Methods for Evaluating the Efficacy and Effectiveness of Direct-to-Consumer Mobile Health Apps: A Scoping Review

Carley Prentice (✉ c_prentice@flo.health)

Flo Health UK Limited

Kimberly Peven

Flo Health UK Limited

Liudmila Zhaunova

Flo Health UK Limited

Vaaruni Nayak

Flo Health UK Limited

Tara Radovic

Flo Health UK Limited

Anna Klepchukova

Flo Health UK Limited

Henry W. W. Potts

Flo Health UK Limited

Sonia Ponzo

Flo Health UK Limited

Research Article

Abstract

Background

There is an evaluation crisis in mobile health (mHealth). The majority of mHealth apps are released with little evidence base. While many agree on the need for comprehensive evaluations to assess the efficacy and effectiveness of mHealth apps, the field is some way from achieving that. This scoping review describes the current state of direct-to-consumer mHealth app evaluations so as to inform how the situation can be improved.

Results

Findings showed a predominance of wellness management apps, focusing on fitness, diet, mental health, or other lifestyle factors. Evaluations were conducted by companies at varied financing stages, with a mix of start-ups, scale-ups, and public companies. Most studies employed full-scale or pilot randomised controlled trial designs.

Conclusions

Participant demographics indicate a need for more inclusive recruitment strategies around ethnicity and gender so as to avoid worsening health inequalities. Measurement tools varied widely, highlighting the lack of standardisation in assessing mHealth apps. Promoting evidence-based practices in digital health should be a priority for organisations in this space.

Background

Mobile health (mHealth) apps have transformed healthcare delivery, with over 350,000 direct-to-consumer (DTC) apps available worldwide [1]. These apps provide individuals with what promises to be convenient, accessible, and personalised health tools. Broadly, they can be classified into two main categories: wellness management and health condition management [1, 2]. Wellness management apps often use behaviour change techniques to track and promote healthy habits related to fitness, diet, sleep, or other lifestyle factors [1–3]. In comparison, health condition management apps are designed to support individuals in self-managing specific, often chronic conditions, such as diabetes, hypertension, or substance addictions [1, 2, 4]. These apps encompass a range of functionalities including tools for self-diagnosis and monitoring, clinical decision support tools for information and guidance from healthcare providers, and therapeutic apps for delivering treatment interventions directly to the user [1, 5, 6]. Thus, individuals looking to turn to technology to manage their health are exposed to a wealth of options.

The mHealth app industry has the potential to enhance population health outcomes and address access disparities [7, 8]. However, there is a concomitant need for rigorous evaluation and evidence-based

approaches. Regulatory bodies and initiatives, such as the US Food and Drug Administration (FDA) Digital Health Center [9, 10], the European Commission eHealth policies [11], and the German Federal Institute for Drugs and Medical Devices (Bundesinstitut für Arzneimittel und Medizinprodukte, BfArM) [12], all play a crucial role in ensuring the safety and real-world effectiveness of mHealth apps and ultimately protecting consumers [13]. While the majority of wellness apps are deemed low risk, health condition management apps are often categorised as medical devices, leading to more stringent regulations and lengthy oversight, which can result in significant delays to the commercialisation of innovative health products [13, 14]. Hence, app developers either find themselves outside of the regulatory landscape, with little guidance on how to conduct evaluations, or under the obligation to embark on an expensive and long regulatory journey [15, 16].

The overall body of clinical evidence on mHealth app effectiveness has been growing, with 1,500 studies published between 2016–2021 [1]. Nevertheless, evidence on DTC health apps is sparse and inconclusive regarding safety and effectiveness [8, 17, 18]. Attempts have been made in recent years by regulatory bodies to streamline the path to commercialisation under their oversight [19]. Further, standardised approaches and reporting guidelines now exist to enhance research efficiency and quality for app developers and scientists evaluating digital health products [20–23]. However, mHealth app research poses challenges such as lengthy study timelines, complex interventions with multiple features, changing app iterations, and a lack of in-house research expertise and funding [8, 16, 20]. Further, resources on the state of the art in the field to guide app developers in making research design choices appropriate for the stage of their product are lacking. A wide variety of study designs exist to evaluate digital health products [21, 24, 25], but it is not always clear which would be optimal in different contexts, with a range of different frameworks available [22]. Additionally, alternative approaches are being considered, like micro-randomised trials, but uptake may be slow [20, 26–28].

The aim of the current scoping review is to summarise research designs and study characteristics used to evaluate and validate commercially available DTC mHealth apps at different company financing stages. Specifically, we focused on evaluations assessing health efficacy or effectiveness in improving health outcomes. By shedding light on the available research methods and evidence base, this scoping review seeks to understand what methods are commonly used, including identifying possible gaps and inconsistencies, thus contributing to the development of a solid framework for evaluating mHealth apps and promoting evidence-based practice and guidance in digital health.

Methods

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) [29] to guide the literature search (Supplementary Table 1).

Search strategy

A structured search was conducted on January 2, 2023 in the databases MEDLINE via PubMed and EMBASE. The search terms "mHealth" and "evaluation methods" were used and enriched with synonyms,

truncations, and Medical Subject Headings (MeSH) (Supplementary Table 2).

Inclusion and exclusion criteria

Only full-text primary research studies published in English and in peer-reviewed journals between January 2017 to January 2023 were included. Studies were included if they evaluated stand-alone mHealth apps targeted at the general public, measured health outcomes, and the evaluated app was downloadable via the Google Play store (with > 50K global downloads). Download cut-off was to ensure relevance to the public. No download cut-off criteria were applied to iOS store downloads as these data are not publically available.

Evaluations lacking a health efficacy or effectiveness (including cost-effectiveness) evaluation component were excluded. mHealth apps that solely send text messages or phone calls as their primary behaviour change modification were excluded. Finally, studies were excluded if the apps were solely a reminder service (including medication, treatment adherence, and appointment), electronic patient portal, or cloud-based personal health record app (Supplementary Table 3).

Data management and selection process

The literature search results were transferred to Zotero reference management software for de-duplication. Three reviewers (CP, KP, VN) pilot-screened the same random sample of 5% of studies at the title and abstract stage to ensure consistency. The remaining studies were randomly distributed between the reviewers. Studies designated a 'maybe' by any reviewer were subsequently checked by all three reviewers collectively.

For full-text screening, 10% of full-text studies were screened by all three reviewers and the results were discussed to ensure consistency and reliability. The remaining studies were randomly distributed between the reviewers. At the full-text screening stage, all identified app titles were searched in the Google Play store to confirm their eligibility.

Data extraction and analysis

A data extraction form was pilot-tested on five randomly selected studies and completed for all included studies (Supplementary Table 4).

The extracted mHealth app characteristics include: app name, number of global Google Play App store downloads, financing stage of company at publication date, current number of employees, founding or launch date of app, mHealth app category (wellness management versus health condition management), mHealth app sub-category, companion device, and regulatory status (when applicable). Global Google Play app store download data was extracted on August 4, 2023 and therefore the number of downloads at the time of study publication (for all included studies) is unknown. Access to historical app download records from Google Play is typically restricted, making it challenging to obtain precise download figures for the study's publication date. Company financing stage at publishing date, current number of

employees, and founding or launch date of app was extracted using Crunchbase Database, a company providing business information about private and public companies (www.crunchbase.com). We were able to determine the company financing stage at study publishing date, however the data related to number of employees at the time of study publication (for all included studies) is unknown.

Apps were dichotomised as wellness management or health condition management. Wellness management apps were further classified into sub-categories based on the health outcome aim of the app. These included diet and nutrition, exercise and fitness, mental health, sleep, children's health, oral hygiene, and skin health. Health condition management apps were further classified into sub-categories based on the aim of the app. These included diagnostics, clinical decision support tools, or therapeutics. Apps with companion medical devices by the same or an alternative company were recorded as well as the FDA regulatory status of the device [10].

