Review Article

Digital Health
Volume 7: 1–13

DOI: 10.1177/20552076211000559
journals.sagepub.com/home/dhj

Systematic review of cost-effectiveness analysis of behavior change communication apps: Assessment of key methods

Cynthia Afedi Hazel[1], Sheana Bull[1], Elizabeth Greenwell[1], Maya Bunik[2], Jini Puma[1] and Marcelo Perraillon[1]

Abstract

Objective: Evidence backing the effectiveness of mobile health technology is growing, and behavior change communication applications (apps) are fast becoming a useful platform for behavioral health programs. However, data to support the cost-effectiveness of these interventions are limited. Suggestions for overcoming the low output of economic data include addressing the methodological challenges for conducting cost-effectiveness analysis of behavior change app programs. This study is a systematic review of cost-effectiveness analyses of behavior change communication apps and a documentation of the reported challenges for investigating their cost-effectiveness.

Materials and methods: Four academic databases: Medline (Ovid), CINAHL, EMBASE and Google Scholar, were searched. Eligibility criteria included original articles that use a cost-effectiveness evaluation method, published between 2008 and 2018, and in the English language.

Results: Out of the 60 potentially eligible studies, 6 used cost-effectiveness analysis method and met the inclusion criteria.

Conclusion: The evidence to support the cost-effectiveness of behavior change communication apps is insufficient, with all studies reporting significant study challenges for estimating program costs and outcomes. The main challenges included limited or lack of cost data, inappropriate cost measures, difficulty with identifying and quantifying app effectiveness, representing app effects as Quality-adjusted Life Years, and aggregating cost and effects into a single quantitative measure like Incremental Cost Effectiveness Ratio. These challenges highlight the need for comprehensive economic evaluation methods that balance app data quality issues with practical concerns. This would likely improve the usefulness of cost-effectiveness data for decisions on adoption, implementation, scalability, sustainability, and the benefits of broader health-care investments.

Keywords

mHealth, digital health, mHealth apps, cost-effectiveness analysis, behavior change communication apps

Submission date: 11 August 2019; Acceptance date: 4 February 2021

Introduction

Mobile health (mHealth) applications (apps) designed for health promotion and disease management are growing in popularity. Currently, there are about 97,000 mHealth apps listed on 62 app stores.[1] Coupled with the extensive market penetration of smartphones, mHealth apps have broadened in capacity, scope and reach.[2] Apps that encourage healthy

[1]Colorado School of Public Health, University of Colorado, Denver, USA
[2]Children's Hospital Colorado, Aurora, USA

Corresponding author:
Cynthia Afedi Hazel, Colorado School of Public Health, University of Colorado, Denver, USA.
Email: cynthiaafedi@gmail.com

behaviors are classified among behavior change communication (BCC) programs. Recent evidence suggests that BCC programs are common and becoming increasingly successful.[3] Apps provide an ideal platform for BCC programs due to their ubiquity, connectivity and increased sophistication.[4,5] Compared to face-to-face programs, BCC apps have the potential to reach a larger number of people at more convenient times.[6] Common examples of BCC app programs include appointment reminders, support for medication adherence, community mobilization, and awareness campaigns.[7]

Effectiveness studies on BCC apps point to effects, albeit small, in increasing access to health information, enhancing clinical services, and improving health behaviors.[8–13] Likewise, a systematic review showed that nearly 90% of BCC apps that targeted physical activity, weight loss, and mental health reported statistically significant behavior change.[8] Despite claims of success and potential for greater impact, evidence on the effectiveness of BCC apps needed for large-scale implementation is still sparse.[3] To date, few replication studies exist, and best strategies for effectiveness assessments and user engagement at scale are insufficient.[8,12,14,15] In part, institutional adoption is necessary to develop better evidence for program effectiveness, implementation and evaluation.[16]

An additional impediment for the adoption and spread of BCC apps is the difficulty in estimating their economic value through cost-effectiveness analysis (CEA).[17–19] CEA studies compare the incremental costs and incremental effects of interventions with the goal of establishing the "economic value" of interventions,[20] or alternatively, whether the incremental effectiveness justifies the additional cost of the apps. While BCC apps hold great promise, their place in behavioral health might depend on whether they are a cost-effective alternative for replacing or enhancing current practice.[21] The costs associated with the development and maintenance of app programs could enhance or impede their use on a broader scale. Furthermore, economic evaluation can play an important role in prioritizing health programs by enabling decision-makers to make better choices between interventions that provide the most cost-effective health outcomes.[22,23] As with the evidence on effectiveness, the economic evaluation literature on BCC apps is also sparse. De la Torre-Diez et al. indicated there were too few CEA studies in the literature.[19] Systematic reviews by Whitten et al. and Badawy et al. also reported a lack of adequate evidence of the extent to which app interventions represent a sensible priority for healthcare investment. Pointing to deficits in the quality of the economic evaluation, the authors also called attention to the shortfalls in study design methods for evaluating apps.[17,18]

The objectives of this paper are to review cost-effectiveness analyses of BCC apps and to describe associated study challenges. Our goal is to inform public health researchers, health experts, and policymakers about the cost-effectiveness landscape of BCC apps, as well as highlight methodological challenges for conducting CEA studies on BCC apps.

Cost-effectiveness analysis

Cost-effectiveness studies compare the costs and outcomes of two or more health interventions. Generally, a new intervention is compared to "usual care" or a "do-nothing" alternative.[24] The incremental cost of the new intervention is compared to the incremental outcomes to calculate the incremental cost-effectiveness ratio (ICER). Outcomes are measured as either "natural units," such as cases averted or cases detected, or in Quality Adjusted Life Years (QALYs). QALYs is often preferred because life years adjusted for quality is a common outcome that can be used to compare disparate interventions. The QALY measure combines quality of life and survival outcomes into one effect, and is expressed by the sum of years lived weighting each year by a quality of life weight between 0 (dead) and 1 (perfect health).[20] Hence, 1QALY equates to one year in perfect health. ICER is calculated by dividing incremental cost by incremental effect (usually in QALYs) to provide a ratio which represents the additional cost per additional unit of health effect.[25] A lower ICER is preferred because it indicates that the incremental cost of an intervention is low per year of life gained adjusted for quality (QALY). The ICER from an intervention is compared to the ICER of other interventions or to a threshold that is used as a reference for value. In the US, by convention, interventions with an ICER below $100,000 per QALY[26] are considered cost-effective, but acceptable thresholds range from $50,000 to $200,000 per QALY.

