the governing body (i.e., the HTA Coordination Group) would continue to adopt the HTA Core Model® to assess health technologies, but it is not clear whether it will be updated to take into consideration the specific features of MMAs compared to all other products (i.e., drugs and medical devices). It would be advisable that the new HTA Coordination Group strictly collaborate with the Medical Device Coordination Group of the EU MDR because the iterative nature of software development makes it unsuitable for a linear, conventional, and assessment approach based up the pre-market and post-market assessment phases. Rather, a circular process that covers different steps within the entire product lifecycle is needed to generate evidence in support of market authorization, coverage, reimbursement, and purchasing decisions. Regulation and assessment are part of a greater innovation chain, which, starting from early prototyping to product commercialization, then translates market responses into conducive inputs to ideation. Considering this, as long as jurisdictions lack a proper systematization of the broader DH landscape, be it in form of taxonomy or mapping, the risk is that mainstream health care sector stakeholders may interpret DH as part of vertical, independent, and disconnected strategies (Labrique et al., 2013). Germany and the UK represent pioneering exceptions in this regard, especially in terms of assessment. To promote further uptake following examples such as these, a multi-dimensional perspective that also involves HTA is necessary to examine the impact of MMAs on all relevant outcomes (Vis et al., 2020).

Among recent contributions, Moshi et al. (2020) developed a dedicated HTA module aimed at making thorough evaluation of MMAs possible, emphasizing the need to adopt a tailored approach with the goal of facilitating their generalized acceptance as part of standard clinical care (Moshi et al., 2020). Similarly, Haverinen et al. (2019) have proposed their Digi-HTA model, pinpointing the domains needed in this assessment, by combining the HTA core processes with DH care services and highlighting the need for agencies and policymakers to develop standardized, evidence-based assessment processes (Haverinen et al., 2019).

Like the two previous contributions, our objective is to support decision-makers, and producers in the introduction of MMAs in clinical care processes. Our recommendations, however, do not specifically aim at adapting current assessment processes and modules to develop fully dedicated ones for MMAs but rather embrace the lifecycle approach that is needed to fruitfully govern them. As such, we reckon that a preliminary requisite should be the definition of a dynamic, unambiguous taxonomy that groups MMAs, depending on prespecified criteria. Ad hoc evidence requirements should then be attached to each of these groups and continuously updated based on technological advancements and MMA updates that result from performance monitoring. It is beyond doubt that MMAs retain a groundbreaking potential for disrupting health care paradigms and nurturing healthier populations, but this materialization should not be taken for granted. Governments and societies at large hold the unique responsibility of making stronger efforts to ensure the sustenance of cost-effective and efficient digital ecosystems. Our 10 recommendations aim at providing decision-makers and all stakeholders with key elements, beyond economic considerations alone, that should be included in any assessment framework for MMAs which are meant to contribute to the development of cost-effective DH. The implementation of the EU MDR in May 2021 together with the EU HTA Regulation represent a precious opportunity for the EU to endorse these principles.

5 | CONCLUSION

The growth of DH technologies, accelerated by the COVID-19 pandemic, has been (although not proportionally) paralleled by the publications of assessment frameworks by regulatory and HTA bodies as well as by HTA scholars. Nevertheless, most of these frameworks tend to cover the broad range of software-based technologies (e.g., SaMDs), thus leaving ample room for ambiguity, uncertainty, and indecisiveness to the applicability of requirements to specific products such as MMAs. Medical mobile apps pres-

ent distinct characteristics that make them different from other MDs and SaMDs (e.g., telemedicine) and need to be considered when their assessment is at stake. A comprehensive framework that covers the full range of the lifecycle of MMAs and all relevant domains, including those less commonly considered, such as environmental impact and equity implications, must be addressed. Our 10 principles can help improve the development of assessment frameworks for MMAs, especially in view of the upcoming HTA Regulation that together with the recent MDR represent an opportunity for the EU to further expand the work done by EUnetHTA.

ACKNOWLEDGMENTS

We acknowledge the contribution of Davide Targa, MSc, in reviewing the entire manuscript and especially the sections concerning the comparative assessment of international policy documents. Francesco Petracca, Oriana Ciani, Maria Cucciniello and Rosanna Tarricone all reported grants from the European Union's Horizon 2020 research and innovation program under grant agreement No. 779306.
Open Access Funding provided by Universita Bocconi within the CRUI-CARE Agreement.