The extracted study characteristics include: country study was conducted in, potential or stated conflict of interest from the team that conducted the study (including external research groups with and without conflicts of interest, internal team, and mixed teams; these were determined based on declared conflicts of interest, author affiliations, author contributions, acknowledgements, and study funding), sample target age group, sample target health condition and/or population, study purpose (focus on health outcomes), study design (as reported by authors), sample size (defined as sample allocated to intervention and when applicable, split by intervention and control group), study intervention length (in months), retention rate (defined as % of study participants who started the intervention or control method and remained until the defined end of the study intervention period), type of intervention (behaviour change technique), methods for controls (when applicable), health outcome measurement instruments, inferential statistical techniques (and related health outcome measured), study results, and distribution of study demographics at baseline (mean (sd) of age and frequency (%) of sex and ethnicity).

Results

We found 2799 articles, of which 47 were included in the review (Fig. 1) [30–76].

mHealth app characteristics by financing stage

The financing stage of the app companies at study publication date (Fig. 2) included 16% early stage startups (Pre-seed, Seed and Series A; 6/38), 29% scale-ups (Series B-F; 11/38), 39% acquired or public (15/38), and 16% of apps developed by universities or government groups (6/38). Companies where financing data was not found (9/47) were excluded from this categorisation.

We found a significant association ($p < 0.001$) between the financing stage and the team conducting the study (Fisher's exact test; Supplementary Table 5), with early-stage startup (n = 6) and scale-up (n = 11) research both being conducted by external groups (no declared conflict of interest; n = 3 and n = 5, respectively) and mixed teams (internal employees and external collaborators; n = 3 and n = 6, respectively); acquired or public companies (n = 15) being exclusively researched by external groups; and

university or government-developed app research (n = 6) being conducted mainly by internal teams (n = 3), with some efforts led by external groups (n = 2) and mixed teams (n = 1).

No association was found between study design and financing stage using FIsher's exact test (p = 0.55; Supplementary Table 6). However, early stage startups were more apt to use pilot (n = 2) or full-scale RCTs (n = 2); scale-ups showed a higher inclination towards pre-post studies (n = 5) and full-scale RCTs (n = 4); acquired or public companies employed more diverse study design choices, including pilot (n = 6) and full-scale (n = 3) RCTs, pre-post studies (n = 4), and alternative designs (including a micro-randomised trial and a non-randomised open label controlled trial; n = 2); and university or government-developed app research teams tended to use pre-post studies (n = 3), full-scale RCTs (n = 2), and a 2x2 randomised, mixed factorial design (n = 1).

mHealth app categories

Most apps (94%, 44/47) fell under the category of wellness management. Among the wellness management apps, 32% (14/44) targeted three or more health outcome sub-categories. The most prevalent subcategories included diet and nutrition (55%, 24/44), exercise and fitness (55%, 24/44), mental health (52%, 23/44), and sleep (41%, 18/44; Fig. 3 and Supplementary Table 7). Of the few health condition management apps (6%, 3/47), all were therapeutic tools designed for self-monitoring specific health conditions (including diabetes, hypertension, and allergic rhinitis; Fig. 3).

Nearly one-third of the wellness management apps (32%, 14/44) were associated with companion devices offered by the same or alternative companies. Of these, eleven apps were integrated with smartwatches, two apps were linked to smart scales, and one app was linked to a monitoring system for urine excretion. Only smart scale devices have received FDA 510k clearance. Two of the three health condition management apps were accompanied by companion devices with FDA 510k clearance. One of these devices is a capillary glucose reader catering to individuals with diabetes [65] and the other is a blood pressure monitor targeting individuals with hypertension [66].

Numerous wellness management studies focused on distinct populations to explore various health outcomes. Twenty-seven percent focused on overweight and obese individuals to evaluate dietary and physical activity improvements (12/44), 16% involved cancer patients, examining enhancements in diet, exercise, mental health, and sleep (7/44), and 11% targeted women with conditions such as breast cancer and postpartum depression (5/44). However, the majority of wellness management apps did not have a target condition or disease in mind and were interested in improving wellness health outcomes in the general population (36%, 16/44).

Study characteristics of the included studies

A variety of research designs were employed in the evaluation of all included mHealth apps (Table. 1 and Fig. 3), with the majority (64%, 30/47) being RCTs. Among these, 56% were full-scale RCTs (17/30), characterised by medium-sized sample groups (median 107, range 28-1573), moderate intervention

durations (median 2.5 months, range 0.3–24.0 months), and relatively high retention rates (mean 79.6%, SD 18.5). Pilot RCTs (37%, 11/30) had smaller samples (median 54, range 25–142), longer intervention durations (median 4.5 months, range 1.4–12.0 months), and higher retention rates (mean 86.3%, SD 9.7). Full-scale and pilot RCTs employed many control methods, including standard care, waitlist (delayed access to treatment), partial access to treatment, alternative treatments, or no treatment. Novel RCT approaches constituted a minor portion (7%, 2/30). A micro-randomized trial featured a large sample size of 1565 participants over a six-month study period, using a partial treatment control group, though retention rate was not reported. The mixed factorial (2x2) study involved a smaller sample of 52 participants for a one-week study period, using an alternative treatment control method, and achieving a 100% retention rate.

Pre-post studies accounted for 32% (15/47), split between non-pilot (40%, 6/15) and pilot (60%, 9/15) studies. The non-pilot pre-post studies featured larger sample sizes (median 129, range 61–416) and longer study durations (median 2.76 months, range 0.69-12.0 months), but had lower retention rates (mean 68.3%, SD 22.3). In comparison, the pilot pre-post studies had smaller sample sizes (median 27, range 8–90) and shorter durations (median 1.8 months, range 1.0-2.8 months), and exhibited higher retention rates (mean 84.6%, SD 16.0). The majority of pre-post studies used a before/after single group design (87%, 13/15), and only two used a non-randomised comparative design (with intervention and control groups).

Finally, of the non-randomised open label trials (4%, 2/47) the sample sizes were 19 and 75, study intervention lengths were 1.8 and 2.76 months, and the retention rates were 45.0% and 87.0%.

Table 1
Frequency table of study designs and associated study characteristics.

Study Designs (n = 47)		n (%)	Median (range) sample sizes	Median (range) study intervention length in months	Mean (SD) retention rate in %
RCT (n = 30)	Full-scale (including 2- and 3-arms)	17 (36.2%)	107 (28-1573)	2.5 (0.3–24.0)	79.6% (18.5)
	Pilot RCT	11 (23.4%)	54 (25–142)	4.5 (1.4–12.0)	86.3% (9.7)
	Micro-randomised trial	1 (2.1%)	1565*	6*	Not reported
	Mixed factorial design (2x2)	1 (2.1%)	52*	0.23*	100%*
Pre-post studies (n = 15)	Pre-post	6 (12.8%)	129 (61–416)	2.76 (0.69–12.0)	68.3% (22.3)
	Pilot pre-post	9 (19.1%)	27 (8–90)	1.8 (1.0-2.8)	84.6% (16.0)
Non-randomised open label trials		2 (4.2%)	19 and 75*	1.8 and 2.76*	45% and 87%*

*For samples less than 2, true values are reported due to inapplicable median (range).

Participant demographics

The studies were conducted in 15 countries (Fig. 3, Supplementary Table 8). The majority (62%, 29/47) of the studies were conducted in the USA. Other countries were represented by one or two studies and no global or multi-country studies were found.

The majority of the studies (73%, 34/47) targeted adults aged 18 years and older, 10% focused on children under 18 years of age (5/47), and the remaining studies (17%; 8/47) focused on adults aged 40 years and older (Fig. 3). Eight studies were gender/sex-specific, with five of them exclusively researching female participants in the context of breast cancer, pre- and post-partum depression, and premenstrual syndrome. Conversely, three studies solely included male participants, focusing on esophageal cancer and obesity. The remaining studies exhibited a wide range in the proportion of female and male participants at baseline, varying from 21–95% and 5–78%, respectively (Fig. 4). Overall, 75% (36/47) of studies included a majority of female participants. Notably, only one study reported inclusion of individuals outside of a sex or gender binary [73].

