Mobile phone interventions to improve health outcomes among patients with chronic diseases: an umbrella review and evidence synthesis from 34 meta-analyses

Shufang Sun, Otto Simonsson, Stephen McGarvey, John Torous, Simon B Goldberg

This umbrella review of 34 meta-analyses, representing 235 randomised controlled trials done across 52 countries and 48 957 participants and ten chronic conditions, aimed to evaluate evidence on the efficacy of mobile phone interventions for populations with chronic diseases. We evaluated the strengths of evidence via the Fusar-Poli and Radua methodology. Compared with usual care, mobile apps had convincing effects on glycated haemoglobin reduction among adults with type 2 diabetes (d=0·44). Highly suggestive effects were found for both text messages and apps on various outcomes, including medication adherence (among patients with HIV in sub-Saharan Africa and people with cardiovascular disease), glucose management in type 2 diabetes, and blood pressure reduction in hypertension. Many effects (42%) were non-significant. Various gaps were identified, such as a scarcity of reporting on moderators and publication bias by meta-analyses, little research in low-income and lower-middle-income countries, and little reporting on adverse events.

Lancet Digit Health 2024; 6: e857–70

Published Online September 26, 2024 https://doi.org/10.1016/S2589-7500(24)00119-5

Department of Behavioral and Social Sciences (S Sun PhD) **and Department of Epidemiology** (Prof S McGarvey PhD), **School of Public Health, Brown University, Providence, RI, USA; International Health Institute, School of Public Health, Brown University, Providence, RI, USA** (S Sun, Prof S McGarvey); **Mindfulness Center, School of Public Health, Brown University, Providence, RI, USA** (S Sun); **Department of Clinical Neuroscience, Karolinska Institutet, Solna, Sweden** (O Simonsson PhD); **Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA** (J Torous MD); **Department of Counseling Psychology, School of Education, University of Wisconsin–Madison, Madison, WI, USA** (S B Goldberg PhD); **Center for Healthy Minds, University of Wisconsin–Madison, Madison, WI, USA** (S B Goldberg)

Correspondence to:
Dr Shufang Sun, Department of Behavioral and Social Sciences, School of Public Health, Brown University, Providence, RI 02903, USA
shufang_sun@brown.edu

Introduction

The global burden of chronic diseases, defined as "conditions that last one year or more and require ongoing medical attention",[1] such as diabetes, hypertension, and HIV or AIDS, is high. In 2023, WHO reported that non-communicable diseases were responsible for seven out of ten deaths worldwide.[2] People in low-income and middle-income countries (LMICs) will have the highest risk of dying from chronic diseases in the next decade.[3,4] In response, in 2015 global leaders endorsed the UN's Sustainable Development Goal 3.4 of a reduction by a third in premature mortality by non-communicable diseases by 2030 through prevention and treatment.[5] To achieve this goal, it is imperative to reach the vast patient populations with chronic diseases and promote optimal health.

In parallel, the past decade has had rapid development in the area of mobile health, defined as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices".[6] In particular, mobile phone interventions (eg, smartphone apps and SMSs) have gained global ubiquity[7] and have the potential to reach and engage patients with chronic diseases, provide health education, monitor symptoms, promote a healthy lifestyle, and support behavioural interventions. Research on mobile phone interventions for patients with chronic disease, with rigorous, gold-standard designs (ie, randomised controlled trials [RCTs]), has rapidly expanded. Mobile phone interventions are increasingly integrated into clinical care in the real world, and this trend has been accelerated by the COVID-19 pandemic.[8] Regulators such as the US Food and Drug Administration and similar organisations worldwide are piloting approaches to evaluate these interventions.[9,10] Thus, there is a need to understand the current risks and benefits to such mobile health interventions. Given a paucity of research coming from mobile health companies themselves,[11] our use of RCTs to benchmark the state of the science is timely and relevant to many stakeholders.

There are several key limitations to the current evidence base that impede dissemination of knowledge about digital mobile phone interventions and policy efforts. First, reviews on mobile phone interventions reach inconsistent conclusions on their effects across chronic health conditions,[12–16] making implementation and dissemination efforts challenging. Second, despite a growing number of RCTs, the quality of the available meta-analyses varies greatly. Methodological limitations, such as combining active (eg, another intervention programme) and inactive (eg, usual care without additional intervention) control conditions, infrequent reporting of moderators, and little assessment of publication bias, can produce skewed results and misleading policy implications.

This Review aims to synthesise the extant literature on the efficacy of mobile phone interventions for patients with chronic diseases through umbrella review methods.[17] This approach can provide a clear summary of rigorously conducted meta-analyses of RCTs designed to test mobile phone interventions for chronic diseases, uncovering the degree of evidence certainty across subcategories (ie, differing types of participants, interventions, comparisons, and outcomes [PICOs]). In an umbrella review of mobile phone interventions for mental health, published in 2022, we found no convincing (class 1) evidence for efficacy across PICOs despite a large RCT base (145 RCTs, n=47 940 participants).[18] Mobile phone interventions for chronic diseases is an older field compared with these interventions for mental health, based on the history of funding and the number of digital health startups.[11] However, existing umbrella reviews have focused on specific conditions (eg, text message interventions for type 2 diabetes).[19–21] A comprehensive umbrella review of mobile phone interventions for chronic conditions is

therefore timely and can provide information on what types of mobile phone interventions are effective for which chronic conditions and specific health outcomes. As mobile phone interventions become increasingly popular, consolidating evidence can offer guidance to the public, researchers, clinicians, and policy makers on the utility of these interventions.

Methods

Protocol and registration

This umbrella review was done in accordance with established guidelines, methods, and practices for conducting umbrella reviews in the medical field.[17] The study was preregistered through the Open Science Framework (https://osf.io/s2t67/). There were three deviations from the preregistration: (1) we did not evaluate attrition due to the scarcity of reporting in meta-analyses; (2) we applied an established umbrella review methodology[17] to evaluate strength of evidence; and (3) meta-analyses that reported effect sizes without restriction to a single chronic condition were excluded (eg, effects on blood pressure from RCTs that included patients with either hypertension or diabetes).

Search strategy and selection criteria

The following search terms were used: ("meta-analy*") AND ("smartphone*" OR "smart phone" OR "mobile phone" OR "cellular phone" OR "cell phone" OR "mobile app*" OR "mobile device" OR "mobile-based" OR "mobile health" OR "mhealth" OR "m-health" OR "iphone" OR "android" OR "tablet"). Six databases were searched: PubMed, CINAHL, PsycINFO, Scopus, Web of Science, and Cochrane Systematic Reviews. We searched databases from inception to June 13, 2022. Only studies reported in English were considered.

Studies were eligible for review if they conducted a meta-analysis, reported meta-analytical outcomes (ie, effect sizes) for mobile phone interventions, focused on populations with chronic diseases (eg, type 1 and type 2 diabetes, hypertension, heart disease, and HIV or AIDS), reported effect sizes derived from k≥4 RCTs, and included comparison conditions that could be categorised as usual care or other specific treatment (eg, usual care plus another behavioural health programme). Usual care conditions included standard usual care (eg, routine medical visits) or augmented usual care (eg, patients were provided a paper handout on medication adherence or received advice to stay active) without mobile phone intervention or another behavioural health programme. Effect sizes needed to be presented with their 95% CIs and to be based on RCTs that did not combine control condition types (eg, a mix of usual care and usual care plus other behavioural health programme conditions), since they address different scientific questions (ie, effects of mobile phone interventions *vs* standard clinical care or effects of mobile phone interventions *vs* another active intervention on top of usual care). If a meta-analysis included more than one condition (eg, type 1 and type 2 diabetes, diabetes, and hypertension), they needed to report effect sizes specific to one type of chronic condition to be included.

Three scientists with PhDs and expertise in conducting systematic reviews (SS, OS, and SBG) independently reviewed abstracts and full texts in duplicate. Disagreements were discussed until a consensus was reached. These authors also extracted the data.

Data analysis

We recorded multiple data items. First, eligible effect sizes and their 95% CIs were extracted, along with the corresponding number of RCTs and participants each effect size represented, heterogeneity (I^2), and results of tests of publication bias. Second, we extracted the results of moderator tests for eligible effect sizes. Third, to summarise findings across PICO subcategories, we coded sample population (eg, older adults or Chinese individuals), clinical condition (eg, type 2 diabetes or hypertension), intervention (eg, mobile apps or SMS), comparison condition (eg, treatment as usual or another active intervention), and outcome (eg, glycated haemoglobin [HbA_{1c}] concentration, medication adherence, or weight).