CONFLICT OF INTEREST

The authors declare they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study. All data were drawn from public sources as reported in the reference list.

ORCID


Rosanna Tarricone https://orcid.org/0000-0002-2009-9357

ENDNOTES

[1] The selected pilot companies are Apple, Fitbit, Johnson & Johnson, Pear Therapeutics, Phosphorus, Roche, Samsung, Tidepool, and Verily.

[2] The pilot is currently limited to SaMD, but the framework is also expected to ultimately comprehend Software in Medical Devices (SiMD) and software that are accessories to hardware MD.

REFERENCES

Adu, M. D., Malabu, U. H., Callander, E. J., Malau-Aduli, A. E., & Malau-Aduli, B. S. (2018). Considerations for the development of mobile phone apps to support diabetes self-management: Systematic review. *JMIR MHealth and UHealth*, *6*(6), e10115.

Akbar, S., Coiera, E., & Magrabi, F. (2020). Safety concerns with consumer-facing mobile health applications and their consequences: A scoping review. *Journal of the American Medical Informatics Association*, *27*(2), 330–340.

Allaert, F. A., Legrand, L., Abdoul Carime, N., & Quantin, C. (2020). Will applications on smartphones allow a generalization of telemedicine? *BMC Medical Informatics and Decision Making*, *20*, 30.

Azad-Khaneghah, P., Neubauer, N., Miguel Cruz, A., & Liu, L. (2020). Mobile health app usability and quality rating scales: A systematic review. *Disability and Rehabilitation: Assistive Technology*, *19*, 1.

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: Systematic Search and panel-based usage analysis. *Journal of Medical Internet Research*, *21*(9), e14567.

BfArM. (2020a). *Digital health applications (DiGA)*. https://www.bfarm.de/EN/MedicalDevices/DiGA/_node.html

BfArM. (2020b). *The Fast-Track Process for DigitalHealth Applications (DiGA) according to Section 139e SGB V. A Guide for Manufacturers, Service Providers and Users*. Accessed 30 June 2021. https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.pdf;jsession%20id=CFCB8E00B0826085FAEDB0A35F406F49.2_cid354?__blob=publicationFile&v=2

BinDhim, N. F., & Trevena, L. (2015). There's an app for that: A guide for healthcare practitioners and researchers on smartphone technology. *Online Journal of Public Health Informatics*, *7*(2).

Callan, A., & O'Shea, E. (2015). Willingness to pay for telecare programmes to support independent living: Results from a contingent valuation study. *Social Science & Medicine*, *124*, 94–102.

Campbell, B., Campbell, M., Dobson, L., Higgins, J., Dillon, B., Marlow, M., & Pomfrett, C. J. (2018). Assessing the value of innovative medical devices and diagnostics: The importance of clear and relevant claims of benefit. *International Journal of Technology Assessment in Health Care*, *34*(4), 419–424.

Cave, A., Kurz, X., & Arlett, P. (2019). Real-world data for regulatory decision making: Challenges and possible solutions for Europe. *Clinical Pharmacology & Therapeutics*, *106*(1), 36–39.

Chevance, G., Hekler, E. B., Efoui-Hess, M., Godino, J., Golaszewski, N., Gualtieri, L., Krause, A., Marrauld, L., Nebeker, C., Perski, O., Simons, D., Taylor, J. C., & Bernard, P. (2020). Digital health at the age of the Anthropocene. *The Lancet Digital Health*, *2*(6), e290–e291.

Cucciniello, M., Petracca, F., Ciani, O., & Tarricone, R. (2021). Development features and study characteristics of mobile health apps in the management of chronic conditions: A systematic review of randomised trials. *npj Digital Medicine*, *4*, 144.

de Batlle, J., Massip, M., Vargiu, E., Nadal, N., Fuentes, A., Bravo, M. O., Colomina, J., Drudis, R., Torra, M., Pallisó, F., Miralles, F., Barbé, F., & Torres, G. (2020). Implementing mobile health–enabled integrated care for complex chronic patients: Patients and professionals' acceptability study. *JMIR MHealth and UHealth*, *8*(11), e22136.

Duan, H., Wang, Z., Ji, Y., Ma, L., Liu, F., Chi, M., Deng, N., & An, J. (2020). Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: Development and deployment. *JMIR MHealth and UHealth*, *8*(2), e14466.