Ethnicity was reported by 58% (27/47) of included studies (Figs. 3 and 5). Sixty-seven percent of these studies reported a majority of White/Caucasian participants (18/27). Two studies conducted in the USA

targeted Hispanic/Latino adults [35, 71], one study conducted in USA researched an underserved community with 95% Black/African participants [66], and one study conducted in Singapore reported all Asian/Asian participants (92% Chinese, 0.6% Malay, 4.5% Indian, 2.9% Other) [72]. Excluding the studies that targeted specific ethnicities, the median (range) representation of all ethnic groups among included studies were: 62% (4%-98%) White/Caucasian, 7% (0%-50%) Black/African descent, 0.4% (0%-17%) Asian/Asian descent, 7% (0%-48%) Hispanic/Latinx, 9% (0%-60%) Biracial/Multiracial, and 0.3% (0%-14%) Indigenous Groups (Fig. 5).

Measurement tools for evaluating mHealth apps

Various measurement tools were used to assess the effectiveness of health outcomes (Supplementary Table 9). Five commonly employed measurement tools were identified: the Short Form Health Survey (SF-12 or SF-35) [77] for measuring health-related quality of life (7/47), the Patient-Reported Outcomes Measurement Information System (PROMIS) [78] for evaluating physical, mental and social health (6/47), the Perceived Stress Scale (PSS) [79] for measuring individual stress levels (6/47), the Five Facet Mindfulness Questionnaire (FFMQ-SF) [80] for assessing the five vital elements of mindfulness (4/47), and the Hospital Anxiety and Depression Scale (HADS) [81] for measuring anxiety and depression among patients in hospital settings (4/47). These measurement tools were employed to evaluate wellness apps and health promotion apps.

Discussion

To our knowledge, this scoping review is the first attempt to summarise the financial stages, study characteristics, and methods for evaluation of popular commercially available mHealth apps. This is particularly important given the varying guidance available to app developers outside of, or in the phases leading up to, regulatory oversight.

We found that most of the studies were conducted by companies at later stages of financing. The finding that scale-ups and acquired/public companies are better positioned to invest in evidence generation is unsurprising. Whilst governmental and international funding schemes exist to support start-ups and small businesses in their evaluation efforts, internal expertise or ability to leverage university collaborations to obtain such funding may be lacking [15]. Further, even though educational resources on the need for iterative evaluations of health products are available, awareness of their implications among app developers is still limited [7, 8, 18]. Working groups and committees including industry, academia and government representatives are needed to build action plans guiding early-stage companies through evaluating their products at cost.

While the majority of wellness apps (with the exception of smart scale companion devices) did not have FDA clearance, a subset of health condition management apps obtained FDA clearance along with associated medical devices. The lack of clear regulatory pathways and guidance for informational apps for health management and tracking explains this finding [10]. Health condition management apps face more stringent regulations and lengthy oversight, and often result in significant delays to

commercialisation [13, 14]. Future research should assess regulatory challenges faced by health condition management apps and provide recommendations on how to implement better strategies for expediting clearance.

The prevalence of full-scale RCTs, even among early-stage startups, is encouraging. Alongside full scale RCTs, pilot RCTs and pre/post studies were also frequent evaluation methods. Even though the reviewed studies often had small sample sizes and short duration, this finding suggests that RCTs may be more feasible for app developers than previously thought. We also observed some novel study designs: a factorial RCT (assess multiple interventions simultaneously) and a micro-randomized trial (involving frequent, small-scale randomizations within individuals' daily lives). Alternative study designs offer promise as substitutes to traditional approaches, potentially allowing app developers to attain robust evidence at reduced costs, shorter timeframes and increased flexibility [20, 28]. However, there was a notable research gap of economic evaluations. Future research should focus on barriers to adoption of novel designs, as well as shed light on the lack of health economics considerations in app evaluations.

The majority of studies in this review, regardless of study design, had study durations of less than 3 months, utilised wait-list controls, and maintained participant retention rates above 65%. Short study durations are limited in understanding sustained health effects and highlight the need for longer follow-up periods. Wait-list controls prevent any long-term follow-up being possible. But finding or developing active controls with comparable characteristics can be a significant challenge for companies, particularly when faced with resource constraints [82]. Nevertheless the wide adoption of wait-list controls may limit the interpretability of the results. Furthermore, participant retention over the course of a study is a critical factor that influences the validity of results, yet limited resources for participant compensation can impact retention. Industry-wide efforts to establish unified strategies for participant recruitment and retention, with potential support from innovative engagement approaches, such as gamification and personalised interventions, could prove instrumental in mitigating attrition rates.

Our findings reveal a distribution of study conductors and highlight the importance of fostering robust academic-industry partnerships [15]. While external research groups may provide more impartial assessments, the absence of an internal team may limit the integration of research findings into the company's development process. A mixed approach, involving both internal teams and external collaborators, can strike a balance between impartial evaluation and in-house expertise, promoting a culture of evidence-based practice within companies. Future studies should investigate effective models for fostering collaboration between internal and external research groups, optimising the integration of study findings within organisations, and further enhancing evidence-based practices.

A lack of ethnic diversity and limited representation of children, seniors, and non-binary individuals in mHealth app research was revealed. Achieving ethical approval and meaningful engagement of these underrepresented groups is a pivotal but challenging step towards enhancing inclusivity and equity of mHealth app access and research. Future initiatives should prioritise development of culturally sensitive recruitment strategies to address the underrepresentation of ethnic minorities and non-binary individuals,

allowing researchers to generalise study findings to diverse populations. Ideally, target users, patient groups, or other stakeholders should be included in research design as early as possible.

This review highlights a notable lack of specificity and standardisation in the measurement tools and outcomes employed across studies, due to the diverse range of health outcomes addressed by mHealth apps. This variability in instruments makes it challenging to compare findings and draw overarching conclusions. The absence of standardisation poses difficulties in aggregating evidence and establishing clear benchmarks for app developers. Collaboration between researchers, regulatory bodies, and industry stakeholders plays a valuable role in developing a set of standardised tools for evaluating specific health outcomes. These tools should be readily accessible to app developers, enabling them to design interventions aligned with accepted measurement standards.

This review has some limitations. Firstly, the review omitted formative evaluations, potentially missing early-stage app development insights. The emphasis on efficacy and effectiveness evaluations excluded exploratory observational studies, qualitative, implementation, and user experience insights. Secondly, the omission of mHealth apps designed for healthcare providers or facilitating interactions between individuals and healthcare providers resulted in the underrepresentation of health condition management apps and the research methodologies commonly employed to evaluate them. This deliberate exclusion, while aimed at maintaining focus on DTC apps, is the reason for relatively few condition management apps identified in this review. The restriction to studies with over 50K Google Play downloads excluded smaller-scale apps, potentially overlooking innovative evaluation approaches. Language bias may exist due to the English language restriction, with a predominance of US-based studies limiting the transferability of findings to diverse global healthcare contexts. It is crucial to acknowledge that varying regulatory requirements exist among different countries, necessitating localised evidence. Furthermore, the complexities associated with the transference of digital health solutions from one nation to another [83] underscore the need for caution when assuming that successful outcomes in the US readily apply to other global healthcare settings. Lastly, we do not know what evaluations have been conducted, but were not published in the academic literature.

Conclusions

In conclusion, the field of digital health is rapidly expanding, posing significant challenges to app developers, regulators and end users. Academics and industry leaders are calling for streamlining and simplifying of current processes to avoid exerting too much pressure on companies who are innovating but cannot afford full-scale evaluations. We note that RCTs are feasible for a range of company types, but that most evaluations are on fairly small samples and with short follow-ups. Attempts to summarise the available evidence, such as this scoping review, together with detailed how-to guidance [21], represent a first step towards helping companies set realistic and feasible plans for ongoing product evaluations.

