Public Health Impact of Scalable Physical Activity Interventions: Insights from Two Field Studies in Switzerland

DISSERTATION
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor of Philosophy in Management

submitted by

Jan-Niklas Kramer

from

Germany

Approved on the application of

Prof. Dr. Elgar Fleisch

and

Prof. Dr. Urte Scholz

Dissertation no. 4997

Difo Druck GmbH, Untersiemau 2020

The University of St. Gallen, School of Management, Economics, Law, Social Sciences, and International Affairs hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, May 18, 2020

The President:

Prof. Dr. Bernhard Ehrenzeller

Acknowledgements

This dissertation is the result of my work at the Institute of Technology Management at the University of St. Gallen (ITEM) in the period from 2016 to 2019, where I was part of the Center for Digital Health Interventions, a joint research initiative of the University of St. Gallen and ETH Zurich. First and foremost, I would like to thank my supervisor Prof. Dr. Elgar Fleisch for giving me the opportunity to pursue a PhD in this unique and highly interdisciplinary environment. His focus on using research as a tool to solve problems that matter guided and shaped my work at this chair and enabled a steep learning curve. I would like to thank my co-supervisor Prof. Dr. Urte Scholz for her guidance, feedback and support of my work. Our discussions, in person or via email, greatly helped me to focus on the aspects of my work that really mattered. I also like to thank Prof. Dr. Tobias Kowatsch for his engagement and support. From him I learned how much can be achieved independently and with hard work. At the University of St. Gallen, Elisabeth Vetsch-Keller and Sonja Baumgartner deserve my deepest thanks for providing tremendous support in dealing with the administrative and often tedious matters.

A massive thank you goes to all my colleagues and friends at the University of St. Gallen and ETH Zurich. Special thanks go to the notorious "Health Angels", Dominik Ruegger, Filipe Barata, Florian Künzler, Iris Shih and Peter Tinschert who shared the same fate that I did and made my time as a PhD truly memorable. Peter Tinschert deserves special recognition for bearing with all my daily struggles. Many other people have supported my work without whom the research reported in this dissertation would not have been possible. These include Matthias Heuberger, Gabriella Chiesa and Niklas Elser who worked relentlessly to bridge the gaps between research and industry, Shawna Smith for her valuable advice on adaptive interventions and micro-randomized trials and Dr. Tobias Bucher for outstanding mentoring and advice.

Finally, I would like to thank my parents Willi and Michaela and my sister Andrea for their support and for encouraging the courageous decision of pursuing a PhD. Most especially, I would like to thank my wonderful girlfriend Sophia for her unconditional love and support over the years.

Jan-Niklas Kramer, St.Gallen, October 2019

Abstract

Physical inactivity is a key risk factor for non-communicable diseases (NCDs). Recent advances in the field of mobile technology have facilitated the design and implementation of scalable physical activity programs and enabled physical activity promotion at the population level. Against this background, this dissertation aimed to investigate the potential of mobile physical activity interventions to serve as an effective tool for the prevention of NCDs. Using a reward-based physical activity program of a Swiss health insurer as a starting point, this objective was accomplished by (1) defining the relevant outcomes and target effect sizes for mobile physical activity interventions (2) evaluating the program with regard to its ability to achieve the defined targets, and (3) revising and optimizing the program. Models of public health impact revealed that mobile physical activity interventions should maximize uptake among the target population and increase physical activity by at least 1,500 steps per day on average. A participation rate of 10% was defined as a lower threshold for intervention uptake. Subsequently, the first empirical study (N = 1,547) revealed that the insurer's program is unlikely to meet this target. Leveraging effects of small financial incentives, the program reached only 6% of the targeted population and incentives did not support behavior change. More than 25% of participants quit the program during the six-months study and the program's requirement to own an activity tracker limited uptake among insurees. Based on these insights, the insurer's program was revised and a smartphone app was developed that combined physical activity monitoring with redesigned financial rewards and a digital physical activity coaching. The second field study (N = 274) revealed that the redesigned financial incentives increased physical activity by almost 800 steps while the digital coaching had no effect. Again, 30% of participants stopped using the app during the eight-week study and only a subset of participants engaged with the app. However, those that did engage with the app seemed to benefit. Participants in both studies were more active and healthier compared to the Swiss population. Collectively, these results suggest that mobile physical activity interventions can outperform traditional physical activity interventions but their contribution to the prevention of NCDs is small. Limited reach, selection effects and high attrition are key factors that prevent these interventions from contributing to public health. Implications for designing interventions and opportunities for future work are discussed.

