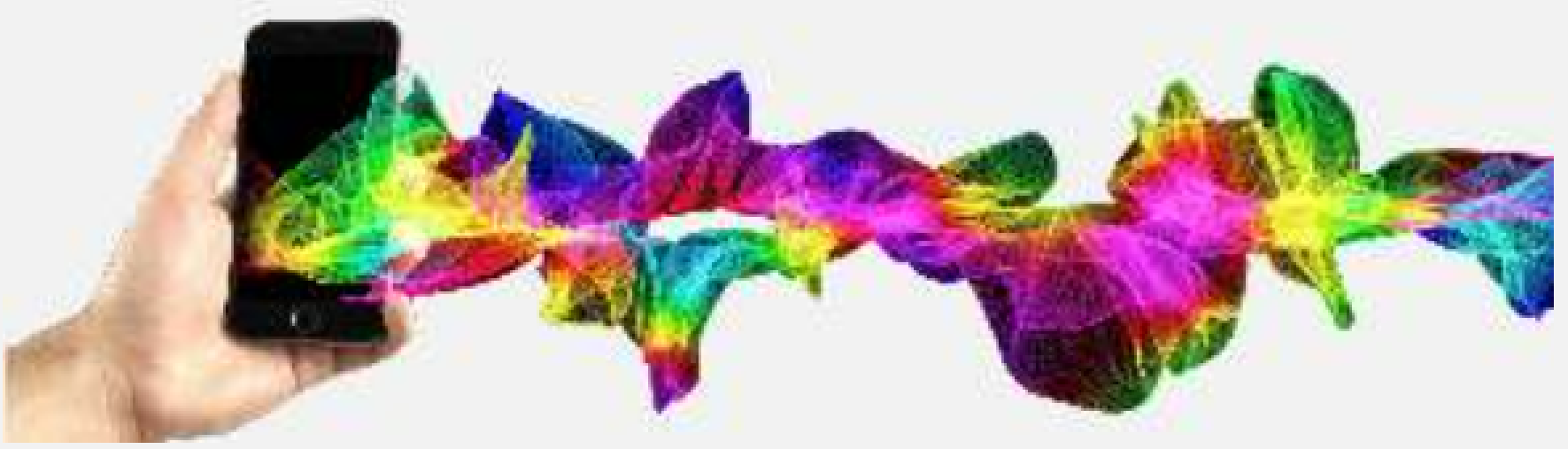
# APLICACIONS MÒBILS EN SALUT: CRITERIS PER AL DESENVOLUPAMENT I AVALUACIÓ DE LA QUALITAT

**Pere Llorens Vernet**

# Aplicacions mòbils en salut: criteris per al desenvolupament i avaluació de la qualitat

Pere Llorens Vernet



TESI DOCTORAL 2021

UNIVERSITAT ROVIRA i VIRGILI

UNIVERSITAT ROVIRA I VIRGILI
APLICACIONS MÒBILS EN SALUT: CRITERIS PER AL DESENVOLUPAMENT I AVALUACIÓ DE LA QUALITAT
Pere Llorens Vernet

Pere Llorens Vernet

# APLICACIONS MÒBILS EN SALUT: CRITERIS PER AL DESENVOLUPAMENT I AVALUACIÓ DE LA QUALITAT

TESI DOCTORAL

Supervisada pel Dr. Jordi Miró Martínez

Departament de Psicologia

UNIVERSITAT ROVIRA i VIRGILI

Tarragona, 2021

UNIVERSITAT ROVIRA I VIRGILI
APLICACIONS MÒBILS EN SALUT: CRITERIS PER AL DESENVOLUPAMENT I AVALUACIÓ DE LA QUALITAT
Pere Llorens Vernet

UNIVERSITAT ROVIRA I VIRGILI
Departament de Psicologia
Carretera de Valls, s/n
43007 Tarragona
Tel. +34 977 55 80 75
Fax +34 977 55 80 68

FAIG CONSTAR que aquest treball, titulat "Aplicacions mòbils en salut: criteris per al desenvolupament i avaluació de la qualitat", que presenta Pere Llorens Vernet per a l'obtenció del títol de Doctor, ha estat realitzat sota la meva direcció al Departament de Psicologia d'aquesta universitat.

Tarragona, 7 de gener de 2021

El director de la tesi doctoral

Dr. Jordi Miró Martínez

UNIVERSITAT ROVIRA I VIRGILI
APLICACIONS MÒBILS EN SALUT: CRITERIS PER AL DESENVOLUPAMENT I AVALUACIÓ DE LA QUALITAT
Pere Llorens Vernet

# Agraïments

Com ha acabat un enginyer telemàtic fent el doctorat en un grup de psicologia? Quan ho explico moltes persones se'm queden mirant amb cara de sorpresa, però tot té una història.

Encara no havia ni acabat la carrera (estava en el meu darrer any) i ni tan sols tenia del tot clar que faria després, però tot va canviar quan em van explicar d'un grup de recerca que buscava persones per desenvolupar aplicacions mòbils de salut. Vaig pensar que seria una bona manera de començar a treballar de lo meu i agafar experiència, amb aquesta idea em vaig ficar en contacte amb ells i hi vaig anar a parlar. Si, era el grup ALGOS, eren totalment desconeguts per mi però que a partir de llavors marcarien una bona part de la meva vida i fins ara. La resta és historia, vaig estar anys fent aquesta mateixa feina fins que vaig fer el salt a realitzar la tesi.

En primer lloc haig d'agrair al Dr. Jordi Miró, director del grup i d'aquesta tesi, per donar-me l'oportunitat d'incorporar-me al grup i poder participar en els projectes de noves tecnologies per a la salut desenvolupats, com també proposar-me realitzar la tesi doctoral. La seva guia durant tot aquest procés han sigut imprescindibles per poder realitzar la tesi i també m'ha ajudat a desenvolupar la meva capacitat com a investigador.

Com no podria ser de cap altra manera, també vull agrair a tots els membres del grup ALGOS que m'han acompanyat al llarg de tots aquests anys. Primer tots els meus companys enginyers en el grup. Començant pel

Roman amb qui he après moltíssimes coses, no podria saber programar tant bé sense haver-te conegut. El Gerard, gran company de la carrera i amb qui vaig entrar al grup, no està malament tenir una cara coneguda quan vas a un lloc nou. I el Youssef, un crac, perfeccionista a més no poder, amb gent com tu al costat un es motiva a millorar. També a les companyes i company de comunicació. L'Anna, una bona companya, encara recordo els temps de barallar-nos amb la primera web del grup i que tot funciones bé. La Lluïsa, amb qui sempre m'he compenetrat molt bé, ets extremadament simpàtica i sempre estàs disposada a donar un cop de mà quan faci falta. I la Mariona i l'Oriol, va ser per poc temps i no ens vam poder veure gaire però almenys vaig poder gaudir de la vostra companyia durant un temps a més de fer una calçotada amb cadascú. Per últim, però no menys important a les psicòlogues i psicòlegs del grup. La Cat, una persona amb un gran cor i molta empatia, sempre tens una paraula amable amb tothom i transmets molta energia. La Rocío, decidida, sempre amb molta iniciativa, treballadora a més no poder, la viva imatge de què amb esforç pots aconseguir tot allò que et proposis. El Santi, no he conegut persona més divertida i alegre, amb tu al costat tots els mals desapareixen, com enyoro els riures que ens fèiem. El Rubén, el meu company de batalla i subministrador oficial de *memes* de la tesi, no hi ha res millor que posar humor als moments durs. Gràcies per estar allí, de veritat, se m'hagués fet molt més difícil arribar fins aquí sense el teu suport. L'Ester, membre del *team* dinar, ets una persona molt decidida, atrevida i que no s'atura davant de res, amb molt caràcter però també amb un gran cor. L'Eli, també membre del *team* dinar, sempre tens les idees molt clares i ets una de les

persones amb més capacitat de saber-se organitzar que he vist. Elena, ets correctíssima per davant de tot, mai tens una paraula més alta que una altra i sempre tractant de ser amable amb tothom, cosa molt complicada de fer. La Lorena, una persona amb molta capacitat d'abstracció i concentració quan es fica a treballar i això és una gran virtut. La Meritxell, reservada al principi però molt simpàtica, amable i treballadora a més no poder. La Jessica, persona amb molta personalitat i que té clars els seus objectius. I la ultimíssima incorporació, el Josep, poc ens coneixem encara però de bones a primeres diria que ets molt simpàtic.

Per altra banda també vull agrair a tota la meva família. Els meu pare Rafael i la meva mare Nuri, la meva germana Yolanda i la seva parella Albert i també el meu germà Rafel i la seva parella Elena. A tots vosaltres gràcies pel vostre suport, afecte i totes les divertides estones que passem junts. Per últim però no menys important la meva gosseta Pruna, per la teva gran companyia i el teu amor incondicional.

També a tots els que m'heu acompanyat durant tot aquest temps i ajudat de forma més directe o indirecte, us dono les gràcies a tots i a totes de tot cor!

Aquesta tesi doctoral s'ha realitzat amb el suport parcial d'una beca predoctoral (2017FI_B_00056, 2018FI_B1_00158 i 2019FI_B2_00151) cofinançada per la Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya, i la Unió Europea (UE) a través del Fons Social Europeu (FSE).

UNIVERSITAT ROVIRA I VIRGILI
APLICACIONS MÒBILS EN SALUT: CRITERIS PER AL DESENVOLUPAMENT I AVALUACIÓ DE LA QUALITAT
Pere Llorens Vernet

# Índex

Prefaci

# Prefaci

# Prefaci

Aquesta tesi doctoral s'ha realitzat en el marc del grup de recerca ALGOS. La seva estructura es divideix en dues parts, una teòrica i una altra pràctica. La primera recull i de forma resumida presenta informació bàsica, la base sobre la qual es va concretar aquesta tesi i la recerca relacionada. La segona presenta els detalls de la recerca realitzada, la essència del com i els resultats obtinguts.

Específicament, la part teòrica detalla l'estat actual dels sistemes de salut i els seus principals desafiaments, així com una visió general del seu futur més immediat. En aquest context, es descriu com s'han introduït les noves tecnologies amb el què s'anomena salut mòbil, i la seva rellevància en aquest sector. Específicament, s'explica l'ús que s'està fent de les aplicacions mòbils de salut, els seus beneficis, com també els inconvenients potencials. Més endavant, es detallen quines mesures i iniciatives han sorgit per tenir un control de la qualitat d'aquestes aplicacions mòbils. Finalment, s'explica la necessitat d'identificar els criteris a seguir i establir una guia general de qualitat per a les aplicacions mòbils de salut.

A la part que aquí anomenen pràctica es presenten els objectius dels estudis que s'han realitzat i, de forma resumida, els mètodes utilitzats. En la secció de resultats es mostren, en la seva totalitat, els diferents estudis que formen part d'aquesta tesi. En la següent secció es valoren i es fa una discussió general dels resultats dels estudis, les principals limitacions i possibles futures línies d'investigació. Finalment, es presenten les principals conclusions que es deriven de la feina feta en aquesta tesi.

# Part teòrica

# 1. Introducció

## 1.1. Sistemes de salut: situació actual i reptes

El sistemes públics de salut de tot el món tenen que fer front a importants reptes, entre ells l'envelliment de la societat, l'increment de les malalties cròniques i l'augment de la població (Ministerio de Sanidad, 2018). Ens trobem en una societat cada vegada més envellida per l'augment de l'esperança de vida, això comporta que, en general, aquesta població més envellida, necessita més atencions per mantenir el seu benestar i una bona qualitat de vida (Roser, Ortiz-Ospina, & Ritchie, 2020). Gràcies a la millora de la sanitat, malalties que abans resultaven mortals ara s'han convertit en cròniques, això fa que les persones directament afectades necessitin una atenció sanitària continua, també més complexa, i sovint força més costosa en termes de despesa econòmica (Hajat & Stein, 2018). Un altre problema afegit és que, en les últimes dècades, hi ha hagut un gran increment en la població, de tal manera que hi ha més persones que fan ús del sistema sanitari (United Nations - Department of Economic and Social Affairs, 2019).

La combinació d'aquestes diferents *forces* ha provocat que, en els últims anys, s'hagi produït un gran increment de la demanda dels serveis sanitaris, en ocasions provocant la seva saturació, com també un augment considerable en la despesa (OECD, 2019; Watt, Charlesworth, & Gershlick, 2019; Xu et al., 2018). I, és així, que els sistemes sanitaris són cada vegada més difícils de mantenir i en ocasions es demostren incapaços d'oferir els serveis de qualitat que la gent necessita (World Health Organization, 2018).

Aquestes problemàtiques també tenen efectes en els propis especialistes sanitaris, fent que aquestes persones demostrin nivells alts d'esgotament (Shanafelt et al., 2012), dificultant que puguin continuar oferint un bon i adequat servei assistencial. Cada regió té les seves particularitats, però cada vegada és més difícil no només millorar l'atenció sanitària sinó el sol fet de mantenir una cobertura sanitària de qualitat (National Academies of Sciences, 2018).

El sector sanitari s'enfronta a un futur cada vegada més incert, ja que per fer-hi front, ha de donar resposta a un conjunt de reptes molt importants que a l'hora s'ha demostrat de molt alta complexitat. Els reptes més urgents que s'ha proposat assolir l'Organització Mundial de la Salut són: (1) aconseguir una atenció sanitària més justa, (2) millor accés als medicaments, (3) invertir en els treballadors sanitaris, (4) donar protagonisme a la salut en el debat sobre el clima, (5) prestar serveis de salut en situacions de crisis, (6) aturar les malalties infeccioses, (7) preparació per a epidèmies, (8) protegir a les persones de productes perillosos, (9) guanyar-se la confiança de la població, (10) protegir els adolescents, (11) protegir els medicaments, (12) mantenir neta l'atenció sanitària i (13) aprofitar les noves tecnologies (World Health Organization, 2019, 2020).

La desigualtat socioeconòmica a nivell mundial afecta en la forma en que s'hi pot fer front per tal de tenir una bona i justa atenció sanitària (Pickett & Wilkinson, 2015). Una de les eines que s'estan fent servir en els

últims anys per tal de millorar aquesta situació són les anomenades noves tecnologies de la informació i de la comunicació.

## 1.2. Tecnologies de la informació i de la comunicació

Les noves tecnologies de la informació i de la comunicació (TIC) fan referència a un concepte dinàmic i es redefineixen constantment però, en general, són tecnologies que emmagatzemen, recuperen, manipulen, transmeten o reben informació electrònicament en forma digital (Zuppo, 2012). L'ús d'aquestes tecnologies en la pràctica sanitària és el que s'anomena salut electrònica o eSalut (en anglès *electronic health* o *eHealth)* (Eysenbach, 2001). No hi ha una única definició per aquest concepte però, en general, aquest camp cobreix tots aquells processos electrònics i digitals en salut (Oh, Rizo, Enkin, & Jadad, 2005). Aquestes noves tecnologies estan demostrant ser realment clau per a l'àmbit sanitari, ja que aporten beneficis en la logística i l'atenció al pacient. Per exemple, contribueixen al facilitar la disponibilitat i l'ús d'una amplia gamma d'eines per millorar el diagnòstic i tractament de les persones (Hanseth & Bygstad, 2015).

Pels beneficis que aporten s'han anat introduint en l'àmbit de la salut (Zheng, Gaff, Smith, & DeLisle, 2014). Sobretot en els últims anys és quan ha sorgit amb molta força un nou camp que fa ús de les tecnologies mòbils, que s'estan establint com una alternativa viable per a poder millorar i fer més accessible l'assistència sanitària (Price et al., 2014).

### 1.2.1. Salut mòbil

L'ús de les tecnologies mòbils per a la gestió de la salut i el benestar de les persones, és el que s'anomena salut mòbil. El concepte de salut mòbil o mSalut (en anglès *mobile health* o *mHealth*) (Adibi, 2015) és definit per l'Organització Mundial de la Salut com "pràctiques mèdiques i de salut pública recolzades per dispositius mòbils, com ara telèfons mòbils, dispositius de monitorització de pacients, assistents digitals personals i altres dispositius sense fil" (World Health Organization, 2011; pàg. 6). Per tant, el camp de l'*mHealth* ha sorgit com un subsegment de l'*eHealth* on s'han implantat l'ús de les tecnologies mòbils en l'àmbit de la salut.

En aquests darrers anys i amb els últims avenços tecnològics van aparèixer els dispositius mòbils. Les estimacions oficials actuals han revelat que hi ha més dispositius mòbils que persones al món i la taxa de penetració a la població continua augmentant (GSMA Intelligence, 2019). Aquesta gran implantació de les tecnologies mòbils, permet disposar d'un nou recurs que ha obert amplíssimes possibilitats. Precisament, l'extensió en l'ús d'aquestes tecnologies en la societat, s'està aprofitant per millorar l'eficiència dels sistemes públics de salut, oferir nous serveis d'assistència sanitària i d'aquesta manera millorar la qualitat de l'atenció que reben les persones (Grady et al., 2018). Aquestes noves eines estan millorant l'assistència sanitària tant en països d'alts ingressos, com també permet oferir serveis sanitaris adequats en països que es troben en vies de desenvolupament o amb infraestructures insuficients (Bastawrous & Armstrong, 2013). Mitjançant l'ús de les aplicacions mòbils que s'han

desenvolupat per als dispositius mòbils és el que permet oferir totes aquestes millores en salut (Bhattacharya, Kumar, Kaushal, & Singh, 2018).

### 1.2.2. Aplicacions mòbils de salut

Les aplicacions mòbils, també anomenades apps, són programes informàtics dissenyats per funcionar i oferir serveis en tot tipus de dispositius mòbils (p.ex., telèfons intel·ligents o tauletes) (Flick & Morehouse, 2011). Les apps de salut són les destinades específicament a oferir utilitats i serveis en l'àmbit de la salut.

Les aplicacions mòbils de salut són una de les principals solucions *mHealth*, arribant a haver-hi apps dirigides a pràcticament tot tipus de problemes de salut (p.ex., dolor, diabetis, alzheimer, càncer; (Kebede & Pischke, 2019; Kessel, Vogel, Schmidt-Graf, & Combs, 2016; Wozney et al., 2018; Zhao, Yoo, Lancey, & Varghese, 2019)). En els últims anys hi ha hagut un gran increment en el nombre d'apps; informes recents senyalen que existirien fins a 325000 apps de salut diferents (Research2Guidance, 2017). Les previsions per als pròxims anys indiquen un gran creixement en el mercat d'apps de salut. Tant és així que l'any 2018, aquest mercat estava valorat en 12400 milions de dòlars i s'espera que per l'any 2025 ja sigui de 236000 milions, amb un CAGR (i.e., taxa de creixement anual compost; (Investopedia, 2019)) del 45% durant aquest període (Research and Markets, 2019).

El gran creixement d'aplicacions mòbils de salut ha multiplicat el seu ús (Accenture, 2018a, 2018b). No solament pacients, també els

mateixos professionals de la salut les estan adoptant i emprant en la seva pràctica clínica (p.ex., per a monitoritzar a distància la salut de pacients crònics; (Chen, Lieffers, Bauman, Hanning, & Allman-Farinelli, 2017; El Amrani, Oude Engberink, Ninot, Hayot, & Carbonnel, 2017). Cada vegada s'estan utilitzant més i fins hi tot s'estan prescrivint apps a pacients per fer tractaments (p.ex., per a realitzar el seguiment del tractament per la diabetis i hipertensió; (Byambasuren, Sanders, Beller, & Glasziou, 2018; Gagnon, Ngangue, Payne-Gagnon, & Desmartis, 2016)). Aquesta proliferació d'apps en combinació amb el seu potencial, està permetent millorar la qualitat de vida de les persones (Pappot et al., 2019; Thurnheer, Gravestock, Pichierri, Steurer, & Burgstaller, 2018).

Existeixen molts tipus d'apps de salut, amb objectius diversos, com també destinades a ser utilitzades per diferents tipus d'usuaris. Aquestes apps solen ser específiques d'una problemàtica, s'adeqüen al seu usuari objectiu i van dirigides exclusivament a pacients i professionals de la salut (p.ex., pacient amb càncer o especialista en pediatria) (Adibi, 2015; World Health Organization, 2011). Però, en general, per les diferents funcions que tenen ens permet fer-ne ús per: (1) cercar i tenir accés a informació, (2) donar suport al diagnòstic i tractament, (3) comunicar-se amb professionals de la salut, (4) monitoritzar a distància el/s problema/es de salut i (5) recollir dades de forma remota (Adibi, 2015; Vital Wave Consulting, 2009).

### 1.2.3. Beneficis en l'ús de les aplicacions de salut

Els sistemes de salut públics de tot el món s'estan beneficiant de l'ús de les aplicacions mòbils (Bhattacharya et al., 2018). El fet més influent

en aquest ús generalitzat, és que les apps de salut han demostrat ser una bona manera de proporcionar serveis sanitaris a persones en llocs remots que tenen dificultats per accedir a aquests serveis (World Health Organization, 2011). També han mostrat alguns avantatges econòmics respecte a les formes tradicionals d'atenció sanitària. Faciliten l'accés a tractaments per a pacients amb ingressos baixos doncs, entre altres coses, redueixen la freqüència de visites a l'especialista (Ernsting et al., 2017; Ventola, 2014). També alleugereix la pressió sobre els sistemes de salut, optimitzant serveis i reduint costs (Adler-Milstein et al., 2013). Per exemple, en el tractament dels pacients crònics, al reduir aquesta necessitat de visites presencials continuades (Balapour, Reychav, Sabherwal, & Azuri, 2019). Les aplicacions mòbils de salut també s'han demostrat que milloren el compliment del tractament, gràcies a que faciliten el seguiment i adherència a les prescripcions terapèutiques (Pérez-Jover, Sala-González, Guilabert, & Mira, 2019). El fet que el pacient pot registrar tota la informació important per al seu tractament a l'instant, permet mostrar les seves dades de salut en temps real, i per tant oferir un millor servei (de la Vega & Miró, 2014; de la Vega, Roset, Galán, & Miró, 2018; Hamine, Gerth-Guyette, Faulx, Green, & Ginsburg, 2015). I, així, es pot fer un seguiment molt més exhaustiu i una monitorització en temps real del pacient, a més de poder donar suport al tractament a distància i amb més rapidesa. Tot això repercuteix en una millora de la qualitat de vida del pacient, ja que rep una millor atenció sanitària (de la Vega et al., 2018; Ernsting et al., 2017; Hamine et al., 2015).

## Introducció

Les aplicacions mòbils de salut també milloren la coordinació assistencial ja que les dades del pacients són més fàcilment accessibles i poden ser compartides entre els diferents professionals i serveis de salut, afavorint una millor atenció al pacient (Bhattacharya et al., 2018; Lu et al., 2018). Tanmateix, també contribueix positivament al sistema públic de salut ja que fent ús d'aquestes tecnologies en el seu dia a dia, li permet ser capaç d'oferir un millor servei, més personalitzat i arribar a més persones (Hanseth & Bygstad, 2015; Ventola, 2014). La Taula 1 presenta de forma resumida els beneficis de les apps de salut.