We evaluated the quality of each meta-analysis with the Assessment of Multiple Systematic Reviews 2 (AMSTAR 2),[22] which aims to provide an understanding of domains of strengths and weaknesses across meta-analyses. We also coded the risk of bias of the primary RCTs (eg, Cochrane) and reports of adverse events using data reported in the meta-analyses. To describe the primary RCTs, we coded the following items: year of publication, sample size, chronic condition, and country where the trial occurred. If an RCT was done across more than one country, all country locations were included and recorded accordingly.

We organised our reporting of results by chronic disease condition and outcomes within each condition, reviewing effect size magnitude and certainty of the evidence separated by population, intervention, and comparison condition. Within each medical condition, we identified representative effect sizes for unique outcomes based on the largest sample, which in theory would provide the most recent and comprehensive evidence, and the most statistically reliable estimate.[17] For instance, among several effect sizes that estimated the effect of mobile apps for reducing HbA_{1c} among people with type 2 diabetes,[23–26] the one with the larger sample number was selected as a representative effect size.[26]

Meta-analyses reported effect sizes as standardised (eg, Cohen's d) and unstandardised (eg, percentage HbA_{1C} reduction) mean differences. When available, we coded standardised effect sizes. Odds ratios were converted to Cohen's d for ease of comparison. For each effect size, we calculated an exact p value with 95% CIs.[27] We interpreted the magnitude of standardised effect sizes (Cohen's d) and heterogeneity (I^2) using established guidelines.[28,29]

We applied a previously proposed umbrella review methodology to evaluate the strength of evidence.[17] Evidence grade was determined for each representative effect size based on the associated sample size, p value, heterogeneity, and presence of publication bias (panel 1). For instance, class 1 or convincing evidence requires a sample of 1000 participants or more, $p<10^{-6}$, $I^2 \leq 50\%$, and no publication bias. Class 2 or highly suggestive evidence requires a sample of 1000 participants or more and $p<10^{-6}$ but other class 2 criteria (ie, heterogeneity or publication bias) are not met. To characterise class 2 effect sizes that did not test for publication bias (which can occur when insufficient studies are available for an adequately powered test[30]), as shown in panel 1, effect sizes without an evaluation of publication bias yet meeting all other criteria for class 1 were categorised as class 2+ or highly suggestive+.

As additional information on evidence certainty, we reported meta-analysts' evaluation of evidence via the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach when it was used. We did not conduct additional GRADE ratings on extracted effect sizes due to the Cochrane guideline on overviews of reviews that suggests the extraction of certainty of evidence from the systematic reviews themselves and little reliability in GRADE ratings when made with meta-analyses.[31]

***Panel 1*: Classification of evidence in meta-analysis for standardised effect sizes**

Convincing evidence (class 1)
- ≥1000 cases
- $p \leq 10^{-6}$ for random effect models
- Low to moderate between-studies heterogeneity ($I^2 \leq 50\%$)
- 95% CIs excluding the null value
- No evidence of publication bias

Highly suggestive evidence (class 2)
- ≥1000 cases
- $p \leq 10^{-6}$ for random effect models
- Class 1 was not met

Suggestive evidence (class 3)
- ≥1000 cases
- $p \leq 10^{-3}$ for random effect models
- Class 2 was not met

Weak evidence (class 4)
- $p \leq 0{\cdot}05$

Non-significant evidence
- $p > 0{\cdot}05$

If effect sizes met all class 1 criteria but did not test for publication bias, they were marked as highly suggestive+ (class 2+).

Results

Study selection

A total of 6982 citations were retrieved, with 34 meta-analyses reporting eligible effect sizes (appendix p 1). Inter-rater reliability for abstract and full-text review was excellent ($\kappa \geq 0{\cdot}75$). The 34 meta-analyses included data from 235 primary RCTs with 48 957 participants. The appendix (pp 4–24) presents meta-analyses that were reviewed in full text and excluded, and reasons for exclusion. A list of reviewed abstracts and reasons for exclusion for full-text review is available online.

Study characteristics

Characteristics of the included meta-analyses are reported in table 1. Meta-analyses included a mean of 11·03 studies (SD=8·23). Meta-analyses examined the following chronic disease categories: type 2 diabetes (n=11 studies, 32%), HIV or AIDS (n=6 studies, 18%), hypertension (n=5 studies, 15%), type 1 or type 2 diabetes (n=3 studies, 9%), cardiovascular disease (n=2 studies, 6%), asthma (n=2 studies, 6%), type 2 diabetes or hypertension (n=1 study, 3%), heart disease (n=1 study, 3%), chronic obstructive pulmonary disease (COPD; n=1 study, 3%), coronary heart disease (n=1 study, 3%), and osteoporosis (n=1 study, 3%). In terms of population, 31 meta-analyses focused exclusively on adults (91%, including one focused on older adults aged ≥60 years), three included both adult and youth (ie, adolescents and children) or adult samples (9%), and none focused exclusively on youth. All meta-analyses were published between 2014 and 2022. Ratings of AMSTAR 2 are reported in the appendix (pp 25–26). Notably, none of the meta-analyses reported funding sources of individual RCTs, and a minority of meta-analyses assessed the potential effect of risk of bias in individual studies on the results (n=2, 6%), accounted for risk of bias when interpreting findings (n=10, 29%), and provided satisfactory explanations regarding heterogeneity (n=14, 41%).

See **Online** for appendix

The 235 primary studies represented in the 34 meta-analyses had a mean sample size of 209·22 participants (SD 339·03). The appendix (pp 27–37) has the primary studies, their full citation, the location (ie, country) of the RCT, and a matrix on their appearance in each meta-analysis. A primary study appeared across all included meta-analyses a mean of 1·57 times (SD=1·29; median=1, range 1–10). These primary studies were published between 2004 and 2021, with 2016 being the median year. The appendix (p 2) presents the regions and country income of primary RCTs. RCTs were based in Asia (n=120, 51%), Europe (n=42, 18%), North America (n=38, 16%), Africa (n=21, 9%), Oceania (n=12, 5%), and South America (n=5, 2%). The figure presents the country locations of the RCTs. The 235 RCTs reported 243 locations (some RCTs involved more than one country), with most evidence coming from China (n=87 studies, 36%) and the USA (n=34 studies, 14%). A total of 240 country income levels for 235 RCTs were documented (some studies involved countries with varied income levels). More than 82% of the trials

For the **reviewed abstracts and reasons for exclusion** see https://osf.io/s2t67/

	Condition	Intervention	Population	Outcomes	Number of included studies	Risk of bias assessment
Al-Arkee et al (2021)[32]	Cardiovascular disease	Mobile apps	Adults	Medication adherence	6	Cochrane
Alhussein and Hadjileontiadis (2022)[33]	Osteoporosis	Mobile apps	Adults	Disability, pain intensity	9	Cochrane
Aminuddin et al (2021)[34]	Diabetes, type 2	Smartphone interventions	Adults	BMI, DBP, HbA_{1c}, self-efficacy, self-care activities, SBP	22	Cochrane
Arambepola et al (2016)[35]	Diabetes, type 2	Automated SMS	Adults	BMI, HbA_{1c}	13	Cochrane
Cai et al (2020)[36]	Diabetes, type 2	Mobile apps	Adults	BMI, bodyweight, waist circumference	14	Cochrane
Cui et al (2016)[37]	Diabetes, type 2	Mobile apps	Adults	HbA_{1c}	6	Cochrane
Daher et al (2017)[38]	HIV	SMS	adults	ART adherence	4	Cochrane
Deng et al (2017)[23]	Diabetes, type 2	Mobile apps	Adults	DBP, HbA_{1c}, HDL, LDL, total cholesterol, triglycerides, weight, SBP	10	Cochrane
El-Gayar et al (2021)[24]	Diabetes, type 1 or type 2	Mobile apps	Adults	HbA_{1c}	24	Cochrane
Gandhi et al (2017)[39]	Cardiovascular disease	SMS and apps	Adults	Adherence to medical therapy, adherence to pharmacological recommendations, reduction in blood pressure, hospital readmission, smoking cessation	15	Cochrane
Han et al (2020)[40]	Hypertension	Mobile apps	Adults, Chinese	DBP, SBP	18	Cochrane
Hou et al (2016)[25]	Diabetes, type 1 or type 2	Mobile apps	Adults	HbA_{1c}	10	NA
Jong et al (2017)[41]	HIV	Mobile phone reminders	Adults and adolescents	Clinical care attendance	5	NA
Liu et al (2020)[26]	Diabetes, type 2 or hypertension (or both)	Mobile apps	Adults	Bodyweight, BMI, DBP, FBG, HbA_{1c}, HDL, SBP, LDL, total cholesterol, waist circumference	21	Cochrane
Manby et al (2022)[42]	HIV	One-way SMS	Adults in sub-Saharan Africa	ART adherence behaviour	6	Cochrane
Mikulski et al (2022)[43]	Hypertension	Mobile apps	Adults	Medication adherence	8	Cochrane
Miller et al (2017)[44]	Asthma	SMS and apps	Adults and youth	Medication adherence, unscheduled visits	4	Cochrane
Enricho Nkhoma et al (2021)[45]	Diabetes, type 2	Mobile apps with DSMES principles	Adults	BMI, HbA_{1c}	4	Cochrane
Saffari et al (2014)[46]	Diabetes, type 2	SMS	Adults	HbA_{1c}	6	Cochrane
Shah et al (2019)[47]	HIV	SMS	Adults	Medication adherence	4	Cochrane
Shaw et al (2020)[48]	Chronic obstructive pulmonary disease	Mobile apps	Adults	Physical function, quality of life	10	Cochrane
Shen et al (2018)[49]	Diabetes, type 2	Mobile technology	Adults	HbA_{1c}	8	Cochrane
Snoswell et al (2021)[50]	Asthma	Mobile apps	Adults and youth	Quality of life	4	NA
Sua et al (2020)[12]	Heart disease	Mobile phone interventions	Adults	Blood pressure, medication adherence	10	Cochrane
Tam et al (2021)[51]	Hypertension	SMS	Adults	Blood pressure	11	Physiotherapy Evidence Database (PEDro)
Tam et al (2022)[52]	Hypertension	SMS	Older people (aged ≥60 years)	Blood pressure	6	Cochrane
Taylor et al (2019)[53]	HIV	SMS	Adults and adolescents	Appointment adherence, HIV adherence pill count, HIV adherence (self-reported)	33	Cochrane
Verma et al (2021)[53]	Diabetes, type 2	SMS	Adults, Asian	Fasting blood glucose, HbA_{1c}	6	Cochrane
Wang et al (2019)[14]	HIV	SMS	Adults	Medication adherence	12	Cochrane