Edbrooke-Childs, J., Edridge, C., Averill, P., Delane, L., Hollis, C., Craven, M. P., Martin, K., Feltham, A., Jeremy, G., Deighton, J., & Wolpert, M. (2019). A feasibility trial of power up: Smartphone app to support patient activation and shared decision making for mental health in young people. *JMIR MHealth and UHealth*, *7*(6), e11677.

European Commission. (2018). *Proposal for a regulation of the European parliament and of the Council on health technology assessment and amending directive 2011/24/EU*.

European Commission. (2020). *The White paper on Artificial Intelligence is a non-industry specific document that discusses the adoption of AI across sectors and activities*. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission. (2021). *NextGenerationEU*. https://ec.europa.eu/info/strategy/eu-budget/eu-borrower-investor-relations/nextgenerationeu_en

European Council. (2021). *Health technology assessment: Informal deal between Council and European parliament*. https://www.consilium.europa.eu/en/press/press-releases/2021/06/22/health-technology-assessment-informal-deal-between-council-and-european-parliament/

European Union. (2017a, May 5). *REGULATION (EU) 2017/745 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 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*. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745&from=EN

European Union. (2017b, May 5). *REGULATION (EU) 2017/746 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU*. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0746&from=it

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research*, *7*(1), e11.

Farao, J., Malila, B., Conrad, N., Mutsvangwa, T., Rangaka, M. X., & Douglas, T. S. (2020). A user-centred design framework for mHealth. *PloS One*, *15*(8), e0237910.

Fawsitt, C. G., Bourke, J., Murphy, A., McElroy, B., Lutomski, J. E., Murphy, R., & Greene, R. A. (2017). A cost-benefit analysis of two alternative models of maternity care in Ireland. *Applied Health Economics and Health Policy*, *15*(6), 785–794.

FDA. (2019a). *Developing a software pre-certification program: A working model*.

FDA. (2019b). *Policy for device software functions and mobile medical applications*. https://www.fda.gov/media/80958/download

FDA. (2020a). *Developing the software precertification program: Summary of learnings and ongoing activities*. https://www.fda.gov/media/142107/download

FDA. (2020b). *Digital health center of excellence*. Www.Fda.Gov. https://www.fda.gov/medical-devices/digital-health-center-excellence

FDA. (2020c). *Digital health innovation action plan*. https://www.fda.gov/media/106331/download

FDA. (2020d). *Digital health software precertification (Pre-Cert) program*. Www.Fda.Gov. https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-program

FDA. (2020e). *Enforcement policy for digital health devices for treating psychiatric disorders during the Coronavirus disease 2019 (COVID-19) public health emergency*. Guidance for IndustryandFood and Drug Administration Staff.

FDA. (2021). *Artificial intelligence and machine learning in software as a medical device*. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

Forsyth, J. R., Chase, H., Roberts, N. W., Armitage, L. C., & Farmer, A. J. (2021). Application of the national Institute for health and care excellence evidence standards framework for digital health technologies in assessing mobile-delivered technologies for the self-management of type 2 diabetes mellitus: Scoping review. *JMIR Diabetes*, *6*(1), e23687.

Global X (ETFs). (2021). *Projected global digital health market size from 2019 to 2025*. Accessed 1 July 2021. https://www.statista.com/statistics/1092869/global-digital-health-market-size-forecast/

Gordon, W. J., Landman, A., Zhang, H., & Bates, D. W. (2020). Beyond validation: Getting health apps into clinical practice. *NPJ Digital Medicine*, *3*(1), 1–6.

Grant, M., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies 2009. *Health Information and Libraries Journal*, *26*(2), 91–108.

Greaves, F., Joshi, I., Campbell, M., Roberts, S., Patel, N., & Powell, J. (2018). What is an appropriate level of evidence for a digital health intervention? *The Lancet*, *392*(10165), 2665–2667.

Grundy, Q., Chiu, K., Held, F., Continella, A., Bero, L., & Holz, R. (2019). Data sharing practices of medicines related apps and the mobile ecosystem: Traffic, content, and network analysis. *BMJ*, *364*, l920.

Guo, Y., Guo, J., Shi, X., Yao, Y., Sun, Y., Xia, Y., Yu, B., Liu, T., Chen, Y., Lip, G. Y. H., & mAF-App II Trial investigators. (2020). Mobile health technology-supported atrial fibrillation screening and integrated care: a report from the mAFA-II trial Long-term Extension Cohort. *European Journal of Internal Medicine*, *82*, 105–111.