Taula 1. Beneficis de l'ús de les aplicacions mòbils de salut

| Categoria | Benefici |
|---|---|
| **Econòmics** | |
| | Accedir a tractaments des de llocs remots |
| | Reduir els desplaçaments tant dels metges com dels pacients |
| | Optimitzar els recursos sanitaris al no ser necessari tantes visites al metge |
| **Execució i adherència al tractament** | |
| | Millorar l'accessibilitat a recursos i informació sobre salut per resoldre dubtes |
| | Facilitar la comunicació pacient-metge |
| | Contribuir al registre de dades de salut en temps real |
| | Contactar amb persones amb la mateixa situació més fàcilment |
| | Realitzar el tractament majoritàriament a distància |
| **Seguiment del pacient** | |

| |
|---|
| Accedir fàcilment a la informació clínica |
| Monitoritzar en temps real l'estat del pacient |
| Contribuir al suport al pacient a distància |

Per tots aquests grans avantatges que ofereixen i les seves funcionalitats, les aplicacions mòbils de salut estant tenint un gran impacte en millorar l'atenció de la salut en el món (Vital Wave Consulting, 2009). Per exemple, un dels camps en el que les apps de salut estant tenint més repercussió és amb les persones amb problemes crònics de salut (un 56% de les apps de salut va dirigit a aquest tipus de pacients (Research2Guidance, 2016)). Segons estudis recents, almenys un 30% de la població pateix algun problema d'aquest mena (p.ex., depressió, diabetis, hipertensió, dolor). A Catalunya, el 38% de la població major de 15 anys n'està afectada (Generalitat de Catalunya. Departament de Salut, 2019), mentre que a Espanya és el 31% dels majors de 16 anys (Instituto Nacional de Estadística - España, 2019). A nivell internacional les dades són molt similars, a la Unió Europea el 33% de les persones amb 16 anys o més (OECD/EU, 2018) i als països que pertanyen a l'Organització de Cooperació i Desenvolupament Econòmic (OECD) és el 31% de les persones amb 15 anys o més (OECD, 2019). L'estimació és que per l'any 2030 hi hagin un total de 52 milions de morts en el món per aquests tipus de malalties cròniques (World Health Organization, 2014).

Malgrat la gran repercussió que estant tenint les apps de salut, fa relativament poc temps que s'estan utilitzant. Tot i que hi ha evidència

abundant que contribueixen a millorar la qualitat de vida, no hi ha encara prou dades sobre els seus efectes a llarg termini (Marcolino et al., 2018). Per tant, encara hi ha plantejats dubtes, raonables, sobre els efectes de les apps en la salut de les persones, sobretot a llarg termini.

### 1.2.4. Risc en l'ús de les aplicacions de salut

La gran popularització dels dispositius mòbils ha permès en els darrers anys, estendre l'ús de les aplicacions mòbils de salut de forma molt considerable en la societat (GSMA Intelligence, 2019; Research2Guidance, 2017), fet que ha portat molts beneficis, com hem vist anteriorment. Però, el fervor tecnològic que desperten aquestes apps provoca -sense voler-ho- que apareguin problemes que poden passar desapercebuts. Pot resultar difícil saber si una app és de qualitat, té les funcions esperades i realment fa el què indica (Larson, 2018). Particularment greu és l'ús d'apps no validades i que promouen propostes que no estan fonamentades en evidències científiques (Akbar, Coiera, & Magrabi, 2020; Lewis & Wyatt, 2014). Això fa que les persones que utilitzen les aplicacions mòbils de salut estiguin exposades a certs riscs que els puguin provocar perjudicis, en ocasions fins i tot importants (p.ex., la falta d'eficàcia del tractament, diagnòstic erroni, proporcionar informació incorrecta, fer un seguiment inadequat de la salut del usuari o un mal ús de les seves dades personals (Akbar et al., 2020; Balapour et al., 2019)). Aquests riscos impliquen tant la integritat de la salut personal com també la seguretat informàtica, privacitat, usabilitat, utilitat i la validesa del

contingut que ofereixen les apps (Kamel Boulos, Brewer, Karimkhani, Buller, & Dellavalle, 2014; Lewis & Wyatt, 2014; Singh et al., 2016).

Hi ha múltiples estudis que han revisat la qualitat de les aplicacions mòbils de salut i les conclusions no són gaire positives en cap cas, detectant molts casos de apps de baixa qualitat. Per exemple, en un estudi sobre l'estat de la privadesa i la seguretat de la informació en les apps, els autors informen que un 96% presenten importants deficiències, fet que suposa un perill potencial per a la protecció de les dades personals dels usuaris (Dehling, Gao, Schneider, & Sunyaev, 2015). Igualment, en un estudi sobre com la falta d'evidències i com aquestes afecten a la qualitat de les apps, es mostren múltiples problemes relacionats amb la seva utilitat i usabilitat (Buijink, Visser, & Marshall, 2013). Una revisió recent sobre les aplicacions mòbils relacionades amb el dolor, va identificar-ne un total de 283 de disponibles en les botigues d'aplicacions mòbils, i la gran majoria d'elles no havien passat proves d'usabilitat i validació (de la Vega & Miró, 2014). En un estudi semblant, en el que es van analitzar apps relacionades amb la gestió de l'ansietat, es va comprovar que en el desenvolupament de la majoria d'aquests apps no hi havia implicats professionals de la salut (Sucala et al., 2017). És significatiu que un dels estudis més exhaustius que s'han publicat fins ara, analitzant les aplicacions mòbils de salut acreditades per la *National Health Service* del Regne Unit, s'haguessin detectat un bon número de deficiències en la major part d'elles (p.ex., sense polítiques de privadesa i transmissió d'informació personal de forma insegura) (Huckvale, Prieto, Tilney, Benghozi, & Car, 2015; Wicks &

Chiauzzi, 2015). A la Taula 2 es presenten de forma resumida els problemes detectats a les apps de salut.

Taula 2. Problemes detectats en les aplicacions mòbils de salut

| **Categoria** | **Problema** |
|---|---|
| **Qualitat del contingut** | |
| | Proporcionar informació incorrecta |
| | Donar informació incompleta |
| | Donar resultats erronis en els càlculs i diagnòstics |
| | Resposta inadequada a les necessitats del consumidor |
| **Usabilitat i utilitat** | |
| | No admet adequadament les tasques dels consumidors amb mancances en les seves funcions |
| | Falta de validació per a l'entrada dels usuaris |
| | Retards en el funcionament |
| | Resposta inadequada als perills per a la salut |
| | Problemes en les notificacions i les alarmes |
| **Procés de desenvolupament** | |
| | Falta d'implicació d'experts |
| | No estar basada en l'evidència disponible |
| | Validació deficient |
| **Seguretat i privadesa** | |
| | Recopilar en excés informació sensible per a l'usuari |
| | Emmagatzemar i transmetre informació de manera insegura ja que no esta xifrada |
| | Transmetre informació sensible a tercers sense consentiment |
| | Falta d'una política de privadesa clara |

## 1.3. Desenvolupament i criteris de qualitat per a les aplicacions de salut

### 1.3.1. Iniciatives pel control de qualitat de les aplicacions

Davant de la baixa qualitat detectada en moltes de les aplicacions mòbils de salut (Wisniewski et al., 2019; Wyatt, 2018) i dels riscs que això suposa per a les persones, diferents institucions han tractat de desenvolupar sistemes per resoldre aquest problema. En general, les iniciatives provenen, (1) d'organitzacions professionals, (2) d'organitzacions reguladores i (3) d'investigadors. En el primer grup trobem entitats que agrupen professionals del sector tecnològic i de salut, que realitzen diferents accions per tal d'establir una sèrie de recomanacions i guies pel bon desenvolupament d'aplicacions de salut. En el segon grup hi ha iniciatives que impulsen els propis estats per tal de desenvolupar unes normatives i establir unes regulacions clares per a les apps. Finalment, les accions empreses per l'últim grup, estan relacionades amb àmbits específics de la recerca, de manera que els i les investigadores s'esmercen intentant desenvolupar instruments per tal de valorar la qualitat de les apps de salut.

Algunes d'aquestes iniciatives sorgides per tal d'afrontar i resoldre la falta de qualitat d'una gran quantitat d'aplicacions mòbils de salut són les següents. En el cas de recomanacions i guies, a Catalunya es va aprovar el *Pla de Mobilitat mHealth.cat* (Tic Salut Social Foundation, 2015) per donar suport al desenvolupament de la salut mòbil, això inclou el portal *AppSalut* (TIC Salut Social, 2018). Aquest portal, és un sistema

d'acreditació de la qualitat per apps ja existents i una guia per crear-ne de noves, relacionades amb els àmbits de la salut i social (Tic Salut Social Foundation, 2018). Hi ha altres guies d'aquest estil que s'han creat i pretenen acreditar les apps de salut, per exemple el *Distintivo AppSaludable* (Agencia de Calidad Sanitaria de Andalucia, 2014) d'Andalusia, *Good practice guidelines on health apps and smart devices (mobile health or mhealth)* (Haute Autorité de Santé, 2016) de França, *Health apps & co: safe digital care products with clearer regulations* (IGES Institut GmbH, 2016) d'Alemanya i *Medical devices: software applications (apps)* (Medicines and Healthcare products Regulatory Agency, 2014) del Regne Unit.

En l'establiment de normatives i regulacions, hi han hagut iniciatives internacionals per tal de poder regular aquest tipus d'apps i disposar d'uns criteris unificats. Per exemple, la Comissió Europea va publicar el *Llibre Verd sobre salut mòbil* (European Commission, 2014), on es defineix el conjunt de pràctiques mèdiques i de salut pública basades en dispositius mòbils que cal seguir en el conjunt de la Unió Europea. Tot això, per tal d'oferir un marc legal en el qual es pugui garantir la qualitat de les apps utilitzades per totes les parts implicades. També, l'agència *Food and Drug Administration* dels Estats Units va publicar la *Policy for Device Software Functions and Mobile Medical Applications* (Food and Drug Administration, 2019) on indica de quina manera es regula les apps de salut (Yetisen et al., 2014). En aquesta regulació es diferència entre les apps clarament mèdiques i la resta, i es concreta que les primeres han de passar

uns controls molt més estrictes. Una altra proposta que segueix una línia molt semblant a l'anterior és la *Regulation of Software as a Medical Device* (Therapeutic Goods Administration (TGA), 2019) d'Austràlia. Aquesta, pretén regular tot aquell programari o apps que siguin dispositius mèdics.

Per altra banda i individualment, els investigadors i els grups de recerca, també estan intentant aportar solucions (Nouri, Kalhori, Ghazisaeedi, Marchand, & Yasini, 2018), creant eines per a mesurar la qualitat de les aplicacions mòbils de salut. Una d'aquestes iniciatives és la d'Stoyanov i col·laboradors (Stoyanov et al., 2015). El seu treball va implicar una revisió dels estudis publicats sobre criteris de qualitat explícits relacionats amb les aplicacions mòbils. Fruit de la seva feina és la *Mobile App Rating Scale*, una escala per a valorar la qualitat de les aplicacions mòbils de salut. Malgrat ser un avenç, la feina d'aquests investigadors planteja algunes limitacions importants, ja que es va desenvolupar des d'una perspectiva molt específica i força limitada. Els autors van utilitzar fonts d'informació sobre apps de salut però van deixar de banda altra informació rellevant (p.ex., normatives i estàndards sobre software de dispositius mèdics). L'ús de fonts d'informació més amplia hauria augmentat la seva validesa i fiabilitat, per tant, no sembla adequada per a tots els casos.

Tot i els grans avenços que s'han fet en els darrers anys en el control de qualitat, encara no hi ha un únic criteri establert per tal de valorar la qualitat de les apps de salut. Com hem vist anteriorment hi ha moltes iniciatives i, encara que tenen coses en comú, també tenen moltes

diferències entre si ja que estan fetes des de punts de vista diferents. En conseqüència no hi han uns criteris clars per establir que una aplicació de salut és segura d'utilitzar i de qualitat. I això dificulta el progrés de forma segura i eficient d'aquesta àrea de coneixement.

### 1.3.2. Necessitat d'una guia general

En un món globalitzat com l'actual, qualsevol persona pot accedir a qualsevol tipus d'aplicació, independentment del país on s'hagi creat, suposant un avantatge però alhora un problema important per l'usuari. Al no haver-hi cap mena de filtre i control de qualitat generalitzat, es posa en risc la salut de l'usuari, si l'aplicació no és de qualitat, fonamentada en evidències científicament contrastades.

Per tal de garantir la seguretat de tots els usuaris i que puguin utilitzar les apps de salut amb plena confiança, s'han de determinar quines característiques ha de tenir una aplicació de salut per considerar-la de qualitat i que sigui fàcilment comprovable pels mateixos usuaris. Per tal d'aconseguir-ho, és necessari disposar d'un estàndard que sigui una guia de qualitat per a les aplicacions mòbils de salut. Aquesta guia, ha de servir tant per a ajudar als professionals de la salut, com als pacients i desenvolupadors a valorar la qualitat de les apps existents com a desenvolupar-ne de noves. Els professionals de la salut han d'estar segurs que les apps que utilitzen no causaran cap perjudici als seus pacients i que no afectaran negativament a la seva pràctica sanitària (Boudreaux et al., 2014; Ventola, 2014). Els pacients necessiten tenir la seguretat que l'aplicació de salut que volen utilitzar, compleix amb uns estàndards de

qualitat mínims per tal que no hi hagi cap mena de riscs per a ells (Larson, 2018; Lewis & Wyatt, 2014). Per últim, els desenvolupadors s'han d'assegurar que les apps que creen puguin ser utilitzades de forma segura, compleixen amb el seu propòsit i siguin de qualitat (Martínez-Pérez, de la Torre-Díez, & López-Coronado, 2015; Wicks & Chiauzzi, 2015).

### 1.3.3. Criteris de qualitat

La guia ha d'establir una sèrie de criteris clarament definits, per tal de determinar amb total certesa la qualitat de les aplicacions mòbils de salut. Com hem vist, ja hi han diferents guies, legislacions i estudis científics que proven d'establir i promoure una sèrie de criteris a seguir o les característiques que han de tenir les aplicacions mòbils de salut. Però cada una d'elles proposa conclusions i solucions molt específiques sobre com ha de ser una aplicació mòbil en salut de qualitat. Per exemple, s'han fet criteris dirigits a un problema de salut específic (p.ex., intervencions d'emergència) (Gaziel-Yablowitz & Schwartz, 2018) o a un col·lectiu en concret (p.ex., adolescents) (Brown, Yen, Rojas, & Schnall, 2013). Per tal de progressar, s'ha d'anar més enllà del que han realitzat estudis anteriors, superant les seves limitacions (p.ex., perspectiva limitada i centrada en un sol punt de vista).

Per obtenir aquest estàndard general i així aconseguir una guia de qualitat, s'han d'identificar quins són els criteris realment importants i imprescindibles per tal de determinar la qualitat de les apps. Però identificar els criteris no seria suficient, tanmateix cal arribar a un consens amb tots els actors rellevants del sector, de manera que d'aquest acord es pugui

derivar i validar la importància del conjunt de criteris identificats. Tenir una guia acordada per totes les parts interessades, facilitaria que fos una eina adoptada i utilitzada per tothom. Aquesta tesi doctoral està relacionada, precisament, amb aquest objectiu, i aspira a donar respostes a algunes de les necessitats detectades i que aquí s'han descrit de forma breu.

# Part Pràctica

# 2. Objectius

A grans trets, aquesta tesi doctoral presenta el treball realitzat per a aconseguir una guia amb la que facilitar el desenvolupament i avaluació d'aplicacions mòbils de salut de qualitat. Assolir aquest objectiu implicava superar tres objectius parcials i específics, que es descriuen a continuació.

## 2.1. Objectiu 1

Concretar un estàndard per a les aplicacions relacionades amb la salut mòbil. En essència, es tractava d'identificar d'entre els conjunt de criteris disponibles aquells que fossin estratègics, per a que es poguessin recomanar i integrar-se en un estàndard general (és a dir, una guia) que permetés ajudar aquest camp a avançar sobre unes bases sòlides.

## 2.2. Objectiu 2

Estudiar la validesa de la *Mobile App Development and Assessment Guide* (MAG). Bàsicament, volíem analitzar la validesa i explorar si calia incloure nous criteris per a millorar la guia. També volíem examinar la importància percebuda d'aquests criteris des dels diferents grups d'interès.

## 2.3. Objectiu 3

Estudiar i comparar la *Mobile App Rating Scale* (MARS) -una de les guies més utilitzades en l'àrea- i la guia creada per nosaltres (*Mobile App Development and Assessment Guide;* MAG). Concretament,

## Objectius

preteníem comparar la fiabilitat entre avaluadors d'ambdós guies, i contrastar la seva fiabilitat en l'avaluació de diferents tipus d'apps.

# 3. Mètodes

Aquesta tesi doctoral inclou tres estudis: una revisió i identificació dels criteris de qualitat existents per integrar-los en una guia (Estudi 1), un estudi Delphi per validar els criteris identificats i examinar la seva importància (Estudi 2) i una anàlisis empírica i comparació de la fiabilitat de la guia creada (Estudi 3).

## 3.1. Participants

En el primer estudi (Llorens-Vernet & Miró, 2020a), vam utilitzar tres estratègies. Primer, vam realitzar una revisió sistemàtica d'estudis publicats utilitzant la següent estratègia de cerca: *(pain OR *ache) AND (smartphone* OR mobile) AND (app OR apps)*, en les bases de dades: Web of Science, Scopus, PubMed, ScienceDirect. En segon lloc, vam cercar recomanacions per a apps de salut d'organitzacions professionals. Per últim, es van revisar les normatives i estàndards en desenvolupament del programari de dispositius mèdics d'organitzacions reguladores. Finalment, es va demanar a un grup de parts interessades (professionals sanitaris, enginyers i usuaris finals d'apps de salut), per fer un anàlisi dels criteris recopilats.

En el segon estudi (Llorens-Vernet & Miró, 2020b), van participar tres grups diferents de persones, tots actors rellevants sobre el tema: (1) professionals sanitaris, (2) desenvolupadors d'aplicacions relacionades amb la salut i (3) usuaris d'aplicacions de salut. Per identificar aquests potencials participants vam utilitzar cinc estratègies: (1) contactar amb

organitzacions de professionals de salut digital, (2) cercar els desenvolupadors d'aquelles aplicacions de salut més valorades, (3) buscar en associacions de pacients usuaris amb experiència d'ús d'aplicacions de salut, (4) difondre l'estudi a través de xarxes socials per reclutar participants i (5) demanar ajuda a investigadors i clínics coneguts per identificar participants. En la primera ronda de l'enquesta hi van participar 42 persones i en la segona ronda 24.

En el tercer estudi (Miró & Llorens-Vernet, 2020), 8 *stakeholders* van avaluar 4 apps utilizant dues guies: la *Mobile App Rating Scale* (MARS; (Stoyanov et al., 2015)) que és una de les escales de qualificació més utilitzades actualment, i la *Mobile App Development and Assessment Guide* (MAG), la guia creada per nosaltres. Les apps que es van seleccionar estaven relacionades amb problemes crònics de salut i es van identificar en les botigues d'aplicacions d'Android i iOS (Google Play i App Store) en la categoria mèdica. Els termes de cerca utilitzats, relacionats amb les 4 majors condicions cròniques de salut, van ser: *"pain", "cancer", "diabetes"* i *"cardiovascular"*. Amb els resultats de la cerca obtinguts, es van descartar aquelles que no tinguessin a veure amb malalties cròniques, es van seleccionar les que eren gratuïtes i estaven en castellà o anglès. Finalment, es van identificar les 4 aplicacions més descarregades (la primera per cada categoria) per ser utilitzades en l'estudi. El grup de revisors seleccionats incloïa investigadors clínics, enginyers, professionals de la salut i usuaris finals que es van encarregar d'avaluar aquest conjunt d'apps.

## 3.2. Procediments i mesures

Per a la realització de la revisió sistemàtica del primer estudi, vam seguir les recomanacions PRISMA (Moher, Liberati, Tetzlaff, Altman, & Group, 2009). Per a la cerca de les recomanacions d'organitzacions professionals i les normatives i estàndards d'organitzacions reguladores, ens vam centrar en aquelles regions on l'ús de les tecnologies mòbils era més elevat. Primer, es van recopilar en una llista els criteris identificats en les tres fonts d'informació. Després, es van agrupar els criteris per temes en funció de les característiques compartides. Com a resultat es va crear una llista de criteris únics, classificats en categories. Finalment es va demanar la col·laboració d'un grup de 18 persones (actors rellevants en el tema), per fer una anàlisi preliminar sobre la comprensibilitat i importància percebuda de tots els criteris recopilats.

En el segon estudi, es va utilitzar el mètode Delphi (Linstone, Turoff, & Helmer, 1975). En aquest cas, va consistir en dues rondes d'enquestes en línia als participants. Les enquestes recollien informació sociodemogràfica dels participants (p.ex., edat, sexe) i una llista d'ítems de la guia per a ser avaluats. Als participants se'ls demanava que afegissin criteris que consideressin importants i que la llista no incloïa.

En el tercer estudi, per contrastar la fiabilitat de la MAG, es van avaluar 4 apps de salut amb aquesta guia i amb la MARS. Les persones encarregades d'avaluar les apps i que van participar en aquest estudi van rebre: (1) el llistat d'apps a analitzar, (2) una enquesta amb cada un dels elements a analitzar utilitzant la MAG i la MARS i (3) les instruccions

especifiques de com havien de procedir. Per a l'avaluació, cada revisor es va descarregar i instal·lar les aplicacions en el seu mòbil o tauleta personal.

Una explicació més detallada dels procediments es pot trobar en cada un dels articles.

# 4. Resultats

Els tres estudis inclosos en aquesta tesi doctoral es detallen a continuació:

- Estudi 1: Llorens-Vernet P, Miró J. Standards for Mobile Health-related Apps: Systematic review and Development of a Guide. JMIR mHealth and uHealth 2020;8(3):e13057. URL:https://mhealth.jmir.org/2020/3/e13057. doi:10.2196/13057. PMID:32130169
- Estudi 2: Llorens-Vernet P, & Miró J. The Mobile App Development and Assessment Guide (MAG): Delphi-Based Validity Study. JMIR mHealth and uHealth 2020;8(7):e17760. URL:https://mhealth.jmir.org/2020/7/e17760. doi:10.2196/17760. PMID: 32735226
- Estudi 3: Miró J, Llorens-Vernet P. On the Assessment of the Quality of mHealth-related Apps: an Interrater Reliability Study of Two Guides. 2020. Article enviat per a publicació.

## 4.1. Estudi 1

Standards for Mobile Health-related Apps: Systematic review and Development of a Guide.