Finally, two important considerations are the analytical perspective and the time horizon. The perspective of a CEA refers to the viewpoint of the evaluation, which is based on the types of costs and health effects an analyst selects to estimate the ICER. The time horizon of a CEA is the follow-up period accounted for in the calculation of costs and outcomes used to compute the economic measures. Since QALYs take into account life expectancy, the time horizon is often long.

CEA follows the "Rule of Reason," which states that the costs and effects expected to be trivially small with little impact on results, be reasonably excluded at the analyst's discretion.[27] While the "Rule of Reason" increases flexibility in conducting CEA, studies involving BCC apps are still challenging using recommended guidelines. A review of the CEA guidelines is therefore important to highlight methodological challenges that decrease the quality and usefulness of CEA evidence for BCC apps. Overcoming these challenges could largely expand the CEA evidence base for apps to make studies more comparable and supportive of healthcare decisions and resource allocation.

The panels' recommendations. In 2016, the Second Panel on Cost-Effectiveness in Health and Medicine in the US (herein, the Second Panel) updated the First Panel's CEA recommendations (APPENDIX C) with the same intent of study comparability. In summary, the Second Panel called on analysts to report costs and effects from societal and healthcare sector perspectives in addition to other analytic perspectives of interest. The Second Panel also recommended that an Impact Inventory be included in CEA reporting. An Impact Inventory is a catalogue of related and unrelated costs and effects of the intervention from societal and healthcare sector perspectives. The Impact Inventory aims to reduce barriers of low transparency and misunderstandings of CEA measures and how they were generated for cost-effectiveness calculations.[26] Finally, it was recommended that CEA results be reported in Incremental Cost-Effectiveness Ratio (ICER), with health effects aggregated into a single measure of QALYs.[20]

Analytic perspective. The challenges faced by healthcare decision makers in separating CEA results from the costs and effects that are solely attributable to the healthcare sector are well documented.[28,29] In part, these challenges are due to the fact that many CEA studies report results from a societal perspective.[30] A societal perspective measures the total cost and effect of an intervention regardless of who incurred them. Conversely, a healthcare sector perspective estimates the costs and effects of the intervention that is attributable to the healthcare sector without consideration for the costs and effects contributed by other sectors.[26] The Second Panel's recommendation for analysts to report results for both societal and healthcare sector perspectives is to increase the relevance and usefulness of CEA data for healthcare decision-making.[31]

Methods

In this review, we: 1) selected relevant studies based on specific inclusion criteria, 2) developed and summarized study characteristics, and 3) assessed CEA studies using the Second Panel's key recommendations for conducting CEA.[26] We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 guidelines (Appendix B).[32]

Inclusion criteria

All cost-effectiveness analysis of BCC apps targeting a specific behavioral health outcome(s) were included in this review. The eligible studies were restricted to English language publications, with publication date ranging from January 1, 2008 to January 17, 2018.

Information source and search strategy

The search strategy was developed by the lead author and independently verified by a research librarian to ensure search comprehensiveness and accuracy of studies retrieved for the review. All discrepancies were discussed and agreed upon with the remaining authors. Information sources included Medline (Ovid), EMBASE, CINAHL (Cumulative Index to Nursing and Allied Health Literature) and Google Scholar. The databases were searched on January 17, 2018 to generate relevant studies using key search terms, such as "mHealth AND app* AND ((Cost* adj3 (Benefit* or Effectiveness or Utilit* or saving* or minimization))". The full search strategy for Medline (Ovid) is shown in Table 1.

Study selection

The search results were exported into EndNote X8 citation manager, and duplicates were removed. The remaining citations were examined for relevant titles and abstracts. Systematic reviews, meta-analyses and non-original research studies were excluded. After the full text review, we further excluded studies that did not provide clear and sufficient information about CEA method or specific behavioral outcome(s) of interest. The references of eligible studies were scrutinized to find additional studies.

Data collection process

Microsoft Excel was used to extract study characteristics (see Appendix A): aim(s), intervention, participants, comparators, research perspective, time horizon, economic method used, outcome(s), outcome measure(s) and data source(s). When a study had both primary and secondary outcomes, the outcome directly related to the economic analysis was documented. Outcome was defined as any objective measure of behavioral outcome, that can be accurately captured by the BCC app.

Results

Study selection

The search strategy initially retrieved 694 studies: 243 in Medline, 80 in CINAHL, and 371 in Embase.

Table 1 Search strategy in medline (ovid).

Number	Search	Results
Query string: *The query string of subheadings (MeSH) and text words used for the database search were, "((Cost* adj3 (Benefit* or Effectiveness or Utilit* or saving* or minimization)) or (Economic adj3 (Evaluation* or analys*)) or (Marginal adj3 Analys*) or exp "Costs and Cost Analysis"/ AND [(((Mobile adj3 Health) or mHealth or m-Health or Telehealth or Tele-health or eHealth or e-Health) and (application or app or apps)) or (exp Telemedicine/ and exp Mobile Applications/) OR (((Mobile or Portable or electronic or software or cell or smartphone* or smart phone* or web-based) adj3 (application* or app or apps)) or mApps or m-Apps or m-App or mApp or m-Application* or mApplication* or exp Mobile Applications/]".*		
1	(Cost* adj3 (Benefit* or Effectiveness or Utilit* or saving* or minimization)) or (Economic adj3 (Evaluation* or analys*)) or (Marginal adj3 Analys*) or exp "Costs and Cost Analysis"/	2,81,955
2	(((Mobile n3 Health) or mHealth or m Health or Telehealth or Tele health or eHealth or e Health) and (application or app or apps)) or (MH "Telehealth+" and MH "Mobile Applications")	2,807
3	((Mobile or Portable or electronic or software or cell or smartphone* or smart phone* or web-based) adj3 (application* or app or apps)) or mApps or m-Apps or m-App or mApp or m-Application* or mApplication* or exp Mobile Applications/	22,819
4	2 OR 3	23,727
5	1 AND 4	349
6	Remove duplicates from 5	304
7	Limit 6 to English language	297
8	Limit 7 to yr="2008 - Current"	243