(Table 1 continues on next page)

	Condition	Intervention	Population	Outcomes	Number of included studies	Risk of bias assessment
(Continued from previous page)						
Wu et al (2017)[54]	Diabetes, type 1 or type 2	Mobile apps	Adults	HbA_{1c}, severe hypoglycaemia	12	Cochrane
Xu and Long (2020)[16]	Hypertension	Mobile apps	Adults	Blood pressure, medication adherence	5	Cochrane
Xu et al (2021)[55]	Coronary heart disease	SMS and apps	adults	BMI, total cholesterol	5	Cochrane
Yang et al (2021)[15]	Diabetes, type 2	WeChat app	adults, Chinese	fasting plasma glucose, 2-h plasma glucose, HbA_{1c}, self-efficacy (in diet, exercise, medication adherence, monitoring blood glucose, and foot care)	38	Cochrane
Zhang et al (2022)[56]	Diabetes, type 2	SMS	adults	HbA_{1c}	6	Cochrane

For meta-analyses that included more than one chronic condition, only those that reported effect sizes corresponding to specific condition were included. NA under the risk of bias column indicates no available risk of bias assessment from the meta-analysis was reported; Cochrane indicates the meta-analysis used Cochrane Risk of Bias tool to assess risk of bias. ART=antiretroviral therapy. DBP=diastolic blood pressure. DSMES=diabetes self-management, education, and support. FBG=fasting plasma glucose. HbA_{1c}=glycated haemoglobin. NA=not available. SBP=systolic blood pressure.

Table 1: **Characteristics of included meta-analyses**

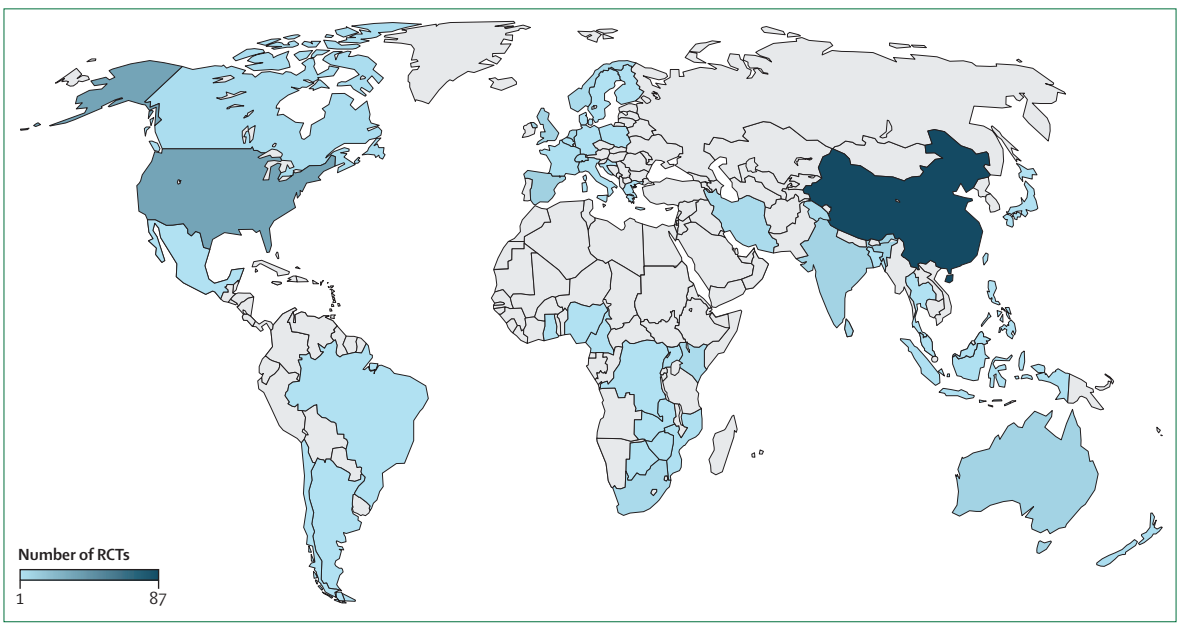

Figure: **Number of primary RCTs mapped by country**
RCT=randomised controlled trial.

occurred in high-income (n=95, 40%) and upper-middle-income countries (n=100, 42%), with only 37 (15%) taking place in lower-middle-income countries and eight (3%) in low-income countries. Interventions lasted a median 6 months (range 0·5–24 months).

Risk of bias within studies

Most meta-analyses (k=30; 88%) used the Cochrane Risk of Bias tool to evaluate RCTs' risk of bias, although one (3%) used the Physiotherapy Evidence Database, and three (9%) did not evaluate the risk of bias (table 1). The appendix (p 3) presents a summary of bias assessment of the primary RCTs from meta-analyses that reported data (ie, 193/235 RCTs; 82%). Incorrect or insufficient masking of personnel and participants (n=125, 53%), incorrect or insufficient masking of study outcome assessment (n=52, 22%), and incomplete outcome data (n=40, 17%) were the three areas with the highest risk of bias. The appendix (pp 38–46) presents available Cochrane Risk of Bias for the primary RCTs.

Risk of bias across studies

Of the 64 representative effect sizes, a total of ten evaluated publication bias (16%) in the

corresponding analysis, in which nine (14%) of them reported no publication bias (eg, from Egger's test).

Representative effect sizes from individual studies

A total of 89 effect sizes were extracted from the 34 eligible meta-analyses (appendix pp 47–50). A total of 64 unique representative effect sizes were identified, including 31 standardised effect sizes and 33 unstandardised effect sizes. Table 2 presents these effect sizes along with corresponding 95% CIs, PICO categories, heterogeneity (I^2), publication bias, and strength of evidence per umbrella review methodology.[17] Additionally, GRADE was reported when it was evaluated in the meta-analysis.

Moderators

We summarise the results from nine moderator tests that were done specific to a chronic disease in table 3. Conditions represented include cardiovascular disease,[39] hypertension,[51,52] and type 2 diabetes.[36] Significant moderators included sample characteristics (eg, age, baseline BMI, baseline HbA_{1c}, or Asian *vs* non-Asian population) and intervention characteristics (eg, study duration, dose, or SMS *vs* non-SMS delivery), although many were not consistently significant across meta-analyses or PICOs. Full details of all moderation tests are presented in the appendix (pp 51–53).