Review

# Standards for Mobile Health–Related Apps: Systematic Review and Development of a Guide

Pere Llorens-Vernet, MSc; Jordi Miró, PhD

Unit for the Study and Treatment of Pain-ALGOS, Department of Psychology, Universitat Rovira i Virgili, Tarragona, Spain

**Corresponding Author:**
Jordi Miró, PhD
Unit for the Study and Treatment of Pain-ALGOS
Department of Psychology
Universitat Rovira i Virgili
Carretera de Valls, s/n
Tarragona, 43007
Spain
Phone: 34 977558179
Email: jordi.miro@urv.cat

## Abstract

**Background:** In recent years, the considerable increase in the number of mobile health (mHealth) apps has made health care more accessible and affordable for all. However, the exponential growth in mHealth solutions has occurred with almost no control or regulation of any kind. Despite some recent initiatives, there is still no specific regulation procedure, accreditation system, or standards to help the development of the apps, mitigate risks, or guarantee quality.

**Objective:** The main aim of this study was to propose a set of criteria for mHealth-related apps on the basis of what is available from published studies, guidelines, and standards in the various areas that are related to health app development.

**Methods:** We used three sources of information to identify the most important criteria. First, we conducted a systematic review of all the studies published on pain-related apps. Second, we searched for health app recommendations on the websites of professional organizations. Third, we looked for standards governing the development of software for medical devices on the specialized websites of regulatory organizations. Then, we grouped and subsumed the criteria we had identified on the basis of their shared characteristics. Finally, the comprehensibility and perceived importance of the resulting criteria were evaluated for face validity with a group of 18 stakeholders.

**Results:** We identified a total of 503 criteria from all sources, which, after close analysis, were grouped into eight different categories, including 36 important criteria for health apps. The resulting categories were *usability, privacy, security, appropriateness and suitability, transparency and content, safety, technical support and updates,* and *technology*. The results of the preliminary analysis showed that the criteria were mostly understood by the group of stakeholders. In addition, they perceived all of them as important.

**Conclusions:** This set of criteria can help health care providers, developers, patients, and other stakeholders to guide the development of mHealth-related apps and, potentially, to measure the quality of an mHealth app.

*(JMIR Mhealth Uhealth 2020;8(3):e13057)* doi: 10.2196/13057

**KEYWORDS**

mHealth; mobile apps; review; medical device; standards

## Introduction

### Background

Public health care systems worldwide are facing major challenges (eg, a shortage of resources and a steady increase in demand), which can make them increasingly unsustainable [1]. It is in this environment that what is known as mobile health (mHealth) is proving to be of key importance [2-4]. In the last few years, mHealth has undergone considerable development because of its potential to make health care more accessible and affordable for all [2,5-7].

However, mHealth solutions have grown exponentially with almost no control or regulation of any kind. In fact, very few of the health apps available have undergone a thorough

validation process, and this causes a lack of confidence among health professionals [8,9]. For example, a recent review of the mobile apps available for chronic pain—which is one of the most prevalent health problems, with an enormous economic cost to individuals, families, and society [10]—highlighted that of the 283 apps available at the time, just a handful had undergone usability and validity tests [7]. This situation has been identified as preventing the field from improving and advancing [9].

In this so-called *strategic field*, progress depends not only on what each research group is doing but also on developing general standards and improving certification procedures [11,12]. There are some local and international initiatives to help in this process. For example, Catalonia approved a strategic action plan to support the development of mHealth (ie, *The Mobility Master Plan: mHealth solutions* [13]), which includes *AppSalut* [14], an accreditation system and guide [15] created to certify the quality of health– and social-related apps. At the international level, the European Commission published a Green Paper on mHealth [16] and launched a public consultation to identify potential barriers to and problems in the development of mHealth. Despite these initiatives, there is still no specific regulation procedure, accreditation system, or standards to help the development of apps, mitigate risks, and guarantee quality.

Therefore, the certification process is weighed down by the lack of clear standards to guide users through the different stages of the process. This is a problem not only for the safety of end users (ie, patients and health care professionals) but also for professional developers. Clearly, having a set of common criteria would be instrumental in helping the field to make progress in a consensual way and overcome potential risks for all stakeholders. There has been one recent attempt to develop a rating scale for mobile apps that could be used to help overcome this problem. Stoyanov et al [17] developed a scale (Mobile App Rating Scale [MARS]) to classify and rate the quality of mHealth apps. This scale was on the basis of a review of the papers published between 2000 and 2013, which contained explicit app-related quality rating criteria. However, this scale was created from a very narrow perspective for assessing already developed apps. That is to say, although the authors used information from studies on existing mobile apps, they failed to include information from other relevant sources that had been used, which would have increased the reliability and validity of their work (eg, standards governing the development of software for health or medical devices). Therefore, this scale does not seem to be suited for use by all stakeholders. Some more recent attempts to provide alternatives to assess mHealth apps also share some of these weaknesses (eg, developed for one specific group of stakeholders and using one specific source of information) [18,19].

### Objectives

The general aim of this study, then, is to go beyond what is already available and provide a standard for mHealth-related apps by studying the published studies, guidelines, and standards available in the field of health app development. In particular, we want to identify a set of criteria that are used, and which of these are strategic, so that they can be recommended and integrated into a general standard (ie, a guide) that can help the field move forward on solid grounds.

## Methods

### Procedure

We used three strategies to identify criteria. First, we conducted a systematic review of all the studies published on pain-related apps. Second, we searched the websites of professional organizations. Finally, we analyzed the standards governing the development of software for medical devices. Although these regulations are not specific for health apps, they can provide information of interest and complement the information collected.

### Information From the Systematic Review

For the systematic review, to address an otherwise unmanageable amount of information, we limited our search to mobile apps related to pain, one of the most prevalent health problems causing millions of visits to health care professionals. In so doing, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines [20] and searched the following databases: Web of Science, Scopus, PubMed, and ScienceDirect. We used the search terms (pain OR *ache) AND (smartphone* OR mobile) AND (app OR apps), and also hand searched the reference lists from relevant articles. Only peer-reviewed articles published in English or Spanish between 2008 (the release date of the first apps stores [21]) and December 2017 were included.

### Information on Websites From Professional Organizations

The second strategy consisted of searching websites from professional organizations that had guidelines and recommendations for health apps. We decided to focus our search on those regions where the mHealth market is most significant and, therefore, where these regulations are most likely to be found. According to a forecast of revenues of the world mHealth market [22], in 2017, the main mHealth markets by regions were Europe, North America, and Asia–Pacific, representing 30%, 28%, and 30% of the world market, respectively. In addition, the 15 most attractive countries for digital health solutions in 2017 were located in the regions mentioned above [23]. Therefore, we limited our search to those three areas. Again, and to make the search feasible, we limited our analysis to those countries that are the main markets in each region. In Europe, we included the United Kingdom and Spain as they had the same importance in terms of mHealth markets; in North America, the United States; and in Asia–Pacific, Australia.

### Information From Standards Governing the Development of Software for Medical Devices

Finally, in our search for information, we also searched for standards governing the development of software for medical devices on the specialized websites of regulatory organizations. To conduct this search, we also focused on the regions and countries where the mHealth market has been shown to be most significant, as described above. In this analysis, we focused the

search on those standards with criteria related to health apps, added to our list only those criteria that are specific to health apps, and left all others out of our scrutiny (eg, protection against radiation and chemical properties).

### Development of a Common Set of Criteria and Categories

We first compiled a list of the criteria identified in (1) published studies, (2) guidelines, and (3) standards governing the development of software for medical devices. Next, we grouped the criteria in categories on the basis of their shared characteristics. That is, each criterion was closely analyzed to identify what its general purpose was (eg, the criterion *the functionality is adapted to the purpose of the app* was considered related to usability, and this opened a group or category that was labeled *usability*). All criteria underwent the same scrutiny. In the case that no category existed, a new one was created and labeled. If the category already existed, then the criterion was subsumed under that existing category. As a result of this analysis, we obtained a list of unique criteria classified into categories according to their similarity.

### Preliminary Analysis of the Set of Criteria

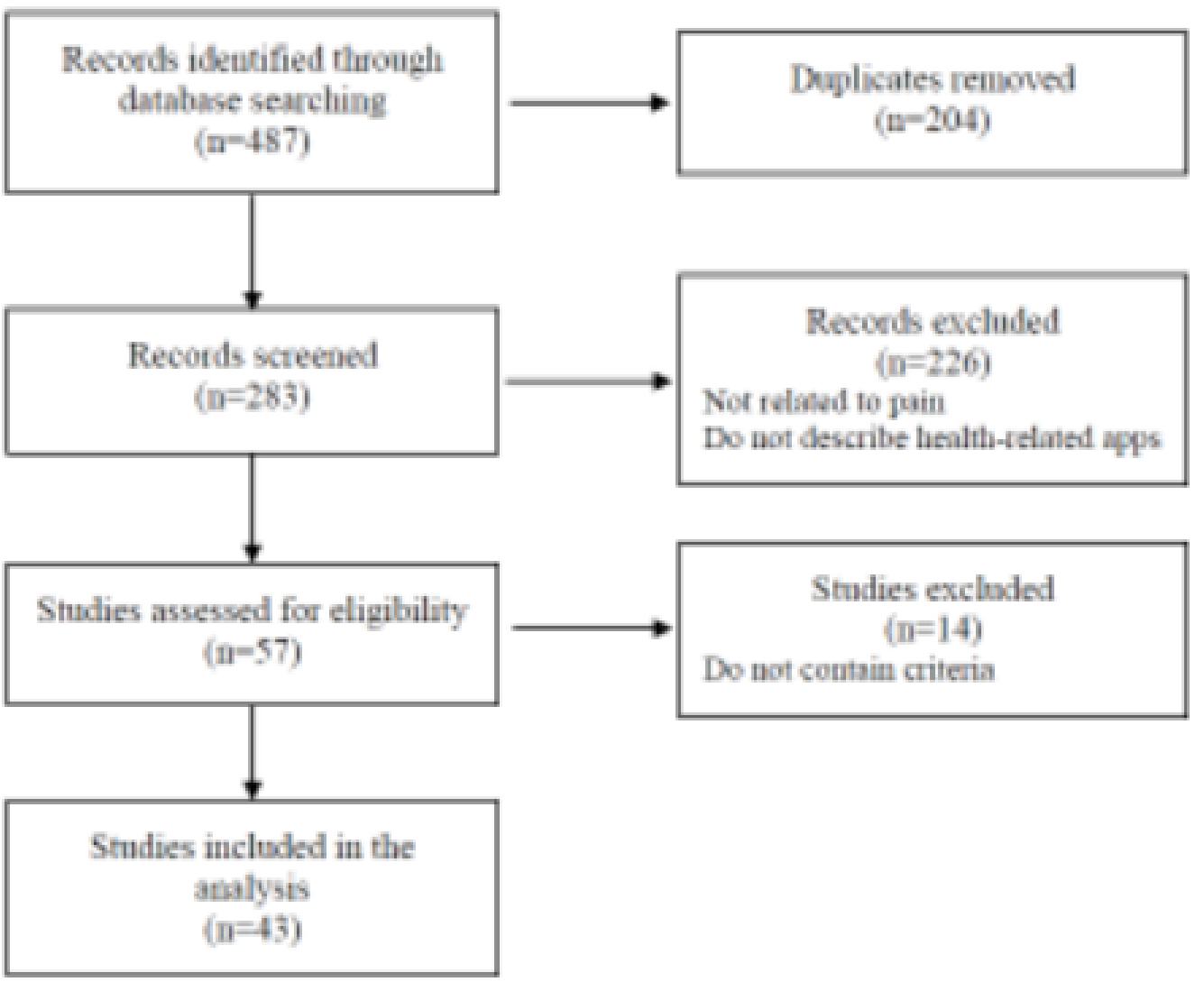
The resulting set of criteria underwent a preliminary analysis of their face validity by asking stakeholders to report on the comprehensibility and perceived importance of all the criteria. Specifically, in this analysis, we requested the collaboration of a group of individuals from different groups of stakeholders (ie, final users, potential patients, health care professionals, and developers or engineers). Final users or potential patients and health care professionals were approached by the authors while at the university hospital (while they were visiting for a health checkup and while at work, respectively). Engineers were professors or technicians working at the university. Before the participation of stakeholders, we first requested and obtained permission from the Ethics Committee of the School of Education Sciences and Psychology for the study procedures. Participants had to sign a consent form. All were asked to respond to two questions in relation to each criterion: (1) "Do you understand the criterion?" and (2) "How important is this criterion for a health-related mobile application?" The first one was responded with *yes*, *no*, or *partially* to the question, whereas the second one was to be responded by providing a number that best represented the importance of the criterion, between 0 (*not important at all*) to 10 (*utmost important*).

## Results

### Information From the Systematic Review

Our review of the scientific databases identified 283 nonduplicated papers. Of these, only 43 were of interest for our purposes. Studies that were not related to pain or that did not describe health-related apps were deemed irrelevant, and not included in the analysis (see Figure 1). In this search, 168 criteria were identified (the full list is provided in Multimedia Appendix 1).

**Figure 1.** Flowchart of systematic review selection process.



### Information From Websites of Professional Organizations

Following the planned strategy, we found just 4 organizations that had developed guidelines and recommendations for health-related apps. Of these, 3 were of national coverage—Andalusian Agency for Healthcare Quality (Spain), TIC Salut Social Foundation (Spain), and National Health Service (United Kingdom)—and 1 of supranational or international coverage, the European Commission (European Union [EU]). No similar information was found in the other searched regions. From each of the guidelines, we collected only the criteria that were specifically related to mobile phone apps and discarded the criteria related to other technologies (eg, wearables and websites): Andalusian Agency for Healthcare Quality, 31 criteria; European Commission, 58 criteria; National Health Service, 78 criteria; and TIC Salut Social Foundation, 115 criteria (see Multimedia Appendix 1).

### Information From Standards Governing the Development of Software For Medical Devices

As planned, in our search of the main mHealth markets, we also looked at the specialized websites of regulatory organizations and searched through standards in the regulations of medical devices. In so doing, we found just two standards that were of interest: (1) *Mobile Medical Applications: Guidance for Food and Drug* (the United States) and (2) *Regulation of medical software andmobile medical "apps"* (Australia). In this analysis, we added to our list only those criteria that were specific to health apps and left all others out of our scrutiny (eg, protection against radiation and chemical properties). We identified 42 and 11 criteria, respectively (see Multimedia Appendix 1).

### Development of a Common Set of Criteria and Categories

Then, the set of criteria were grouped in categories according to their similarity. That is to say, the criteria of the same class were grouped and subsumed together (see Table 1), resulting in eight categories. The categories were the following: *usability* (this includes criteria that are related to user experience), which contained eight criteria; *privacy* (ie, criteria related to data protection, compliance with the law, and treatment of users' data), which contained six criteria; *security* (ie, criteria related to cybersecurity, encryption mechanisms for the storage and transmission of data, and measures against vulnerabilities), which contained four criteria; *appropriateness and suitability* (ie, criteria related to the adaptation of the app for the benefit of the targeted user), which contained three criteria; *transparency and content* (ie, criteria related to the sharing of information in relation to the development of the app), which contained five criteria; *safety* (ie, criteria related to the identification and prevention of harm to end users), which contained two criteria; *technical support and updates* (ie, criteria related to helping the user to solve problems in using the app), which contained four criteria; and *technology* (ie, criteria related to the proper functioning of the app), which contained four criteria (see Table 1).

### Preliminary Analysis of the Set of Criteria

A total of 18 individuals participated: 7 final users or potential patients, 6 health care professionals, and 5 developers, all of whom were approached and consented. Participants' age ranged from 18 to 53 years, with an equal distribution of females and males in the sample. At the time of participation, all were attending school or working. They all had experience with mobile phones and in using mobile apps.

The results of this analysis are summarized in Table 1, which includes information about the percentage of participants within each group that understood the criteria, and the mean of the perceived importance of each one.

The criteria were understood by most of the participants in the three groups. However, at least one participant in one or more groups reported being unsure about the exact meaning. All the issues were related to the use of technical vocabulary or lack of some very specific (technical) knowledge; nevertheless, with additional explanation, the issues were solved. In addition, all criteria were perceived as important; 7 (on a 0-10 numerical rating scale) was the lowest rating received by any criterion, and most ratings were between 8 and 10 (Table 1 summarizes the information).

**Table 1.** Comprehensibility and perceived importance of the criterion by stakeholders.

| Category and criterion | Comprehension, n (%) | | | Perceived importance (0-10) | | |
|---|---|---|---|---|---|---|
| | Patients (N=7) | Clinicians (N=6) | Engineers (N=5) | Patients | Clinicians | Engineers |
| **Usability** | | | | | | |
| The app has been tested by potential users before being made available to the public. | 7 (100) | 5 (83) | 5 (100) | 8.9 | 8.6 | 9.4 |
| It has instructions or some kind of assistance for use. | 6 (86) | 6 (100) | 5 (100) | 10 | 8.5 | 7 |
| It is easy to use (ie, navigation is intuitive). | 7 (100) | 6 (100) | 5 (100) | 9.6 | 9.3 | 9 |
| It follows the recommendations, patterns, and directives in the official manuals of the different operating systems (Android, iOS, or others). | 7 (100) | 5.5 (92) | 4 (80) | 7.5 | 7 | 7.9 |
| The interface design follows the same pattern. That is, all graphic elements (typographies, icons, and buttons) have a consistent appearance. The function of each element (navigation menu, lists, and photo gallery) is clearly identified. | 7 (100) | 6 (100) | 5 (100) | 9 | 8.3 | 8.8 |
| The functionality is adapted to the purpose of the app. | 7 (100) | 5.5 (92) | 5 (100) | 9.6 | 8.5 | 8.8 |
| The information of the app must be able to be accessed in the shortest possible time. All users must be able to access all resources regardless of their capabilities. | 7 (100) | 5.5 (92) | 5 (100) | 8.3 | 8.6 | 8.4 |
| The app can be consulted in more than one language. All languages adapt appropriately to the content interface. | 7 (100) | 5.5 (92) | 5 (100) | 8.3 | 7.4 | 7.2 |
| **Privacy** | | | | | | |
| The app gives information about the terms and conditions of purchases in the app and personal data recorded. | 7 (100) | 6 (100) | 5 (100) | 9.5 | 8.5 | 9 |
| It gives information about the kind of user data to be collected and the reason (the app must only ask for user data that is essential for the app to operate). It gives information about access policies and data treatment and ensures the right of access to recorded information. It describes the maintenance policy and the data erasure procedure. It gives information about possible commercial agreements with third parties. | 7 (100) | 5.5 (92) | 5 (100) | 9.6 | 8.8 | 9.4 |
| It guarantees the privacy of the information recorded. It requires users to give their express consent. It warns of the risks of using the app. | 7 (100) | 6 (100) | 5 (100) | 9.9 | 9.2 | 9 |
| It tells users when it accesses other resources of the device, such as their accounts or their social network profile. | 7 (100) | 6 (100) | 5 (100) | 9.3 | 8.3 | 9.4 |
| It takes measures to protect minors in accordance with the current legislation. | 7 (100) | 5.5 (92) | 5 (100) | 8.7 | 9.7 | 8.2 |
| Confidential user data are protected and anonymized, and there is a privacy mechanism so that users can control their data. | 7 (100) | 6 (100) | 5 (100) | 9.6 | 9.2 | 9.4 |
| **Security** | | | | | | |
| The app has encryption mechanisms for storing, collecting, and exchanging information. It has password management mechanisms. | 7 (100) | 6 (100) | 5 (100) | 9.9 | 8.3 | 8.6 |
| The cloud services used have the relevant security measures. It states the terms and conditions of cloud services. | 7 (100) | 6 (100) | 5 (100) | 9.5 | 8.2 | 8.4 |
| The authorization and authentication mechanisms protect the users' credentials and gives access to their data. It limits access to data that is only necessary for the user. | 7 (100) | 6 (100) | 5 (100) | 9.6 | 7.5 | 9.6 |
| It detects and identifies cybersecurity vulnerabilities, possible threats, and the risk of being exploited. It applies the appropriate security measures to cybersecurity vulnerabilities in the face of possible threats. | 7 (100) | 5.5 (92) | 5 (100) | 9.3 | 8.3 | 8.4 |
| **Appropriateness and suitability** | | | | | | |
| The end users for whom the app is designed are explicitly indicated or actually intuitable (the name identifies the app) to the audience to whom it is set out. | 7 (100) | 6 (100) | 5 (100) | 8.2 | 7.7 | 8.4 |

| Category and criterion | Comprehension, n (%) | | | Perceived importance (0-10) | | |
|---|---|---|---|---|---|---|
| | Patients (N=7) | Clinicians (N=6) | Engineers (N=5) | Patients | Clinicians | Engineers |
| The benefits and advantages of using the app are explained. | 7 (100) | 6 (100) | 5 (100) | 9 | 7 | 7.5 |
| The app has been validated or created by experts (eg, a group of specialized professionals, a health organization, or a scientific society). | 6.5 (93) | 5.5 (92) | 5 (100) | 8 | 9.7 | 9 |
| **Transparency and content** | | | | | | |
| The app identifies the authors of the content and their professional qualifications. | 7 (100) | 6 (100) | 5 (100) | 8 | 8.5 | 8.4 |
| It gives transparent information about the owners' identity and location. | 7 (100) | 6 (100) | 5 (100) | 7.2 | 8.2 | 8.4 |
| It gives information about its sources of funding, promotion and sponsorship, and possible conflicts of interests. Any third parties or organizations who have contributed to the app development are clearly identified. | 7 (100) | 6 (100) | 5 (100) | 7.5 | 7 | 7 |
| It uses scientific evidence to guarantee the quality of the content. It is based on ethical principles and values. | 7 (100) | 6 (100) | 5 (100) | 10 | 9.3 | 9 |
| The sources of the information are indicated. Concise information is given about the procedure used to select the content. | 7 (100) | 6 (100) | 5 (100) | 8.2 | 7 | 8 |
| **Safety** | | | | | | |
| The possible risks to users are identified. Users are warned that the app does not intend to replace the services provided by a professional. | 7 (100) | 6 (100) | 5 (100) | 9.5 | 8.9 | 8.6 |
| Potential risks for users caused by bad usage or possible adverse effects are explained. | 7 (100) | 6 (100) | 5 (100) | 8.5 | 8.5 | 8.6 |
| **Technical support and updates** | | | | | | |
| It gives a warning if updates modify or affect how the app functions. It gives a warning if updates can influence insensitive data. | 7 (100) | 6 (100) | 5 (100) | 8.5 | 7.2 | 7 |
| Frequent security updates are guaranteed. Every time an update of a third-party component is published, the change is inspected, and the risk evaluated. | 7 (100) | 5.5 (92) | 4.5 (90) | 8.2 | 8.2 | 7 |
| The frequency with which the content of the app is revised or updated is shown. | 7 (100) | 6 (100) | 4.5 (90) | 7.7 | 7 | 7 |
| Users have support mechanisms (email, phone, and contact form) for solving doubts, problems, or issues related to the health content, and technical support. | 7 (100) | 6 (100) | 5 (100) | 9.2 | 9 | 8.4 |
| **Technology** | | | | | | |
| It works correctly. It does not fail during use (eg, blocks). Functions are correctly retrieved after context changes (eg, switch to another app and return), external interruptions (eg, incoming calls or messages), and switching off the terminal. | 7 (100) | 6 (100) | 5 (100) | 9.5 | 8.5 | 9 |
| It does not waste resources excessively: battery, central processing unit, memory, data, or network. | 7 (100) | 6 (100) | 5 (100) | 8.9 | 7.8 | 8.2 |
| It can work in flight mode and deal with network delays and any loss of connection. | 7 (100) | 5.5 (92) | 5 (100) | 7.9 | 7.2 | 7 |
| It supports multiple versions of data structures or formats (eg, to support different operating systems). | 6.5 (93) | 5.5 (92) | 4.5 (90) | 8 | 7 | 7 |

## Discussion

### Principal Findings

To the best of our knowledge, this study is the first one to provide a guide to help with the design, development, and analysis of mHealth-related apps, in the form of a list of criteria and categories. This guide is based on an in-depth analysis of criteria that have been described in published studies on pain-related mHealth apps, guidelines, and best practices, as designated on the websites of professional and regulatory organizations from the most significant regions and countries of the world mHealth market.