Declarations

References

1. Aitken M, Nass D. Digital Health Trends 2021 [Internet]. IQVIA; 2021 Jul. Available from: https://www.iqvia.com/insights/the-iqvia-institute/reports/digital-health-trends-2021

2. Kao C-K, Liebovitz DM. Consumer Mobile Health Apps: Current State, Barriers, and Future Directions. PM&R. 2017;9:S106–15.

3. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating Digital Health Interventions. Am J Prev Med. 2016;51:843–51.

4. Scott IA, Scuffham P, Gupta D, Harch TM, Borchi J, Richards B. Going digital: a narrative overview of the effects, quality and utility of mobile apps in chronic disease self-management. Aust Health Rev. 2020;44:62.

5. Agarwal P, Gordon D, Griffith J, Kithulegoda N, Witteman HO, Sacha Bhatia R, et al. Assessing the quality of mobile applications in chronic disease management: a scoping review. Npj Digit Med. 2021;4:46.

6. Cucciniello M, Petracca F, Ciani O, Tarricone R. Development features and study characteristics of mobile health apps in the management of chronic conditions: a systematic review of randomised trials. Npj Digit Med. 2021;4:144.

7. Ku JP, Sim I. Mobile Health: making the leap to research and clinics. Npj Digit Med. 2021;4:83.

8. Cohen AB, Mathews SC, Dorsey ER, Bates DW, Safavi K. Direct-to-consumer digital health. Lancet Digit Health. 2020;2:e163–5.

9. Digital Health Center of Excellence | FDA [Internet]. [cited 2023 Dec 7]. Available from: https://www.fda.gov/medical-devices/digital-health-center-excellence

10. Policy for Device Software Functions and Mobile Medical Applications [Internet]. U.S. Department of Health and Human Services Food and Drug Administration Center for Devices and Radiological Health Center for Biologics Evaluation and Research; 2022 [cited 2023 Jul 11]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications

11. European Commission. Green Paper on Mobile Health (mHealth) [Internet]. 2014 [cited 2023 Nov 22]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52014DC0219

12. BfArM - Homepage [Internet]. [cited 2023 Dec 7]. Available from: https://www.bfarm.de/EN/Home/_node.html

13. Torous J, Stern AD, Bourgeois FT. Regulatory considerations to keep pace with innovation in digital health products. Npj Digit Med. 2022;5:121.

14. Watson A, Chapman R, Shafai G, Maricich YA. FDA regulations and prescription digital therapeutics: Evolving with the technologies they regulate. Front Digit Health. 2023;5:1086219.

15. Njoku C, Green Hofer S, Sathyamoorthy G, Patel N, Potts HW. The role of accelerator programmes in supporting the adoption of digital health technologies: A qualitative study of the perspectives of small- and medium-sized enterprises. Digit Health. 2023;9:205520762311733.

16. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions—A call for innovative evidence generation approaches. Npj Digit Med. 2020;3:110.

17. Millenson ML, Baldwin JL, Zipperer L, Singh H. Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis. Diagnosis. 2018;5:95–105.

18. Safavi K, Mathews SC, Bates DW, Dorsey ER, Cohen AB. Top-Funded Digital Health Companies And Their Impact On High-Burden, High-Cost Conditions. Health Aff (Millwood). 2019;38:115–23.

19. Essén A, Stern AD, Haase CB, Car J, Greaves F, Paparova D, et al. Health app policy: international comparison of nine countries' approaches. Npj Digit Med. 2022;5:31.

20. Wilson K, Bell C, Wilson L, Witteman H. Agile research to complement agile development: a proposal for an mHealth research lifecycle. Npj Digit Med. 2018;1:46.

21. Potts H, Death F, Bondaronek P, Gomes M, Raftery J. Evaluating digital health products [Internet]. GOV.UK. 2020 [cited 2023 Jul 11]. Available from: https://www.gov.uk/government/collections/evaluating-digital-health-products

22. Silberman J, Wicks P, Patel S, Sarlati S, Park S, Korolev IO, et al. Rigorous and rapid evidence assessment in digital health with the evidence DEFINED framework. Npj Digit Med. 2023;6:101.

23. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. Npj Digit Med. 2019;2:38.

24. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health. 2018;4:205520761877032.

25. Karpathakis K, Libow G, Potts HWW, Dixon S, Greaves F, Murray E. An Evaluation Service for Digital Public Health Interventions: User-Centered Design Approach. J Med Internet Res. 2021;23:e28356.

26. Bell L, Garnett C, Bao Y, Cheng Z, Qian T, Perski O, et al. How Notifications Affect Engagement With a Behavior Change App: Results From a Micro-Randomized Trial. JMIR MHealth UHealth. 2023;11:e38342.

27. Bell L. Designing Randomised Trials to Improve Engagement through Optimising the Notification Policy of a Behaviour Change App [Doctor of Philosophy Thesis]. London School of Hygiene and Tropical Medicine (University of London); 2023.

28. Pham Q, Wiljer D, Cafazzo JA. Beyond the Randomized Controlled Trial: A Review of Alternatives in mHealth Clinical Trial Methods. JMIR MHealth UHealth. 2016;4:e107.

29. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018;169:467–73.

30. Cairo J, Williams L, Bray L, Goetzke K, Perez AC. Evaluation of a Mobile Health Intervention to Improve Wellness Outcomes for Breast Cancer Survivors. J Patient-Centered Res Rev. 2020;7:313–22.

31. Patel ML, Hopkins CM, Bennett GG. Early weight loss in a standalone mHealth intervention predicting treatment success. Obes Sci Pract. 2019;5:231–7.

32. Slazus C, Ebrahim Z, Koen N. Mobile health apps: An assessment of needs, perceptions, usability, and efficacy in changing dietary choices. Nutrition. 2022;101:111690.

33. Domin A, Uslu A, Schulz A, Ouzzahra Y, Vögele C. A Theory-Informed, Personalized mHealth Intervention for Adolescents (Mobile App for Physical Activity): Development and Pilot Study. JMIR Form Res [Internet]. 2022 [cited 2023 Jul 11];6. Available from: https://formative.jmir.org/2022/6/e35118

34. Saleh ZT, Elshatarat RA, Elhefnawy KA, Helmi Elneblawi N, Abu Raddaha AH, Al-Za'areer MS, et al. Effect of a Home-Based Mobile Health App Intervention on Physical Activity Levels in Patients With Heart Failure: A Randomized Controlled Trial. J Cardiovasc Nurs. 2022;38:128–39.

35. Rowland SA, Ramos AK, Trinidad N, Quintero S, Johnson-Beller R, Struwe L, et al. mHealth Intervention to Improve Cardiometabolic Health in Rural Hispanic Adults: A Pilot Study. J Cardiovasc Nurs. 2022;37:439–45.

36. Li J, Hodgson N, Lyons MM, Chen K-C, Yu F, Gooneratne NS. A personalized behavioral intervention implementing mHealth technologies for older adults: A pilot feasibility study. Geriatr Nur (Lond). 2020;41:313–9.

37. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the Mindfulness Meditation Mobile App "Calm" to Reduce Stress Among College Students: Randomized Controlled Trial. JMIR MHealth UHealth. 2019;7:e14273.

38. Kubo A, Kurtovich E, McGinnis M, Aghaee S, Altschuler A, Quesenberry C, et al. Pilot pragmatic randomized trial of mHealth mindfulness-based intervention for advanced cancer patients and their informal caregivers. Psychooncology. 2020;pon.5557.

39. Hamamatsu Y, Ide H, Kakinuma M, Furui Y. Maintaining Physical Activity Level Through Team-Based Walking With a Mobile Health Intervention: Cross-Sectional Observational Study. JMIR MHealth UHealth. 2020;8:e16159.