Adverse events

Only two of 34 meta-analyses (6%) reported adverse events. In a meta-analysis focused on type 1 and type 2 diabetes,[54] risk of severe hypoglycaemia from four trials, including three RCTs on type 1 diabetes and one on type 2 diabetes (risk ratio 1·07; 95% CI 0·23–5·09), and overall hypoglycaemia from three trials, including two RCTs on type 1 diabetes and one on type 2 diabetes (1·62; 0·48–5·40), did not differ between mobile phone and usual care conditions. A meta-analysis that focused on type 2 diabetes for Chinese adults[15] noted that incidence of hypoglycaemia after 6 months from five primary studies and incidence of diabetic complications (eg, diabetic neuropathy or diabetic ketoacidosis) in three primary studies was lower in the mobile phone intervention group than usual care group, although an effect size was not reported.

Discussion

This umbrella review examined the effectiveness of mobile phone interventions across populations with chronic diseases. We analysed 34 meta-analyses of 235 RCTs that included a wide range of chronic health conditions (eg, diabetes, cardiovascular conditions, and HIV or AIDS) and represented 48 957 participants in 52 countries. Major findings are summarised in panel 2.

Among 64 representative effect sizes, only one was convincing (class 1), four were highly suggestive+ (class 2+), five were highly suggestive (class 3), four were suggestive (class 4), 23 were weak (class 5), and 27 were non-significant. The convincing effect, requiring a large sample ($n \geq 1000$), low p value ($p \leq 10^{-6}$), no publication bias, and low heterogeneity ($I^2 \leq 50\%$), was the effect of mobile apps in reducing HbA_{1c} among adults with type 2 diabetes. This effect was of moderate magnitude (d –0·44; 95% CI –0·59 to –0·29). The four highly suggestive+ effects met all requirements for the convincing category (class 1), but publication bias was untested. These effects involved apps and SMS, and outcomes on medication adherence (SMS interventions for HIV and AIDS in sub-Saharan Africa[42] and app interventions for cardiovascular disease[16]), reduction in HbA_{1C} (in type 2 diabetes[34]), and blood pressure reduction (systolic blood pressure [SBP] reduction via SMS interventions in older adults[52]). Similarly, the five highly suggestive effects were the effect of SMS and app interventions for HbA_{1c} reduction (in type 2 diabetes[15]), 2 h plasma glucose and fasting glucose (in type 2 diabetes[15]), blood pressure (in hypertension[40]), and therapy adherence outcomes (in HIV or AIDS[39]).[12,32,48] These findings support the conclusion that mobile phone interventions could be especially effective on outcomes that can be facilitated and modified by self-management behaviours, such as medication adherence, glucose management, and blood pressure control.

Almost half (42%) of the effects were non-significant. Although mobile apps might be effective for glucose management for patients with type 2 diabetes,[26,34] effects on BMI[34–36] and bodyweight[36] were non-significant. Similarly, one-way SMS had highly suggestive+ effects for antiretroviral therapy adherence among adults with HIV in sub-Saharan Africa,[42] yet two effect sizes for SMS for medication adherence among people with HIV (without a specific regional focus) were non-significant (despite large sample sizes, n >1000).[13,14] Taylor and colleagues suggested that findings might vary based on measurement type (eg, suggestive effect for self-reported adherence, yet non-significant effect for adherence by pill count),[13] which underscores the need for multimodal assessment in future RCTs. Notably, non-significant results do not necessarily mean no efficacy. Moreover, significant effects in one domain do not necessarily mean effects were reliably larger than another domain. Non-significant results could be related to smaller study sizes (eg, for asthma, COPD, and osteoporosis); this highlights the need for larger clinical trials.

Moderator tests can reveal characteristics of the intervention and participants that affect efficacy and so provide a more clinically relevant understanding of outcomes. Unfortunately, of the 92 effect sizes reviewed, moderators were tested for only nine (10%). There was evidence for variation in effect sizes across levels of some moderators examined. Significant moderators included intervention type (eg, SMS delivery had better outcomes than apps on self-efficacy for adults with type 2 diabetes than usual care[34]) and frequency (SMS interventions that delivered messages >1 per week compared with

	Outcome	Population	Intervention	Comparison group	Number of studies	Number of combined participants	Effect size	95% CI	p value	I^2	Publication bias	Strength of evidence	GRADE
Asthma													
Miller et al (2017)[44]	Medication adherence	Adults and youth	SMS and apps	Paper-based intervention	4	450	0·16	−0·03 to 0·34	>0·05	<0·01	No	Non-significant	..
Snoswell et al (2021)[59]	Quality of life	Adults and youth	Mobile apps	Usual care	4	478	0·30	−0·16 to 0·76	>0·05	36	NA	Non-significant	..
Miller et al (2017)[44]	Unscheduled visits	Adults and youth	SMS and apps	Paper-based intervention	4	443	−0·49	−0·27 to 1·26	>0·05	91	Unclear	Non-significant	..
Cardiovascular disease													
Gandhi et al (2017)[39]	Adherence to pharmacological and non-pharmacological therapies	Adults	SMS and apps	Usual care	15	3937	0·74	0·53 to 0·96	$\leq 10^{-6}$	63	NA	Highly suggestive	..
Gandhi et al (2017)[39]	Hospital readmission	Adults	SMS and apps	Usual care	5	2436	−0·04	−1·55 to 1·48	>0·05	NA	NA	Non-significant	..
Al-Arkee et al (2021)[32]	Medication adherence	Adults	Mobile apps	Usual care	6	820	MD 0·90	0·03 to 1·78	<0·05	93	NA	Weak	..
Gandhi et al (2017)[39]	Medication adherence	Adults	SMS and apps	Usual care	9	1355	0·83	0·48 to 1·18	$\leq 10^{-3}$	61	NA	Suggestive	..
Gandhi et al (2017)[39]	Smoking cessation	Adults	SMS and apps	Usual care	6	3313	0·19	−0·31 to 0·69	>0·05	NA	NA	Non-significant	..
COPD													
Shaw et al (2020)[48]	Physical function	Adults	Mobile apps	Usual care	4	526	MD 8·38	−4·40 to 21·17	>0·05	52	NA	Non-significant	..
Shaw et al (2020)[48]	Quality of life	Adults	Mobile apps	Usual care	8	604	0·40	−0·05 to 0·86	>0·05	83	NA	Non-significant	..
Coronary heart disease													
Xu et al (2021)[55]	BMI	Adults	SMS and apps	Usual care	4	1718	MD −1·71	−2·66 to −0·77	$\leq 10^{-3}$	95	NA	Suggestive	Moderate
Xu et al (2021)[55]	Total cholesterol	Adults	SMS and apps	Usual care	4	979	MD −0·65	−0·88 to −0·42	$\leq 10^{-6}$	95	NA	Weak	Moderate
Diabetes, type 1													
El-Gayar et al (2021)[24]	HbA_{1c}	Adults	Mobile apps	Usual care	6	532	−0·38	−0·63 to −0·12	<0·05	34	NA	Weak	..
Diabetes, type 2													
Yang et al (2021)[15]	2 h plasma glucose	Chinese adults	WeChat apps	Usual care	31	5345	MD −1·91	−2·35 to −1·48	$\leq 10^{-6}$	96	NA	Highly suggestive	..
Zhang et al (2022)[56]	HbA_{1c}	Adults	SMS	Usual care	6	1682	MD −0·09	−0·35 to −0·11	>0·05	65	NA	Non-significant	
Enricho Nkhoma et al (2021)[45]	HbA_{1c}	Adults	Mobile apps with DSMES principles	Usual care	4	555	MD −0·31	−0·48 to −0·15	$\leq 10^{-3}$	0	NA	Weak	..
Verma et al (2021)[53]	HbA_{1c}	Asian adults	SMS	Usual care	5	819	MD −0·58	−1·03 to −0·13	<0·05	84	NA	Weak	..
Yang et al (2021)[15]	HbA_{1c}	Chinese adults	WeChat apps	Usual care	32	5214	MD −1·07	−1·27 to −0·86	$\leq 10^{-6}$	95	NA	Highly suggestive	..
Aminuddin et al (2021)[34]	HbA_{1c}	Adults	Apps and SMS	Usual care	18	1980	MD −0·55	−0·69 to −0·40	$\leq 10^{-6}$	38	NA	Highly suggestive+	..
Liu et al (2020)[26]	HbA_{1c}	Adults	Mobile apps	Usual care	21	1671	−0·44	−0·59 to −0·29	$\leq 10^{-6}$	50	No	Convincing	Low
Aminuddin et al (2021)[34]	BMI	Adults	Apps and SMS	Usual care	9	967	MD −0·23	−0·76 to 0·29	>0·05	7	NA	Non-significant	..
Cai et al (2020)[36]	BMI	Adults	Mobile apps	Usual care	9	1605	MD −0·08	−0·41 to 0·25	>0·05	64	No	Non-significant	..
Enricho Nkhoma et al (2021)[45]	BMI	Adults	Mobile apps with DSMES principles	Usual care	4	554	MD −0·28	−0·55 to −0·02	<0·05	0	NA	Weak	..
Arambepola et al (2016)[35]	BMI	Adults	SMS	Usual care	5	406	MD −0·25	−1·02 to 0·52	>0·05	0	No	Non-significant	..
Cai et al (2020)[36]	Bodyweight	Adults	Mobile apps	Usual care	9	785	MD −0·84	−1·51 to −0·17	<0·05	49	No	Weak	..