In this study, we identified 36 criteria that are important to the design, development, and analysis of mHealth-related apps,

which were grouped and subsumed into eight categories according to their similarity: (1) *usability* (ie, the app must be adapted to the targeted population), (2) *privacy* (ie, compliance with the law and treatment of users' data), (3) *security* (ie, data protection, authorization mechanisms, and detection of vulnerability), (4) *appropriateness and suitability* (ie, the benefits and advantages for the end users are explained), (5) *transparency and content* (ie, scientific evidence and sources information), (6) *safety* (ie, the potentiality of risk to end users), (7) *technical support and updates* (ie, there is a policy about the maintenance of the app after it has been launched), and (8) *technology* (ie, the app works smoothly and does not fail abruptly).

In addition, this set of criteria underwent a test, and the preliminary data have shown that the criteria are understood by potential users. Furthermore, they have been reported to be of high importance by the group of stakeholders. Of particular importance (ie, a criterion that was valued as 9 or higher by all stakeholders groups on a 0-10 numerical rating scale) were the following: (1) *It is easy to use* (ie, navigation is intuitive); (2) *It guarantees the privacy of the information recorded. It requires users to give their express consent. It warns of the risks of using the app*; (3) *Confidential user data is protected and anonymized, and there is a privacy mechanism so that users can control their data*; and (4) *It uses scientific evidence to guarantee the quality of the contents. It is based on ethical principles and values*.

Our work improves previous proposals as it brings together information from a variety of internationally relevant sources (ie, research studies, data from websites of professional organizations, and standards governing the development of software for health or medical devices), whereas available ones have been developed narrowly, mostly using just one source (eg, studies on mobile apps [17]), sometimes using data of unknown scientific value (ie, mobile apps available on Web-based stores that have not undergone usability or validity studies [19]). This might be responsible, at least in part, for missing information in available guides. For example, in the case of the MARS [17], which is one of the most used rating systems, authors have failed to include some very basic items on their scale. Of particular concern are the issues of privacy and security of users' information, which are not on the scale. The protection of users' information is mandatory by law, so it is fundamental for all scales to include this as part of an integral evaluation of a mobile app. Likewise, the scale attaches little importance to whether an app is evidence-based or trialed in well-controlled studies. For example, a recent study that used MARS [24] to assess the quality of pain-related mobile apps showed that of the 18 apps, the 2 that had been scientifically tested were given the worst scores on the scale, and 1 of these had already been awarded a seal of quality from a public agency. It does seem that with MARS, the so-called commercial apps are better rated than those that have been scientifically tested and shown to provide valid and reliable information. This goes against the current trend in the area, which is seeking apps that have been scientifically tested and designed on the basis of evidence [25-27]. Furthermore, Salazar et al [24] showed that when MARS is used, an app developed with a highly specific objective in mind (eg, to measure pain intensity) will show lower scores (and will, therefore, be assumed to provide worse measurements) simply because of its specificity. Finally, the questions on the rating scale developed by Stoyanov et al [17] were mostly written to be answered by end users and require responses that are highly subjective or cannot be answered by a person who is not an expert in the field (eg, "Is app content correct, well written, and relevant to the goal or topic of the app?").

In addition, the preliminary data on the comprehension of the criteria showed that they can be understood by different profiles of stakeholders, as intended. However, a few of them reported having problems with some criteria, which were solved after giving additional explanations. Therefore, it is important that the information is presented with the least technical wording possible to facilitate comprehension. Nevertheless, additional studies with more participants to validate and extend the findings are warranted.

The resulting guide with this set of criteria describes the standard to follow, identifies the main categories of criteria, and provides stakeholders with a systematic approach by which they can determine the general requirements of a mobile app if it is to be considered of high quality. An app that meets these criteria is one that will provide users with the greatest security and confidence in performance and the objectives being fulfilled.

### Limitations

This study has limitations that should be considered when interpreting the results. First, our search strategy was limited to papers written in English or Spanish, pain-related apps, and guidelines and standards published in specific regions and countries. We made these choices because it was what we could feasibly do, but we cannot be certain that we have included *all* the important criteria. For example, some issues could be seen as more important by developers of pain-related apps compared with developers of apps related to sexual health (eg, pain-related apps are biased toward treatment rather than diagnosis; pain-related apps may primarily be targeted at the patient, rather than health professionals or carers). We analyzed the studies on pain-related apps, and the information was combined with that from the most important markets for mHealth apps and on guidelines and standards available, as a way to complement each other and solve the potential limitations. Nevertheless, the final result of our analysis is limited in ways that we cannot completely foresee. Therefore, future studies on the validity and reliability of this set of criteria are warranted. Second, the comprehension test was conducted with a small group of 18 individuals from three groups of stakeholders. Although the number of participants was enough for a preliminary analysis, the sample is not representative. Thus, additional studies, including samples with more participants, are needed. Despite these limitations, this study provides important new information to help advance the field.

### Conclusions

This set of criteria can be readily used by health care providers, engineers and developers, researchers, patients, and regulators. The data have shown them to be comprehensible and of importance for a group of stakeholders. Nevertheless, future

studies will have to empirically test the validity, reliability, and suitability of this set of criteria. Furthermore, they should be analyzed in terms of their significance to all stakeholders so that the set of criteria could also be used as a guide to the quality of the apps by all interested parties.

## Acknowledgments

This work was partly supported by grants from the Spanish Ministry of Economy and Competitiveness (RTI2018-09870-B-I00; RED2018-102546-T), the European Regional Development Fund, and Obra Social de Caixabank. JM's work is supported by Fundación Grünenthal. PL benefitted from a predoctoral fellowship (2019 FI_B2 00151) cofinanced by the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya, the EU, and the European Social Fund.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Full list of criteria.

[DOC File , 438 KB-Multimedia Appendix 1]

## References

1. OECD. Health at a Glance 2017 OECD Indicators. Paris, France: OECD; 2017.
2. World Health Organization. mHealth: New Horizons for Health through Mobile Technologies. Geneve: World Health Organization; 2011.
3. Wang Y, Sun L, Hou J. Hierarchical medical system based on big data and mobile internet: a new strategic choice in health care. JMIR Med Inform 2017 Aug 8;5(3):e22 [FREE Full text] [doi: 10.2196/medinform.6799] [Medline: 28790024]
4. Hanseth O, Bygstad B. Flexible generification: ICT standardization strategies and service innovation in health care. Eur J Inf Syst 2015;24(6):645-663. [doi: 10.1057/ejis.2015.1]
5. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. Pharm Ther 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
6. Ernsting C, Dombrowski SU, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 5;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
7. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
8. Germanakos P, Mourlas C, Samaras G. A Mobile Agent Approach for Ubiquitous and Personalized eHealth Information Systems. 2005 Presented at: Proc Work ?Personalization e-Health? 10th Int Conf User Model ;67?70; 2005; Berlin.
9. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 7;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
10. Miró J. Dolor Crónico. Procedimientos de evaluación e intervención psicológica. Bilbao: Desclée de Brower; 2003.
11. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
12. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
13. Tic Salut Social Foundation. 2015. The Mobility Master Plan: mHealth solutions URL: https://ticsalutsocial.cat/en/actualitat/the-mobility-master-plan-mhealth-solutions/ [accessed 2018-11-12] [WebCite Cache ID 73sMqAseU]
14. Tic Salut Social Foundation. 2015. AppSalut URL: https://ticsalutsocial.cat/en/projects/mhealth-en/appsalut-marketplace/ [accessed 2018-11-12] [WebCite Cache ID 73sLiv7ZO]
15. Tic Salut Social Foundation. 2018. Criteria Guide of Accreditation Process for Health and Wellness Apps URL: https://ticsalutsocial.cat/en/resources/mhealth-en/apps-accreditation-guidebooks/ [accessed 2018-11-12] [WebCite Cache ID 73sMDaN6Y]
16. European Commission. 2014 Apr 10. Green Paper on mobile Health ('mHealth') URL: https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth [accessed 2020-01-17] [WebCite Cache ID 73sKpDXVn]
17. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

18. Brown W, Yen PY, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
19. Gaziel-Yablowitz M, Schwartz DG. A review and assessment framework for mobile-based emergency intervention apps. ACM Comput Surv 2018;51(1):1-32. [doi: 10.1145/3145846]
20. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
21. Wikipedia. Mobile app URL: https://en.wikipedia.org/wiki/Mobile_app [accessed 2018-11-12] [WebCite Cache ID 73sLCDxl2]
22. Statista. Global Mobile Health Market in 2017, by Region URL: https://www.statista.com/statistics/218852/global-mobile-health-market-share-forecast-by-region/ [accessed 2018-11-12] [WebCite Cache ID 73sLFZA7t]
23. Research2Guidance. mHealth Economics 2017 – Current Status and Future Trends in Mobile Health URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2020-01-17] [WebCite Cache ID 73sK8Aedl]
24. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
25. Buijink AW, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
26. Larson RS. A path to better-quality mHealth apps. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e10414 [FREE Full text] [doi: 10.2196/10414] [Medline: 30061091]
27. Jonassaint CR, Rao N, Sciuto A, Switzer GE, de Castro L, Kato GJ, et al. Abstract animations for the communication and assessment of pain in adults: cross-sectional feasibility study. J Med Internet Res 2018 Aug 3;20(8):e10056 [FREE Full text] [doi: 10.2196/10056] [Medline: 30076127]

**Abbreviations**

**EU:** European Union
**MARS:** Mobile App Rating Scale
**mHealth:** mobile health

*Edited by G Eysenbach; submitted 07.12.18; peer-reviewed by H de Sola, N West, HT Yang, A Salazar, D Schwartz, S Wali, S Zheng; comments to author 14.04.19; revised version received 06.06.19; accepted 23.07.19; published 03.03.20*

*Please cite as:*
*Llorens-Vernet P, Miró J*
*Standards for Mobile Health–Related Apps: Systematic Review and Development of a Guide*
*JMIR Mhealth Uhealth 2020;8(3):e13057*
*URL: https://mhealth.jmir.org/2020/3/e13057*
*doi: 10.2196/13057*
*PMID:*

## 4.2. Estudi 2

The Mobile App Development and Assessment Guide (MAG): Delphi-Based Validity Study.

Original Paper

# The Mobile App Development and Assessment Guide (MAG): Delphi-Based Validity Study

Pere Llorens-Vernet*, MSc; Jordi Miró*, PhD

Unit for the Study and Treatment of Pain-ALGOS, Research Center for Behavior Assessment, Universitat Rovira i Virgili, Tarragona, Spain
*all authors contributed equally

**Corresponding Author:**
Jordi Miró, PhD
Unit for the Study and Treatment of Pain-ALGOS
Research Center for Behavior Assessment
Universitat Rovira i Virgili
Carretera de Valls, s/n
Tarragona, 43007
Spain
Phone: 34 +34649298921
Email: jordi.miro@urv.cat

## Abstract

**Background:** In recent years, there has been an exponential growth of mobile health (mHealth)–related apps. This has occurred in a somewhat unsupervised manner. Therefore, having a set of criteria that could be used by all stakeholders to guide the development process and the assessment of the quality of the apps is of most importance.

**Objective:** The aim of this paper is to study the validity of the Mobile App Development and Assessment Guide (MAG), a guide recently created to help stakeholders develop and assess mobile health apps.

**Methods:** To conduct a validation process of the MAG, we used the Delphi method to reach a consensus among participating stakeholders. We identified 158 potential participants: 45 patients as potential end users, 41 health care professionals, and 72 developers. We sent participants an online survey and asked them to rate how important they considered each item in the guide to be on a scale from 0 to 10. Two rounds were enough to reach consensus.

**Results:** In the first round, almost one-third (n=42) of those invited participated, and half of those (n=24) also participated in the second round. Most items in the guide were found to be important to a quality mHealth-related app; a total of 48 criteria were established as important. "Privacy," "security," and "usability" were the categories that included most of the important criteria.

**Conclusions:** The data supports the validity of the MAG. In addition, the findings identified the criteria that stakeholders consider to be most important. The MAG will help advance the field by providing developers, health care professionals, and end users with a valid guide so that they can develop and identify mHealth-related apps that are of quality.

*(JMIR Mhealth Uhealth 2020;8(7):e17760)* doi: 10.2196/17760

**KEYWORDS**
assessment; Delphi method; MAG; mobile apps; mobile health; validity; guide

## Introduction

Mobile apps are increasingly being used for health care [1-3]. The implementation of mobile devices such as phones, patient monitoring devices or personal digital assistants, and wireless devices has proven that they can be used for improving communication between patients and health professionals [4], and improving adherence to treatment [5]. Importantly, recent reports have suggested that smartphones have become the most popular technology among physicians [6,7]. In addition, there has been a sharp increase in the use of these technologies by the general population. For example, official estimates indicate that in 2019 a total of 65% of people had a smartphone, and by 2025, this figure will have increased to 80% [8].

However, this increase in use has occurred in a somewhat unsupervised manner; that is to say, it has not been regulated or supervised in any way. In addition, a large number of mobile health (mHealth) apps have been developed without any rigorous scientific basis [9,10] or having undergone any validation process, thus undermining the confidence of both patients and

health care professionals [11]. Moreover, information privacy practices are not transparent to users and, in many cases, are absent, opaque, or irrelevant [12]. Finally, there is mounting evidence to show that this lack of control and development without guidance is placing consumers at risk [13].

In an attempt to solve this problem, and guarantee the quality of existing and future health apps, various government-related initiatives have been taken at the regional level (eg, the proposal "AppSalut" [14,15] in Catalonia and the "AppSaludable Quality Seal" [16] in Andalusia, Spain), the national level (eg, "Good practice guidelines on health apps and smart devices [mobile health or mhealth]" [17] in France; "Health apps & co: safe digital care products with clearer regulations" [18] in Germany; "Medical devices: software applications [apps]" [19] in the United Kingdom; "Policy for Device Software Functions and Mobile Medical Applications" [20] in the United States; "Regulation of Software as a Medical Device" [21,22] in Australia), and the international level, such as the "Green Paper on mobile health" by the European Commission [23]. In general, these initiatives provide recommendations and regulations to establish how health apps should be and guarantee their quality. However, they show important differences on the key criteria. For example, "Appsalut" emphasizes usability issues [14,15], while "Regulation of Software as a Medical Device" emphasizes safety [21,22] as it equates health apps with medical devices. Clinicians and researchers have also attempted to provide specific solutions to this major problem [24]. For example, Stoyanov and colleagues [25] developed a scale to classify and rate the quality of mHealth apps. There have also been other attempts to provide alternatives for assessing mHealth apps (eg, [26,27]), each one of which suggests its own quality criteria. All these attempts have positive and negative characteristics. A major limitation common to many of these initiatives is that they have been created from one narrow perspective and focusing on, for example, a specific health problem or intervention such as emergency interventions [27] or a stakeholder such as adolescents [26]. In addition, some of them have been created from a specific perspective, for example, taking into account usability issues rather than safety. Thus, there is no common set of criteria that can be used by all stakeholders to guide the development process and the assessment of the apps' quality.

Recently, to help overcome these limitations, we conducted a study to develop such a guide: the Mobile App Development and Assessment Guide (MAG) [28]. We studied the guidelines, frameworks, and standards available in the field of health app development, with a particular focus on the world regions where the mHealth market was most significant, and pinpointed the criteria that could be recommended as a general standard. We suggested a guide containing 36 criteria that could be used by stakeholders. Our preliminary study showed that stakeholders found them to be important. They also found them easy to understand and use [28].

However, that study had some limitations. Most importantly, although the criteria identified underwent a preliminary analysis of comprehensibility and importance by a selected group of stakeholders (ie, health care experts, engineers, and potential patients), they did not undergo a validation process. Therefore, to address this issue, here we use the Delphi method [29,30] to analyze the validity of the MAG. By using this method, we also want to explore whether new criteria could be included to improve the guide. We also want to examine the importance of these criteria as perceived by the stakeholders.

## Methods

### Procedure

The Delphi method was created for people to reach consensus by answering questions in an iterative process [29]. Although the traditional Delphi process has an open initial phase [29], in this study we use a modified Delphi process, which provides a common starting point for discussion. This modified Delphi method is widely used, as it saves time and does not interfere with the original tenets of the method that participants can give suggestions and inputs at any stage [31]. It has been shown that results from Delphi-based studies offer more accurate answers to difficult questions than other prognostication techniques [32]. This modified Delphi method and the judgment of people are acknowledged as legitimate and valid inputs to generate forecasts, and have been used in many different areas to reach consensus on such strategic issues as the identification of health care quality indicators [33]; predictors of chronic pain and disability in children [34]; predictors of chronic pain in adults with cancer [35]; the needs of adolescents and young adult cancer survivors [36]; and, in the mHealth field, to develop an assessment framework for electronic mental health apps [37].

### Participants

Our goal was to recruit 30 stakeholders, as this number has been shown to be sufficient for this kind of study [38,39], from any of the following groups: (1) health care professionals, (2) developers of health-related apps, and (3) users of health apps.

To identify potential participants and ensure an appropriate panel of stakeholders, we adopted five strategies: (1) we searched for national (Spain) and international organizations or associations of digital health professionals to make contact with health professionals knowledgeable about the topic; (2) we searched for medical health apps in the app stores of the main smartphone systems (Android and iOS), identified the most downloaded and best rated apps, and searched for their developers to ensure the participation of experienced individuals; (3) we searched for national (Spain) and international organizations to recruit patients with experience in the use of health-related apps or with an interest in this area; (4) we made a general call through the social networks of our research group to increase the likelihood of recruiting participants who satisfied the inclusion requirements; and (5) we asked researchers and clinicians who we personally knew were experts in the field to participate and help us identify other potential participants.

We identified 158 potential participants from Europe, Asia, Australia, and North and South America. They were multidisciplinary and included health care professionals, patients as potential end users, and developers.

### Survey Content and Procedure

We developed a list of items on the basis of the criteria in the MAG [28]. Some of the criteria were broad and encompassed several issues and characteristics, so we broke them down into specific items to facilitate the comprehensibility and accuracy of responses. For example, the original criterion "The app can be consulted in more than one language. All languages adapt appropriately to the content interface" was divided into two items: "The app can be consulted in more than one language" and "All languages adapt appropriately to the content interface."

When the set of items was ready, it was moved to an online survey so that it could be distributed to participants more easily. Potential participants received an email with explanations about the study and a link to the survey. All the information was available in Spanish and English.

The survey included some sociodemographic questions and 56 items for rating, which were grouped in the same categories as the original guide, such as usability [28]. On a numerical rating scale from 0 (not important at all) to 10 (extremely important), participants had to report how important they considered each one to be for the quality of a health-related mobile app. Participants were also given instructions to include any other item they felt was important and missing from the original list. Like the original items, these new items were also rated. Participants were informed that only the criteria that received a score of 9 or higher from at least 75% of the participants would be included in the final list of criteria that a health-related app should meet. The rest were discarded.

Study data were collected and managed using LimeSurvey tools (LimeSurvey GmbH) [40]. We computed means and standard deviations of the demographics to describe the sample of participants. We used paired *t* tests (two-tailed) to study potential differences in the variance of the items between rounds and of the potential changes in the age or sex of participants. To study the consensus, mean, standard deviation, 95% confidence interval (with the lower and upper values for each item), and significance level ($P<.05$) were computed. All data analyses were performed using SPSS v.25 for Windows (IBM Corp).

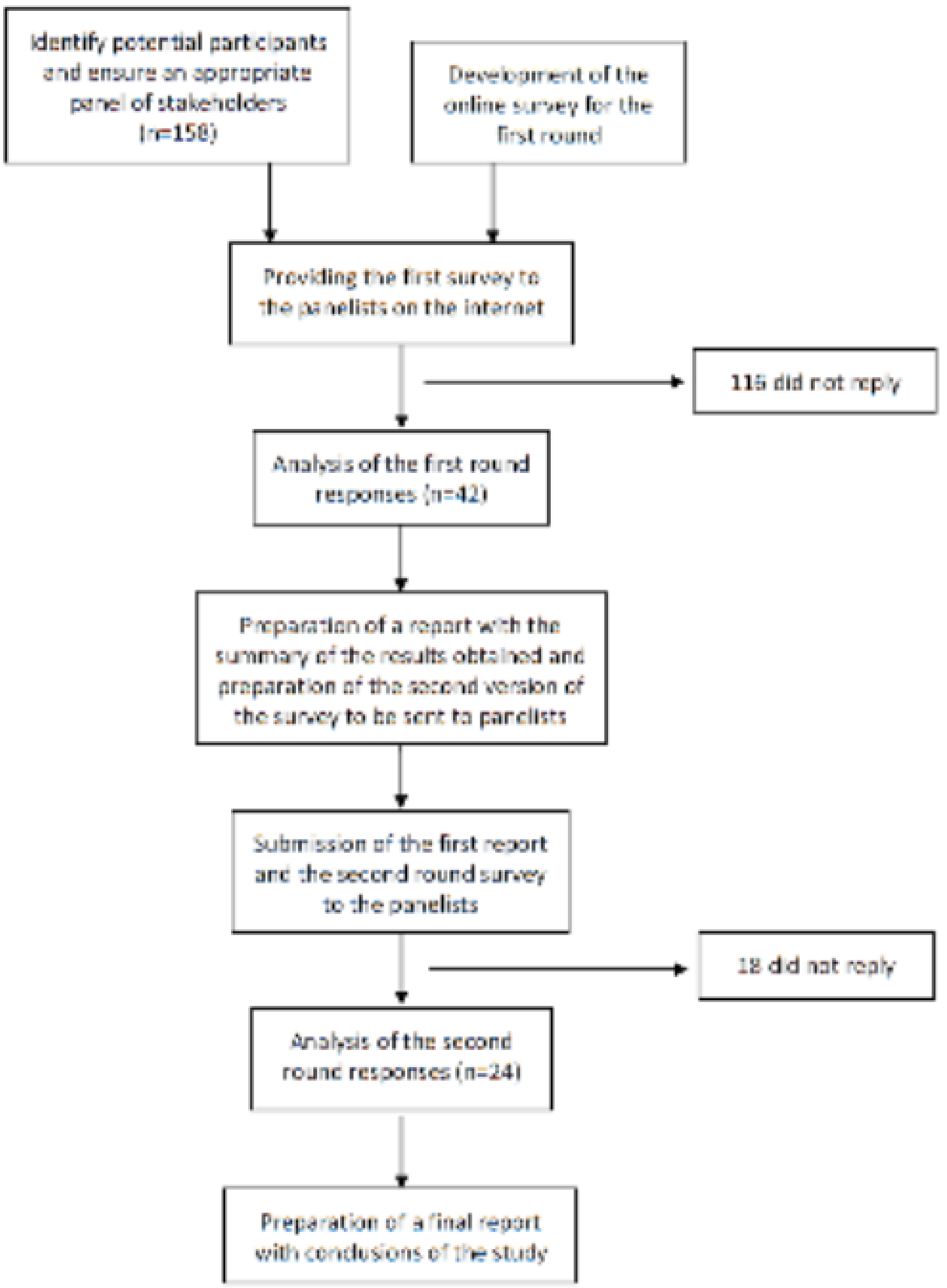
### Delphi Rounds

In the first round, the survey was sent to 158 potential participants: 45 patients as potential end users, 41 health care professionals, and 72 developers. They were informed about the study and the requirements to participate. Participants were given 3 weeks to respond, during which time a reminder was sent each week to maximize the involvement of as many participants as possible. The survey took approximately 15 minutes. The answers were analyzed, and some new items were added in response to the suggestions of the participants.