Zusammenfassung

Körperliche Inaktivität ist ein bekannter Risikofaktor für nichtübertragbare Krankheiten (NCDs). Fortschritte im Bereich mobiler Sensorik haben es möglich gemacht, skalierbare Bewegungsinterventionen zu entwickeln, die eine Steigerung der körperlichen Aktivität auf Bevölkerungsebene ermöglichen. Vor diesem Hintergrund untersucht diese Arbeit das Potenzial skalierbarer Bewegungsinterventionen, dem Auftreten von NCDs wirksam zu vorzubeugen. Ausgehend von einem bereits entwickelten Bewegungsprogramm eines Schweizer Krankenversicherers werden (1) die relevanten Outcomes und Effektstärken für skalierbare Bewegungsinterventionen definiert, (2) das Bewegungsprogramm im Hinblick auf diese Ziele evaluiert und (3) das Bewegungsprogramm überarbeitet und optimiert. Effektszenarien zeigen, dass skalierbare Bewegungsinterventionen ihre Reichweite maximieren und körperliche Aktivität um mindestens 1.500 Schritte pro Tag steigern sollten, um NCDs wirksam vorzubeugen. Eine Teilnahmequote von 10% wurde als minimale Reichweite definiert. Die erste empirische Studie (N = 1.547) zeigte, dass das Bewegungsprogramm dieses Ziel wahrscheinlich nicht erreicht. Auch unter dem Einsatz finanzieller Anreize erreichte das Programm nur 6% der Zielpopulation. Die Voraussetzung, einen Aktivitäts-Tracker zu besitzen, um am Programm teilnehmen zu können, schränkte die Teilnahme stark ein. Darüber hinaus beendeten mehr als 25% der Versicherten ihre aktive Teilnahme am Programm während der sechsmonatigen Studie. Basierend auf diesen Erkenntnissen wurde das Programm überarbeitet. Zu diesem Zweck wurde eine Smartphone-App entwickelt, welche neu gestaltete finanzielle Anreize mit einem digitalen Bewegungscoaching kombiniert. In einer zweiten Feldstudie (N = 274) konnten die neu gestalteten finanziellen Anreize die körperliche Aktivität der Teilnehmer um fast 800 Schritte erhöhten, während das digitale Coaching keine Effekte erzielte. Erneut stellten 30% der Teilnehmer die Nutzung der App während der Studie ein. Die Teilnehmer beider Studien waren aktiver und gesünder als die Schweizer Bevölkerung. Zusammenfassend deuten diese Ergebnisse darauf hin, dass skalierbare Bewegungsinterventionen zwar bessere Ergebnisse liefern als klassische Interventionen, der Nutzen im Bereich Prävention aber begrenzt ist. Fehlende Reichweite, Selektionseffekte und ein hoher Drop-out der Teilnehmer verhindern aktuell das Erreichen der notwendigen Effektgrössen. Implikationen für Interventionsdesign sowie Ansatzpunkte für weitere Forschung werden diskutiert.

Previous Publications

This dissertation contains material, which has already been published or has recently been submitted for publication by myself and colleagues as scientific articles in peer-reviewed journals or conference proceedings. The core findings of this dissertation are built upon two studies, which are described in chapter 4 and chapter 6 and are reported in the following publications:

Kramer, J.-N. & Kowatsch, T. (2017). Using Feedback to Promote Physical Activity: the Role of the Feedback Sign. *Journal of Medical Internet Research, 19*(6), e192.