Duplicates were removed, followed by studies without relevant titles and abstracts which resulted in 60 potentially eligible studies. The full text of the 60 potentially eligible studies were scrutinized, 51 of which were excluded. Nine studies were original research that used an economic evaluation method to assess a BCC app intervention. A search through the references of the nine studies identified 1 additional eligible study. Of these studies, 6 used a CEA method. Figure 1 shows the flow diagram for the final study selection.

Study characteristics

The 6 CEA studies were published between 2012 and 2017. Van Reijen M, et al. evaluated the "Strengthen your Ankle" mobile app, which provides written, visual, and verbal instructions as well as a calendar function, to help athletes prevent the recurrence of ankle sprains.[33] Martín JAC, et al. evaluated "CardioManager", a medical app which provides heart disease patients with the necessary behavioral and clinical guidelines to enable them to self-manage and monitor their own health conditions.[34] The "Tät" app evaluated by Sjöström M, et al. was developed as a first line treatment for stress urinary incontinence based on self-management, which also provides patients with instructions for pelvic floor muscle training.[35] Dahlberg K, et al. evaluated the "Recovery Assessment by Phone Points (RAPP)" app, which is a post-operative recovery monitoring program that allows patients to report progress of recovery directly to healthcare professionals after day surgery.[36] The "Bring-your-own-device (BYOD)" mobile app evaluated by Armstrong K, at al. supports post-operative care in breast reconstruction patients by providing patients with validated quality of recovery questionnaires and surgical site photo submissions. Results from the questionnaire enables health professionals to detect post-operative complications and eliminate in-person follow-up care.[5] Ryan D, et al. evaluated the "t+ Asthma" app that enables twice daily transmission of symptoms, drug use, and peak flow, and prompts patients about agreed action plans. An incursion into the red or amber zones triggers a contact by an asthma nurse the next day.[21]

The CEA studies were analyzed using the Second Panel's recommendations and the study characteristics were described in Table 2. The BCC apps used in four studies (66.7%) were focused on patient adherence to behavioral and clinical guidelines for self-monitoring,

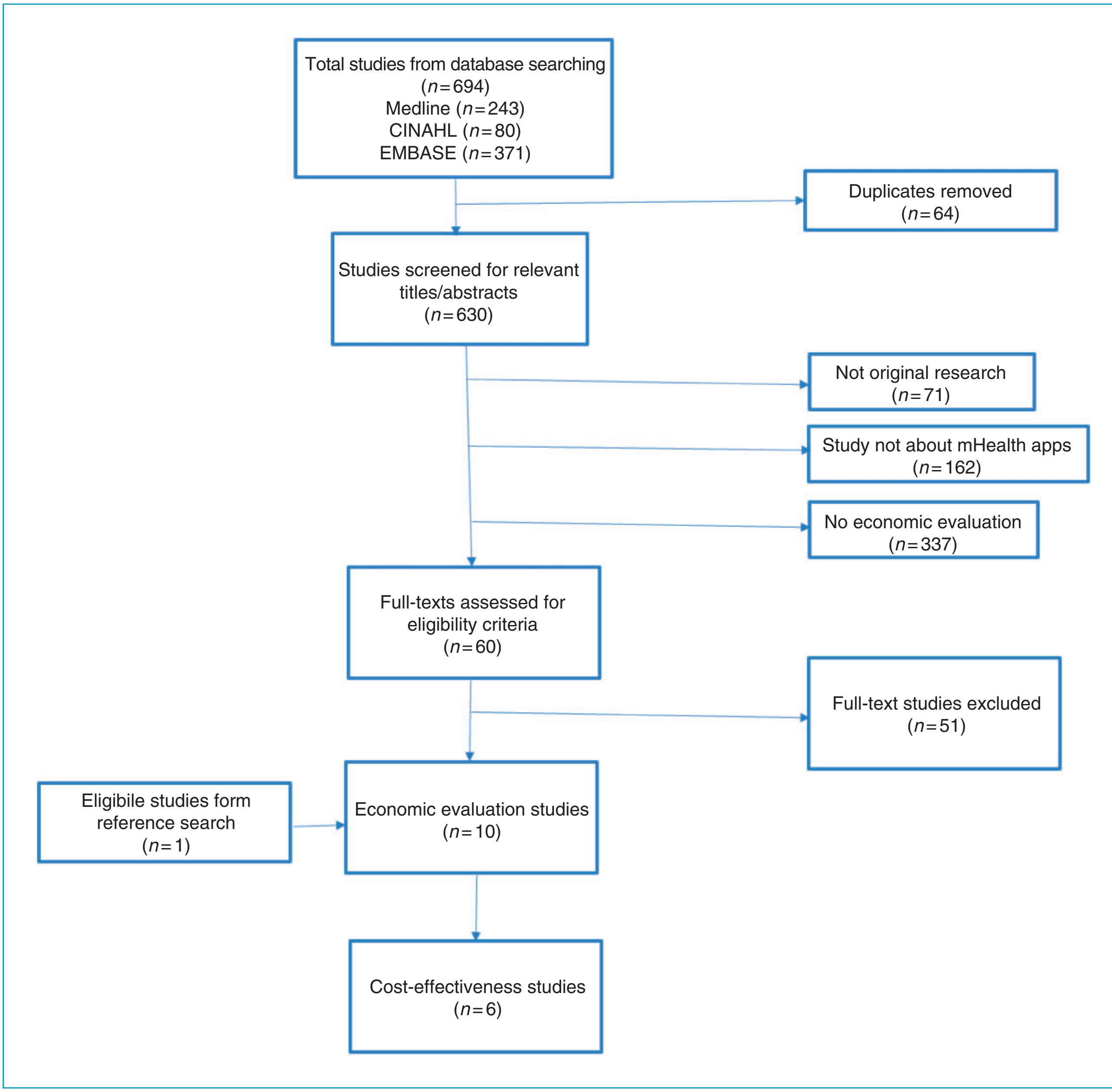

Figure 1. Flow chart showing final study selection.

while the apps in the remaining two studies required patient feedback to enable healthcare professionals conduct risk assessments. Two studies (33.3%) each reported a societal or a healthcare sector perspective. One study (16.7%) reported both a societal and healthcare sector perspective, and one study (16.7%) reported a health service provider perspective. The data sources for conducting the CEA varied across studies; four studies (66.7%) used questionnaires in addition to other primary data sources.