In the second round, the results of the first iteration were sent by email to all the participants who had provided responses to the initial survey so that they could see their position and the position of the group as a whole on the items, as well as the level of agreement among the participants. This information was given so that participants in the group could re-examine their initial responses, in light of the group's opinion. Participants in this second round were asked to respond to the revised survey. Again, they were given 3 weeks to answer. The Delphi methodology requires that this procedure be repeated until participants' responses reach stability or when a point of diminishing returns is reached [39].

The stability of responses was the criterion used to identify that consensus had been reached on any given question [30,38]. In this study, stability was reached after two rounds (see the Results section), which is consistent with the findings of previous Delphi studies (eg, [35,41]). We considered that consensus was reached on a particular criterion when 75% of the participants rated it with at least a 9 [34]. If a criterion was rated with a 9 or more by at least 90% of the participants, we considered it to be of key importance for an mHealth-related app. The results only showed statistically significant differences for two items (see the Results section). Thus, given the stability of the responses, we decided to stop the iteration process after round 2. Figure 1 describes the steps of the study.

**Figure 1.** Steps in the Delphi poll.




## Results

### Round 1

Of the 158 stakeholders invited, 42 (27%) responded to the first round. The demographic characteristics of the participants in each round are shown in Table 1. There were no statistically significant differences in terms of sex or age between those who were invited and those who participated. Only a small increase in the mean age of participants and female participation was detected between rounds (see Table 1).

Multimedia Appendix 1 summarizes all the information about participants' responses to the initial 56 items.

To determine consensus on the items, we examined the percentage of participants who agreed on their importance. Items with an agreement ≥75% of participants were considered to have reached consensus. We also used confidence intervals,

instead of discrete estimation, because they have less error (see [34] for a similar procedure). Out of the total 56 items, participants reached consensus on the importance of 32 (57%) of the items (ie, at least 75% of the participating stakeholders rated their importance with a 9 or higher on the 0-10 scale).

In this first round, participants added 36 new items to the original list. As previously described, in response to participants' suggestions, changes were made to items 3, 51, 68, and 73. In addition, items 33 and 69 were divided in two, as several participants considered that they included two different clauses (see Multimedia Appendix 1).

**Table 1.** Sample characteristics in each round.

| Characteristics | Invited participants (N=158) | Participants round 1 (n=42) | Participants round 2 (n=24) |
|---|---|---|---|
| **Sex, n (%)**[a] | | | |
| Male | 17 (42.5) | 15 (44.12) | 9 (37.5) |
| Female | 23 (57.5) | 19 (55.88) | 15 (62.5) |
| **Stakeholder group, n (%)** | | | |
| Health care professionals | 45 (28.48) | 16 (38.1) | 9 (37.5) |
| Developers | 41 (25.95) | 14 (33.33) | 8 (33.33) |
| Final users | 72 (45.57) | 12 (28.57) | 7 (29.17) |
| Age (years), mean[a] | 39.82 | 41.39 | 43.08 |
| **Citizenship, n (%)**[a] | | | |
| Spain | 78 (49.37) | 28 (82.35) | 20 (83.33) |
| United States | 13 (8.23) | 1 (2.94) | 1 (4.17) |
| Argentina | 1 (0.63) | 1 (2.94) | 1 (4.17) |
| Italy | 1 (0.63) | 1 (2.94) | 1 (4.17) |
| Australia | 6 (3.8) | 1 (2.94) | 1 (4.17) |
| United Kingdom | 15 (9.49) | 2 (5.88) | 0 (0.00) |
| Canada | 8 (5.06) | 0 (0.00) | 0 (0.00) |
| Brazil | 3 (1.9) | 0 (0.00) | 0 (0.00) |
| India | 6 (3.8) | 0 (0.00) | 0 (0.00) |
| Indonesia | 1 (0.63) | 0 (0.00) | 0 (0.00) |
| Germany | 5 (3.16) | 0 (0.00) | 0 (0.00) |
| Poland | 2 (1.27) | 0 (0.00) | 0 (0.00) |
| Ireland | 2 (1.27) | 0 (0.00) | 0 (0.00) |
| Turkey | 1 (0.63) | 0 (0.00) | 0 (0.00) |
| Kazakhstan | 1 (0.63) | 0 (0.00) | 0 (0.00) |
| Switzerland | 3 (1.9) | 0 (0.00) | 0 (0.00) |
| Colombia | 1 (0.63) | 0 (0.00) | 0 (0.00) |
| Romania | 1 (0.63) | 0 (0.00) | 0 (0.00) |
| Hungary | 1 (0.63) | 0 (0.00) | 0 (0.00) |
| Lithuania | 1 (0.63) | 0 (0.00) | 0 (0.00) |
| Bulgaria | 2 (1.27) | 0 (0.00) | 0 (0.00) |
| Greece | 1 (0.63) | 0 (0.00) | 0 (0.00) |
| Sweden | 2 (1.27) | 0 (0.00) | 0 (0.00) |
| Israel | 1 (0.63) | 0 (0.00) | 0 (0.00) |
| Belgium | 2 (1.27) | 0 (0.00) | 0 (0.00) |

[a]Information was not available for all participants.

### Round 2

Of the 42 that participated in the first round, 24 (57%) of the stakeholders participated in the second round. Out of the total 92 items, a total of 48 items (52%) were rated with a 9 or more by at least 75% of the participants (see Table 2 below).

The consensus on the importance of the 32 items in round 1 was maintained in round 2, except for item 69, which fell below the criteria of 75% agreement. On the other hand, items 8, 32, 46, and 47 did not reach consensus in round 1 but did in round 2. Consensus was also reached on the importance of 14 of the new items suggested by participants. Of all the items, 9 were particularly important (ie, at least 90% of participants rated their importance with a 9 or higher).

By categories, the number of items for which consensus was reached were *usability*: 8 of 18 items (44% of the total in the category); *privacy*: 14 of 19 items (74%); *security*: 9 of 13 items (69%); *appropriateness and suitability*: 2 of 5 items (40%); *transparency and content*: 2 of 11 items (18%); *safety*: 7 of 8 items (88%), *technical support and updates*: 2 of 9 items (22%); and *technology*: 4 of 9 items (44%).

**Table 2.** Items that reached consensus about their importance.

| Category and item | Round 1 (n=42) | | | Round 2 (n=24) | | |
|---|---|---|---|---|---|---|
| | Consensus, n (%) | Mean (SD) | 95% CI | Consensus, n (%) | Mean (SD) | 95% CI |
| **Usability** | | | | | | |
| 1. The app has been tested by potential users before being made available to the public. | 33 (78.57) | 9.14 (1.92) | 8.56-9.72 | 20 (83.33) | 9.21 (1.61) | 8.56-9.85 |
| 2. It is easy to use (ie, navigation is intuitive). | 39 (92.86) | 9.67 (0.61) | 9.48-9.85 | 21 (87.50) | 9.50 (0.72) | 9.21-9.79 |
| 3. Functionality is adapted to the purpose of the app. | 40 (95.24) | 9.74 (0.63) | 9.55-9.93 | 21 (87.50) | 9.42 (0.93) | 9.05-9.79 |
| 4. Functionality is adjusted according to the profile and needs of the targeted user. | —[a] | — | — | 19 (79.17) | 9.08 (0.93) | 8.71-9.45 |
| 5. Access is adapted for people with disabilities. | — | — | — | 19 (79.17) | 9.00 (1.41) | 8.43-9.57 |
| 6. It complies with regulatory accessibility standards. | — | — | — | 18 (75.00) | 9.17 (1.13) | 8.71-9.62 |
| 7. The language used makes it accessible to any user. | — | — | — | 19 (79.17) | 9.04 (1.55) | 8.42-9.66 |
| 8. All users have access to all resources regardless of their capabilities. | 28 (66.67) | 8.60 (1.67) | 8.09-9.10 | 18 (75.00) | 9.08 (1.06) | 8.66-9.51 |
| **Privacy** | | | | | | |
| 9. The app gives information about the terms and conditions of purchases in the app. | 35 (83.33) | 9.29 (1.90) | 8.71-9.86 | 21 (87.50) | 9.50 (1.18) | 9.03-9.97 |
| 10. It must only ask for user data that is essential for the app to operate. | 34 (80.95) | 9.26 (1.53) | 8.80-9.72 | 18 (75.00) | 8.92 (2.04) | 8.10-9.73 |
| 11. It gives information about access policies and data processing, and ensures the right of access to recorded information. | 34 (80.95) | 9.02 (2.25) | 8.34-9.70 | 18 (75.00) | 9.38 (0.97) | 8.99-9.76 |
| 12. It gives information about possible commercial agreements with third parties. | 32 (76.19) | 8.79 (2.54) | 8.02-9.55 | 20 (83.33) | 9.17 (2.12) | 8.32-10.01 |
| 13. It clearly allows the user the option of non-transfer of data to third parties or for commercial purposes. | — | — | — | 23 (95.83) | 9.71 (0.55) | 9.49-9.93 |
| 14. It guarantees the privacy of the information recorded. | 39 (92.86) | 9.71 (0.77) | 9.48-9.95 | 20 (83.33) | 9.46 (1.10) | 9.02-9.90 |
| 15. It requires users to give their express consent. | 36 (85.71) | 9.12 (2.19) | 8.46-9.78 | 19 (79.17) | 9.38 (1.01) | 8.97-9.78 |
| 16. It warns of the risks of using the app. | 36 (85.71) | 9.33 (1.86) | 8.77-9.89 | 19 (79.17) | 9.25 (1.19) | 8.77-9.73 |
| 17. It tells users when it accesses other resources on the mobile device such as their accounts or their social network profiles. | 36 (85.71) | 9.33 (1.76) | 8.80-9.87 | 22 (91.67) | 9.71 (0.75) | 9.41-10.01 |
| 18. It takes measures to protect minors in accordance with current legislation. | 38 (90.48) | 9.43 (1.74) | 8.90-9.96 | 23 (95.83) | 9.79 (0.51) | 9.59-10.00 |
| 19. Confidential user data is protected and anonymized, and there is a privacy mechanism so that users can control their data. | 38 (90.48) | 9.60 (1.06) | 9.27-9.92 | 21 (87.50) | 9.46 (1.18) | 8.99-9.93 |
| 20. It offers to erase the data when the service is finished. | — | — | — | 19 (79.17) | 9.04 (1.68) | 8.37-9.71 |
| 21. It gives information about privacy policies in a simple and understandable way. | — | — | — | 20 (83.33) | 9.33 (1.09) | 8.90-9.77 |
| 22. It complies with all current privacy laws. | — | — | — | 22 (91.67) | 9.54 (1.28) | 9.03-10.06 |
| **Security** | | | | | | |
| 23. The app has encryption mechanisms for storing, collecting, and exchanging information. | 35 (83.33) | 9.40 (1.33) | 9.00-9.81 | 19 (79.17) | 9.13 (1.57) | 8.50-9.75 |
| 24. It has password management mechanisms. | 33 (78.57) | 9.05 (1.71) | 8.53-9.56 | 19 (79.17) | 9.04 (1.90) | 8.28-9.80 |

| Category and item | Round 1 (n=42) | | | Round 2 (n=24) | | |
|---|---|---|---|---|---|---|
| | Consensus, n (%) | Mean (SD) | 95% CI | Consensus, n (%) | Mean (SD) | 95% CI |
| 25. It states the terms and conditions of cloud services. | 32 (76.19) | 8.93 (2.23) | 8.25-9.60 | 19 (79.17) | 9.29 (1.08) | 8.86-9.72 |
| 26. The cloud services used have the relevant security measures. | 36 (85.71) | 9.40 (1.47) | 8.96-9.85 | 21 (87.50) | 9.29 (1.60) | 8.65-9.93 |
| 27. The authorization and authentication mechanisms protect users' credentials and allow access to their data. | 37 (88.10) | 9.57 (1.02) | 9.26-9.88 | 21 (87.50) | 9.42 (1.21) | 8.93-9.90 |
| 28. It limits access to data that is only necessary for the user. | 33 (78.57) | 8.98 (2.25) | 8.30-9.66 | 19 (79.17) | 8.96 (2.10) | 8.12-9.80 |
| 29. It detects and identifies cybersecurity vulnerabilities, possible threats, and the risk of being exploited. | 36 (85.71) | 9.33 (1.76) | 8.80-9.87 | 18 (75.00) | 8.96 (2.16) | 8.10-9.82 |
| 30. It applies the appropriate security measures to cybersecurity vulnerabilities in the face of possible threats to reduce the risk of being exploited. | 35 (83.33) | 9.62 (0.82) | 9.37-9.87 | 19 (79.17) | 9.38 (0.92) | 9.01-9.74 |
| 31. It informs users of the possible risks associated with the app's use of personal data. | — | — | — | 20 (83.33) | 9.25 (1.11) | 8.80-9.70 |
| **Appropriateness and suitability** | | | | | | |
| 32. The benefits and advantages of using the app are explained. | 31 (73.81) | 8.95 (1.58) | 8.48-9.43 | 18 (75.00) | 9.08 (1.53) | 8.47-9.70 |
| 33. Experts have participated in the development of the app (for example, specialized professionals, health organizations, scientific societies, or specialized external organizations). | 35 (83.33) | 9.52 (1.02) | 9.22-9.83 | 21 (87.50) | 9.58 (0.72) | 9.30-9.87 |
| **Transparency and content** | | | | | | |
| 34. It uses scientific evidence to guarantee the quality of the content. | 36 (85.71) | 9.60 (0.86) | 9.34-9.85 | 20 (83.33) | 9.46 (0.78) | 9.15-9.77 |
| 35. It is based on ethical principles and values. | 39 (92.86) | 9.71 (0.77) | 9.48-9.95 | 22 (91.67) | 9.75 (0.61) | 9.51-9.99 |
| **Safety** | | | | | | |
| 36. The possible risks to users are identified. | 36 (85.71) | 9.45 (1.15) | 9.10-9.80 | 20 (83.33) | 9.46 (0.88) | 9.10-9.81 |
| 37. It ensures that there are no adverse effects. | — | — | — | 18 (75.00) | 8.92 (2.12) | 8.07-9.77 |
| 38. It complies with regulatory standards as a medical device. | — | — | — | 22 (91.67) | 9.46 (1.67) | 8.79-10.13 |
| 39. Users are warned when adverse events are identified so they can delete the app and avoid potential risks. | — | — | — | 18 (75.00) | 8.83 (1.93) | 8.06-9.60 |
| 40. Users are warned that the app is not meant to replace the services provided by a professional. | 40 (95.24) | 9.74 (0.63) | 9.55-9.93 | 22 (91.67) | 9.67 (0.64) | 9.41-9.92 |
| 41. It recommends always consulting a specialist in case of doubt. | — | — | — | 22 (91.67) | 9.33 (1.66) | 8.67-10.00 |
| 42. Potential risks for users caused by incorrect use and possible adverse effects are explained. | 34 (80.95) | 9.48 (0.92) | 9.20-9.75 | 20 (83.33) | 9.38 (1.17) | 8.91-9.84 |
| **Technical support and updates** | | | | | | |
| 43. It gives a warning if updates can influence insensitive data (changes the use of the data or different data is collected). | 32 (76.19) | 8.90 (2.07) | 8.28-9.53 | 19 (79.17) | 9.17 (1.40) | 8.60-9.73 |
| 44. Every time an update of a third-party component is published, the change is inspected and the risk evaluated. | 33 (78.57) | 8.98 (1.81) | 8.43-9.52 | 20 (83.33) | 8.96 (1.97) | 8.17-9.75 |
| **Technology** | | | | | | |

| Category and item | Round 1 (n=42) | | | Round 2 (n=24) | | |
|---|---|---|---|---|---|---|
| | Consensus, n (%) | Mean (SD) | 95% CI | Consensus, n (%) | Mean (SD) | 95% CI |
| 45. It works correctly. It does not fail during use (blocks, etc). | 36 (85.71) | 9.36 (1.23) | 8.99-9.73 | 23 (95.83) | 9.75 (0.53) | 9.54-9.96 |
| 46. Functions are correctly retrieved after context changes (switch to another app and return, etc), external interruptions (incoming calls or messages, etc), and switching off the terminal. | 30 (71.43) | 8.93 (1.50) | 8.47-9.38 | 21 (87.50) | 9.46 (0.83) | 9.13-9.79 |
| 47. It does not waste resources excessively: battery, central processing unit, memory, data, network, etc. | 29 (69.05) | 8.88 (1.48) | 8.43-9.33 | 19 (79.17) | 9.25 (1.11) | 8.80-9.70 |
| 48. It has a data recovery system in case of loss. | — | — | — | 19 (79.17) | 8.67 (2.28) | 7.76-9.58 |

[a]These items were not available in round 1.

## Discussion

### Main Findings

The key finding from this study is that the MAG [28] is a valid tool to help guide the development of health-related mobile apps and assess their quality. The findings also indicate the items that are important to a health-related mobile app (the MAG is provided with this article; see Multimedia Appendix 2).

The data showed that 48 items on the MAG were considered to be of high importance (ie, they were rated with at least a 9 on a 0-10 numerical rating scale by at least 75% of the participants). Most of the items belonged to the categories *privacy* and *security*, thus showing that these are the issues that most concern stakeholders when assessing the quality of health mobile apps. In particular, the following items reached a consensus of 90%: *it clearly allows the user the option of nontransfer of data to third parties or for commercial purposes* (item 13); *it tells users when it accesses other resources on the mobile device, such as their accounts or their social network profiles* (item 17); *it takes measures to protect minors in accordance with current legislation* (item 18); *it complies with all current privacy laws* (item 22); *it is based on ethical principles and values* (item 35); *it complies with regulatory standards as a medical device* (item 38); *users are warned that the app is not meant to replace the services provided by a professional* (item 40); *it recommends always consulting a specialist in case of doubt* (item 41); and *it works correctly, it does not fail during use (blocks, etc*; item 45).

Our work adds to previous proposals of quality guides or checklists by studying the validity of MAG, a comprehensive guide developed by Llorens-Vernet and Miró [28]. This guide was found to be a significant improvement on existing guides, as it had been developed with a comprehensive focus from a variety of sources (ie, research studies, recommendations from professional organizations, and standards governing the development of software for health or medical devices) and an international perspective (ie, resources used came from several regions worldwide). In addition, the guide was created to be of help to all stakeholders and not limited to a specific health problem.

### Future Research

Additional research is needed to establish the applicability of the MAG as a guide for health-related mobile app development. Future studies will have to test the MAG with real apps and check their functionality and usability among the different stakeholders who are interested in using it. Furthermore, studies to determine the relative importance of the items and the reliability and suitability of the guide in assessing mobile apps are also warranted. In this regard, a user version of the MAG will be developed to study the association between the quality of the user experience and the score in MAG. In the future, it is highly likely that additional items or criteria will be required to be able to look into the new functions and actions included in mobile apps. Thus, revised and updated versions of the MAG are to be expected.

### Limitations

The results of this study should be interpreted in the light of some limitations. The first of these is the representativeness of the participants. Although participants were an international sample of stakeholders, most of them were individuals living in Spain. We do not know if the results would have been the same with other experts. Nevertheless, for the most part, the group included individuals with extensive experience (in clinical work, research, and development), which suggests that their assessments are relevant and of good quality. Second, the number of participating experts changed from round 1 to round 2. However, this is quite normal and to be expected in all Delphi polls [23,28]. Although we cannot be certain that the results would have been the same had all participants in round 1 also responded to round 2, it is quite probable, as the differences between the rounds were minimal. Finally, the number of participants in each round was appropriate for the objectives (a minimum of 7 and maximum of 30 participants is recommended for studies like this; see [39,42]).

### Conclusions

Despite the limitations, the results of this study will help advance the field by providing developers, health care professionals, and end users with a valid guide (the MAG) for developing and identifying quality mHealth-related apps. The data shows that the stakeholders reached a consensus on 48 items distributed in 8 categories to establish them as the important criteria for health apps.

The MAG provides stakeholders with a valid tool for systematically reviewing health-related mobile apps. The MAG can be readily used to develop new apps by pointing to the general requirements that a mobile app ought to have if it is to be of high quality. Furthermore, the guide can help to rate existing apps and identify those that are of most interest on the basis of quality criteria. The apps that meet most of the criteria will give users the confidence that their objectives will be fulfilled. It can be used to provide a checklist for the evaluation and development of quality health apps.