Kramer, J.-N., Tinschert, P., Scholz, U., Fleisch, E., & Kowatsch, T. (2019). A Cluster-randomized Trial on Small Incentives to Promote Physical Activity. *American Journal of Preventive Medicine*, *56*(2), e45-e54.

Kramer, J.-N., Künzler, F., Mishra, V., Kotz, D., Smith, S. N., Scholz. U., Fleisch, E. & Kowatsch, T. (under review). Which Components of a Smartphone Walking App Help Users to Reach Personalized Step Goals? Results from an Optimization Trial. *Annals of Behavioral Medicine.*

The rationale, design and methodology of these studies are described in detail in two corresponding study protocols, which are published:

Kowatsch, T., **Kramer, J.-N.**, Kehr, F., Wahle, F., Elser, N. & Fleisch, E. (2016). Effects of Charitable Versus Monetary Incentives on the Acceptance of and Adherence to a Pedometer-Based Health Intervention: Study Protocol and Baseline Characteristics of a Cluster-Randomized Controlled Trial. *JMIR Research Protocols, 5*(3)*,* e181.

Kramer, J.-N., Künzler, F., Mishra, V., Presset, B., Kotz, D., Smith, S. N., Scholz. U. & Kowatsch, T. (2019). Investigating Intervention Components and Exploring States of Receptivity for a Smartphone App to Promote Physical Activity: Protocol of a Microrandomized Trial. *JMIR Research Protocols, 8*(1)*,* e11540.

Further, results from the analysis of engagement metrics described in chapter 6 have been presented at a scientific conference. The abstract of this conference presentation is published online:

Kramer, J.-N., Künzler, F., Tinschert, P. & Kowatsch, T. (2019, February). *Trajectories of Engagement with a Digital Physical Activity Coach: Secondary Analysis of a Micro-Randomized Trial*. Paper presented at the 10th Scientific Meeting of the International Society for Research on Internet Interventions (ISRII), Auckland, New

Zealand. Abstract retrieved from: https://uoaevents.eventsair.com/QuickEventWebsitePortal/isrii2019/ online-programme/Agenda

I hereby declare that most of the content of the abovementioned manuscripts, including those without first authorship, has been written by myself. Of course, my co-authors contributed substantially to these manuscripts with their feedback, reviews, edits and changes. Accordingly, parts of this dissertation can bear strong resemblance or correspond literally to my own previously published work. The findings presented in the abovementioned papers do not contribute to dissertations that are written by some of my co-authors. However, additional data collected in the two studies that are not relevant for the findings of this dissertation contribute to research projects of my co-authors. Consequently, descriptions of study design and procedures in this dissertation may inadvertedly show similarities to corresponding descriptions in dissertations of my co-authors.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMSTAR	Assessing the Methodological Quality of Systematic Reviews
API	Application Programming Interface
BAG	Federal Ministry of Health
BCT	Behavior Change Technique
BREQ-2	Behavioral Regulation in Exercise Questionnaire 2
CDHI	Center for Digital Health Interventions
CHD	Coronary Heart Disease
CI	Confidence Interval
GEE	Generalized Estimating Equation
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
HBCP	Human Behavior Change Project
HR	Hazard Ratio
IF	Potential Impact Fraction
JITAI	Just-in-time Adaptive Intervention
MOST	Multiphase Optimization Strategy
MET	Metabolic Equivalent
NCD	Non-communicable Disease

OR	Odds Ratio
RCT	Randomized Controlled Trial
RQ	Research Question
RR	Risk Ratio
SOBC	Science of Behavior Change
T2DM	Type-2 Diabetes Mellitus
WHO	World Health Organisation
XML	Extensible Markup Language

Chapter 1

Introduction

"Lack of activity destroys the good condition of every human being, while movement and methodical physical exercise save it and preserve it."

Plato, *Theaetetus*

This chapter outlines the general motivation of this dissertation and derives the research questions that this dissertation seeks to answer. Further, the methodological approach to answering the research questions is briefly described. The chapter concludes with an outline of the remainder of the dissertation.