40. Ramsey SE, Ames EG, Uber J, Habib S, Clark S, Waldrop D. A Preliminary Test of an mHealth Facilitated Health Coaching Intervention to Improve Medication Adherence among Persons Living with HIV. AIDS Behav. 2021;25:3782–97.

41. Kwan RY, Lee D, Lee PH, Tse M, Cheung DS, Thiamwong L, et al. Effects of an mHealth Brisk Walking Intervention on Increasing Physical Activity in Older People With Cognitive Frailty: Pilot Randomized Controlled Trial. JMIR MHealth UHealth. 2020;8:e16596.

42. Chow EJ, Doody DR, Di C, Armenian SH, Baker KS, Bricker JB, et al. Feasibility of a behavioral intervention using mobile health applications to reduce cardiovascular risk factors in cancer survivors: a pilot randomized controlled trial. J Cancer Surviv. 2021;15:554–63.

43. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: A pilot study. Pediatr Blood Cancer. 2017;64:e26660.

44. Rosen KD, Paniagua SM, Kazanis W, Jones S, Potter JS. Quality of life among women diagnosed with breast Cancer: A randomized waitlist controlled trial of commercially available mobile app-delivered mindfulness training. Psychooncology. 2018;27:2023–30.

45. Yuan NP, Brooks AJ, Burke MK, Crocker R, Stoner GM, Cook P, et al. My Wellness Coach: evaluation of a mobile app designed to promote integrative health among underserved populations. Transl Behav Med. 2022;12:752–60.

46. Brinker TJ, Faria BL, De Faria OM, Klode J, Schadendorf D, Utikal JS, et al. Effect of a Face-Aging Mobile App–Based Intervention on Skin Cancer Protection Behavior in Secondary Schools in Brazil: A Cluster-Randomized Clinical Trial. JAMA Dermatol. 2020;156:737–45.

47. Kubo A, Aghaee S, Kurtovich EM, Nkemere L, Quesenberry CP, McGinnis MK, et al. mHealth Mindfulness Intervention for Women with Moderate-to-Moderately-Severe Antenatal Depressive Symptoms: a Pilot Study Within an Integrated Health Care System. Mindfulness. 2021;12:1387–97.

48. Hernández-Reyes A, Cámara-Martos F, Molina-Luque R, Moreno-Rojas R. Effect of an mHealth Intervention Using a Pedometer App With Full In-Person Counseling on Body Composition of Overweight Adults: Randomized Controlled Weight Loss Trial. JMIR MHealth UHealth. 2020;8:e16999.

49. Miyazato M, Ashikari A, Nakamura K, Nakamura T, Yamashiro K, Uema T, et al. Effect of a mobile digital intervention to enhance physical activity in individuals with metabolic disorders on voiding patterns measured by 24-h voided volume monitoring system: Kumejima Digital Health Project (KDHP). Int Urol Nephrol. 2021;53:1497–505.

50. Yang K, Oh D, Noh JM, Yoon HG, Sun J-M, Kim HK, et al. Feasibility of an Interactive Health Coaching Mobile App to Prevent Malnutrition and Muscle Loss in Esophageal Cancer Patients Receiving

Neoadjuvant Concurrent Chemoradiotherapy: Prospective Pilot Study. J Med Internet Res. 2021;23:e28695.

51. Eisenhauer CM, Brito F, Kupzyk K, Yoder A, Almeida F, Beller RJ, et al. Mobile health assisted self-monitoring is acceptable for supporting weight loss in rural men: a pragmatic randomized controlled feasibility trial. BMC Public Health. 2021;21:1568.

52. Al-Nawaiseh HK, McIntosh WA, McKyer LJ. An-m-Health Intervention Using Smartphone App to Improve Physical Activity in College Students: A Randomized Controlled Trial. Int J Environ Res Public Health. 2022;19:7228.

53. Avalos LA, Aghaee S, Kurtovich E, Quesenberry Jr C, Nkemere L, McGinnis MK, et al. A Mobile Health Mindfulness Intervention for Women With Moderate to Moderately Severe Postpartum Depressive Symptoms: Feasibility Study. JMIR Ment Health. 2020;7:e17405.

54. NeCamp T, Sen S, Frank E, Walton MA, Ionides EL, Fang Y, et al. Assessing Real-Time Moderation for Developing Adaptive Mobile Health Interventions for Medical Interns: Micro-Randomized Trial. J Med Internet Res. 2020;22:e15033.

55. Moore H, White MJ, Finlayson G, King N. Exploring acute and non-specific effects of mobile app-based response inhibition training on food evaluation and intake. Appetite. 2022;178:106181.

56. Mangieri CW, Johnson RJ, Sweeney LB, Choi YU, Wood JC. Mobile health applications enhance weight loss efficacy following bariatric surgery. Obes Res Clin Pract. 2019;13:176–9.

57. Turner-McGrievy GM, Crimarco A, Wilcox S, Boutté AK, Hutto BE, Muth ER, et al. The role of self-efficacy and information processing in weight loss during an mHealth behavioral intervention. Digit Health. 2020;6:205520762097675.

58. Worthen-Chaudhari L, McGonigal J, Logan K, Bockbrader MA, Yeates KO, Mysiw WJ. Reducing concussion symptoms among teenage youth: Evaluation of a mobile health app. Brain Inj. 2017;31:1279–86.

59. Win H, Russell S, Wertheim BC, Maizes V, Crocker R, Brooks AJ, et al. Mobile App Intervention on Reducing the Myeloproliferative Neoplasm Symptom Burden: Pilot Feasibility and Acceptability Study. JMIR Form Res. 2022;6:e33581.

60. Kozlov E, McDarby M, Pagano I, Llaneza D, Owen J, Duberstein P. The feasibility, acceptability, and preliminary efficacy of an mHealth mindfulness therapy for caregivers of adults with cognitive impairment. Aging Ment Health. 2022;26:1963–70.

61. Spring B, Pellegrini C, McFadden HG, Pfammatter AF, Stump TK, Siddique J, et al. Multicomponent mHealth Intervention for Large, Sustained Change in Multiple Diet and Activity Risk Behaviors: The Make Better Choices 2 Randomized Controlled Trial. J Med Internet Res. 2018;20:e10528.

62. Blair CK, Harding E, Wiggins C, Kang H, Schwartz M, Tarnower A, et al. A Home-Based Mobile Health Intervention to Replace Sedentary Time With Light Physical Activity in Older Cancer Survivors: Randomized Controlled Pilot Trial. JMIR Cancer. 2021;7:e18819.

63. Elicherla SR, Bandi S, Nuvvula S, Challa RS, Saikiran KV, Priyanka VJ. Comparative evaluation of the effectiveness of a mobile app (Little Lovely Dentist) and the tell-show-do technique in the

management of dental anxiety and fear: a randomized controlled trial. J Dent Anesth Pain Med. 2019;19:369–78.

64. Ju H, Kang E, Kim Y, Ko H, Cho B. The Effectiveness of a Mobile Health Care App and Human Coaching Program in Primary Care Clinics: Pilot Multicenter Real-World Study. JMIR MHealth UHealth. 2022;10:e34531.

65. Oliveira FMD, Calliari LEP, Feder CKR, Almeida MFOD, Pereira MV, Alves MTTDAF, et al. Efficacy of a glucose meter connected to a mobile app on glycemic control and adherence to self-care tasks in patients with T1DM and LADA: a parallel-group, open-label, clinical treatment trial. Arch Endocrinol Metab [Internet]. 2021 [cited 2023 Jul 11];65. Available from: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S2359-39972021005003206&lng=en&nrm=iso

66. Zha P, Qureshi R, Porter S, Chao Y-Y, Pacquiao D, Chase S, et al. Utilizing a Mobile Health Intervention to Manage Hypertension in an Underserved Community. West J Nurs Res. 2019;42:201–9.