## Acknowledgments

This work was partly supported by grants from the Spanish Ministry of Economy, Industry and Competitiveness (RTI2018-09870-B-I00; RED2018-102546-T); the European Regional Development Fund; and Obra Social de Caixabank. JM's work is supported by Fundación Grünenthal. PL-V benefited from a predoctoral fellowship (2019 FI_B2 00151) cofinanced by the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya, the European Union, and the European Social Fund.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Importance of the items in each round.
[DOCX File , 53 KB-Multimedia Appendix 1]

## Multimedia Appendix 2

The Mobile App Development and Assessment Guide.
[DOCX File , 22 KB-Multimedia Appendix 2]

## References

1. Accenture. Meet today's healthcare team: patients + doctors + machines. Accenture. 2018. URL: https://www.accenture.com/_acnmedia/PDF-71/Accenture-Health-Meet-Todays-Healthcare-Team-Patients-Doctors-Machines.pdf [accessed 2020-07-14]
2. The growing value of digital health. IQVIA. 2017. URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [accessed 2020-07-14]
3. Al Bawab AQ, AlQahtani F, McElnay J. Health care apps reported in newspapers: content analysis. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e10237 [FREE Full text] [doi: 10.2196/10237] [Medline: 30348632]
4. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
5. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile apps for increasing treatment adherence: systematic review. J Med Internet Res 2019 Jun 18;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]
6. The socio-economic impact of mobile health. Group TBC. 2012. URL: https://www.telenor.com/wp-content/uploads/2012/05/BCG-Telenor-Mobile-Health-Report-May-20121.pdf [accessed 2020-07-14]
7. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
8. The Mobile Economy. GSMA. 2020. URL: https://www.gsma.com/r/mobileeconomy/ [accessed 2020-07-14]
9. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
10. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312. [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
11. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 07;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
12. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
13. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
14. AppSalut Marketplace. TIC Salut Social. 2018. URL: https://ticsalutsocial.cat/en/projectes/mhealth-en/appsalut-marketplace/ [accessed 2020-07-14]

15. López Seguí F, Pratdepàdua Bufill C, Rius Soler A, de San Pedro M, López Truño B, Aguiló Laine A, et al. Prescription and integration of accredited mobile apps in Catalan health and social care: protocol for the AppSalut site design. JMIR Res Protoc 2018 Dec 21;7(12):e11414 [FREE Full text] [doi: 10.2196/11414] [Medline: 30578234]
16. Distintivo AppSaludable. Agencia de Calidad Sanitaria de Andalucia. 2014. URL: http://www.calidadappsalud.com/listado-completo-recomendaciones-app-salud/ [accessed 2020-07-14]
17. Good practice guidelines on health apps and smart devices (mobile health or mhealth). Haute Autorité de Santé. 2016. URL: https://www.has-sante.fr/jcms/c_2681915/en/good-practice-guidelines-on-health-apps-and-smart-devices-mobile-health-or-mhealth [accessed 2020-07-14]
18. Health apps & co: safe digital care products with clearer regulations. IGES Institut GmbH. 2016. URL: https://www.iges.com/kunden/gesundheit/forschungsergebnisse/2016/digitalisierung-iii/index_ger.html [accessed 2020-07-14]
19. Medical devices: software applications (apps). Medicines and Healthcare products Regulatory Agency. 2014. URL: https://www.gov.uk/government/publications/medical-devices-software-applications-apps [accessed 2020-07-14]
20. Policy for device software functions and mobile medical applications. US Food and Drug Administration. 2019. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications [accessed 2020-07-14]
21. Regulation of software as a medical device. Therapeutic Goods Administration. 2019. URL: https://www.tga.gov.au/regulation-software-medical-device [accessed 2020-07-14]
22. Therapeutic G. Therapeutic goods legislation amendment (2019 measures no.1) regulations 2019. Therapeutic Goods Administration. 2019. URL: https://www.tga.gov.au/therapeutic-goods-legislation-amendment-2019-measures-no1-regulations-2019 [accessed 2020-07-14]
23. Green paper on mobile health ("mhealth"). European Commission. 2014. URL: https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth [accessed 2020-07-14]
24. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098. [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
25. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
26. Brown W, Yen P, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
27. Gaziel-Yablowitz M, Schwartz DG. A review and assessment framework for mobile-based emergency intervention apps. ACM Comput Surv 2018 Apr 14;51(1):1-32. [doi: 10.1145/3145846]
28. Llorens-Vernet P, Miró J. Standards for mobile health-related apps: systematic review and development of a guide. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e13057 [FREE Full text] [doi: 10.2196/13057] [Medline: 32130169]
29. Linstone H, Turoff M, Helmer O. The Delphi Method Techniques and Applications. London: Addison-Wesley; 1975.
30. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. Inf Manage 2004 Dec;42(1):15-29. [doi: 10.1016/j.im.2003.11.002]
31. Landeta J. Current validity of the Delphi method in social sciences. Technological Forecasting Soc Change 2006 Jun;73(5):467-482. [doi: 10.1016/j.techfore.2005.09.002]
32. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. J Clin Epidemiol 2003;56(12):1150-1156. [doi: 10.1016/s0895-4356(03)00211-7] [Medline: 14680664]
33. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One 2011;6(6):e20476. [doi: 10.1371/journal.pone.0020476] [Medline: 21694759]
34. Miró J, Huguet A, Nieto R. Predictive factors of chronic pediatric pain and disability: a Delphi poll. J Pain 2007 Oct;8(10):774-792. [doi: 10.1016/j.jpain.2007.04.009] [Medline: 17627893]
35. Expósito-Vizcaíno S, Burjalés-Martí D, Miró J. Factores psicosociales en el dolor crónico oncológico: un estudio Delphi. Rev Soc Esp Dolor 2018;26(1):21-30. [doi: 10.20986/resed.2018.3658/2018]
36. Galán S, de la Vega R, Tomé Pires C, Racine M, Solé E, Jensen MP, et al. What are the needs of adolescents and young adults after a cancer treatment? A Delphi study. Eur J Cancer Care (Engl) 2017 Mar;26(2). [doi: 10.1111/ecc.12488] [Medline: 27007727]
37. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-Mental health apps in Canada: results of a modified Delphi process. JMIR Mhealth Uhealth 2018 Jul 09;6(7):e10016 [FREE Full text] [doi: 10.2196/10016] [Medline: 29986846]
38. Landeta J. El método Delphi: una técnica de previsión para la incertidumbre. Barcelona: Ariel; 1999:8434428369.
39. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000 Oct;32(4):1008-1015. [Medline: 11095242]

40. LimeSurvey: an open source survey tool. LimeSurvey GmbH. Hamburg, Germany URL: https://www.limesurvey.org/ [accessed 2020-07-14]
41. Miró J, Nieto R, Huguet A. Predictive factors of chronic pain and disability in whiplash: a Delphi poll. Eur J Pain 2008 Jan;12(1):30-47. [doi: 10.1016/j.ejpain.2007.02.004] [Medline: 17400491]
42. Varela-Ruiz M, Díaz-Bravo L, García-Durán R. Descripción y usos del método Delphi en investigaciones del área de la salud. Investigación en Educación Médica 2012;1(2):90-95.

**Abbreviations**

**MAG:** Mobile App Development and Assessment Guide
**mHealth:** mobile health

*Edited by G Eysenbach; submitted 10.01.20; peer-reviewed by A García-Palacios, R Warty, T van Mierlo, M Khalemsky; comments to author 03.03.20; revised version received 28.04.20; accepted 03.06.20; published 31.07.20*

*Please cite as:*
*Llorens-Vernet P, Miró J*
*The Mobile App Development and Assessment Guide (MAG): Delphi-Based Validity Study*
*JMIR Mhealth Uhealth 2020;8(7):e17760*
*URL: http://mhealth.jmir.org/2020/7/e17760/*
*doi: 10.2196/17760*
*PMID:*

## 4.3. Estudi 3

On the Assessment of the Quality of mHealth-related Apps: an Interrater Reliability Study of Two Guides.

# On the Assessment of the Quality of mHealth-related Apps: an Interrater Reliability Study of Two Guides

Jordi Miró and Pere Llorens-Vernet

Universitat Rovira i Virgili, Unit for the Study and Treatment of Pain – ALGOS, Research Center for Behavior Assessment (CRAMC), Institut d'Investigació Sanitària Pere Virgili, Catalonia, Spain

**Correspondence:**

Jordi Miró
Departament de Psicología
Universitat Rovira i Virgili
Carretera de Valls, s/n
43007 Tarragona
Spain

jordi.miro@urv.cat

Tel: +34-977558179
Fax: +34-977558088

**Running title:** Interrater Reliability of mHealth Apps Guides

## Abstract

**Background:** There are a huge number of health-related apps available, and the numbers are growing fast. However, many of them have been developed without any kind of quality control. In an attempt to contribute to the development of high quality apps and enable existing ones to be assessed, several guides have been developed.

**Objective:** The main aim of this study was to study the interrater reliability of a new guide – the Mobile App Development and Assessment Guide (MAG) – and compare it with one of the most used guides in the field, the Mobile App Rating Scale (MARS).

**Methods:** In order to study the interrater reliability of MAG and MARS, we evaluated the four most downloaded chronic health apps for Android and IOS devices. A group of eight reviewers, which included different types of stakeholders such as clinical researchers, engineers, health-care professionals, and end-users as potential patients, independently evaluated the quality of the apps using MAG and the MARS. We used Krippendorff's alpha to calculate the interrater reliability.

**Results:** Only a few categories of MAG and MARS demonstrated a high interrater reliability. Although MAG was found to be superior, there was considerable variation in the scores between the different types of reviewer. The categories with the highest interrater reliability in MAG were “Security” (alpha = 0.78) and “Privacy” (alpha = 0.73).

**Conclusions:** This study shows that some categories of MAG have significant interrater reliability and that the MAG scores are better than the MARS scores. However, there is great variability in the responses, which seems to be associated with subjective interpretation by the reviewers.

**Keywords:** mHealth; mobile health; mobile apps; evaluation studies, rating; interrater reliability; MARS; MAG

## Introduction

In recent years, there has been an explosion of interest in the use of mobile devices (e.g., smartphones, tablets) [1], alongside huge advances in the development of health-related mobile apps [2]. For example, a total of 325,000 different

health-related apps have recently been reported to be available [3]). There are mobile apps for virtually all kinds of health conditions: for example, chronic pain [4,5], cancer [6], diabetes [7] and cardiovascular diseases [8]. This growth has brought considerable benefits not only to patients but also to society at large, and at multiple levels. For example, health-related apps help to: (1) improve treatment management, (2) facilitate patient-doctor communication, (3) monitor the patient's condition in real time and (4) improve accessibility to treatment [9–12]. But there are also a number of caveats, mostly related to the somewhat unsupervised and unregulated nature of the process. And it has been suggested that the fact that the field is evolving without much scientific support or guidance [13] not only acts as a barrier to improvement [14] but also, and more importantly, can potentially put an individual's health at risk [15]. Some of the main problems related to health apps are: (1) faulty reminders that make proper treatment follow-up difficult (e.g., the instructions on when to do an activity or take medication are not correct [16]); (2) lack of health-expert involvement [17]; (3) inappropriate response to consumer needs (e.g., bipolar disorder apps

failing to provide any response when asked about extreme mood swings or suicidal ideation [18]); and (4) incorrect medication doses (e.g., incorrect calculation of insulin dose from blood glucose values [19]).

In order to overcome the issues health-related apps are facing, some rating scales and guides have been developed (e.g., [20,21]). One of the first was the Mobile App Rating Scale (MARS; [22]). It is one of the most used rating scales to measure the quality of health-related apps [23–27]. However, the MARS was created from a narrow perspective [28–30] on the basis of analyzing studies on existing mobile applications, and leaving out information from other relevant sources (e.g., standards governing the design of software for medical devices).

Recently, the Mobile App Development and Assessment Guide (MAG;(Llorens-Vernet & Miró, 2020a)) was created to address the problems observed in the guides available (but not current key concerns such as privacy and security), and to help assess health-related apps and guide stakeholders in the development of new quality apps. MAG was developed using data from all potential relevant sources,

and a representative sample of the guidelines, frameworks and standards in the field of health apps development. MAG has been acknowledged as a good quality guide by an international and interdisciplinary group of stakeholders (Llorens-Vernet & Miró, 2020b).

These guides are important in the field as they provide quality scores that are key to identifying the best apps available and distinguishing them from the poorly designed ones. However, there is little data on the comparative value and consistency of the very few guides there are. The field would benefit considerably from studies that guide the development of new apps and comparatively assess the quality of existing ones.

The main objective of this research was to study and compare MAG and MARS. More specifically, we aimed to compare the interrater reliability of the two measures. We also focused on whether the interrater reliability of the measures is consistent across multiple types of apps and stakeholders.

## Methods

### *Apps selection process*

In order to evaluate the interrater reliability of MAG and MARS across different types of app, we evaluated the top four search results for chronic health conditions in the medical category of the Apple and Android stores (i.e., App Store and Google Play). The search and selection of the apps was conducted in October 2020.

The inclusion criteria were as follows: the app had to be focused on a chronic health condition, in English or Spanish and free to download. We selected chronic health conditions because it is one of the domains in which health apps are becoming more relevant (56% of health apps are intended for this kind of patient [32]). Reports by governmental agencies indicate that they are a major health problem that affects 31% of the population (e.g., [33–36]). In addition, chronic health conditions are the leading cause of death and disability in both the developed and developing world in the global burden of disease equation. The most important chronic health conditions are: low back pain and headache, neoplasms,

diabetes and kidney diseases, and cardiovascular diseases [37–40]. We used the following search terms, which are related to the top four chronic health conditions in the Global Burden of Disease study [41]: “pain", “cancer”, “diabetes”, and “cardiovascular”.

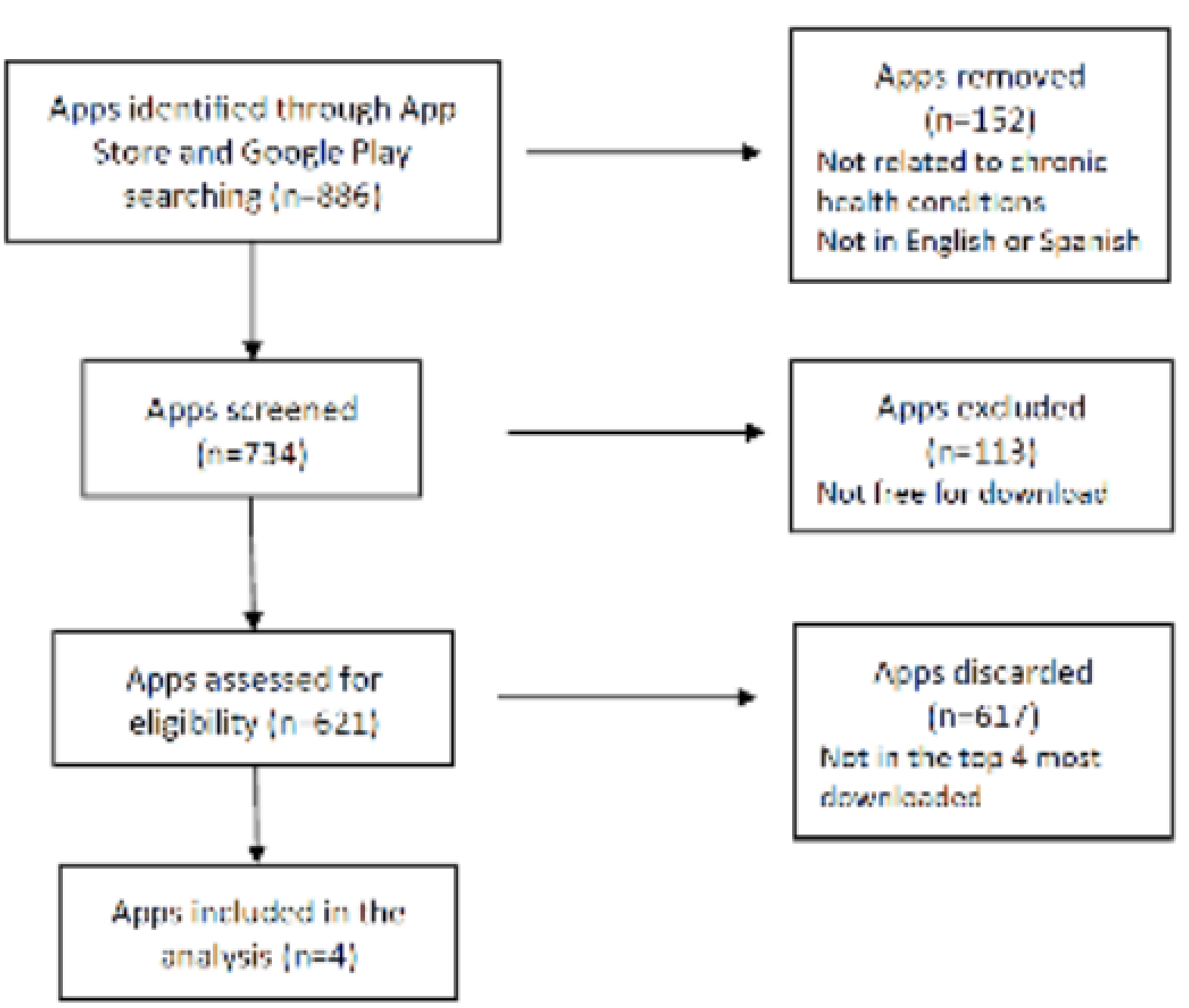



Figure 1. Steps in the apps selection process

Following this process we identified the top four most downloaded apps (one for each chronic health condition), which we then used in this study.

*Apps evaluation process*

The apps were rated by eight reviewers during the months of October and November in the year 2020. The reviewers were a group of stakeholders that included clinical researchers, engineers, health-care professionals, and end-users as potential patients. Reviewers received (1) the list of apps, (2) a survey including the items of MAG and MARS to be evaluated, and (3) specific instructions as to how to proceed with the review and evaluation of the apps. In order to avoid potential interferences and help reviewers to work independently, and in line with similar studies (e.g., [42]), they were not given any other suggestions, indications or training about the procedure.

For the evaluation, all reviewers downloaded and installed the apps on their personal mobile device. Then, they reviewed each of the apps using the specific criteria in MAG and MARS. In their assessment, the reviewers were instructed to only take into account the content and information provided

within the app itself and the stores (i.e., App Store and Google Play). This included websites, scientific studies and other external references as long as they were suggested or mentioned explicitly within the app or the stores. Like similar successful procedures, although reviewers spent several minutes examining the apps, they were not instructed to use them realistically [42]. The objective of this activity and procedure was that they would evaluate the apps in the same way as experts who do not need them would.

MAG [31] has 48 items grouped into 8 categories or domains: *usability*, *privacy*, *security*, *appropriateness and suitability*, *transparency and content*, *safety*, *technical support and updates* and *technology*. The reviewers used each of the items in the categories to assess the quality of the apps, and checked if the apps met those characteristics and functions or not (1=yes; 0=no).

The MARS [22] has 23 items that are grouped into 5 categories: *engagement*, *functionality*, *aesthetics*, *information quality*, and *subjective quality*. It also has six items, which are *app-specific,* that can be adapted to include or exclude specific information on the topic of interest. For example, these items

have been used to assess the perceived effects on the user's knowledge, attitudes, and intentions to change as well as the likelihood of changing the identified targeted behaviors in a study of mobile apps supporting heart failure symptom monitoring and self-care management [23]). In this study we discarded these *app-specific* items. When using the MARS, the reviewers used each of the items to assess the quality of the apps, and scored them using a 5-point rating scale (1-Inadequate, 2-Poor, 3-Acceptable, 4-Good, 5-Excellent).

*Data analysis*

In order to study and compare the interrater reliability of MAG and MARS, we calculated Krippendorff's alpha [43,44] for every category in the two guides, for each kind of reviewer and every app, separately and combined. Krippendorff's coefficient has been found to be superior to Cohen's coefficient, and can be used with an unlimited number of reviewers [45–47]. An alpha >.667 indicated agreement [44]. A negative alpha indicated that agreement was less than could be expected by chance. All data analyses were performed using SPSS v.26 for Windows using the Kalpha macro [48].

## Results

A total of eight reviewers rated the four apps using the MAG and MARS guides. The mobile apps included in the analysis were: “Manage My Pain” (i.e., pain); “BELONG Beating Cancer Together” (i.e., cancer); “mySugr - Diabetes App & Blood Sugar Tracker” (i.e., Diabetes); and “ASCVD Risk Estimator Plus” (i.e., cardiovascular diseases).

The group of reviewers included two clinical researchers, two engineers, two health-care professionals, and two end-users as potential patients. Complete responses were provided for almost all criteria and apps, although a small number of criteria showed a percentage of data completeness that ranged from 78% to 97% (e.g., “It has password management mechanisms”; see Multimedia Appendix 1). Tables 1 and 2 show the interrater reliability coefficients by categories, and overall, for both guides.

In MAG, the reviewers’ scores for several categories complied with the criteria. The highest interrater reliability scores were for the categories “*Privacy*” (Engineers: p = 0.73) and “*Security*” (Engineers: p = 0.78; and Health care

professionals: p = 0.76). In addition, two other categories, "*Usability*" and "*Safety*", were very close to compliance (Health care professionals: p = 0.62 and 0.61, respectively). The total interrater reliability of MAG (i.e., for all categories) was 0.45 (see Table 1).

Table 1. Interrater reliability scores when reviewers used MAG

| | Reliability | | | | |
|---|---|---|---|---|---|
| Category | Clinical researchers | Engineers | Health care professionals | End-users | Aggregate |
| **Usability** | 0.28 | 0.28 | 0.62 | 0.45 | 0.38 |
| **Privacy** | 0.36 | 0.73 | 0.42 | 0.43 | 0.45 |
| **Security** | 0.18 | 0.78 | 0.76 | 0.26 | 0.47 |
| **Appropriateness and suitability** | 0.38 | 0 | -0.15 | 0 | 0.25 |
| **Transparency and content** | 0 | 1 | -0.40 | -0.36 | 0.15 |
| **Safety** | 0.59 | 0.51 | 0.61 | -0.23 | 0.33 |
| **Technical support and updates** | 0.38 | 1 | 1 | 0.76 | 0.30 |
| **Technology** | 0.44 | 0.45 | -0.05 | 0.45 | 0.39 |
| **Total** | 0.40 | 0.66 | 0.55 | 0.29 | 0.45 |

In MARS, none of the reviewers' scores or the aggregate scores complied with the criteria. The categories with the highest interrater index were "*Engagement*" and "*Subjective*" with an overall alpha coefficient of 0.43 in both cases. The total interrater reliability of MARS (i.e., for all categories) was 0.29 (see Table 2).

## Resultats

Table 2. Interrater reliability scores when reviewers used MARS

| | Reliability | | | | |
|---|---|---|---|---|---|
| Category | Clinical researchers | Engineers | Health care professionals | End-users | Aggregate |
| **Engagement** | 0.18 | 0.50 | 0.53 | 0.41 | 0.43 |
| **Functionality** | 0.24 | 0.52 | 0.40 | -0.38 | 0.19 |
| **Aesthetics** | 0.42 | 0.26 | 0.23 | -0.14 | 0.17 |
| **Information** | 0.03 | 0.08 | 0.05 | -0.09 | 0.06 |
| **Subjective** | 0.57 | 0.41 | -0.08 | 0.54 | 0.43 |
| **Total** | 0.27 | 0.41 | 0.25 | 0.19 | 0.29 |

Tables 3 and 4 show the interrater reliability scores for each mobile app assessed using the MAG and MARS guides. As can be seen, none of the scores complied with the criteria overall or in any category. Nevertheless, the highest interrater reliability scores were for the MAG guide.