Five studies (83.3%) discounted the costs used in the analysis, but only one study (16.7%) reported a discount rate of 3%. The cost measure was consistent across studies, with the predominant outcome measure as QALYs in five studies (83.3%). Two studies (33.3%) reported the cost-effectiveness results in ICER, whilst the remaining used other cost-effectiveness measures. All studies performed at least one sensitivity analysis, 5 (83.3%). No study reported a cost-effectiveness threshold for concluding the app's cost-effectiveness, and only four studies (66.7%) addressed the implications of the costs and outcomes for scaling the app intervention beyond the trial.

Table 2. Characteristics of CEA studies.

CEA characteristics	Number of studies (%)
Focus of intervention	
Patient adherence to behavioral and clinical guidelines for self-management	4 (66.7)
Provide feedback to healthcare professionals for risk assessment/reduction	2 (33.3)
Clearly stated reference case perspective	
Yes	6 (100.0)
No	0 (0.0)
Reference case perspective	
Societal	2 (33.3)
Healthcare Sector	2 (33.3)
Health Service Provider	1 (16.7)
Societal and Healthcare Sector	1 (16.7)
Data Source (quantifying cost and outcomes)	
Questionnaire	4 (66.7)
Other	6 (100)
Discounted cost and/or outcomes	
Yes	5 (83.3)
No	1 (16.7)
Outcome measure	
Quality Adjusted Life Years (QALYs)	5 (83.3)
Other	1 (16.7)
Cost-effectiveness results reported as	
Incremental cost-effectiveness ratio (ICER)	2 (33.3)
Other	4 (66.7)
Use of cost-effectiveness threshold	
Yes	0 (0.0)
No	6 (100.0)
Sensitivity analysis	
One-way	5 (83.3)
Two-way	1 (16.7)

(continued)

Table 2. Continued.

CEA characteristics	Number of studies (%)
Multi-way	1 (16.7)
Probabilistic	2 (33.3)
Multiple analysis	2 (33.3)
Addressed cost and outcome implications for scale	
Yes	4 (66.7)
No	2 (33.3)

All studies reported significant methodological challenges in conducting CEA of the BCC apps. Limitations varied from lack of data to difficulties in valuing confounding effects. A summary of the limitations outlined in the CEA studies are categorized in Table 3.

Discussion

Our findings show a paucity of CEA studies of BCC apps and various challenges for conducting CEA studies involving BCC apps. This results suggest that the overall economic benefits of mHealth apps is often assumed and have thus far not been substantiated with sufficient empirical evidence. We hypothesized linkages between the challenges for conducting CEA that ultimately leads to poor-quality evidence and low comparability across CEA studies involving BCC apps (Figure 2). Barriers include data quality and availability. Inadequate cost and outcome measures, and insufficient data sources also represent important quality considerations which could ultimately affect ICER estimations. Overcoming these barriers can increase the transparency, consistency and transferability of CEA findings.[27]

Challenges for conducting CEA of BCC apps

Program effectiveness. Most studies used Randomized Controlled Trials (RCT) to estimate program effectiveness while others relied on effectiveness evidence from the literature or no evidence at all. Randomization is a major defense against unequal distribution of confounders. Thus, RCT remains the gold-standard for assessing effectiveness and efficacy in many mHealth interventions.[37] Restricting effectiveness assessments of BCC apps to RCTs is, however, problematic because it limits the scope of the evidence base by overlooking additional evidence from non-RCT methods. Furthermore, RCTs have a time lag of approximately 5.5 years between initiation of subject recruitment and publication of results which increases the risk for technology obsolescence of app functions.[37] While there is a need to advocate for the recognition of non-RCT methods, rigorous approaches that balance fidelity with practicality are needed to ensure that quantifying app effects in economic evaluations is feasible. Three key methodological approaches for overcoming this barrier includes: 1) causal inference from observational data,[38,39] 2) obtaining multiple repeated measures on a few participants, rather than a few measures on many participants; and 3) model-based designs for adaptive interventions, which requires an understanding of within-subject differences and the effect of mediating variables on health outcomes.[37] These approaches will likely reduce the cost and time for data collection, and accommodate the continuously evolving technology in BCC app designs.

Attention should be given to small sample size and high losses to follow-up which can increase the uncertainty in cost and effect estimations, and introduce errors into scale analysis.[7] Methods for determining an accurate sample size varied among the studies, with sample size ranging from 123 to 1000 participants. For example, Dahlberg et al. used sample size of 1000 participants based on QALY weights in patients with an asymptomatic gallstone disease and a surgical scar, and assumptions of detecting a difference of 0.03 in QALYs between the groups.[36] Conversely, Sjöström et al. used a convenience sample of 123, who were community dwelling women with stress urinary incontinence recruited via a website.[35] Inconsistencies in calculating sample size for BCC app evaluation is challenging because a small sample size could indicate lower cost-savings, which may be higher with a larger

Table 3. Limitations in the CEA studies.