67. Fanning J, Brooks AK, Hsieh KL, Kershner K, Furlipa J, Nicklas BJ, et al. The Effects of a Pain Management-Focused Mobile Health Behavior Intervention on Older Adults' Self-efficacy, Satisfaction with Functioning, and Quality of Life: a Randomized Pilot Trial. Int J Behav Med. 2022;29:240–6.

68. Mazaheri Asadi D, Zahedi Tajrishi K, Gharaei B. Mindfulness Training Intervention With the Persian Version of the Mindfulness Training Mobile App for Premenstrual Syndrome: A Randomized Controlled Trial. Front Psychiatry. 2022;13:922360.

69. Laird B, Puzia M, Larkey L, Ehlers D, Huberty J. A Mobile App for Stress Management in Middle-Aged Men and Women (Calm): Feasibility Randomized Controlled Trial. JMIR Form Res. 2022;6:e30294.

70. Zahid T, Alyafi R, Bantan N, Alzahrani R, Elfirt E. Comparison of Effectiveness of Mobile App versus Conventional Educational Lectures on Oral Hygiene Knowledge and Behavior of High School Students in Saudi Arabia. Patient Prefer Adherence. 2020;Volume 14:1901–9.

71. Garcia DO, Valdez LA, Aceves B, Bell ML, Rabe BA, Villavicencio EA, et al. mHealth-Supported Gender- and Culturally Sensitive Weight Loss Intervention for Hispanic Men With Overweight and Obesity: Single-Arm Pilot Study. JMIR Form Res. 2022;6:e37637.

72. Ong WY, Sündermann O. Efficacy of the Mental Health App "Intellect" to Improve Body Image and Self-compassion in Young Adults: A Randomized Controlled Trial With a 4-Week Follow-up. JMIR MHealth UHealth. 2022;10:e41800.

73. Toh SHY, Tan JHY, Kosasih FR, Sündermann O. Efficacy of the Mental Health App Intellect to Reduce Stress: Randomized Controlled Trial With a 1-Month Follow-up. JMIR Form Res. 2022;6:e40723.

74. Huberty JL, Espel-Huynh HM, Neher TL, Puzia ME. Testing the Pragmatic Effectiveness of a Consumer-Based Mindfulness Mobile App in the Workplace: Randomized Controlled Trial. JMIR MHealth UHealth. 2022;10:e38903.

75. Sysko R, Michaelides A, Costello K, Herron DM, Hildebrandt T. An Initial Test of the Efficacy of a Digital Health Intervention for Bariatric Surgery Candidates. Obes Surg. 2022;32:3641–9.

76. Giuliano AFM, Buquicchio R, Patella V, Bedbrook A, Bousquet J, Senna G, et al. Rediscovering Allergic Rhinitis: The Use of a Novel mHealth Solution to Describe and Monitor Health-Related Quality of Life in Elderly Patients. Int Arch Allergy Immunol. 2022;183:1178–88.

77. Item Short Form Survey Instrument (SF-36) [Internet]. RAND Corp. Object. Anal. Eff. Solut. [cited 2023 Jul 17]. Available from: https://www.rand.org/health-care/surveys_tools/mos/36-item-short-form/survey-instrument.html

78. What is PROMIS? [Internet]. PROMIS Health Organ. [cited 2023 Jul 17]. Available from: https://www.promishealth.org/57461-2/

79. Perceived Stress Scale [Internet]. State of New Hampshire Employee Assistance Program; Available from: https://www.das.nh.gov/wellness/Docs%5CPercieved%20Stress%20Scale.pdf

80. Madhuleena RC. The Five Facet Mindfulness Questionnaire (FFMQ) [Internet]. PositivePsychology.com. 2019 [cited 2023 Jul 17]. Available from: https://positivepsychology.com/five-facet-mindfulness-questionnaire-ffmq/

81. Stern AF. The Hospital Anxiety and Depression Scale. Occup Med. 2014;64:393–4.

82. 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.

83. Pham Q, Wong D, Pfisterer KJ, Aleman D, Bansback N, Cafazzo JA, et al. The Complexity of Transferring Remote Monitoring and Virtual Care Technology Between Countries: Lessons From an International Workshop. J Med Internet Res. 2023;25:e46873.
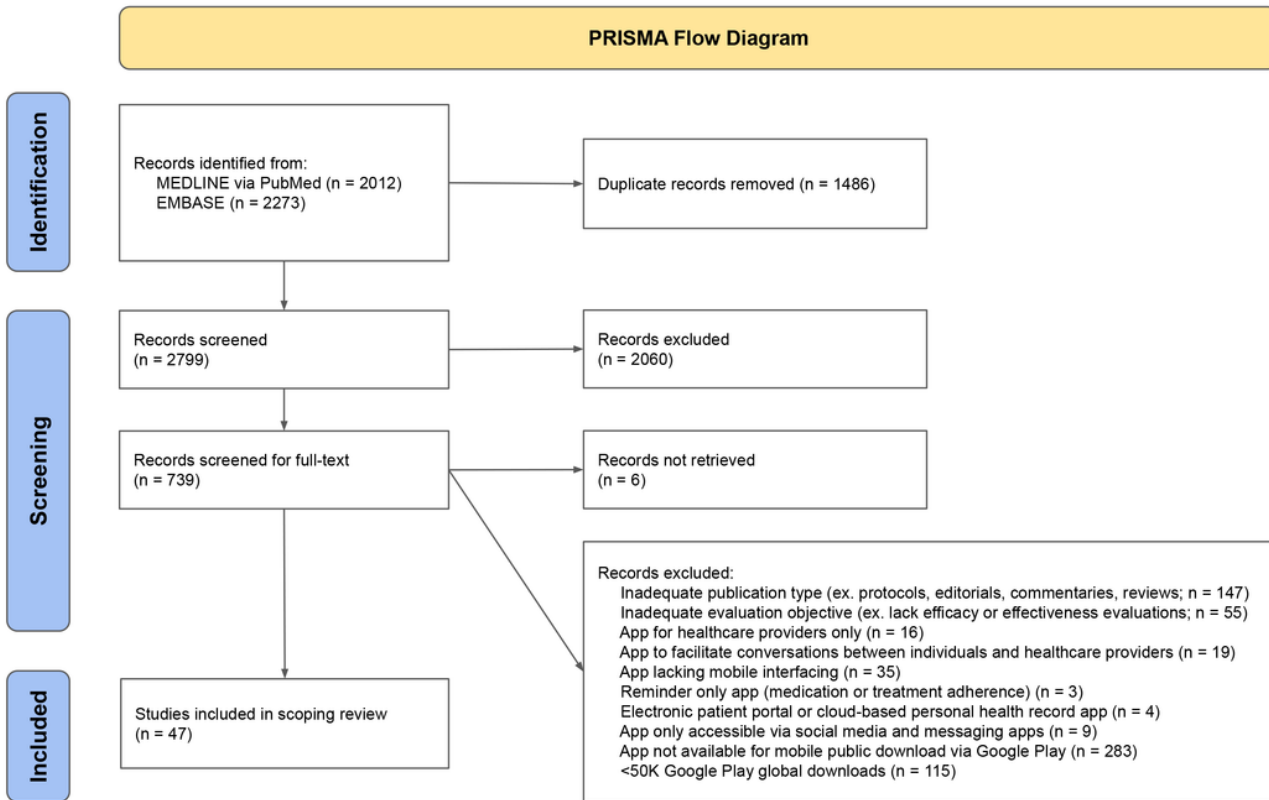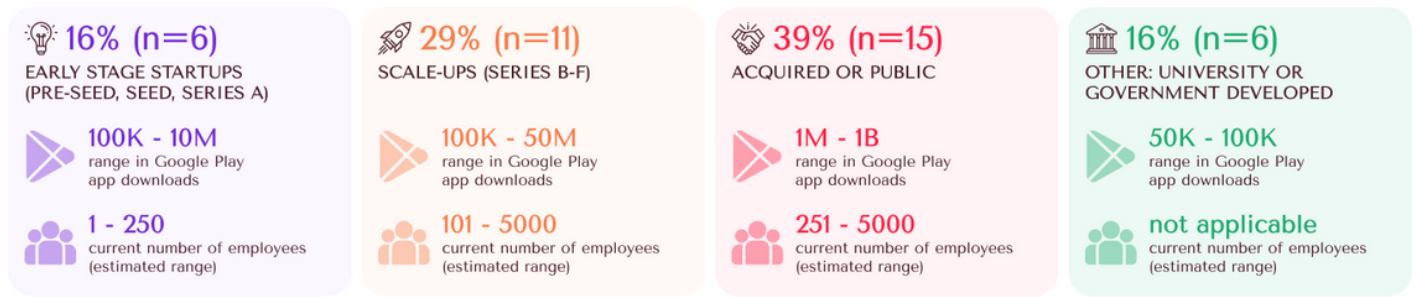
Figures