Table 3. Interrater reliability scores for apps when reviewers used MAG

| | Reliability | | | |
|---|---|---|---|---|
| Category | Manage My Pain | BELONG Beating Cancer Together | mySugr - Diabetes App & Blood Sugar Tracker | ASCVD Risk Estimator Plus |
| **Usability** | 0.58 | 0.49 | 0.27 | 0.15 |
| **Privacy** | 0.47 | 0.38 | 0.28 | 0.20 |
| **Security** | 0.44 | 0.18 | 0.42 | 0.32 |
| **Appropriateness and suitability** | 1 | 0.42 | 0 | -0.04 |
| **Transparency and content** | 0.08 | -0.08 | -0.06 | 0.00 |
| **Safety** | 0 | 0.47 | 0.33 | 0.21 |
| **Technical support and updates** | 0.10 | 0.57 | 0.16 | 0.10 |
| **Technology** | 0.17 | 0.36 | 0.12 | 0.45 |
| **Total** | 0.53 | 0.42 | 0.32 | 0.35 |

Table 4. Interrater reliability scores for apps when reviewers used MARS

| | Reliability | | | |
|---|---|---|---|---|
| Category | Manage My Pain | BELONG Beating Cancer Together | mySugr - Diabetes App & Blood Sugar Tracker | ASCVD Risk Estimator Plus |
| **Engagement** | 0.31 | 0.24 | -0.10 | 0.18 |
| **Functionality** | 0.27 | 0.05 | -0.02 | 0.16 |
| **Aesthetics** | -0.05 | -0.03 | -0.07 | 0.12 |
| **Information** | -0.08 | 0.08 | -0.03 | 0.09 |
| **Subjective** | 0.55 | 0.44 | 0.16 | 0.14 |
| **Total** | 0.20 | 0.18 | 0.01 | 0.42 |

Additional supplementary information is also provided on the interrater reliability scores for each item (see Multimedia Appendix 1).

## Discussion

### *Principal findings*

This research is the first to measure the interrater reliability of the Mobile App Development and Assessment Guide (MAG; [13,31]). We used MAG to study four mobile health-related apps, and compared the results with those obtained with the Mobile App Rating Scale (MARS; [22], one of the most extensively used guides in the field.

The data showed that few categories had a high interrater reliability. This finding is similar to that of other

studies (e.g., [26,42,46]) that have analyzed this type of guide. Taken as a whole, the findings demonstrate that it is difficult for reviewers to rate the apps in the same or similar way. A potential explanation for this finding is that reviewers do not interact with the apps in the same way, so they display different responses and functions [46]. Therefore, it is unlikely that reviewers will detect all app functions, which leads to differences in the ratings because they might not be assessing exactly the same. Support for this explanation can be found in the fact that the most objective categories evaluated, those which require less subjective interpretation by the reviewer (e.g., "*Privacy*", "*Security*"), are the ones with the highest interrater reliabilities. This finding is similar to the one reported by Powell and colleagues who detected that the less judgment required by reviewers, the higher the reliability [42].

Another important finding of this study was that interrater reliability scores for MAG were better than for MARS. Importantly, some of the MAG categories with the highest interrater reliability are not included in MARS (e.g., "*Privacy*", "*Security*", "*Technology*"). These are issues that have grown in importance in the field in recent years.

It should also be noted that some MAG categories showed a higher interrater reliability than others, but there was considerable variation in the scores between the types of reviewer. This finding suggests that the differences in the interrater reliability scores are related to such individual characteristics of the reviewers as background or training. This could help explain, in part at least, why engineers showed the highest reliability scores in the category of “Security”, as this is an important issue that is currently a matter of key interest in the training of engineers but not in the case of clinical researchers. And it implies that reviewers from different backgrounds are required to assess apps and that reviewers need to be trained. However, it is also possible that the low interrater reliability scores were not only reviewer-related, but also app-related. That is, although we selected the four most downloaded apps, they may not have been quality apps or easy to assess (e.g., the functions or properties of the apps were not easy to find or identify). In support of this explanation, some items were not answered by any reviewers in either of the guides (e.g., “It has a data recovery system in case of loss”; “It is based on ethical principles and values”). Finally, another

non-exclusive explanation for these results could be guide-related. The fact that the categories which required less interpretation (e.g., "*Security*") were the ones with the highest interrater reliability would support this explanation. This suggests that the guides must be improved.

The differences in interrater reliability, and more importantly the lower scores found, suggest that there is a very important underlying problem which is indicative of the difficulty of creating a good guide to help in the development and assessment of health-related apps. On the basis of the results of this study and others (e.g., [42,46]), users of health-related apps should use and interpret the results of quality assessments with caution.

The assessment of the quality of health-related apps is very important. Therefore, we must continue working on improving the way assessments are conducted. This may not only require improving the available guides, but also working with specialized centers and trained reviewers.

*Future research*

Studies are needed to help improve available guides that are psychometrically sound so future research should

focus on how to improve and empirically test interrater reliability. For example, studies should examine whether giving reviewers additional training is enough, or how reviewers' knowledge and assessment skills can best be improved. They should also establish whether the quality of health-related apps should be assessed by reviewers with different qualifications, training and background. Additional research would also help to determine whether understandable and well-defined criteria can improve interrater reliability above and beyond the improvement in reviewer training. Moreover, and specifically in relation with MAG, additional research with more apps of different types is also warranted. This would help ascertain whether and how different types of app influence the reviewers' evaluations.

*Limitations*

This study has a number of limitations that should be taken into account when interpreting the results. First, we studied the interrater reliability of MAG when it was used to evaluate apps that were available in both Android and IOS. Although the applications are generally the same on both platforms, there may be small differences that influence the

user's experience or performance when using different platforms and devices. For example, the amount of information displayed may differ due to the size of the screen, or the position and size of some elements (e.g., buttons, menu). Second, we used a very limited number of apps. We selected the most downloaded ones, as we thought they would be of better quality and therefore easier for reviewers to assess. However, they may not be of quality or representative of health-related apps, and so may not be suitable for an accurate study of the interrater reliability of the guides. Third, during the period of time that the apps were being assessed, they may have been updated or modified, which would have had an unknown impact on the results of the assessments. Finally, although individuals from different groups participated, they may be not representative. Even though they were extremely knowledgeable in their respective areas they may or may not be the best individuals for assessing the quality of the apps, as none of them had received any training. Future studies, then, should examine whether training can help improve the reviewers' assessment and the interrater reliability.

*Conclusions*

Despite the limitations of the study, our findings provide new and important information about MAG. Of particular consequence is that several categories in MAG have a significant interrater reliability. In addition, the data shows that the scores are better than the ones provided by MARS, the most commonly used guide in the area.

**Acknowledgements**

This work was partly supported by grants from the Spanish Ministry of Economy, Industry and Competitiveness (RTI2018-09870-B-I00; RED2018-102546-T), and the European Regional Development Fund (ERDF). JM's work is supported by Fundación Grünenthal. PL benefitted from a predoctoral fellowship (2019 FI_B2 00151) co-financed by the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya, the European Union (EU) and the European Social Fund (ESF).

**Conflicts of interest**

The authors declare no financial or other relationships that might lead to a conflict of interest related to this study.

## References

1. GSMA Intelligence. The Mobile Economy 2019. 2019. https://www.gsmaintelligence.com/research/?file=b9a6e6202ee1d5f787cfebb95d3639c5&download
2. Statista. App stores: number of apps in leading app stores 2020. 2020. https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/
3. Research2Guidance. mHealth Economics 2017 – Current Status and Future Trends in Mobile Health. 2017. https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health
4. de la Vega R, Roset R, Castarlenas E, Sánchez-Rodríguez E, Solé E, Miró J. Development and Testing of Painometer: A Smartphone App to Assess Pain Intensity. J Pain 2014;15(10):1001–1007. PMID:24854065
5. de la Vega R, Roset R, Galán S, Miró J. Fibroline : A mobile app for improving the quality of life of young

people with fibromyalgia. J Health Psychol 2018;23(1):67–78. PMID:27225285

6. Kessel KA, Vogel MME, Schmidt-Graf F, Combs SE. Mobile Apps in Oncology: A Survey on Health Care Professionals' Attitude Toward Telemedicine, mHealth, and Oncological Apps. J Med Internet Res 2016;18(11):e312. PMID:27884810

7. Kebede MM, Pischke CR. Popular Diabetes Apps and the Impact of Diabetes App Use on Self-Care Behaviour: A Survey Among the Digital Community of Persons With Diabetes on Social Media. Front Endocrinol (Lausanne) 2019;10(1):135. PMID:30881349

8. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: Systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol. 2018. p. 505–521. PMID:29313363

9. World Health Organization. mHealth: New horizons for health through mobile technologies, Global Observatory for eHealth series - Volume 3. World Health

Organization; 2011. www.who.int/goe/publications/ehealth_series_vol3/en/

10. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile Apps for Increasing Treatment Adherence: Systematic Review. J Med Internet Res 2019;21(6):e12505. PMID:31215517

11. Ernsting C, Dombrowski SU, Oedekoven M, O'Sullivan JL, Kanzler E, Kuhlmey A, Gellert P. Using smartphones and health apps to change and manage health behaviors: A population-based survey. J Med Internet Res 2017;19(4):e101. PMID:28381394

12. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52. PMID:25803266

13. Llorens-Vernet P, Miró J. Standards for Mobile Health-related Apps: Systematic review and Development of a Guide. JMIR mHealth uHealth 2020;8(3):e13057. PMID:32130169

14. de la Vega R, Miró J. mHealth: A Strategic Field without

a Solid Scientific Soul. A Systematic Review of Pain-Related Apps. E. N, editor. PLoS One 2014;9(7):e101312. PMID:24999983

15. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Informatics Assoc 2020;27(2):330–340. PMID:31599936

16. Park JYE, Li J, Howren A, Tsao NW, De Vera M. Mobile Phone Apps Targeting Medication Adherence: Quality Assessment and Content Analysis of User Reviews. JMIR mHealth uHealth 2019;7(1):e11919. PMID:30702435

17. Subhi Y, Bube SH, Rolskov Bojsen S, Skou Thomsen AS, Konge L. Expert Involvement and Adherence to Medical Evidence in Medical Mobile Phone Apps: A Systematic Review. JMIR mHealth uHealth 2015;3(3):e79. PMID:26215371

18. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: A systematic review of features and content quality. J Med Internet Res 2015;17(8):e198. PMID:26283290

19. Huckvale K, Adomaviciute S, Prieto JT, Leow MKS, Car J. Smartphone apps for calculating insulin dose: A systematic assessment. BMC Med 2015;13(1):1–10. PMID:25943590

20. Medicines and Healthcare products Regulatory Agency. Medical devices: software applications (apps). 2014. https://www.gov.uk/government/publications/medical-devices-software-applications-apps

21. Haute Autorité de Santé. Good practice guidelines on health apps and smart devices (mobile health or mhealth). 2016. https://www.has-sante.fr/jcms/c_2681915/en/good-practice-guidelines-on-health-apps-and-smart-devices-mobile-health-or-mhealth

22. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. JMIR mHealth uHealth 2015;3(1):e27. PMID:25760773

23. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and Analysis of

Existing Mobile Phone Apps to Support Heart Failure Symptom Monitoring and Self-Care Management Using the Mobile Application Rating Scale (MARS). JMIR mHealth uHealth 2016;4(2):e74. PMID:27302310

24. Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone Apps Targeting Alcohol and Illicit Substance Use: Systematic Search in in Commercial App Stores and Critical Content Analysis. JMIR mHealth uHealth 2019;7(4):e11831. PMID:31008713

25. Mehdi M, Stach M, Riha C, Neff P, Dode A, Pryss R, Schlee W, Reichert M, Hauck FJ. Smartphone and Mobile Health Apps for Tinnitus: Systematic Identification, Analysis, and Assessment. JMIR mHealth uHealth 2020;8(8):e21767. PMID:32808939

26. Knitza J, Tascilar K, Messner EM, Meyer M, Vossen D, Pulla A, Bosch P, Kittler J, Kleyer A, Sewerin P, Mucke J, Haase I, Simon D, Krusche M. German mobile apps in rheumatology: Review and analysis using the mobile application rating scale (MARS). J Med Internet Res 2019;21(8):e14991. PMID:31381501

27. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ.

Apps for People With Rheumatoid Arthritis to Monitor Their Disease Activity: A Review of Apps for Best Practice and Quality. JMIR mHealth uHealth 2017;5(2):e7. PMID:28223263

28. Nouri R, Kalhori SRN, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: A systematic review. J Am Med Informatics Assoc 2018;25(8):1089–1098. PMID:29788283

29. Siddique AB, Krebs M, Alvarez S, Greenspan I, Patel A, Kinsolving J, Koizumi N. Mobile apps for the care management of chronic kidney and end-stage renal diseases: Systematic search. J Med Internet Res 2019;21(9):e12604. PMID:31486408

30. Levine DM, Co Z, Newmark LP, Groisser AR, Holmgren AJ, Haas JS, Bates DW. Design and testing of a mobile health application rating tool. npj Digit Med 2020;3(74). PMID:32509971

31. Llorens-Vernet P, Miró J. The Mobile App Development and Assessment Guide (MAG): Delphi-Based Validity Study. JMIR mHealth uHealth 2020;8(7):e17760. PMID:32735226

32. Research2Guidance. mHealth Economics 2016 – Current Status and Trends of the mHealth App Market. 2016. https://research2guidance.com/product/mhealth-app-developer-economics-2016/

33. Centers for Disease Control and Prevention. Chronic Diseases in America. 2020. https://www.cdc.gov/chronicdisease/resources/infographic/chronic-diseases.htm

34. Australian Institute of Health and Welfare. Chronic disease Overview. 2019. https://www.aihw.gov.au/reports-data/health-conditions-disability-deaths/chronic-disease/overview

35. European Commission. EU research on chronic diseases. 2014. https://ec.europa.eu/info/research-and-innovation/research-area/health-research-and-innovation/chronic-diseases_en

36. OECD. Health at a Glance 2019. 2019. doi: 10.1787/4dd50c09-en

37. Institute for Health Metrics and Evaluation (IHME). Findings from the Global Burden of Disease Study 2017. 2018.

http://www.healthdata.org/sites/default/files/files/policy_report/2019/GBD_2017_Booklet.pdf

38. Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, … Murray, CJL. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1736–1788. PMID:30496103

39. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, … Murray, CJL. Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1789–1858. PMID:30496104

40. Kyu HH, Abate D, Abate KH, Abay S M, Abbafati C, Abbasi N, … Murray, CJL. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017.

Lancet 2018;392(10159):1859–1922. PMID:30415748

41. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Data Resources. Inst Heal Metrics Eval. 2018 [cited 2020 Oct 1]. http://ghdx.healthdata.org/gbd-2017

42. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, Landman AB. Interrater Reliability of mHealth App Rating Measures: Analysis of Top Depression and Smoking Cessation Apps. JMIR mHealth uHealth 2016;4(1):e15. PMID:26863986

43. Krippendorff K. Computing Krippendorff's Alpha-Reliability. 2011. https://repository.upenn.edu/asc_papers/43

44. Krippendorff K. Content Analysis: An Introduction to Its Methodology (4th edition). Sage Publ. Thousand Oaks, California; 2018. ISBN:9781506395661

45. Karlsson D, Gøeg K, Orman H, Højen A. Semantic Krippendorff's α for measuring inter-rater agreement in SNOMED CT coding studies. Stud Health Technol Inform 2014;205:151–155. PMID:25160164

46. McKay FH, Slykerman S, Dunn M. The App Behavior

Change Scale: Creation of a Scale to Assess the Potential of Apps to Promote Behavior Change. JMIR mHealth uHealth 2019;7(1):e11130. PMID:30681967

47. Volkmann N, Stracke J, Kemper N. Evaluation of a gait scoring system for cattle by using cluster analysis and Krippendorff's α reliability. Vet Rec 2019;184(7):220. PMID:30455190

48. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Communication Methods and Measures. 2007;1:77–89. doi: 10.1080/19312450709336664

# 5. Discussió

Aquesta tesi doctoral, a través de la realització de 3 estudis consecutius, proposa una guia que faciliti l'avaluació de la qualitat de les apps de salut disponibles i el desenvolupament de noves apps de salut de qualitat. El treball realitzat s'ha concretat en tres fases: en (1) estudiar els criteris més importants per a identificar i desenvolupar apps mòbils de salut per tal d'integrar-los en una guia que permetés superar els problemes existents, (2) estudiar la validesa dels criteris i de la nova guia creada i (3) contrastar la fiabilitat de la guia.

Aquest treball es va plantejar per la falta d'un estàndard i d'un consens general de com han de ser les apps mòbils de salut. Si bé diferents grups d'experts i institucions han creat guies, ho han fet seguint les seves pròpies pautes amb un coneixement limitat del seu àmbit d'actuació, per exemple, pensant en criteris específics per a problemes de salut o col·lectius de pacients concrets (Brown et al., 2013; Gaziel-Yablowitz & Schwartz, 2018). Aquesta diversitat de criteris ha provocat certa inseguretat a usuaris i experts al no disposar d'un referent clar de com hauria de ser una app de salut de qualitat (Akbar et al., 2020; Peng, Kanthawala, Yuan, & Hussain, 2016).

Els estudis que configuren aquesta tesi, donen resposta al problema i contribueixen a donar solució, encara que sigui parcialment i provisionalment, a la creixent presència d'apps de dubtosa qualitat (Wisniewski et al., 2019; Wyatt, 2018). La nova guia que es proposa en

aquesta tesi (MAG) es va fer a partir de fonts d'informació recollides d'arreu del món i validada per un conjunt d'actors rellevants, multidisciplinari i de procedència internacional. Així, doncs, supera guies semblants existents doncs té en compte els aspectes de les apps més rellevants, dels que actualment són considerats pertinents, i inclou tots els criteris possibles. Comptar amb tan ampli ventall de criteris en una sola guia fa que es pugui utilitzar per a desenvolupar o avaluar apps de qualsevol tipus.

Aquesta guia ha estat sotmesa a diferents proves de comprensió i fiabilitat, per tal d'assegurar que estava a l'alçada de les necessitats del que es requereix d'una eina com aquesta. I els estudis realitzats mostren que supera allò que caldria exigir a una guia. Primer, per què s'ha verificat que els criteris inclosos en la guia són comprensibles i, per tant podrien ser utilitzats per la majoria de persones. Segon, per què s'ha pogut comprovar que els criteris inclosos a la guia són importants per a identificar i garantir la qualitat de les apps. I tercer, per què malgrat les dificultats observades, alguns dels criteris que inclou la MAG demostren una alta fiabilitat entre jutges. De fet, després de comparar la MAG amb la MARS, que és la guia més utilitzada en aquest àmbit, s'observa que la MAG mostraria detalls pels quals se la podria valorar d'una forma més positiva.

Fins on sabem, aquest treball és el primer que proporciona una guia completa en tots els àmbits, per ajudar en el desenvolupament i avaluació d'apps mòbils relacionades amb la salut, en forma de llista de criteris i categories. Per exemple, incorpora criteris relacionats amb la usabilitat (és a dir, l'aplicació s'ha d'adaptar a la població objectiu),

privacitat (és a dir, compliment de la llei i el tractament de les dades dels usuaris), com també relacionats amb la seguretat (és a dir, la potencialitat del risc per als usuaris finals), entre altres molts criteris. Amb aquesta varietat de criteris, s'ha aconseguit que la MAG avaluï des de totes les possibles perspectives i valori tot allò rellevant per a una app de salut de qualitat.

Malgrat el progrés que significa la MAG i la feina feta pel seu desenvolupament, el treball realitzat no està exempt de limitacions. Primer, en el recull de fonts d'informació internacionals, ens vam centrar solament en aquelles regions i països en que l'ús de les tecnologies mòbils era molt extens. I això per limitar una informació que d'altra forma no hauria estat possible d'analitzar. Per tant, podria ser que algun criteri important no s'hagués inclòs. Tot i que sembla poc probable que algun criteri veritablement important s'hagués escapat, doncs, per una banda, la informació provenia dels països amb una major implantació de la tecnologia mòbil que segurament tenen una més llarga i reeixida tradició en aquest camp i, per altra banda, per què a les persones que participaven en els nostres estudis se'ls demanava que afegissin qualsevol criteri rellevant que consideressin que faltava. Segon, algunes persones a les que es va proposar participar en les estudis no ho van voler fer. Desconeixem si qui no va voler participar és molt diferent de les persones que sí van col·laborar, en aspectes rellevants. Sigui com sigui, els grups de participants superaven amb escreix els criteris metodològics exigibles, en els que hi eren presents les sensibilitats i interessos dels actors més

rellevants (això és, grups multidisciplinaris d'investigadors, usuaris potencials, desenvolupadors d'apps, especialistes sanitaris potencials prescriptors d'apps). Més encara, eren grups de configuració internacional, constituïts amb persones molt expertes amb una àmplia experiència en el seu camp. Tercer, i pel que fa a l'avaluació de la fiabilitat, vam triar un número limitat d'apps que tot i ser les més descarregades, podrien no ser de qualitat. Tanmateix, malgrat que van participar *stakeholders* ben diferents, fet que garantia incloure diferents sensibilitats, no tenien cap mena d'experiència en l'avaluació d'apps, i malgrat que se'ls van donar instruccions molt precises de què calia fer i com, la manca d'experiència en aquesta activitat podria haver influït en els resultats.

Malgrat que els resultats dels estudis realitzats indiquen que la MAG és una guia comprensible, i més fiable que anteriors guies, cal considerar aquests resultats com de naturalesa preliminar. I, així, cal nous estudis amb els que determinar quina és la seva idoneïtat, és a dir, si es pot utilitzar per a desenvolupar i avaluar qualsevol tipus d'app de salut mantenint els mateixos resultats de qualitat. Un altre aspecte a estudiar és la importància relativa dels diferents criteris inclosos. Es a dir, tot i que hem verificat la importància dels criteris que inclou la MAG, caldria estudiar quins d'ells tenen un major pes en la definició de la qualitat i si aquesta variació depèn del tipus d'app de la que es tracti. D'aquesta manera es podria millorar la precisió en l'avaluació de les apps, a més de poder prioritzar quins aspectes de les apps necessiten més atenció al moment de desenvolupar-les. També seria d'interès desenvolupar una versió d'usuari

de la MAG, d'aquesta manera es podria estudiar l'associació entre la qualitat de l'experiència d'usuari i la puntuació a la MAG, i una versió abreujada per tal de facilitar la feina dels futurs potencials usuaris i millorar l'experiència d'ús. Igualment, futures investigacions hauran d'estudiar si l'entrenament dels avaluadors millora els resultats de les avaluacions. Finalment, val a dir que la MAG serà una eina en revisió i millora constant, doncs és altament probable que en un futur proper convingui modificar o afegir criteris per poder avaluar les noves funcionalitats de les apps de salut i adaptar-se a les exigències d'un món que està en constant desenvolupament.

La MAG es pot utilitzar per dos objectius diferents, per comprovar la qualitat d'una app ja existent o com una guia per crear una nova app, tant per un motiu com per l'altre el procés d'ús és similar.

La MAG s'ha creat perquè no tingui expressions o terminologies excessivament tècniques i, per tant, no hi hagi dificultat per a la seva comprensió. Tot i així, considerant els resultats del nostre treball, abans de fer-ne ús, es recomana tenir uns mínims coneixements en el camp de la salut mòbil per poder-la utilitzar correctament i amb la màxima eficàcia.