(Table 2 continues on next page)

	Outcome	Population	Intervention	Comparison group	Number of studies	Number of participants	Effect size	95% CI	p value	I^2	Publication bias	Strength of evidence	GRADE
(Continued from previous page)													
Aminuddin et al (2021)[34]	DBP	Adults	SMS and apps	Usual care	9	904	MD 0·49	−1·67 to 0·68	>0·05	0	NA	Non-significant	..
Deng et al (2017)[23]	DBP	Adults	Mobile apps	Usual care	6	544	MD 1·48	−3·04 to 0·09	>0·05	0	NA	Non-significant	..
Yang et al (2021)[15]	Diet score	Chinese adults	WeChat apps	Usual care	6	640	MD 1·31	0·86 to 1·77	≤10^{-6}	83	NA	Weak	..
Yang et al (2021)[15]	Exercise score	Chinese adults	WeChat apps	Usual care	6	640	MD 1·92	1·40 to 2·44	≤10^{-6}	76	NA	Weak	..
Verma et al (2021)[53]	Fasting glucose	Asian adults	SMS	Usual care	4	1335	MD −5·84	−17·03 to 5·35	>0·05	60	NA	Non-significant	..
Yang et al (2021)[15]	Fasting glucose	Chinese adults	WeChat apps	Usual care	34	5606	MD −1·36	−1·62 to −1·10	≤10^{-6}	96	NA	Highly suggestive	..
Yang et al (2021)[15]	Foot care score	Chinese adults	WeChat apps	Usual care	6	640	MD 1·71	1·34 to 2·08	≤10^{-6}	68	NA	Weak	..
Deng et al (2017)[23]	HDL	Adults	Mobile apps	Usual care	5	533	0·13	−0·15 to 0·40	>0·05	56	NA	Non-significant	..
Deng et al (2017)[23]	LDL	Adults	Mobile apps	Usual care	5	528	−0·14	−0·42 to 0·14	>0·05	55	NA	Non-significant	..
Yang et al (2021)[15]	Medication adherence	Chinese adults	WeChat apps	Usual care	6	640	MD 1·45	0·97 to 1·94	≤10^{-6}	88	NA	Weak	..
Yang et al (2021)[15]	Monitoring blood glucose	Chinese adults	WeChat apps	Usual care	6	640	MD 1·17	0·51 to 1·83	≤10^{-3}	91	NA	Weak	..
Aminuddin et al (2021)[34]	SBP	Adults	SMS and apps	Usual care	9	904	MD −1·17	−3·17 to 0·83	>0·05	0	NA	Non-significant	..
Deng et al (2017)[23]	SBP	Adults	Mobile apps	Usual care	6	544	MD −2·53	−4·89 to −0·17	<0·05	0	NA	Weak	..
Aminuddin et al (2021)[34]	Self-efficacy	Adults	SMS and apps	Usual care	6	682	0·98	0·42 to 1·55	≤10^{-3}	91	NA	Weak	..
Deng et al (2017)[23]	Total cholesterol	Adults	Mobile apps	Usual care	4	499	−0·14	−0·51 to 0·22	<0·05	73	NA	Non-significant	..
Deng et al (2017)[23]	Triglycerides	Adults	Mobile apps	Usual care	5	530	−0·24	−0·42 to −0·06	<0·05	0	NA	Weak	..
Cai et al (2020)[36]	Waist circumference	Adults	Mobile apps	Usual care	5	618	MD −1·35	−2·16 to −0·55	<0·05	8	No	Weak	..
Heart disease													
Sua et al (2020)[12]	DBP	Adults	SMS and apps	Usual care	8	1417	MD −1·99	−3·20 to −0·78	<0·05	12	NA	Weak	..
Sua et al (2020)[12]	Medication adherence	Adults	SMS and apps	Usual care	4	598	0·72	−0·32 to 1·75	>0·05	97	NA	Non-significant	..
Sua et al (2020)[12]	SBP	Adults	SMS and apps	Usual care	8	1417	MD −1·08	−5·51 to 3·35	>0·05	77	NA	Non-significant	..
HIV or AIDS													
Manby et al (2022)[42]	ART adherence behaviour	Adults in sub-Saharan Africa	One-way SMS	Usual care	10	1833	0·22	0·08 to 0·36	≤10^{-6}	14	NA	Highly suggestive+	..
Jong et al (2017)[41]	Clinical attendance	Adults	SMS	Usual care	5	1135	0·39	0·06 to 0·73	<0·05	73	No	Weak	..
Wang et al (2019)[14]	ART adherence by objective measures	Adults	SMS	Usual care	12	2008	0·15	−0·10 to0·40	>0·05	NA	NA	Non-significant	..
Taylor et al (2019)[13]	ART adherence tracked: MEMS pill count	Adults	SMS	Usual care	10	1782	0·09	−0·03 to 0·20	>0·05	0	NA	Non-significant	Very low
Taylor et al (2019)[13]	Medication adherence: self-report	Adults	SMS	Usual care	5	1037	0·27	0·14 to 0·41	≤10^{-3}	0	NA	Suggestive	Very low
Taylor et al (2019)[13]	Viral suppression	Adults	SMS	Usual care	5	1576	0·04	−0·12 to 0·20	>0·05	22	NA	Non-significant	..

(Table 2 continues on next page)

	Outcome	Population	Intervention	Comparison group	Number of studies	Number of participants	Effect size	95% CI	p value	I^2	Publication bias	Strength of evidence	GRADE
(Continued from previous page)													
Hypertension													
Tam et al (2021)[51]	Blood pressure control	Adults	SMS	Usual care	6	1275	0·43	0·05 to 0·81	<0·05	84	NA	Weak	..
Tam et al (2021)[51]	DBP	Adults	SMS	Usual care	10	3146	0·06	−0·13 to 0·25	>0·05	83	NA	Non-significant	..
Tam et al (2022)[52]	DBP	Older adults (aged ≥60 years)	SMS	Usual care or health education	6	1103	MD −1·47	−4·52 to 1·59	>0·05	85	NA	Non-significant	..
Han et al (2020)[40]	DBP	Chinese adults	Mobile apps	Usual care	18	2965	MD −6·67	−8·92 to −4·41	≤10^{-6}	96	No	Highly suggestive	..
Xu and Long (2020)[16]	Medication adherence	Adults	Mobile apps	Usual care or SMBP	4	1109	0·38	0·26 to 0·50	≤10^{-6}	0	NA	Highly suggestive+	..
Mikulski et al (2022)[43]	Medication adherence measured by MMAS	Adults	Mobile apps	Usual care	4	683	0·65	0·09 to 1·20	<0·05	88	NA	Weak	..
Mikulski et al (2022)[43]	Medication adherence, measured by events of full or global adherence, calculated with counting	Adults	Mobile apps	Usual care	4	1021	0·72	0·17 to 1·27	<0·05	71	NA	Weak	..
Tam et al (2021)[51]	SBP	Adults	SMS	Usual care	11	4518	0·13	0·03 to 0·23	<0·05	58	NA	Weak	..
Xu and Long (2020)[16]	SBP	Adults	Mobile apps	Usual care or SMBP	4	1059	MD −2·31	−5·06 to 0·44	>0·05	62	NA	Non-significant	..
Tam et al (2022)[52]	SBP	Older people (aged ≥60 years)	SMS	Usual care or health education	6	1103	MD −6·11	−8·45 to −3·78	≤10^{-6}	36	NA	Highly suggestive+	..
Han et al (2020)[40]	SBP	Chinese adults	Mobile apps	Usual care	18	2965	MD −8·12	−11·47 to −4·77	≤10^{-3}	97	No	Suggestive	..
Osteoporosis													
Alhussein and Hadjileontiadis (2022)[33]	Disability	Adults	Mobile apps	Usual care	6	688	−0·77	−0·05 to 1·59	>0·05	94	NA	Non-significant	..
Alhussein and Hadjileontiadis (2022)[33]	Pain intensity	Adults	Mobile apps	Usual care	8	906	−1·09	0·45 to 1·68	≤10^{-3}	93	NA	Weak	..