1.1 Motivation

In 1953, when Jeremiah Morris published his analysis of staff statistics from the London Transport Executive and the General Post Office in the medical journal *The Lancet* (J N Morris, Heady, Raffle, Roberts, & Parks, 1953), he likely could not have anticipated the full impact of his publication on the fields of medicine and public health. In his article, Morris compared the detailed recorded mortality rates for drivers and conductors of the famous red double-decker busses in London between 1949 and 1952 (the design of the double-decker busses at the time separated the bus driver from the passengers, thus, busses required a two-person crew: the driver and the conductor, who sold tickets, supervised loading and unloading, and communicated with passengers). Morris discovered that mortality rates due to cardiovascular disease were lower among conductors, who were quite active, than among the sedentary bus drivers. The same pattern emerged for active postmen compared to inactive telephonists (switchboard operators) at the General Post Office. Although hypothesized for a long time (cf. the introductory quote of Plato), J N Morris et al. (1953) is the earliest empirical record that demonstrated a positive relationship between physical activity and health. It led to more than six decades of intensive research that investigated the role of physical activity for health. More specifically, following the publication of Morris' findings, physical activity researchers over the second half of the 20th century were determined to answer three important questions: (1) What are the health-related benefits of physical activity?; (2) How much physical activity is necessary to achieve these benefits?; (3) How can we best support people in reaching the recommended levels of physical activity?

In 2012, *The Lancet*, the very same journal that published Morris' original article, published a series of papers aimed at summarizing the empirical evidence that had accumulated over time pertaining to these three questions. This paper series was updated once more in 2016. The papers included provide an impressive account of the positive effects of physical activity and, correspondingly, the negative effects of physical inactivity on health. Lee et al. (2012) estimate that physical inactivity causes between 6% and 10% of cases of the four major non-communicable diseases (NCDs), coronary heart disease, type 2 diabetes and breast and colon cancer, which are the leading cause of mortality and altogether account for more than 50% of deaths worldwide (World Health Organization [WHO], 2014). These findings make physical inactivity a health risk of comparable importance to established risk factors, such as smoking and obesity. Other authors found that physical activity can play a major role in the prevention of depression (Choi et al., 2019) and has positive effects on mood, functional capacity and health-related quality of life in certain populations (Penedo & Dahn, 2005). In addition, Ding et al. (2016) estimated the economic burden of physical inactivity and concluded that in 2013 alone the world economy spent $67.5 billion on healthcare expenditures and productivity losses due to insufficient physical activity. This number is equal to the gross domestic product of Costa Rica, which was ranked 80 out of 193 countries in the same year. Thus promoting physical activity has emerged as a key strategy for the prevention of NCDs because of its immense potential to affect a variety of health outcomes and lower worldwide healthcare expenditure.

Over the years, researchers have documented an inverse and curvilinear relationship between physical activity and health with the greatest benefits occurring for inactive individuals who become more active. A review of 73 studies within a scientific report used to shape the US physical activity guidelines concluded that 2 to 2.5 hours of moderate activity or walking per week are necessary to achieve significant health benefits, although inactive populations can achieve health benefits at lower activity levels (Physical Activity Guidelines Advisory Committee, 2008). This led the authors to endorse a "some is good; more is better" message (p. G1-14). The results of this report have largely been adopted by other national and international physical activity recommendations (e.g. WHO, 2010), which today typically promote a minimum recommendation of 150 minutes of moderate-to-vigorous aerobic physical activity per week[1]. This recommendation also helped to identify inactive individuals and populations, i.e. those who typically perform less than the recommended level of physical activity. Current estimates reveal that 23% of the adult population worldwide does not meet the recommended physical activity levels (Sallis et al., 2016), illustrating the urgent need for effective physical activity interventions.