Study limitation	Number of studies affected (%)
Omitted healthcare sector, societal or both perspectives.	5 (83.3)
Difficulty with estimating results in ICER	2 (33.3)
No use of cost-effectiveness threshold(s)	6 (100%)
Lack of comparative cost-effectiveness data; false equivalency assumptions	3 (50.0)
Small sample size and/or losses to follow-up	4 (66.7)
Lack of (or limited access to) data	2 (33.3)
Difficulty with costing (e.g. micro versus macro)	4 (66.7)
Difficulty in estimating QALYs (or outcome effects or confounding effects)	5 (83.3)
Data extrapolation/modeling difficulty (assuming cost and outcome consistency over time)	1 (16.7)

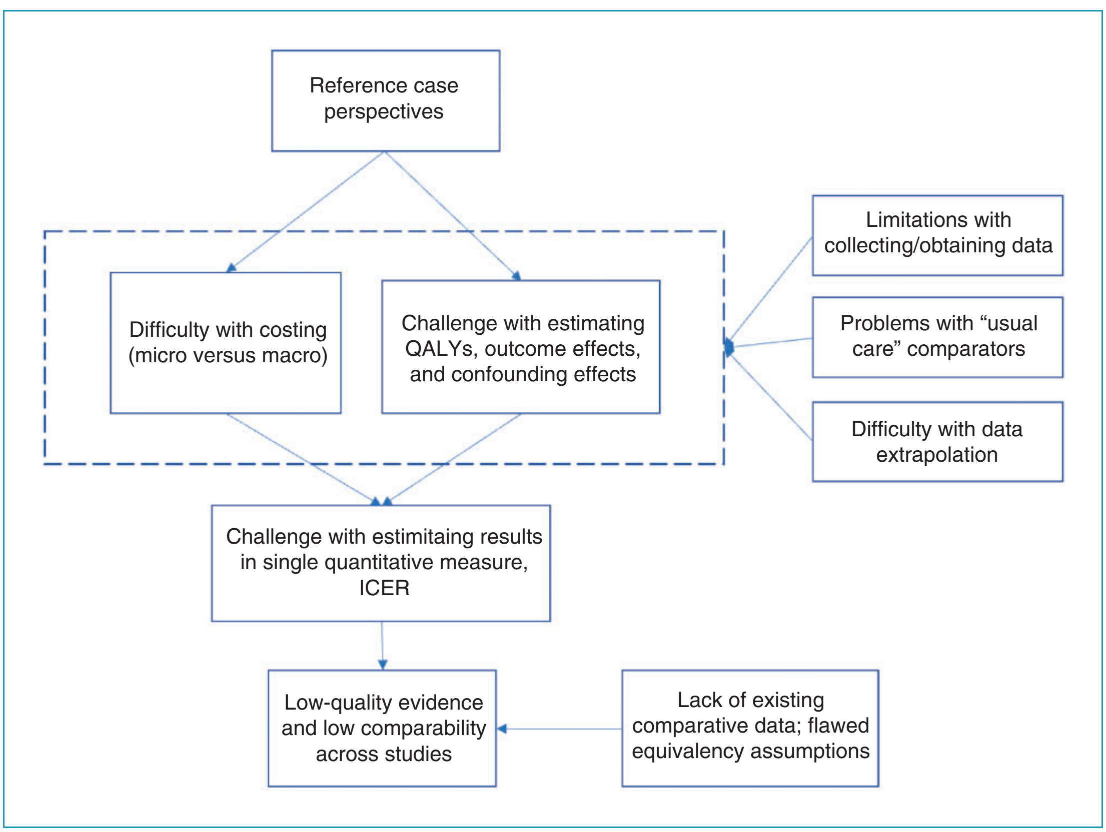

Figure 2. Hypothesized linkages between limitations for conducting CEA of BCC Apps.

sample size. Additionally, a small sample could represent a lack of user volume, which could indicate that more utilization may increase cost.[7] This problem was exemplified by Luxton et al. who used a sample size of 1000 participants to conclude that there was no cost-savings to using the "Breath2Relax" app. However, the app became less expensive compared to in-office treatment at approximately 1600 users.[40] An appropriate sample size enhances economies of scale analysis, which is important for determining the costs and outcomes associated with the intervention if scaled to a larger population post-trial.

Analytical perspective. One study (16.7%) considered both societal and healthcare sector perspectives, while the remaining studies each considered a single perspective. To conduct CEA from a societal perspective, the Second Panel recommended the inclusion of an Impact Inventory.[27] Though most of the studies predated this recommendation, our assessment was based on whether all relevant costs and effects were included in the studies' analyses. The review showed that all studies had deficits in their costs and outcome estimations. While a narrower perspective accounts for costs and outcomes attributable to specific decision makers, it also represents a missed opportunity for evaluating the full cost and effects of the intervention. It was however uncertain whether a societal perspective which considered costs and outcomes regardless of the payer and beneficiary was better justified for any of the BCC apps. For example, Cano Martin et al. evaluated the CardioManager app intervention which required significant user input but only considered a healthcare sector perspective which ignored time costs attributable to the user.[41] A healthcare sector perspective was advantageous in this case due to its relevance to budget holders, but the omission of costs and outcomes attributable to the user constituted incomplete data which was important for decisions regarding large-scale implementation.

It is important to note that BCC apps vary widely in interventions, app features, and the targeted stages of behavior change.[3] In many cases, BCC apps are used outside the formal healthcare sector with cost savings accrued to the broader society.[7] This assertion was corroborated by Armstrong et al. who found that the higher cost of in-person follow-up care was spread between the healthcare system and patient, but the patient reaped the majority of cost-savings from participating in a mobile app follow-up care.[5] Furthermore, health interventions can have multiple unintended effects in targeted and untargeted participants over a short period of time.[27] Thus, identifying and quantifying health costs and outcomes in different sectors can be challenging from a societal perspective. In cases where the "Rule of Reason" is applied, still identifying and valuing certain app costs and effects before ultimately judging that their exclusion has no significant impact on results can be time-consuming. While the Second Panel acknowledges these challenges, a more feasible measurement strategy for quantifying the external effects of BCC apps is yet to be identified.