Figure 1

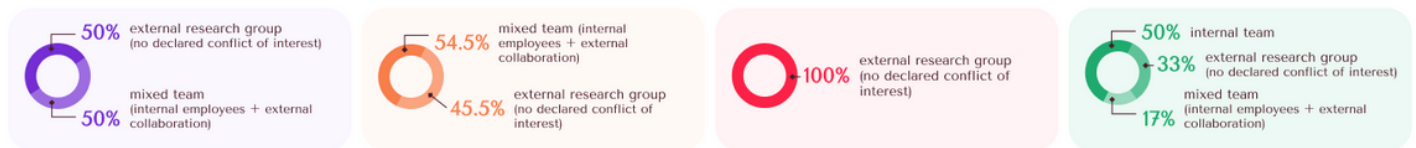**PRISMA flow chart.** Search and study selection process for this review.

mHealth

DTC APP EVALUATION METHODOLOGY BY FINANCING STAGE OF COMPANY

Financing stage of company at study publish date

16% (n=6)
EARLY STAGE STARTUPS
(PRE-SEED, SEED, SERIES A)

100K - 10M
range in Google Play
app downloads

1 - 250
current number of employees
(estimated range)

29% (n=11)
SCALE-UPS (SERIES B-F)

100K - 50M
range in Google Play
app downloads

101 - 5000
current number of employees
(estimated range)

39% (n=15)
ACQUIRED OR PUBLIC

1M - 1B
range in Google Play
app downloads

251 - 5000
current number of employees
(estimated range)

16% (n=6)
OTHER: UNIVERSITY OR
GOVERNMENT DEVELOPED

50K - 100K
range in Google Play
app downloads

not applicable
current number of employees
(estimated range)

Study conducted by

50% external research group
(no declared conflict of interest)

50% mixed team
(internal employees + external
collaboration)

54.5% mixed team (internal
employees + external
collaboration)

45.5% external research group
(no declared conflict of
interest)

100% external research group
(no declared conflict of
interest)

50% internal team

33% external research group
(no declared conflict of interest)

17% mixed team
(internal employees + external
collaboration)

Study Designs

33% full scale RCT

33% pilot RCT

17% non-randomised open label
trial

17% pre-post study

36% full scale RCT

18% pilot RCT

46% pre-post study

20% full scale RCT

40% pilot RCT

7% micro randomised trial

7% non-randomised open label trial

27% pre-post study

33% full scale RCT

17% 2x2 randomised, mixed
factorial design

50% pre-post study

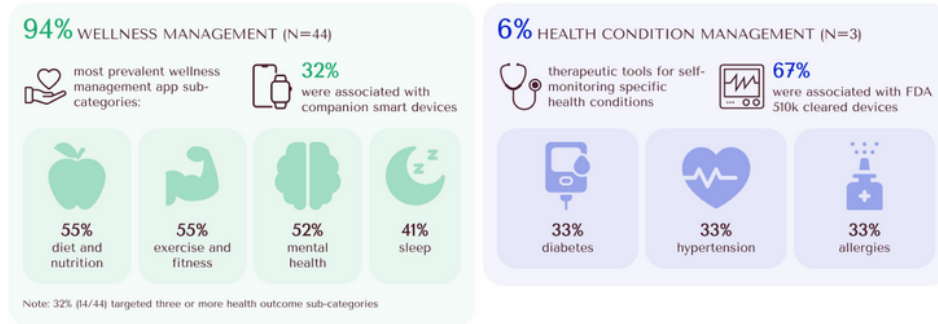**Figure 2**

Infographic of mHealth DTC app evaluation methodology. Grouped by financing stage of company at study publish date.
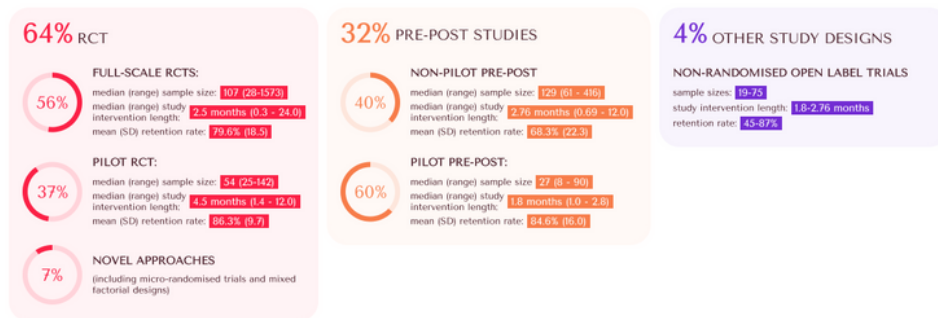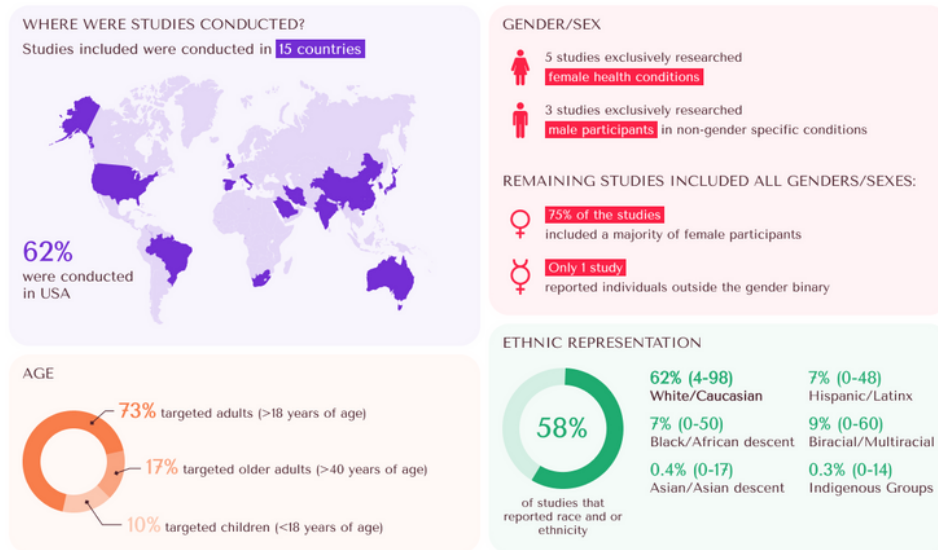
Figure 3