# 6. Conclusions

Les principals conclusions d'aquesta tesi doctoral són les següents:

1. Els criteris de qualitat identificats i que inclou la *Mobile App Development and Assessment Guide* (MAG) poden ser utilitzats per proveïdors de salut, enginyers i desenvolupadors, investigadors, pacients i reguladors, i han demostrat ser fàcilment comprensibles.
2. Els 48 criteris més importants estan relacionats amb les temàtiques de Privacitat, Ciberseguretat, Usabilitat, Seguretat, Tecnologia, Adequació i idoneïtat, Transparència i contingut i Suport tècnic i actualitzacions.
3. La MAG és una eina comprensible, i vàlida tant per avaluar de forma sistemàtica la qualitat de les aplicacions de salut com per guiar el desenvolupament de noves aplicacions rigoroses i de qualitat en salut.
4. Malgrat que de forma preliminar, els estudis realitzats mostren que la MAG seria una millor eina per avaluar la qualitat de les apps i per a guiar el desenvolupament de noves, en base a criteris de qualitat.

# 7. Referències

Accenture. (2018a). *2018 Consumer survey on Digital Health*. Retrieved from https://www.accenture.com/t20180306t103559z__w__/us-en/_acnmedia/pdf-71/accenture-health-2018-consumer-survey-digital-health.pdf

Accenture. (2018b). *Consumer Survey on Digital Health in Australia*. Retrieved from https://www.accenture.com/_acnmedia/pdf-84/accenture-health-meet-todays-healthcare-team-patients-doctors-machines.pdf

Adibi, S. (2015). *Mobile Health: A Technology Road Map* (Vol. 5). https://doi.org/10.1007/978-3-319-12817-7

Adler-Milstein, J., Salzberg, C., Franz, C., Orav, E. J., Newhouse, J. P., & Bates, D. W. (2013). Effect of electronic health records on health care costs: Longitudinal comparative evidence from community practices. *Annals of Internal Medicine*, *159*(2), 97–104. https://doi.org/10.7326/0003-4819-159-2-201307160-00004

Agencia de Calidad Sanitaria de Andalucia. (2014). Distintivo AppSaludable. Retrieved from http://www.calidadappsalud.com/listado-completo-recomendaciones-app-salud/

Akbar, S., Coiera, E., & Magrabi, F. (2020). Safety concerns with consumer-facing mobile health applications and their consequences: a scoping

## Referències

review. *Journal of the American Medical Informatics Association*, *27*(2), 330–340. https://doi.org/10.1093/jamia/ocz175

Balapour, A., Reychav, I., Sabherwal, R., & Azuri, J. (2019). Mobile technology identity and self-efficacy: Implications for the adoption of clinically supported mobile health apps. *International Journal of Information Management*, *49*, 58–68. https://doi.org/10.1016/j.ijinfomgt.2019.03.005

Bastawrous, A., & Armstrong, M. J. (2013). Mobile health use in low-and high-income countries: An overview of the peer-reviewed literature. *Journal of the Royal Society of Medicine*. https://doi.org/10.1177/0141076812472620

Bhattacharya, S., Kumar, A., Kaushal, V., & Singh, A. (2018). Applications of m-Health and e-Health in Public Health sector: The Challenges and Opportunities. *A Multifaceted Peer Reviewed Journal in the Field of Medicine and Public Health International Journal of Medicine and Public Health*, *8*(2), 56–63. https://doi.org/10.5530/ijmedph.2018.2.12

Boudreaux, E. D., Waring, M. E., Hayes, R. B., Sadasivam, R. S., Mullen, S., & Pagoto, S. (2014). Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. *Translational Behavioral Medicine*, *4*(4), 363–371. https://doi.org/10.1007/s13142-014-0293-9

Brown, W., Yen, P.-Y., Rojas, M., & Schnall, R. (2013). Assessment of the

## Referències

Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. *Journal of Biomedical Informatics*, *46*(6), 1080–1087. https://doi.org/10.1016/j.jbi.2013.08.001

Buijink, A. W. G., Visser, B. J., & Marshall, L. (2013). Medical apps for smartphones: lack of evidence undermines quality and safety. *Evidence Based Medicine*, *18*(3), 90 LP – 92. https://doi.org/10.1136/eb-2012-100885

Byambasuren, O., Sanders, S., Beller, E., & Glasziou, P. (2018). Prescribable mHealth apps identified from an overview of systematic reviews. *Npj Digital Medicine*, *1*(1). https://doi.org/10.1038/s41746-018-0021-9

Chen, J., Lieffers, J., Bauman, A., Hanning, R., & Allman-Farinelli, M. (2017). The use of smartphone health apps and other mobile health (mHealth) technologies in dietetic practice: a three country study. *Journal of Human Nutrition and Dietetics*, *30*(4), 439–452. https://doi.org/10.1111/jhn.12446

de la Vega, R., & Miró, J. (2014). mHealth: A Strategic Field without a Solid Scientific Soul. A Systematic Review of Pain-Related Apps. *PLoS ONE*, *9*(7), e101312. https://doi.org/10.1371/journal.pone.0101312

de la Vega, R., Roset, R., Galán, S., & Miró, J. (2018). Fibroline : A mobile app for improving the quality of life of young people with fibromyalgia.

## Referències

*Journal of Health Psychology*, *23*(1), 67–78. https://doi.org/10.1177/1359105316650509

Dehling, T., Gao, F., Schneider, S., & Sunyaev, A. (2015). Exploring the Far Side of Mobile Health: Information Security and Privacy of Mobile Health Apps on iOS and Android. *JMIR MHealth and UHealth*, *3*(1), e8. https://doi.org/10.2196/mhealth.3672

El Amrani, L., Oude Engberink, A., Ninot, G., Hayot, M., & Carbonnel, F. (2017). Connected Health Devices for Health Care in French General Medicine Practice: Cross-Sectional Study. *JMIR MHealth and UHealth*, *5*(12), e193. https://doi.org/10.2196/mhealth.7427

Ernsting, C., Dombrowski, S. U., Oedekoven, M., O'Sullivan, J. L., Kanzler, E., Kuhlmey, A., & Gellert, P. (2017). Using smartphones and health apps to change and manage health behaviors: A population-based survey. *Journal of Medical Internet Research*, *19*(4), e101. https://doi.org/10.2196/jmir.6838

European Commission. (2014). Green Paper on mobile health. Retrieved from https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth

Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research*, *3*(2), 1–5. https://doi.org/10.2196/jmir.3.2.e20

Flick, T., & Morehouse, J. (2011). Mobile Applications and Devices. In *Securing the Smart Grid* (pp. 179–193). Elsevier.

Referències

https://doi.org/10.1016/b978-1-59749-570-7.00010-8

Food and Drug Administration. (2019). Policy for Device Software Functions and Mobile Medical Applications | FDA. Retrieved from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications

Gagnon, M. P., Ngangue, P., Payne-Gagnon, J., & Desmartis, M. (2016). M-Health adoption by healthcare professionals: A systematic review. *Journal of the American Medical Informatics Association*, *23*(1), 212–220. https://doi.org/10.1093/jamia/ocv052

Gaziel-Yablowitz, M., & Schwartz, D. G. (2018). A Review and Assessment Framework for Mobile-Based Emergency Intervention Apps. *ACM Computing Surveys*, *51*(1), 1–32. https://doi.org/10.1145/3145846

Generalitat de Catalunya. Departament de Salut. (2019). *Enquesta de salut de Catalunya*. Retrieved from http://salutweb.gencat.cat/web/.content/_departament/estadistiques-sanitaries/enquestes/Enquesta-de-salut-de-Catalunya/Resultats-de-lenquesta-de-salut-de-Catalunya/documents/2018-resultats/resum-executiu-esca-2018.pdf

Grady, A., Yoong, S., Sutherland, R., Lee, H., Nathan, N., & Wolfenden, L. (2018). Improving the public health impact of eHealth and mHealth interventions. *Australian and New Zealand Journal of Public Health*,

Referències

*42*(2), 118–119. https://doi.org/10.1111/1753-6405.12771

GSMA Intelligence. (2019). The Mobile Economy 2019. Retrieved from https://www.gsmaintelligence.com/research/?file=b9a6e6202ee1d5f787cfebb95d3639c5&download

Hajat, C., & Stein, E. (2018). The global burden of multiple chronic conditions: A narrative review. *Preventive Medicine Reports*. https://doi.org/10.1016/j.pmedr.2018.10.008

Hamine, S., Gerth-Guyette, E., Faulx, D., Green, B. B., & Ginsburg, A. S. (2015). Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. *Journal of Medical Internet Research*, *17*(2), e52. https://doi.org/10.2196/jmir.3951

Hanseth, O., & Bygstad, B. (2015). Flexible generification: ICT Standardization Strategies and Service Innovation in Health Care. *European Journal of Information Systems*, *24*(6), 645–663. https://doi.org/10.1057/ejis.2015.1

Haute Autorité de Santé. (2016). Good practice guidelines on health apps and smart devices (mobile health or mhealth). Retrieved from https://www.has-sante.fr/jcms/c_2681915/en/good-practice-guidelines-on-health-apps-and-smart-devices-mobile-health-or-mhealth

Huckvale, K., Prieto, J. T., Tilney, M., Benghozi, P. J., & Car, J. (2015).

## Referències

Unaddressed privacy risks in accredited health and wellness apps: A cross-sectional systematic assessment. *BMC Medicine*, *13*(1), 1–13. https://doi.org/10.1186/s12916-015-0444-y

IGES Institut GmbH. (2016). Health apps & co: safe digital care products with clearer regulations. Retrieved from https://www.iges.com/clients/health/forschungsergebnisse/2016/health-apps-ii/index_eng.html

Instituto Nacional de Estadística - España. (2019). Península Ibérica en cifras 2018. Retrieved from https://www.ine.es/prodyser/pubweb/pin/pin2018/

Investopedia. (2019). Compound Annual Growth Rate – CAGR Definition. Retrieved from https://www.investopedia.com/terms/c/cagr.asp

Kamel Boulos, M. N., Brewer, A. C., Karimkhani, C., Buller, D. B., & Dellavalle, R. P. (2014). Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online Journal of Public Health Informatics*, *5*(3), 229. https://doi.org/10.5210/ojphi.v5i3.4814

Kebede, M. M., & Pischke, C. R. (2019). Popular Diabetes Apps and the Impact of Diabetes App Use on Self-Care Behaviour: A Survey Among the Digital Community of Persons With Diabetes on Social Media. *Frontiers in Endocrinology*, *10*(1), 135. https://doi.org/10.3389/fendo.2019.00135

## Referències

Kessel, K. A., Vogel, M. M. E., Schmidt-Graf, F., & Combs, S. E. (2016). Mobile Apps in Oncology: A Survey on Health Care Professionals' Attitude Toward Telemedicine, mHealth, and Oncological Apps. *Journal of Medical Internet Research*, *18*(11), e312. https://doi.org/10.2196/jmir.6399

Larson, R. S. (2018). A Path to Better-Quality mHealth Apps. *JMIR MHealth and UHealth*, *6*(7), e10414. https://doi.org/10.2196/10414

Lewis, T. L., & Wyatt, J. C. (2014). mHealth and mobile medical apps: A framework to assess risk and promote safer use. *Journal of Medical Internet Research*, *16*(9), e210. https://doi.org/10.2196/jmir.3133

Linstone, H. A., Turoff, M., & Helmer, O. (1975). *The Delphi Method Techniques and Applications*. Retrieved from https://web.njit.edu/~turoff/pubs/delphibook/delphibook.pdf

Llorens-Vernet, P., & Miró, J. (2020a). Standards for Mobile Health-related Apps: Systematic review and Development of a Guide. *JMIR MHealth and UHealth*, *8*(3), e13057. https://doi.org/10.2196/13057

Llorens-Vernet, P., & Miró, J. (2020b). The Mobile App Development and Assessment Guide (MAG): Delphi-Based Validity Study. *JMIR MHealth and UHealth*, *8*(7), e17760. https://doi.org/10.2196/17760

Lu, C., Hu, Y., Xie, J., Fu, Q., Leigh, I., Governor, S., & Wang, G. (2018). The use of mobile health applications to improve patient experience: cross-sectional study in chinese public hospitals. *JMIR MHealth and*

Referències

*UHealth*, *6*(5). https://doi.org/10.2196/mhealth.9145

Marcolino, M. S., Oliveira, J. A. Q., D'Agostino, M., Ribeiro, A. L., Alkmim, M. B. M., & Novillo-Ortiz, D. (2018). The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. *JMIR MHealth and UHealth*, *6*(1), e23. https://doi.org/10.2196/mhealth.8873

Martínez-Pérez, B., de la Torre-Díez, I., & López-Coronado, M. (2015). Privacy and Security in Mobile Health Apps: A Review and Recommendations. *Journal of Medical Systems*, *39*(1). https://doi.org/10.1007/s10916-014-0181-3

Medicines and Healthcare products Regulatory Agency. (2014). Medical devices: software applications (apps). Retrieved from https://www.gov.uk/government/publications/medical-devices-software-applications-apps

Ministerio de Sanidad. (2018). *Nota Técnica Encuesta Nacional de Salud. España 2017 Principales resultados.* Retrieved from https://www.mscbs.gob.es/estadEstudios/estadisticas/encuestaNacional/encuestaNac2017/ENSE2017_notatecnica.pdf

Miró, J., & Llorens-Vernet, P. (2020). On the Assessment of the Quality of mHealth-related Apps: an Interrater Reliability Study of Two Guides. Article submitted for publication.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009).

## Referències

Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

National Academies of Sciences, E. and M. (2018). *Crossing the Global Quality Chasm: Improving Health Care Worldwide*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25152

Nouri, R., Kalhori, S. R. N., Ghazisaeedi, M., Marchand, G., & Yasini, M. (2018). Criteria for assessing the quality of mHealth apps: A systematic review. *Journal of the American Medical Informatics Association*, *25*(8), 1089–1098. https://doi.org/10.1093/jamia/ocy050

OECD/EU. (2018). *Health at a Glance: Europe 2018 State of Health in the EU Cycle*. https://doi.org/10.1787/health_glance_eur-2018-en

OECD. (2019). *Health at a Glance 2019*. https://doi.org/10.1787/4dd50c09-en

Oh, H., Rizo, C., Enkin, M., & Jadad, A. (2005). What Is eHealth (3): A Systematic Review of Published Definitions. *J Med Internet Res*, *7*(1), e1. https://doi.org/10.2196/jmir.7.1.e1

Pappot, H., Assam Taarnhøj, G., Elsbernd, A., Hjerming, M., Hanghøj, S., Jensen, M., & Boisen, K. A. (2019). Health-Related Quality of Life Before and After Use of a Smartphone App for Adolescents and Young Adults With Cancer: Pre-Post Interventional Study. *JMIR MHealth and UHealth*, *7*(10), e13829. https://doi.org/10.2196/13829

## Referències

Peng, W., Kanthawala, S., Yuan, S., & Hussain, S. A. (2016). A qualitative study of user perceptions of mobile health apps. *BMC Public Health*, *16*(1), 1–11. https://doi.org/10.1186/s12889-016-3808-0

Pérez-Jover, V., Sala-González, M., Guilabert, M., & Mira, J. J. (2019). Mobile Apps for Increasing Treatment Adherence: Systematic Review. *Journal of Medical Internet Research*, *21*(6), e12505. https://doi.org/10.2196/12505

Pickett, K. E., & Wilkinson, R. G. (2015). Income inequality and health: A causal review. *Social Science and Medicine*. https://doi.org/10.1016/j.socscimed.2014.12.031

Price, M., Yuen, E. K., Goetter, E. M., Herbert, J. D., Forman, E. M., Acierno, R., & Ruggiero, K. J. (2014). mHealth: A mechanism to deliver more accessible, more effective mental health care. *Clinical Psychology and Psychotherapy*, *21*(5), 427–436. https://doi.org/10.1002/cpp.1855

Research and Markets. (2019). *mHealth Apps Market Size, Share & Trends Analysis Report By Type (Fitness, Lifestyle Management, Nutrition & Diet, Women's Health, Medication Adherence, Healthcare Providers/Payers), And Segment Forecasts, 2019 - 2026*. Retrieved from https://www.researchandmarkets.com/reports/4396364/mhealth-apps-market-size-share-and-trends?utm_source=GNDIY&utm_medium=PressRelease&utm_cod

## Referències

e=mt9s62&utm_campaign=1288277+-+mHealth+Apps+Market+to+Reach+%2524236+Billion+by+2026%252C+Growing+at+a+CAGR+of+44.7%252

Research2Guidance. (2016). *mHealth Economics 2016 – Current Status and Trends of the mHealth App Market*. Retrieved from https://research2guidance.com/product/mhealth-app-developer-economics-2016/

Research2Guidance. (2017). mHealth Economics 2017 – Current Status and Future Trends in Mobile Health. Retrieved from https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/

Roser, M., Ortiz-Ospina, E., & Ritchie, H. (2020). Life Expectancy. Retrieved from https://ourworldindata.org/life-expectancy

Shanafelt, T. D., Boone, S., Tan, L., Dyrbye, L. N., Sotile, W., Satele, D., … Oreskovich, M. R. (2012). Burnout and Satisfaction With Work-Life Balance Among US Physicians Relative to the General US Population. *Archives of Internal Medicine*, *172*(18), 1377–1385. https://doi.org/10.1001/archinternmed.2012.3199

Singh, K., Drouin, K., Newmark, L. P., Jae, H. L., Faxvaag, A., Rozenblum, R., … Bates, D. W. (2016). Many mobile health apps target high-need, high-cost populations, but gaps remain. *Health Affairs*, *35*(12), 2310–2318. https://doi.org/10.1377/hlthaff.2016.0578

## Referències

Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. *JMIR MHealth and UHealth*, *3*(1), e27. https://doi.org/10.2196/mhealth.3422

Sucala, M., Cuijpers, P., Muench, F., Cardoș, R., Soflau, R., Dobrean, A., … David, D. (2017). Anxiety: There is an app for that. A systematic review of anxiety apps. *Depression and Anxiety*. https://doi.org/10.1002/da.22654

Therapeutic Goods Administration (TGA). (2019). Regulation of Software as a Medical Device. Retrieved from https://www.tga.gov.au/regulation-software-medical-device

Thurnheer, S. E., Gravestock, I., Pichierri, G., Steurer, J., & Burgstaller, J. M. (2018). Benefits of Mobile Apps in Pain Management: Systematic Review. *JMIR MHealth and UHealth*, *6*(10), e11231. https://doi.org/10.2196/11231

TIC Salut Social. (2018). AppSalut Marketplace | TIC Salut Social. Retrieved from https://ticsalutsocial.cat/en/projectes/mhealth-en/appsalut-marketplace/

Tic Salut Social Foundation. (2015). The Mobility Master Plan: mHealth solutions. 2015. Retrieved from http://www.webcitation.org/73sMqAseU

Tic Salut Social Foundation. (2018). Criteria Guide of Accreditation Process

Referències

for Health and Wellness Apps. 2018. Retrieved from http://www.webcitation.org/73sMDaN6Y

United Nations - Department of Economic and Social Affairs. (2019). *World Population Prospects 2019 Highlights*. Retrieved from https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf

Ventola, C. L. (2014). Mobile devices and apps for health care professionals: uses and benefits. *P & T : A Peer-Reviewed Journal for Formulary Management*, *39*(5), 356–364.

Vital Wave Consulting. (2009). *mHealth for Development: The Opportunity of Mobile Technology for Healthcare in the Developing World*. Retrieved from www.unfoundation.org/vodafone.

Watt, T., Charlesworth, A., & Gershlick, B. (2019). Health and care spending and its value, past, present and future. *Future Healthcare Journal*, *6*(2), 99–105. https://doi.org/10.7861/futurehosp.6-2-99

Wicks, P., & Chiauzzi, E. (2015). "Trust but verify" - five approaches to ensure safe medical apps. *BMC Medicine*, *13*(1), 1–5. https://doi.org/10.1186/s12916-015-0451-z

Wisniewski, H., Liu, G., Henson, P., Vaidyam, A., Hajratalli, N. K., Onnela, J. P., & Torous, J. (2019). Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. *Evidence-Based Mental Health*, *22*(1), 4–9. https://doi.org/10.1136/ebmental-2018-

Referències

300069

World Health Organization. (2011). mHealth: New horizons for health through mobile technologies, Global Observatory for eHealth series - Volume 3. Retrieved from www.who.int/goe/publications/ehealth_series_vol3/en/

World Health Organization. (2014). *Global Status Report On Noncommunicable Diseases 2014*. Retrieved from https://www.who.int/nmh/publications/ncd-status-report-2014/en/

World Health Organization. (2019). *Thirteenth general programme of work 2019–2023*. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/324775/WHO-PRP-18.1-eng.pdf

World Health Organization. (2020). Urgent health challenges for the next decade. Retrieved from https://www.who.int/news-room/photo-story/photo-story-detail/urgent-health-challenges-for-the-next-decade

World Health Organization, O. & I. B. for R. and D. (2018). *Delivering quality health services: a global imperative for universal health coverage*. World Health Organization. Retrieved from https://apps.who.int/iris/handle/10665/272465

Wozney, L., Freitas de Souza, L. M., Kervin, E., Queluz, F., McGrath, P. J., & Keefe, J. (2018). Commercially Available Mobile Apps for

## Referències

Caregivers of People With Alzheimer Disease or Other Related Dementias: Systematic Search. *JMIR Aging*, *1*(2), e12274. https://doi.org/10.2196/12274

Wyatt, J. C. (2018). How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? *BMC Medicine*. https://doi.org/10.1186/s12916-018-1211-7

Xu, K., Soucat, A., Kutzin, J., Brindley, C., Maele, N. Vande, Touré, H., … Siroka, A. (2018). *Public Spending on Health: A Closer Look at Global Trends*. Retrieved from https://www.who.int/health_financing/documents/health-expenditure-report-2018/en/

Yetisen, A. K., Martinez-Hurtado, J. L., Da Cruz Vasconcellos, F., Simsekler, M. C. E., Akram, M. S., & Lowe, C. R. (2014). The regulation of mobile medical applications. *Lab on a Chip*. https://doi.org/10.1039/c3lc51235e

Zhao, P., Yoo, I., Lancey, R., & Varghese, E. (2019). Mobile applications for pain management: An app analysis for clinical usage. *BMC Medical Informatics and Decision Making*, *19*(1), 106. https://doi.org/10.1186/s12911-019-0827-7

Zheng, H., Gaff, H., Smith, G., & DeLisle, S. (2014). Epidemic surveillance using an electronic medical record: An empiric approach to performance improvement. *PLoS ONE*, *9*(7).

## Referències

https://doi.org/10.1371/journal.pone.0100845

Zuppo, C. M. (2012). Defining ICT in a Boundaryless World: The Development of a Working Hierarchy. *International Journal of Managing Information Technology (IJMIT)*, *4*(3), 13–22. https://doi.org/10.5121/ijmit.2012.4302

UNIVERSITAT ROVIRA I VIRGILI
APLICACIONS MÒBILS EN SALUT: CRITERIS PER AL DESENVOLUPAMENT I AVALUACIÓ DE LA QUALITAT
Pere Llorens Vernet