Consequently, researchers have developed and tested interventions aimed at helping people change their behavior and become more active. These interventions typically consist of combinations of group-based, telephone or face-to-face counselling, supervised or unsupervised physical activity sessions, educational and motivational material, and self-monitoring. However, although positive examples do exist, meta-analyses of physical activity interventions indicate that the average effects of interventions are small (Conn, Hafdahl, & Mehr, 2011; Foster, Hillsdon, Thorogood, Kaur, & Wedatilake, 2005). For example, a meta-analysis of 358 study reports comprising 99,011 participants estimated a standardized mean intervention effect of $d = 0.19$ in healthy adults (Conn et al., 2011), which can be considered a small effect according to the guidance on effect sizes by Cohen (1988). Some reviews suggest that interventions can produce somewhat greater effects in children (Beets, Beighle, Erwin, & Huberty, 2009) or obese populations (Gourlan, Trouilloud, & Sarrazin, 2011). What is more, even if researchers develop powerful interventions, these are rarely implemented and scaled-up in practice: a 2016 review could identify only 16 researcher-led interventions with findings published in the peer-reviewed literature that were implemented and scaled-up in practice (Reis et al., 2016). This is not comparable to the large number of physical activity interventions that are evaluated each year by scientists. This number steadily increased from around 50 randomized controlled trials (RCTs) in

[1] Moderate-to-vigorous physical activity (MVPA) is defined as physical activity of 3 to 6 metabolic equivalents (MET), i.e. activities that burn 3 to 6 times as much as energy as sitting quietly (Ainsworth et al., 2011). These are activities that, for most people, increase breathing and heart rate while still allowing them to have a conversation. Examples include brisk walking or bicycling.

2000 to more than 500 RCTs in the year 2015 (Dimensions, 2019)[2]. This lack of translation into practice may be attributed (at least in part) to the large amount of resources these evaluated interventions demand, which makes them difficult to deliver and scale-up under real-world conditions (Foster et al., 2005). Thus, research on earlier physical activity interventions illustrates low average intervention effects and a substantial research-to-practice gap.

The conventional approach to designing physical activity interventions was fundamentally challenged in the year 2013, when Apple introduced the iPhone 5s and Google released version 4.4 of its Android smartphone operating system. Both companies introduced new technologies that allowed their smartphones to deal with continuous collection and processing of sensor data without severely impacting the smartphone's battery consumption. While Apple used a separate processor to deal with continuous sensor data (Sumra, 2013), Google shifted from continuous processing of sensor events to batch-processing, i.e. processing of data in pre-defined time intervals (Google, 2013). Both innovations enabled continuous processing of the smartphone's accelerometer data and equipped millions of people with ubiquitous activity monitors. Apple and Google provided application programming interfaces (APIs) that enabled app developers to access pre-classified raw sensor data, most notably the number of steps individuals walked, which could be detected easily and with acceptable accuracy (Case, Burwick, Volpp, & Patel, 2015). As a result, smartphone apps aimed at measuring, monitoring and increasing physical activity entered the app stores, and by 2017 the number of apps in the category "exercise and fitness" amounted to more than 95,000 (Aitken, 2017).

This development coincided with – and possibly interacted with – other trends that promoted awareness of physical activity and large-scale adoption of physical activity monitoring. These trends primarily include: the rise of wearable devices (i.e. clips, smartwatches and wristbands that connect to smartphone apps and are capable of tracking physical activity); the quantified-self movement, an international community in which people generate personal insights through extensive tracking of their own physiological, behavioral or environmental information (Nafus & Sherman, 2014); and

[2] These numbers were calculated using the Dimensions database and provide rough estimates only. Dimensions is a scholarly search engine (similar to Google Scholar) that provides advanced search metrics. Titles and abstracts of scientific publications were searched using the following search string: "physical activity" AND "randomized controlled trial" NOT "review" NOT "meta-analysis".