Haute Authorité de Santé. (2016). *Good practice guidelines on health apps and smart devices (mobile health or mHealth)*. https://www.has-sante.fr/upload/docs/application/pdf/2017-03/dir1/good_practice_guidelines_on_health_apps_and_smart_devices_mobile_health_or_mhealth.pdf

Haverinen, J., Keränen, N., Falkenbach, P., Maijala, A., Kolehmainen, T., & Reponen, J. (2019). Digi-HTA: Health technology assessment framework for digital healthcare services. *Finnish Journal of EHealth and EWelfare*, *11*(4), 326–341.

Hazel, C. A., Bull, S., Greenwell, E., Bunik, M., Puma, J., & Perraillon, M. (2021). Systematic review of cost-effectiveness analysis of behavior change communication apps: Assessment of key methods. *Digital Health*, *7*, 20552076211000560.

Hensher, M., Cooper, P., Dona, S. W. A., Angeles, M. R., Nguyen, D., Heynsbergh, N., Chatterton, M. L., & Peeters, A. (2021). Scoping review: Development and assessment of evaluation frameworks of mobile health apps for recommendations to consumers. *Journal of the American Medical Informatics Association*, *28*(6), 1318–1329.

Huckvale, K., Prieto, J. T., Tilney, M., Benghozi, P.-J., & Car, J. (2015). Unaddressed privacy risks in accredited health and wellness apps: A cross-sectional systematic assessment. *BMC Medicine*, *13*, 214.

Hussain, M., Al-Haiqi, A., Zaidan, A. A., Zaidan, B. B., Kiah, M., Iqbal, S., Iqbal, S., & Abdulnabi, M. (2018). A security framework for mHealth apps on Android platform. *Computers & Security*, *75*, 191–217.

IMDRF. (2014). *Software as a medical device": Possible framework for risk categorization and corresponding considerations*.

IMDRF. (2015). *Software as a medical device (SaMD): Application of quality management system*. https://www.imdrf.org/consultations/software-medical-device-samd-application-quality-management-system

IMDRF. (2017). *Software as a medical device (SaMD): Clinical evaluation*. https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-170921-samd-n41-clinical-evaluation_1.pdf

IMDRF. (2021). *Artificial intelligence medical devices (AIMD) work item*. https://www.imdrf.org/working-groups/artificial-intelligence-medical-devices

Iribarren, S. J., Akande, T. O., Kamp, K. J., Barry, D., Kader, Y. G., & Suelzer, E. (2021). Effectiveness of mobile apps to promote health and manage disease: Systematic review and meta-analysis of randomized controlled trials. *JMIR MHealth and UHealth*, *9*(1), e21563.

Joiner, K. A., & Lusch, R. F. (2016). Evolving to a new service-dominant logic for health care. *Innovation and Entrepreneurship in Health*, *3*, 25–33.

Kadakia, K., Patel, B., & Shah, A. (2020). Advancing digital health: FDA innovation during COVID-19. *Npj Digital Medicine*, *3*(1), 1–3.

Kasperbauer, T. J., & Wright, D. E. (2020). Expanded FDA regulation of health and wellness apps. *Bioethics*, *34*(3), 235–241.

Khirasaria, R., Singh, V., & Batta, A. (2020). Exploring digital therapeutics: The next paradigm of modern health-care industry. *Perspectives in Clinical Research*, *11*(2), 54.

Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., & Murphy, S. A. (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology*, *34*(S), 1220–1228.

Kouliaridis, V., Kambourakis, G., Chatzoglou, E., Geneiatakis, D., & Wang, H. (2021). Dissecting contact tracing apps in the Android platform. *Plos One*, *16*(5), e0251867.

Krueger, R. A., & Casey, M. A. (2014). *Designing and conducting focus group interviews. Social analysis selected tools and techniques*. The World Bank.

Ku, J. P., & Sim, I. (2021). Mobile health: Making the leap to research and clinics. *NPJ Digital Medicine*, *4*(1), 1–4.

Labrique, A. B., Vasudevan, L., Kochi, E., Fabricant, R., & Mehl, G. (2013). mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. *Global Health: Science and Practice*, *1*(2), 160–171.