Quantifying outcomes (effects). It is critical to accurately identify and quantify intervention outcomes in CEA studies.[27] Outcome measures varied across the studies, and the clinical/behavioral endpoints were inconsistent among the BCC apps. Consistent with the Second Panel's recommendation, the predominant outcome measure used in the studies (83.3%) was QALYs. The QALY end-point combines quality of life (QoL) and survival outcome data, which provides comparability between studies.[26] However, the QoL measure is a multi-dimensional health function usually obtained through preference weights associated with specific health states. Additionally, QALYs sum the total years lived, weighting each year by a quality of life weight between 0(dead) and 1(perfect health).[42] Obtaining preference weights for specific health states through a BCC app is complex, due to the need to collect detailed preference data while the app is in use. For example, Dahlberg et al. in evaluating the RAPP app did not find any difference in QALYs gained between the groups because the SF-6D measure used in estimating the QALYs functions were not affected by user follow-up routines.[36] Overcoming this limitation is challenging because the endpoint for most BCC apps are intermediate outcomes which may not always be clinically quantifiable. Furthermore, estimating QALYs from intermediate outcomes requires assumptions which are not always backed with empirical data.

An additional impediment to estimating intermediate outcomes in BCC apps is the fact that they do not always lead to clinical or behavioral endpoints. Intermediate outcomes are common among health interventions aimed at health promotion and disease prevention because studies to estimate changes in long-term outcomes are costly and time-consuming to conduct.[42] As noted by the Second Panel, intermediate outcomes allows for studies with smaller sample sizes and shorter time frames because they are more likely to occur much sooner than the final desired clinical outcome(s).[27] Furthermore, insufficient data remains a significant barrier to estimating the desired clinical endpoints from intermediate outcomes in BCC apps. Three proposed strategies for overcoming this barrier are to: 1) make the case that intermediate outcomes of BCC apps have value and clinical relevance in their own right, 2) prove that the link between an intermediate and final outcome (clinically-relevant endpoint) have been adequately established by previous research, or 3) address the uncertainties surrounding the link between intermediate and final outcomes in the analysis.[42] Despite these proposed strategies, questions about the usefulness and comparability of intermediate outcomes in CEA studies linger.

Health interventions could have confounding effects that need to be separated and quantified in order to estimate the true effects of the intervention.

However, BCC apps can be confounded by external effects that are not always separable or quantifiable. For example, Reijen et al. noted that many athletes using the "Strengthen your Ankle" app may have already performed some sort of neuromuscular training (NMT) prior to participating in the trial, which may have reduced their initial risk for developing a recurrent ankle sprain.[33] Quantifying the effects of the previous NMT to estimate the true effect of the app was impossible, although a previous NMT had significant impacts on the outcome of using the app. Methodological guidelines for accurately accounting for BCC app confounders is warranted to increase the transparency in estimating BCC app effects.

Quantifying costs. The Second Panel identified two types of costs associated with health interventions: 1) costs for production, delivery and consumption of the intervention, and 2) costs used or saved as a consequence of the intervention. Time is an important consideration in valuing short and long-term costs. With BCC apps, it is not feasible to wait for lifetime data to validate cost-effectiveness. Hence, the need for modelling techniques that can simulate costs and outcome effects over long periods beyond the trial.[7] To overcome the uncertainties of modeling, some analysts make costs and outcome consistency assumptions which may be flawed (Figure 2). This example was shown by Sjöström et al. who assumed that costs and utility weights measured at 3-month follow-up would remain constant over the year based on previous studies of internet-based pelvic floor muscle training where improvements after 3 months were maintained after 1 and 2 years.[35] While this assumption may be reasonable, it does not account for the difference in user engagement or dependence on human resources. Consequently, sensitivity analyses are necessary to evaluate cost and effect uncertainties, the omission of which constitutes a bias. Consistent with the Second Panel's recommendation, all the studies performed at least one sensitivity analysis to address uncertainties.

BCC apps can also have effects outside the health sector from changes in individual behaviors, and the associated costs or savings can accrue to the broader society.[27] Identifying all costs attributable to the BCC app is challenging, and inconsistencies arise when decisions on what costs to include is at the analyst's discretion based on the "Rule of Reason". The major limitations in costing cited by Armstrong et al. were micro versus macro costing, and inadequate justifications for including present and future costs resulting from the app.[5] Micro-costing is a cost estimation method that includes the precise estimation of every input in the intervention, whereas macro-costing only considers broader-level costing whose omission will influence the overall cost. An appropriate costing method is important for deciding which cost data to include in order to increase transparency and transferability of CEA findings. Cost data used in the studies included estimates from accounting departments, patient cost databases, insurance billing codes and business plans. These data sources were inconsistent across studies, which could lead to incorrect assumptions in estimating the costs and effects of different BCC apps.

Lack of comparative data in CEA studies can result in low quality of cost measures (Figure 2).[43] International health technology assessment agencies for economic evaluations recommend that health technology interventions be compared with their "usual care" alternatives to increase transparency in assessments.[44] However, poor comparative data exacerbates problems with comparing BCC apps with their usual care comparators. This is because usual care treatments vary substantially in behavioral health interventions due to the lack of a 'gold-standard' treatment for specific health behaviors.[35] Furthermore, cost-effectiveness data on "usual care" are scarce due to limited number of studies, and equivalency assumptions between interventions evaluated at different times should likewise be used with caution. This challenge is exemplified by Armstrong et al, who assumed equivalency in the effectiveness of mobile app and in-person follow-up care based on observational studies of telephone follow-up care from similar ambulatory surgical patients.[5] Since no RCT had demonstrated equal effectiveness between mobile app and in-person follow-up care, the BCC app could only be considered cost-effective in this context given the stated assumptions. The difficulty with generalizing evidence from evaluations in specific contexts, highlights the importance with balancing local applicability of BCC apps with the growing globalization of mHealth apps. It is imperative that assessments for scale recognize that a BCC app may be cost-effective in one context but highly expensive or ineffective in another context where access and resources are limited.