step-based physical activity recommendations, such as the goal of taking 10,000 steps per day, which was popularized in the media and endorsed by health organizations, for example, in Australia, the US and Japan (Tudor-Locke et al., 2011). The adoption of wearables, in particular, has continued to grow rapidly, and the number of activity trackers sold is expected to almost double in the coming years, from 55 million devices sold in 2016 to 105 million devices in 2022 (Statista, 2017). Currently, around 10% of the Swiss population reports owning an activity tracker (BVDW, 2016). Different businesses have further facilitated the growing adoption of physical activity apps and wearables by rolling out comprehensive physical activity promotion programs designed around apps and activity trackers. Operators of such programs primarily include wearable manufacturers (such as Fitbit or Garmin) – which hope to boost sales of their devices – or employers and insurance companies, whose goal is to improve the health of employees and insurees in order to reduce productivity losses and healthcare costs respectively (Comstock, 2014).

Of course, this development has not gone unnoticed by the research community. Over the last years, the academic field of mobile health (mHealth) has emerged. This field investigates "the use of mobile telecommunication technologies for the delivery of health care and in support of wellness" (Steinhubl, Muse & Topol, 2013, p. E1) and its emergence has been accompanied by the publication of dedicated scientific journals (e.g. JMIR mHealth & uHealth). Researchers noted that mobile technologies, such as smartphones and wearables, indeed hold great potential for physical activity interventions. First and foremost, they enable effortless and continuous monitoring of and automatized feedback on physical activity, which researchers have identified as key factors in the process of behavior change (Harkin et al., 2016; Michie, Abraham, Whittington, McAteer, & Gupta, 2009). Second, unlike non-mobile interventions, smartphones and wearables allow for the provision of interventions and support largely independent of time and location, thereby simplifying intervention access. Third, mobile devices, and smartphones in particular, can actively or passively enrich behavioral data with contextual meta-data, such as the user's location, activity, weather, mood or stress. These contextual annotations can subsequently be used to design adaptive interventions by tailoring intervention delivery to the user's situation and thus maximizing effectiveness (Nahum-Shani et al., 2016). Fourth, given the widespread adoption of smartphones and activity trackers among the population, as well as the number of existing programs offered by businesses, mobile technologies might offer an opportunity to develop scalable interventions that can be implemented through a manageable amount of effort and resources. Thus, the rise of physical activity smartphone apps and wearables opens up the opportunity to develop new, scalable physical activity interventions that seem destined to overcome the limitations of traditional interventions, such as limited effectiveness and low adoption in practice.

Because these interventions can potentially serve a large number of people, they could be a powerful tool to fight the rising prevalence of the major NCDs. Indeed, the need for large-scale NCD-prevention programs has been repeatedly identified (Burke et al., 2015; Kvedar, Fogel, Elenko, & Zohar, 2016; WHO, 2014) and the implementation of large-scale prevention programs enabled by mobile technology is already being pushed forward. Just recently, the government of Singapore announced a partnership with the wearable manufacturer Fitbit to provide its citizens with activity trackers at no cost on the condition that they sign up to Fitbit's digital coaching service (Somauroo, 2019). Earlier examples include community-wide 10,000 steps projects in Belgium (De Cocker, De Bourdeaudhuij, Brown, & Cardon, 2007) and Australia (Brown, Eakin, Mummery, & Trost, 2003), which used various strategies to support citizens in reaching the activity goal of 10,000 steps per day, including providing pedometers at low cost to participants. Clearly, there is a need to evaluate whether such innovative programs can indeed contribute to the prevention of NCDs, to identify what challenges these programs need to overcome and how they can best be designed in order to serve as powerful NCD prevention tools.

Against this background, this dissertation examines the potential public health impact of scalable physical activity interventions that use mobile technologies[3]. Public health impact is broadly defined in the literature as an intervention's effect on the health of a population (WHO, 1999) and typically quantified as the reduction of disease or mortality burden that can be expected following the exposure of the population to the intervention (Mindell, Ison, & Joffe, 2003). This dissertation focuses in particular on the ability of scalable physical activity interventions to effectively prevent major physical-activity-related NCDs, such as cardiovascular diseases and diabetes. Scalable physical activity interventions are understood as physical activity interventions using mobile technologies (e.g. wearables or smartphones) that can be rolled out to a growing number of people without a considerable increase in resources (sometimes referred to as

[3] Focusing on scalable physical activity interventions enabled by mobile technologies, the terms *scalable physical activity intervention* and *mobile physical activity intervention* are used interchangeably throughout this dissertation.

standalone interventions). To judge the potential public health impact of scalable physical activity interventions, a novel physical activity program from Switzerland is used as an example.