Lagan, S., Aquino, P., Emerson, M. R., Fortuna, K., Walker, R., & Torous, J. (2020). Actionable health app evaluation: Translating expert frameworks into objective metrics. *NPJ Digital Medicine*, *3*(1), 1–8.

Lagan, S., Sandler, L., & Torous, J. (2021). Evaluating evaluation frameworks: A scoping review of frameworks for assessing health apps. *BMJ Open*, *11*(3), e047001.

Lehne, M., Sass, J., Essenwanger, A., Schepers, J., & Thun, S. (2019). Why digital medicine depends on interoperability. *NPJ Digital Medicine*, *2*(1), 1–5.

Leigh, S., & Ashall-Payne, L. (2019). The role of health-care providers in mHealth adoption. *The Lancet Digital Health*, *1*(2), e58–e59.

Llorens-Vernet, P., & Miró, J. (2020). Standards for mobile health–related apps: Systematic review and development of a guide. *JMIR MHealth and UHealth*, *8*(3), e13057.

Makri, A. (2019). Bridging the digital divide in health care. *The Lancet Digital Health*, *1*(5), e204e205.

Mathews, S. C., McShea, M. J., Hanley, C. L., Ravitz, A., Labrique, A. B., & Cohen, A. B. (2019). Digital health: A path to validation. *NPJ Digital Medicine*, *2*(1), 1–9.

McCurdie, T., Taneva, S., Casselman, M., Yeung, M., McDaniel, C., Ho, W., & Cafazzo, J. (2012). mHealth consumer apps: the case for user-centered design. *Biomedical Instrumentation & Technology*, *46*(s2), 49–56.

Medical Device Coordination Group. (2019). *Guidance on qualification and classification of software in regulation (EU) 2017/745 – MDR and regulation (EU) 2017/746 – IVDR*. file:///C:/Users/Tarricone/Downloads/mdcg_2019_11_guidance_qualification_classification_software.pdf

Medical Device Coordination Group. (2020). *Guidance on clinical evaluation (MDR)/performance evaluation (IVDR) of medical device software*. https://ec.europa.eu/health/system/files/2020-09/md_mdcg_2020_1_guidance_clinic_eva_md_software_en_0.pdf

Medicine and Healthcare Products Regulatory Agency. (2021). *Guidance: Medical device stand-alone software including apps (including IVDMDs)*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/999908/Software_flow_chart_Ed_1-08b-IVD.pdf

MedTech Reimbursement Consulting. (2020). *First health app in the French LPPR list for add-on reimbursement*. https://mtrconsult.com/news/first-health-app-french-lppr-list-add-reimbursement

Meyerowitz-Katz, G., Ravi, S., Arnolda, L., Feng, X., Maberly, G., & Astell-Burt, T. (2020). Rates of attrition and dropout in app-based interventions for chronic disease: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *22*(9), e20283.

Michie, S., Yardley, L., West, R., Patrick, K., & Greaves, F. (2017). Developing and evaluating digital interventions to promote behavior change in health and health care: Recommendations resulting from an international workshop. *Journal of Medical Internet Research*, *19*(6), e7126.

Morley, J., & Floridi, L. (2019). Enabling digital health companionship is better than empowerment. *The Lancet Digital Health*, *1*(4), e155–e156.

Moshi, M. R., Tooher, R., & Merlin, T. (2018). Suitability of current evaluation frameworks for use in the health technology assessment of mobile medical applications: A systematic review. *International Journal of Technology Assessment in Health Care*, *34*(5), 464–475.

Moshi, M. R., Tooher, R., & Merlin, T. (2020). Development of a health technology assessment module for evaluating mobile medical applications. *International Journal of Technology Assessment in Health Care*, *36*(3), 252–261.

Muehlematter, U. J., Daniore, P., & Vokinger, K. N. (2021). Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): A comparative analysis. *The Lancet Digital Health*.

NHSX. (2021). *Digital technology assessment criteria for health and social care*. https://www.nhsx.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac/

NICE. (2018). *Evidence standards framework for digital health technologies*. https://www.nice.org.uk/corporate/ecd7

NICE. (2021). *Evidence standards framework for digital health technologies*. Updated information https://www.nice.org.uk/corporate/ecd7/chapter/update-information

Nwe, K., Larsen, M. E., Nelissen, N., & Wong, D. C.-W. (2020). Medical mobile app classification using the national Institute for health and care excellence evidence standards framework for digital health technologies: Interrater reliability study. *Journal of Medical Internet Research*, *22*(6), e17457.