Estimating incremental cost-effectiveness ratio (ICER) in BCC apps. An appropriate comparison between two health interventions should be in terms of ICER.[42] Consistent with this recommendation, disaggregated cost and outcome measures are to be summarized into a single quantitative measure like ICER.[27] A review of the studies showed no uniformity in estimating single quantitative measures for cost-effectiveness. For example, Van Reijen et al. expressed the cost-effectiveness of the Strengthen your Ankle app in "cost per injury incidence density",[33] whilst Amstrong et al. expressed cost-effectiveness of the BYOD app in "incremental net benefit".[5] Likewise, Ryan et al. estimated the cost-effectiveness of the t+ Asthma app in "cost per mini-asthma quality of life".[21] While estimating cost per QALY for BCC apps is arduous for analysts due to intermediate outcomes, the lack of a single quantitative measures for BCC apps has far reaching implications for decisions regarding adoption and scale. It also raises questions about the appropriate use of economic evaluation data in healthcare such as comparability with other studies, BCC app pricing, consumer willingness to pay, and potential health insurance reimbursements.

Another concern with estimating ICER lies with the appropriateness in using an ICER threshold, in which a cut-off point is considered for the cost per QALY ratio. The examples above illustrate the dilemma of cost-QALY thresholds when many BCC apps produce intermediate outcomes. The lower the ratio of a cost per QALY, the more cost-effective the health intervention is. Generally, the ICER has no theoretical or empirical basis, but values ranging from $50,000 to $100,000 are a considerable threshold for cost-effectiveness in the US.[45] However, it is unclear whether the use of ICER thresholds in estimating the cost-effectiveness of BCC apps encapsulates all the relevant criteria a decision maker may require for adoption and implementation. While the ICER allows for easy comparison across studies, its applicability for resource allocation is not always straightforward. For example, BCC apps have the potential to reduce health inequities by increasing healthcare access to hard-to-reach populations. However, the use of costs and QALYs in estimating their cost-effectiveness does not automatically consider equity in healthcare access, which may be an important decision for adoption. Methods for incorporating important behavioral variables in estimating the cost-effectiveness of BCC apps are yet to be addressed.

Limitations

This study provides an important overview of the cost-effectiveness landscape of BCC apps, and highlights methodological challenges for conduction CEA. The methodological challenges highlighted may not be generalizable to all types of mHealth apps, although it can serve as a guide. We searched through four major databases to retrieve the final selected studies, and therefore it is possible that a few key studies may have been missed. Articles not published in English were also omitted.

Conclusion

Institutionalizing BCC apps into routine behavioral health intervention could likely increase their population health effects. However, evidence backing the economic value of BCC apps is limited. In this paper, we sought to highlight the economic evaluation characteristics of BCC apps and report limitations for conducting CEA using the Second Panel's key guidelines. The review showed that economic evaluation studies on BCC apps attempt to quantify the costs and behavioral outcomes associated with using apps, but the majority of studies were fraught with study design challenges stemming from limited cost data and intermediate effects which sometimes have no clinically-relevant endpoints. Studies that use CEA methods also face challenges in adhering to recommended guidelines provided by the Second Panel. While economic evaluations are necessary to generate the evidence needed for scaling BCC apps, it is believed that more attention should first be given to the challenges in adhering to rigorous methodological guidelines. Furthermore, practical approaches that make conducting economic evaluations of BCC apps more feasible should be identified and prioritized. For example, economic evaluation methods should be linked to a set of appropriate measurement strategies and relevant data sources that factors in the unique attributes of BCC apps including intermediate outcomes and limited cost data. Only then can analysts pursue more complex economic assessments that can increase the transparency, quality and usefulness of the evidence. Until economic evaluations of BCC apps are conducted in accordance with methods that balance data completeness with practical concerns, decisions made on this body of evidence will be regarded with considerable caution.

Contributorship: CH researched the literature and conceived the study. MP, EG, and SB were involved in protocol development, gaining ethical approval and data analysis. CH wrote the first draft of the manuscript, and all authors reviewed and edited the manuscript for intellectual content and approved the final version of the manuscript. CH led the review efforts and MP gave the final approval.

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: The Colorado Multiple Institutional Review Board (COMIRB) approved this study (Protocol ID: 17-0347).

Funding: The author(s) received no financial support for the research, authorship, and/or publication of this article.

Guarantor: CH.

Peer review: Ivo Vlaev has reviewed this manuscript.

ORCID iD: Cynthia Afedi Hazel https://orcid.org/0000-0001-5757-250X

Supplemental material: Supplemental material for this article is available online.