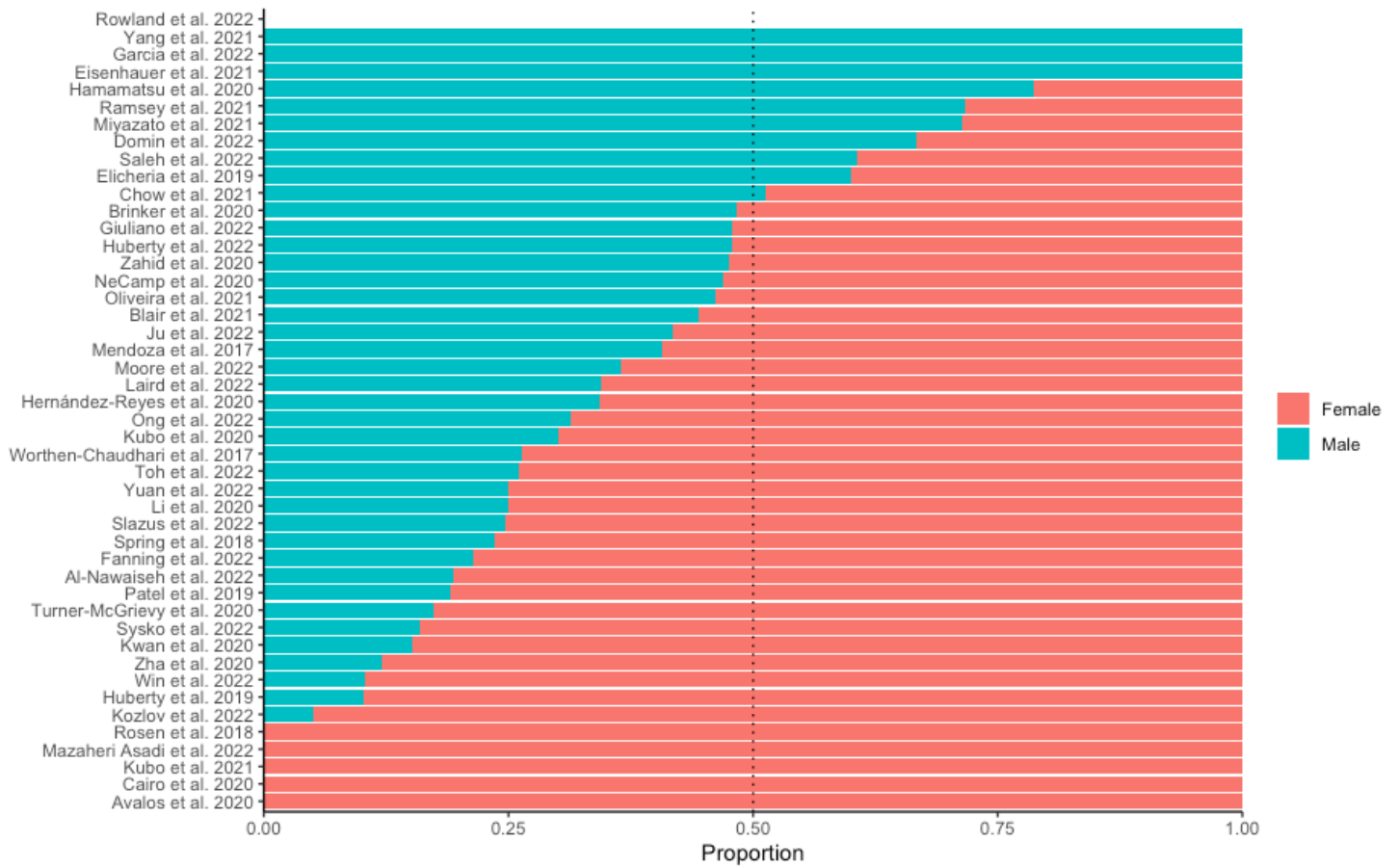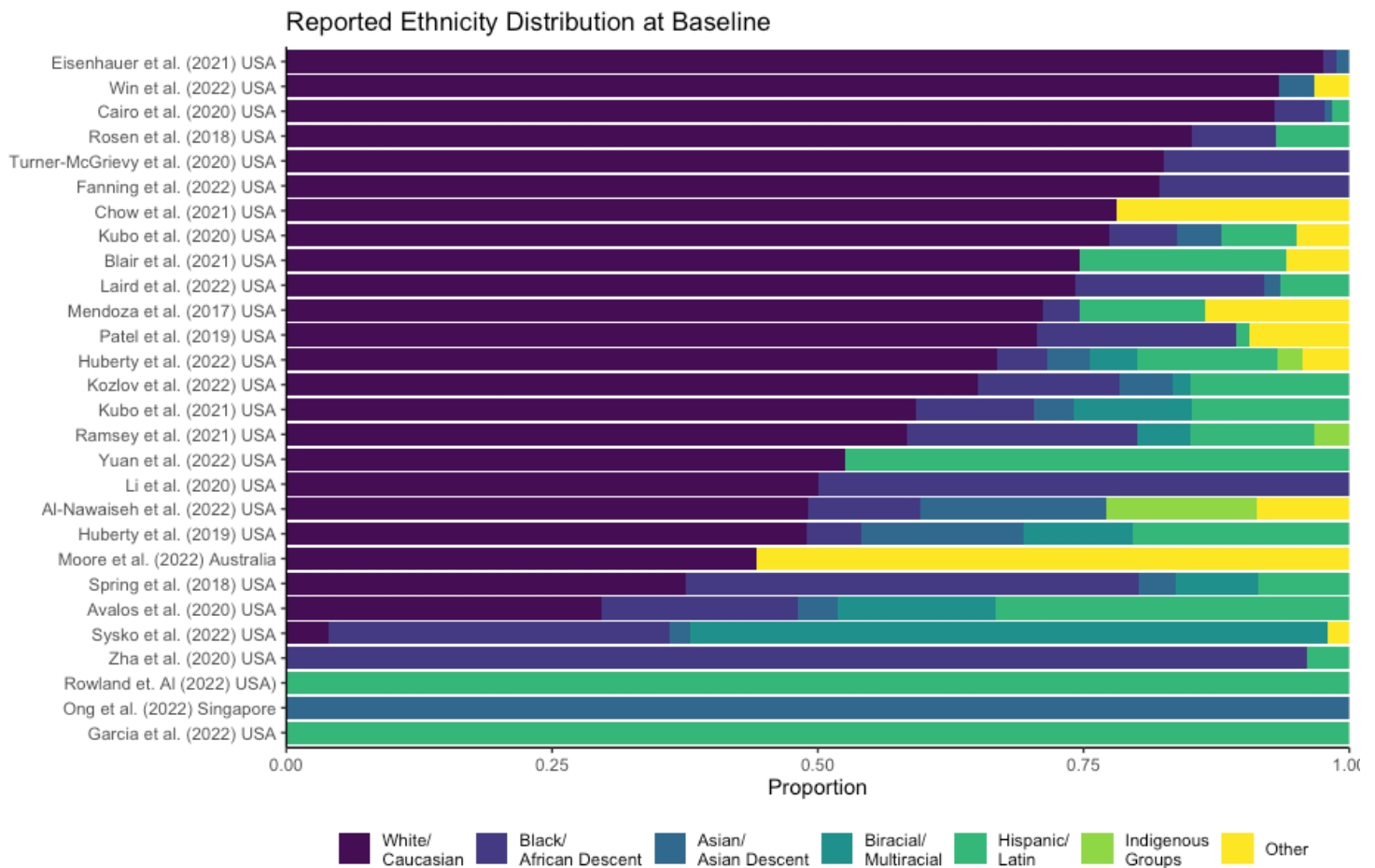Infographic of mHealth DTC app evaluation methodology characteristics.

Figure 4

Reported sex distribution at baseline.

Figure 5

Reported ethnicity distribution from included studies at baseline.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SMTable1PRISMAScRFillableChecklist.pdf
- SMTable2SearchStrategy.pdf
- SMTable3InclusionandExclusioncriteria.pdf
- SMTable4AllStudyCharacteristicsoftheIncludedStudies.xlsx
- SMTable5Financingstage.pdf
- SMTable6Financingstagestudydesign.pdf
- SMTable7Wellnessmanagement.pdf
- SMTable8Country.pdf
- SMTable9Measurementtoolscommonlyused.pdf