Pandey, N., & Pal, A. (2020). Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice. *International Journal of Information Management*, *55*, 102171.

Petracca, F., Ciani, O., Cucciniello, M., & Tarricone, R. (2020). Harnessing digital health technologies during and after the COVID-19 pandemic: Context matters. *Journal of Medical Internet Research*, *22*(12), e21815.

Pham, Q., Wiljer, D., & Cafazzo, J. A. (2016). Beyond the randomized controlled trial: A review of alternatives in mHealth clinical trial methods. *JMIR MHealth and UHealth*, *4*(3), e5720.

Rahimi, K. (2019). Digital health and the elusive quest for cost savings. *The Lancet Digital Health*, *1*(3), e108–e109.

Schnall, R., Rojas, M., Bakken, S., Brown, W., Carballo-Dieguez, A., Carry, M., Gelaude, D., Mosley, J. P., & Travers, J. (2016). A user-centered model for designing consumer mobile health (mHealth) applications (apps). *Journal of Biomedical Informatics*, *60*, 243–251.

Shariful Islam, S. M., Lechner, A., Ferrari, U., Seissler, J., Holle, R., & Niessen, L. W. (2016). Mobile phone use and willingness to pay for SMS for diabetes in Bangladesh. *Journal of Public Health*, *38*(1), 163–169.

Somers, C., Grieve, E., Lennon, M., Bouamrane, M.-M., Mair, F. S., & McIntosh, E. (2019). Valuing mobile health: An open-ended contingent valuation survey of a national digital health program. *JMIR MHealth and UHealth*, *7*(1), e3.

Tarricone, R., Ciani, O., Torbica, A., Brouwer, W., Chaloutsos, G., Drummond, M. F., Martelli, N., Persson, U., Leidl, R., Levin, L., Sampietro-Colom, L., & Taylor, R. S. (2020). Lifecycle evidence requirements for high-risk implantable medical devices: A European perspective. *Expert Review of Medical Devices*, *17*(10), 993–1006.

Tarricone, R., Petracca, F., Ciani, O., & Cucciniello, M. (2021). Distinguishing features in the assessment of mHealth apps. *Expert Review of Pharmacoeconomics & Outcomes Research*, *1–6*.

The Lancet Digital Health. (2021). Technology: The nexus of climate change and COVID-19? *The Lancet Digital Health*, *3*(2), e67.

Tomes, N. (2007). Patient empowerment and the dilemmas of late-modern medicalisation. *The Lancet*, *369*(9562), 698–700.

Van Velthoven, M. H., Smith, J., Wells, G., & Brindley, D. (2018). Digital health app development standards: A systematic review protocol. *BMJ Open*, *8*(8), e022969.

Vis, C., Bührmann, L., Riper, H., & Ossebaard, H. C. (2020). Health technology assessment frameworks for eHealth: A systematic review. *International Journal of Technology Assessment in Health Care*, *36*(3), 204–216.

Watts, G. (2020). COVID-19 and the digital divide in the UK. *The Lancet Digital Health*, *2*(8), e395–e396.

Whitehead, L., & Seaton, P. (2016). The effectiveness of self-management mobile phone and tablet apps in long-term condition management: A systematic review. *Journal of Medical Internet Research*, *18*(5), e97.

WHO. (2016). *Monitoring and evaluating digital health interventions: A practical guide to conducting research and assessment*. Accessed 2 June 2021. https://apps.who.int/iris/handle/10665/252183

WHO. (2018). *Classification of digital health interventions v1.0. A shared language to describe the uses of digital technology for health*. https://www.who.int/reproductivehealth/publications/mhealth/classification-digital-health-interventions/en/

WHO. (2019). *Recommendations on digital interventions for health system strengthening*.

WHO. (2021). *Ethics and governance of artificial intelligence for health*. https://www.who.int/publications/i/item/9789240029200

Zhou, L., Bao, J., Setiawan, I. M. A., Saptono, A., & Parmanto, B. (2019). The mHealth APP usability questionnaire (MAUQ): Development and validation study. *JMIR MHealth and UHealth*, *7*(4), e11500.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Tarricone, R., Petracca, F., Cucciniello, M., & Ciani, O. (2022). Recommendations for developing a lifecycle, multidimensional assessment framework for mobile medical apps. *Health Economics*, 1–25. https://doi.org/10.1002/hec.4505