The directions of unstandardised and standardised effect sizes are presented in their original format. The exact number of p values for representative effect sizes can be viewed through the Open Science Framework repository. ART=antiretroviral therapy. DBP=diastolic blood pressure. DSMES=diabetes self-management education and support. GRADE=Grading of Recommendations Assessment, Development and Evaluation. HbA_{1c}=glycated haemoglobin. HDL=high-density lipoprotein. I^2=heterogeneity. LDL=low-density lipoprotein. MEMS=Medication Event Monitoring System. MD=mean difference. MMAS=Morisky Medication Adherence Scale. NA=not available. SBP=systolic blood pressure. SMBP=self-measure blood pressure.

Table 2: **Representative effect sizes across PICO categories**

	Condition	Population	Intervention	Outcome	Moderators tested	Significance	Description (for significant effects)
Gandhi et al (2017)[39]	Cardiovascular disease	Adults	SMS and apps	Medication adherence	Publication language	Not significant	..
Aminuddin et al (2021)[34]	Diabetes, type 2	Adults	SMS and apps	Self-efficacy	Delivery method; study duration; baseline HbA_{1c}	All three moderators were significant	Patients who had SMS delivery, <6 months of study duration, and <8% baseline HbA_{1c} had larger effect sizes on self-efficacy improvement
Cai et al (2020)[36]	Diabetes, type 2	Adults	Mobile apps	BMI	Baseline BMI; Asian *vs* non-Asian population; app functionalities (exercise recording, diet recording, weight recording, and glucose recording); sample age; glycaemic control; proportion of male to female participants; diabetes duration; intervention duration	Age was a significant moderator; all other moderators were not significant	Age was associated with BMI changes (p=0·03, populations with older age had a bigger BMI reduction)
Cai et al (2020)[36]	Diabetes, type 2	Adults	Mobile apps	Bodyweight	Baseline BMI; Asian *vs* non-Asian population; app functionalities (exercise recording, diet recording, weight recording, glucose recording); sample age; glycaemic control; proportion of male to female participants; diabetes duration; intervention duration	Baseline BMI and ethnicity were significant moderators; all others were not significant	Patients with obesity (BMI >30, compared with those BMI ≤30, p=0·001) and non-Asian population (compared with Asian, p=0·001) had higher bodyweight reduction
Cai et al (2020)[36]	Diabetes, type 2	Adults	Mobile apps	Waist circumference	Baseline BMI; Asian *vs* non-Asian population; app functionalities (exercise recording, diet recording, weight recording, glucose recording); sample age; glycaemic control; proportion of male to female participants; diabetes duration; intervention duration	None were significant	..
Tam et al (2021)[51]	Hypertension	Adults	SMS	DBP reduction	Trial duration (≤6 months or ≥7 months); SMS intervention characteristics, including directionality (one-way or two-way), frequency (>1 per week or ≤1 per week), and with or without health education content	Frequency of SMS was a significant moderator; all others were not significant	SMS interventions that had frequency of >1 per week were more effective in DBP reduction compared with those with a frequency of ≤1 per week (subgroup difference p=0·01)
Tam et al (2021)[51]	Hypertension	Adults	SMS	SBP reduction	Trial duration (≤6 months or ≥7 months); SMS intervention characteristics, including directionality (one-way or two-way), frequency (>1 per week or ≤1 per week), and with or without health education content	Frequency of SMS was a significant moderator; all others were not significant	SMS interventions that had frequency of >1 per week were more effective in SBP reduction compared with those with a frequency of ≤1 per week (subgroup difference p=0·02)
Tam et al (2022)[52]	Hypertension	Older adults	SMS	DBP reduction	Trial duration (3 months or 6 months); SMS intervention characteristics, including frequency (>1 per week or ≤1 per week), and with or without health education content	None were significant	..
Tam et al (2022)[52]	Hypertension	Older adults	SMS	SBP reduction	Trial duration (3 months or 6 months); SMS intervention characteristics, including frequency (>1 per week or ≤1 per week), and with or without health education content	None were significant	..

DBP=diastolic blood pressure. HbA_{1c}=glycated haemoglobin. SBP=systolic blood pressure.

Table 3: **Summary of moderator tests in included meta-analyses**

≤1 per week were more effective in diastolic blood pressure [DBP] and SBP reduction than usual care).[51] Some sample characteristics also showed moderating effects (eg, age, ethnicity, and baseline biomarkers).[36] Mobile phone interventions probably do not work equally for everyone even within the same PICO. These approaches have the advantage that they can, in theory, be customised to better support the unique needs of each person at scale. Testing moderators in meta-analyses (especially in individual patient data meta-analyses),[57] conducting pragmatic clinical trials, and engaging in dissemination and implementation-focused studies could all help optimise delivery.

Close to half of the evidence came from the literature focused on type 2 diabetes (29 of 64 representative effect sizes; 45%), reflecting the uneven level of research maturity across chronic diseases. The effect sizes for type 2 diabetes outcomes are similar to previous umbrella reviews,[20,21] yet the representative effect sizes and heterogeneity tended to be smaller in our review, with a larger number of studies available in the current analysed literature. Given that previous summaries of the area of mobile phone health interventions focused on diabetes-related conditions, this umbrella review provides key information on the utility of mobile health interventions across conditions, medical outcomes, and populations.

Results also highlight an uneven distribution of study settings. The majority of primary RCTs were based in high-income (40%) and upper-middle-income (42%) countries, with fewer than 20% conducted in LMICs. More than 85% of the evidence comes from Asia, Europe, and North America, with little research in the global south. As low-income countries have been experiencing an epidemiological transition from communicable diseases to non-communicable diseases,[4] these findings represent an important missed opportunity to address the global burden of chronic diseases.[4,58] Epidemiological data suggest there will be drastic increases in disability and premature death due to chronic diseases by 2040 in LMICs and the global south (eg, Ethiopia, Bangladesh, Myanmar, and Brazil), regions that also lack resources and preparedness.[4] Investment in research for these regions, infrastructure (eg, satellite connectivity, free internet hotspots, and low-cost mobile phones), and education (eg, digital literacy training) will support the success of mobile phone interventions in areas where the effect could be greatest.

As mobile phone interventions become increasingly used in routine medical care, evaluating safety is essential. A review of safety concerns with consumer-facing health apps also suggests the need for rigorous and standardised reporting of adverse events.[59] However, only two meta-analyses in our review evaluated adverse events, and both focused on diabetes.[15,54] Compared with usual care, neither meta-analysis found increased risk for adverse events with mobile phone interventions, and Yang and colleagues provided a narrative summary showing that there was a lower incidence of medical complications for patients in mobile app interventions,[15] indicating the potential of mobile phone interventions in preventing some negative outcomes. However, the scarcity of data on adverse events makes it challenging to assess the clinical risks and benefits of mobile phone interventions.

There are several limitations to the current review and gaps in the meta-analytical and RCT literature (panel 2). First, as an umbrella review, our evaluation of evidence is limited by the available meta-analyses. Although we evaluated 34 meta-analyses, there is an uneven maturity in evidence across chronic diseases. For instance, results for several conditions that had few participants included in the primary RCTs (eg, osteoporosis and coronary heart disease) were graded as weak. The effects of interventions should be further evaluated when larger clinical trials and updated meta-analyses are available. Second, to avoid combining comparison conditions, we excluded effect sizes that mixed control types (ie, combining active and inactive control condition types). Future meta-analyses should carefully review and categorise comparison conditions to support a clear understanding of the efficacy of mobile phone interventions in the context of specific comparisons.[60] Third, the scarce testing for publication bias could have caused an underestimation of evidence certainty for some effects. Although it is suggested that publication bias should be routinely done in meta-analyses, issues such as small sample size and substantial heterogeneity can negatively affect the power to detect publication bias.[61,62] Fourth, we used the Fusar-Poli and Radua[17] method for this umbrella review, which does not consider risk of bias and "optimal size information

***Panel 2*: Key gaps identified in the umbrella review and potential solutions**

Meta-analytical level

- Mix of control conditions (eg, usual care *vs* other active interventions) in analysis: do separate analyses to calculate aggregated effect sizes for RCTs that compared mHealth to usual care or augmented usual care *vs* those that compared mHealth to other active interventions (eg, other mobile phone programmes or health education interventions); use a typology system[60] to code the strength of control condition in large meta-analysis with variations of primary RCTs.
- Scarcity of publication bias testing: test and report publication bias (eg, Egger's test) in all meta-analyses.
- Little testing for moderators: explore moderators in future meta-analyses with enough studies, including sample demographics (age, gender or sex, race and ethnicity, education, etc); sample disease characteristics at baseline (eg, duration of disease and relevant baseline health characteristics such as weight); study characteristics such as region (eg, LMIC *vs* non-LMIC settings) and study duration; intervention characteristics including use of theory, features such as motivation, self-monitoring, goal setting, and dosing or frequency.
- Little evaluation on AEs: code RCT studies' reporting of AEs (eg, whether AEs were evaluated and the ratio it was reported to not reported) and include AEs in systematic reviews and meta-analyses.