References

1. Smith AUS. *Smartphone use in 2015*. USA: Pew Research Center: Internet, Science & Tech. 2015.
2. Flores Mateo G, Granado-Font E, Ferré-Grau C, et al. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. *J Med Internet Res* 2015; 17: e253.
3. Levine R, Corbacio A, Konopka S, et al. *mHealth compendium*. Vol. 5. Africa: African Strategies for Health, Management Sciences for Health, 2015.
4. Hale K, Capra S and Bauer J. A framework to assist health professionals in recommending high-quality apps for supporting chronic disease self-management: illustrative assessment of type 2 diabetes apps. *JMIR mHealth uHealth* 2015; 3: e87.
5. Armstrong KA, Semple JL and Coyte PC. Replacing ambulatory surgical follow-up visits with mobile app home monitoring: modeling cost-effective scenarios. *J Med Internet Res* 2014; 16: e213.
6. van der Hout A, van Uden-Kraan CF, Witte BI, et al. Efficacy, cost-utility and reach of an eHealth self-management application 'oncokompas' that helps cancer survivors to obtain optimal supportive care: study protocol for a randomised controlled trial. *Trials [Electronic Resource]* 2017; 18: 228.
7. Iribarren SJ, Cato K, Falzon L, et al. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. *PLoS One* 2017; 12: e0170581.
8. Payne HE, Lister C, West JH, et al. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. *JMIR mHealth Uhealth* 2015; 3: e20.
9. Patrick K, Raab F, Adams MA, et al. A text message-based intervention for weight loss: randomized controlled trial. *J Med Internet Res* 2009; 11: e1.
10. Azar KMJ, Lesser LI, Laing BY, et al. Mobile applications for weight management: theory-based content analysis. *Am J Prev Med* 2013; 45: 583–589.
11. Voth EC, Oelke ND and Jung ME. A theory-based exercise app to enhance exercise adherence: a pilot study. *JMIR mHealth Uhealth* 2016; 4: e62.
12. Free C, Phillips G, Watson L, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. *PLoS Med* 2013; 10: e1001363.
13. Siwicki B. Mobile apps emerging as essential population health tools. *Healthcare IT News*. 2016. March 15.
14. Maher CA, Lewis LK, Ferrar K, et al. Are health behavior change interventions that use online social networks effective? A systematic review. *J Med Internet Res* 2014; 16: e40.
15. Woldaregay AZ, Issom DZ, Henriksen A, et al. Motivational Factors for User Engagement with mHealth Apps. *Stud Health Technol Inform* 2018; 249: 151–157. PMID: 29866972.
16. Koenker R and Hallock KF. Quantile regression. *J Econ Perspectives* 2001; 15: 143–156.
17. Whitten PS, Mair FS, Haycox A, et al. Systematic review of cost effectiveness studies of telemedicine interventions. *BMJ* 2002; 324: 1434–1437.
18. Badawy SM and Kuhns LM. Economic evaluation of text-messaging and smartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. *JMIR mHealth Uhealth* 2016; 4: e121.
19. de la Torre-Díez I, López-Coronado M, Vaca C, et al. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. *Telemed J E Health* 2015; 21: 81–85.
20. Neumann PJ, Sanders GD, Russell LB, Siegel JE, and Ganiats TG. *Cost-effectiveness in health and medicine*. 2nd ed. Oxford: Oxford University Press, 2016, p.536.
21. Ryan D, Price D, Musgrave SD, et al. Clinical and cost effectiveness of mobile phone supported self monitoring of asthma: multicentre randomised controlled trial. *BMJ* 2012; 344: e1756.
22. Ross J. The use of economic evaluation in health care: Australian decision makers' perceptions. *Health Policy* 1995; 31: 103–110.
23. Rabarison KM, Bish CL, Massoudi MS, et al. Economic evaluation enhances public health decision making. *Front Public Health* 2015; 3: 164.
24. Hoffmann C and Graf von der Schulenburg JM. The influence of economic evaluation studies on decision making. A European survey. The EUROMET group. *Health Policy (Amsterdam, Netherlands)* 2000; 52: 179–192.
25. YHEC. *Incremental cost-effectiveness ratio (ICER)*. UK: YHEC, 2016.
26. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA* 2016; 316: 1093–1103.
27. Neumann PJ, Sanders GD, Russell LB, et al. *Cost-effectiveness in health and medicine*. 2nd ed. Oxford: Oxford University Press, 2017.
28. Velden MEv, Severens JL and Novak A. Economic evaluations of healthcare programmes and decision making. *PharmacoEconomics* 2005; 23: 1075–1082.
29. Sloan FA, Whetten-Goldstein K and Wilson A. Hospital pharmacy decisions, cost containment, and the use of cost-effectiveness analysis. *Soc Sci Med (1982)* 1997; 45: 523–533.
30. Duthie T, Trueman P, Chancellor J, et al. Research into the use of health economics in decision making in the United Kingdom – phase II. *Health Policy* 1999; 46: 143–157.
31. Frick KD. Aspects of cost-effectiveness analysis for managerial decision making. *J Nurse Pract* 2012; 8: 145–149.
32. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6: e1000097.

33. Van Reijen M, Vriend I, van Mechelen W, et al. Preventing recurrent ankle sprains: is the use of an app more cost-effective than a printed booklet? Results of a RCT. *Scand J Med Sci Sports* 2017; 19: 19.
34. Martin EG, Helbig N and Shah NR. Liberating data to transform health care: New York's open data experience. *JAMA* 2014; 311: 2481–2482.
35. Sjostrom M, Lindholm L and Samuelsson E. Mobile app for treatment of stress urinary incontinence: a cost-Effectiveness analysis. *J Med Internet Res* 2017; 19: e154.
36. Dahlberg K, Philipsson A, Hagberg L, et al. Cost-effectiveness of a systematic e-assessed follow-up of postoperative recovery after day surgery: a multicentre randomized trial. *Br J Anaesth* 2017; 119: 1039–1046.
37. Kumar S, Nilsen WJ, Abernethy A, et al. Mobile health technology evaluation: the mHealth evidence workshop. *Am J Prev Med* 2013; 45: 228–236.
38. Dobkin HE and Zirwas M. Price discrepancies among mobile medication applications for common dermatologic prescriptions: observational cost analysis. *J Am Acad Dermatol* 2017; 77: 1181–1182.
39. Newbould J, Abel G and Ball S. Evaluation of telephone first approach to demand management in English general practice: observational study. *BMJ* 2017; 358: j4197.
40. Luxton DD, Hansen RN and Stanfill K. Mobile app self-care versus in-office care for stress reduction: a cost minimization analysis. *J Telemed Telecare* 2014; 20: 431–435.
41. Cano Martin JA, Martinez-Perez B, de la Torre-Diez I, et al. Economic impact assessment from the use of a mobile app for the self-management of heart diseases by patients with heart failure in a Spanish region. *J Med Syst* 2014; 38: 96.
42. Drummond MF, Sculpher MJ, Claxton K, et al. *Methods for the economic evaluation of health care programmes*. 4th ed. Oxford: Oxford University Press, 2015, p.464.
43. Thompson-Coon J, Abdul-Rahman A-K, Whear R, et al. Telephone consultations in place of face to face outpatient consultations for patients discharged from hospital following surgery: a systematic review. *BMC Health Serv Res* 2013; 13: 128.
44. Mathes T, Jacobs E, Morfeld J-C, et al. Methods of international health technology assessment agencies for economic evaluations – a comparative analysis. *BMC Health Serv Res* 2013; 13: 371.
45. Nuijten MJC and Dubois DJ. Cost-utility analysis: current methodological issues and future perspectives. *Front Pharmacol* 2011; 2: 29.