Primary RCT study level

- Little evidence in low-income and lower-middle-income countries and the global south: allocate research funds in these regions; foster research networks related to mHealth for addressing the burden of global chronic disease in LMICs and the global south; establish task forces or commissions to accomplish research goals in this area.
- Scarcity of research focused on youth: encourage RCT research on mHealth for youth affected by chronic disease (eg, asthma, diabetes type 1, and epilepsy).
- Little research focused on several chronic conditions (eg, cancer, stroke, chronic lung disease, Alzheimer's disease, and chronic kidney disease): encourage research development on mHealth for patients with these conditions and understand their efficacy in relevant outcomes (eg, health outcomes, self-management, and quality of life).
- Little assessment of long-term effect of mobile phone interventions: design RCT studies with follow-up assessments that go beyond post-intervention timepoint and include long-term assessments (eg, a year and beyond).
- Need to increase RCT quality and reduce risk of bias: report study protocols on allocation concealment and masking of outcome assessment; masking of personnel might be possible when compared with other active interventions; preregister RCTs and all outcomes; use intent-to-treat analysis.
- Little reporting on AEs: gather data on AEs during and after the RCT and describe incidences of them, including those associated with the intervention or app and phone usage; incidence of medical complications and other adverse medical outcomes should also be recorded and reported to evaluate the safety of mobile phone interventions and their potential in preventing complications.

AE=adverse event. LMIC=low-income and middle-income country. mHealth=mobile health. RCT=randomised controlled trial.

criterion" when evaluating the strength of evidence from meta-analyses.[63] We could not convert mean differences to standardised mean differences in instances in which SDs were not reported. Although mean differences might have more clinical relevance (eg, percentage change in HbA_{1c}), the reporting of only mean differences introduces challenges in interpreting the magnitude of effects. Fifth, the search strategy could have been more exhaustive. For instance, we did not use Medical Subject Headings terms. In addition, as we used "meta-analy*" as one of the terms to identify studies, reviews that conducted meta-analyses but did not report this term in the title or abstract could have been missed. Only meta-analyses available in English were included, which might have not captured the full literature. Sixth, few meta-analyses tested moderators. Future meta-analyses will ideally examine various aspects of the sample and intervention characteristics, including the use of theory in intervention development as moderators. Relatedly, as mobile phone interventions can involve multiple components (eg, in-person counselling, phone calls, and supplemental materials), it is possible that these features affect results yet were not captured in moderator tests. In addition, little reporting on adverse events limits understanding of the safety of mobile phone interventions. Various gaps also exist on the primary RCT level, including (as noted previously) little evidence in LMICs and the global south, research focused on youth, and other prevalent chronic diseases (eg, cancer, stroke, or chronic kidney disease). There is also a scarcity of knowledge on the sustainability of mobile phone interventions' effects, which warrants RCTs assessing outcomes in long-term follow-ups (eg, 1 year or longer).

For the **Open Science Framework repository** see https://osf.io/s2t67/

This umbrella review included ten chronic conditions and a wide range of physical (eg, blood pressure, weight, and HbA_{1c}), behavioural (eg, medication adherence), and psychological (eg, quality of life) health outcomes. Strengths of this approach include capturing a broad range of conditions and health outcomes, categorisation of the strength of the evidence, and assessment of moderators. In summary, current evidence suggests that mobile phone interventions could support various health outcomes amenable to self-management (eg, medication adherence). The magnitude of effects tends to be moderate compared with usual care alone. In real-world clinical care, the add-on effect of mobile phone interventions in combination with routine medical care could make clinically relevant differences for patients. This possibility presents a promising view of mobile phone interventions in the future of medical care for patients with chronic diseases.

Contributors

SS conceptualised the study, prepared study protocol, did the coding, conducted data analysis, wrote the initial draft, and reviewed and edited the manuscript. OS did the coding and reviewed and edited the manuscript. SM reviewed and edited the manuscript. JT conceptualised the study and reviewed and edited the manuscript. SBG conceptualised the study, did the initial search, did the coding, managed the repository of data and preregistration, and reviewed and edited the manuscript. SS, OS, and SBG accessed and verified the data. All authors had full access to all the data in the study and are collectively responsible for the decision to submit for publication.

Declaration of interests

SS has received research grants from the US National Institutes of Health and the Mind & Life Institute. OS has received research grants from Ekhaga Foundation and Olle Engkvist Foundation, has once received a payment from Mindfully Sweden for educational content, and was a cofounder of Eudelics AB. STM has received research grants from the National Institutes of Health. JT is an unpaid scientific adviser for Precision Mental Wellness with stock options. SBG has received research grants from the US National Center for Complementary and Integrative Health, the Hope for Depression Research Foundation, the Brain and Behavior Research Foundation, the Defense Advanced Research Projects Agency, the Center for Healthy Minds, and the University of Wisconsin–Madison; payments for reviewing grants from the National Institutes of Health and the Patient-Centered Outcomes Research Institute; and payments for delivering lectures from Chemnitz University of Technology and Veterans Affairs Canada.

Acknowledgments

Efforts on this research by authors are supported by National Institutes of Health grants R24AT012845 (to SS and SBG), K23AT011173 (to SS), and K23AT010879 (to SBG). Funders had no role in the study design, data collection, data coding and analysis, data interpretation, writing of the report, or decision to publish. Authors are solely responsible for the content of this Review and the decision for submission. The paper does not represent the official views of the National Institutes of Health. We would like to thank Srideepti Marada, Jane Vaillant, Claire Pisani, and Mamaswatsi Kopeka for their work in coding primary RCT studies. Coding of all abstracts, full texts, and extracted data of meta-analyses are available on the depository through Open Science Framework (https://osf.io/s2t67/).

References

1 CDC. About chronic diseases. July 21, 2022. https://www.cdc.gov/chronic-disease/about/?CDC_AAref_Val=https://www.cdc.gov/chronicdisease/about/index.htm (accessed Sept 9, 2022).
2 WHO. Noncommunicable diseases. Sept 16, 2023. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases (accessed March 15, 2024).
3 NCD Countdown 2030 collaborators. NCD Countdown 2030: efficient pathways and strategic investments to accelerate progress towards the Sustainable Development Goal target 3.4 in low-income and middle-income countries. *Lancet* 2022; **399:** 1266–78.
4 Bollyky TJ, Templin T, Cohen M, Dieleman JL. Lower-income countries that face the most rapid shift in noncommunicable disease burden are also the least prepared. *Health Aff* 2017; **36:** 1866–75.
5 NCD Countdown 2030 collaborators. NCD Countdown 2030: pathways to achieving Sustainable Development Goal target 3.4. *Lancet* 2020; **396:** 918–34.
6 WHO Global Observatory for eHealth. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. 2011. https://iris.who.int/handle/10665/44607 (accessed Sept 9, 2022).
7 Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. *J Biomed Inform* 2012; **45:** 184–98.
8 Al Knawy B, McKillop MM, Abduljawad J, et al. Successfully implementing digital health to ensure future global health security during pandemics: a consensus statement. *JAMA Netw Open* 2022; **5:** e220214.
9 Alon N, Stern AD, Torous J. Assessing the food and drug administration's risk-based framework for software precertification with top health apps in the United States: quality improvement study. *JMIR Mhealth Uhealth* 2020; **8:** e20482.

10 Stern AD, Brönneke J, Debatin JF, et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. *Lancet Digit Health* 2022; **4:** e200–06.

11 Day S, Shah V, Kaganoff S, Powelson S, Mathews SC. Assessing the clinical robustness of digital health startups: cross-sectional observational analysis. *J Med Internet Res* 2022; **24:** e37677.

12 Sua YS, Jiang Y, Thompson DR, Wang W. Effectiveness of mobile phone-based self-management interventions for medication adherence and change in blood pressure in patients with coronary heart disease: a systematic review and meta-analysis. *Eur J Cardiovasc Nurs* 2020; **19:** 192–200.

13 Taylor D, Lunny C, Lolić P, et al. Effectiveness of text messaging interventions on prevention, detection, treatment, and knowledge outcomes for sexually transmitted infections (STIs)/HIV: a systematic review and meta-analysis. *Syst Rev* 2019; **8:** 12.

14 Wang Z, Zhu Y, Cui L, Qu B. Electronic health interventions to improve adherence to antiretroviral therapy in people living with HIV: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2019; **7:** e14404.

15 Yang J, Yang H, Wang Z, et al. Self-management among type 2 diabetes patients via the WeChat application: a systematic review and meta-analysis. *J Clin Pharm Ther* 2021; **46:** 4–16.

16 Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2020; **8:** e21759.

17 Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. *Evid Based Ment Health* 2018; **21:** 95–100.

18 Goldberg SB, Lam SU, Simonsson O, Torous J, Sun S. Mobile phone-based interventions for mental health: a systematic meta-review of 14 meta-analyses of randomized controlled trials. *PLOS Digit Health* 2022; **1:** e0000002.

19 Timpel P, Oswald S, Schwarz PEH, Harst L. Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: an umbrella review of systematic reviews and meta-analyses. *J Med Internet Res* 2020; **22:** e16791.

20 Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. *PLoS One* 2017; **12:** e0173160.

21 Whittemore R, Siverly L, Wischik DL, Whitehouse CR. An umbrella review of text message programs for adults with type 2 diabetes. *Diabetes Educ* 2020; **46:** 514–26.

22 Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017; **358:** j4008.

23 Deng N, He H, Wang Y, Zhang X. Effects of mobile phone management applications for biological and biochemical parameters in patients with type 2 diabetes mellitus: a meta-analysis. *Chin J Evid-Based Med* 2017; **17:** 1167–73 (in Chinese).

24 El-Gayar O, Ofori M, Nawar N. On the efficacy of behavior change techniques in mHealth for self-management of diabetes: a meta-analysis. *J Biomed Inform* 2021; **119:** 103839.

25 Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA_{1c}) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. *Diabetes Care* 2016; **39:** 2089–95.

26 Liu K, Xie Z, Or CK. Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials. *JMIR Mhealth Uhealth* 2020; **8:** e15779.

27 Altman DG, Bland JM. How to obtain the p value from a confidence interval. *BMJ* 2011; **343:** d2304.

28 Cohen J. Statistical power analysis for the behavioral sciences, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers, 1988.

29 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327:** 557–60.

30 Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. *Psychol Methods* 2004; **9:** 426–45.

31 Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Syst Rev* 2014; **3:** 82.

32 Al-Arkee S, Mason J, Lane DA, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. *J Med Internet Res* 2021; **23:** e24190.

33 Alhussein G, Hadjileontiadis L. Digital health technologies for long-term self-management of osteoporosis: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2022; **10:** e32557.

34 Aminuddin HB, Jiao N, Jiang Y, Hong J, Wang W. Effectiveness of smartphone-based self-management interventions on self-efficacy, self-care activities, health-related quality of life and clinical outcomes in patients with type 2 diabetes: a systematic review and meta-analysis. *Int J Nurs Stud* 2021; **116:** 103286.

35 Arambepola C, Ricci-Cabello I, Manikavasagam P, Roberts N, French DP, Farmer A. The impact of automated brief messages promoting lifestyle changes delivered via mobile devices to people with type 2 diabetes: a systematic literature review and meta-analysis of controlled trials. *J Med Internet Res* 2016; **18:** e86.

36 Cai X, Qiu S, Luo D, Wang L, Lu Y, Li M. Mobile application interventions and weight loss in type 2 diabetes: a meta-analysis. *Obesity* 2020; **28:** 502–09.

37 Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. *PLoS One* 2016; **11:** e0166718.

38 Daher J, Vijh R, Linthwaite B, et al. Do digital innovations for HIV and sexually transmitted infections work? Results from a systematic review (1996–2017). *BMJ Open* 2017; **7:** e017604.

39 Gandhi S, Chen S, Hong L, et al. Effect of mobile health interventions on the secondary prevention of cardiovascular disease: systematic review and meta-analysis. *Can J Cardiol* 2017; **33:** 219–31.

40 Han H, Guo W, Lu Y, Wang M. Effect of mobile applications on blood pressure control and their development in China: a systematic review and meta-analysis. *Public Health* 2020; **185:** 356–63.

41 Jong S, Cuca Y, Thompson LM. Meta-analysis of mobile phone reminders on HIV patients' retention to care. *J Mob Technol Med* 2017; **6:** 5–18.

42 Manby L, Aicken C, Delgrange M, Bailey JV. Effectiveness of eHealth interventions for HIV prevention and management in sub-Saharan Africa: systematic review and meta-analyses. *AIDS Behav* 2022; **26:** 457–69.

43 Mikulski BS, Bellei EA, Biduski D, De Marchi ACB. Mobile health applications and medication adherence of patients with hypertension: a systematic review and meta-analysis. *Am J Prev Med* 2022; **62**: 626–34.

44 Miller L, Schüz B, Walters J, Walters EH. Mobile technology interventions for asthma self-management: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2017; **5:** e57.

45 Enricho Nkhoma D, Jenya Soko C, Joseph Banda K, Greenfield D, Li YJ, Iqbal U. Impact of DSMES app interventions on medication adherence in type 2 diabetes mellitus: systematic review and meta-analysis. *BMJ Health Care Inform* 2021; **28:** 28.

46 Saffari M, Ghanizadeh G, Koenig HG. Health education via mobile text messaging for glycemic control in adults with type 2 diabetes: a systematic review and meta-analysis. *Prim Care Diabetes* 2014; **8:** 275–85.

47 Shah R, Watson J, Free C. A systematic review and meta-analysis in the effectiveness of mobile phone interventions used to improve adherence to antiretroviral therapy in HIV infection. *BMC Public Health* 2019; **19:** 915.

48 Shaw G, Whelan ME, Armitage LC, Roberts N, Farmer AJ. Are COPD self-management mobile applications effective? A systematic review and meta-analysis. *NPJ Prim Care Respir Med* 2020; **30:** 11.

49 Shen Y, Wang F, Zhang X, et al. Effectiveness of internet-based interventions on glycemic control in patients with type 2 diabetes: meta-analysis of randomized controlled trials. *J Med Internet Res* 2018; **20:** e172.

50 Snoswell C, Rahja M, Lalor A. A systematic review and meta-analysis of change in health-related quality of life for interactive telehealth interventions for patients with asthma. *Value Health* 2021; **24:** 291–302.

51 Tam HL, Wong EML, Cheung K, Chung SF. Effectiveness of text messaging interventions on blood pressure control among patients with hypertension: systematic review of randomized controlled trials. *JMIR Mhealth Uhealth* 2021; **9:** e24527.

52 Tam HL, Leung LYL, Wong EML, Cheung K, Chan ASW. Integration of text messaging interventions into hypertension management among older adults: a systematic review and meta-analysis. *Worldviews Evid Based Nurs* 2022; **19:** 16–27.

53 Verma D, Bahurupi Y, Kant R, Singh M, Aggarwal P, Saxena V. Effect of mHealth interventions on glycemic control and HbA_{1c} improvement among type II diabetes patients in Asian population: a systematic review and meta-analysis. *Indian J Endocrinol Metab* 2021; **25:** 484–92.

54 Wu Y, Yao X, Vespasiani G, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. *JMIR Mhealth Uhealth* 2017; **5:** e35.

55 Xu Y, Ye H, Zhu Y, Du S, Xu G, Wang Q. The efficacy of mobile health in alleviating risk factors related to the occurrence and development of coronary heart disease: a systematic review and meta-analysis. *Clin Cardiol* 2021; **44:** 609–19.

56 Zhang A, Wang J, Wan X, et al. A meta-analysis of the effectiveness of telemedicine in glycemic management among patients with type 2 diabetes in primary care. *Int J Environ Res Public Health* 2022; **19:** 4173.

57 Oxman AD, Clarke MJ, Stewart LA. From science to practice. Meta-analyses using individual patient data are needed. *JAMA* 1995; **274:** 845–46.

58 Hajat C, Stein E. The global burden of multiple chronic conditions: a narrative review. *Prev Med Rep* 2018; **12:** 284–93.

59 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; **27:** 330–40.

60 Goldberg SB, Sun S, Carlbring P, Torous J. Selecting and describing control conditions in mobile health randomized controlled trials: a proposed typology. *NPJ Digit Med* 2023; **6:** 181.

61 Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000; **53:** 1119–29.

62 Shi L, Lin L. The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. *Medicine (Baltimore)* 2019; **98:** e15987.

63 Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. *J Clin Epidemiol* 2011; **64:** 1283–93.