1.2 Approach & Research Questions

This dissertation and the corresponding research activities were conducted as part of a collaboration between the Center for Digital Health Interventions (CDHI) and a Swiss health insurance company that developed a mobile physical activity promotion program in 2015. The research presented in this dissertation accompanies the implementation of this physical activity program by answering four distinct research questions. Answering these research questions reveals important insights regarding strengths and challenges of mobile physical activity interventions which are the basis to judge their potential public health impact. The research in this dissertation is guided by the Multiphase Optimization Strategy (MOST; Collins, 2018), a framework for the development of effective and scalable behavior change interventions that divides the intervention development process into three phases, preparation, optimization and evaluation (Collins, 2018). The research presented in this dissertation is summarized in Figure 1-1.

This dissertation starts by asking for the requirements that need to be fulfilled by mobile physical activity interventions in order to contribute to disease prevention. This includes the definition of relevant outcomes as well as target effect sizes for each outcome that should be achieved by the intervention so that a contribution to public health can be considered probable. Correspondingly, the first research question (RQ) can be formulated as follows:

RQ 1: What are relevant outcomes and target effect sizes for scalable physical activity interventions?

This research question is crucial as its answer provides the benchmark against which the insurer's program is evaluated in the subsequent steps. To answer this research question, this dissertation adopts a public health perspective by drawing on the literature surrounding the concept of public health impact in particular the RE-AIM framework (Glasgow, Vogt, & Boles, 1999) and the potential impact fraction (Morgenstern & Bursic, 1982). RE-AIM is a widely adopted evaluation framework that proposes several important outcomes for health promotion interventions. The potential impact fraction is a measure developed by epidemiologists to estimate the number of incident cases of a disease that can potentially be prevented by an intervention. Both concepts help defining relevant outcomes and target effect sizes for mobile physical activity interventions.

As a next step, the insurer's physical activity program is analyzed in detail regarding its potential to meet the required target effects by answering the following research question:

RQ 2: To what degree can the insurer's physical activity program meet the defined target effects?

This research question is answered on a conceptual and empirical level. On a conceptual level, the program's components are linked to the target outcomes via mechanisms that have been identified by health behavior change theory. This description of how the program's components are thought to affect the program's target outcomes, also known as the conceptual model of the intervention (Collins, 2018), helps to understand the mechanisms via which the program could, in principle, meet its targets. The conceptual model is complemented by conducting systematic reviews of the scientific literature on mobile technologies and financial rewards, the two main components of the insurer's program. These reviews quantify the effect sizes that can be expected from the program's components based on past research and help to determine whether the program can meet its target effects.

After the targets for the insurer's program have been defined and the potential of the program has been evaluated, two empirical field studies directly address open questions regarding the program's design. This research corresponds to the optimization phase in the MOST framework (Collins, 2018). Building on differences between the program's target effects and effect sizes reported in the literature, the first empirical study investigates, whether the program's small financial incentives can produce sufficient effects for the program to meet the targets for scalable physical activity interventions:

RQ 3: Can small incentives increase participation and subsequent behavior change in a scalable physical activity intervention?

To answer the third research question, a cluster-randomized trial is conducted (N = 1547) that evaluates the effects of the program's financial incentives and an alternative incentive strategy within a three-month pilot phase of the program. Further, two surveys assess barriers for participation and participant's perception of the program.

As a last step, the results of the first field study inform a revision of the insurer's program. Drawing on theories of behavior change and insights from the field of human-computer-interaction, a smartphone-based intervention is developed that combines redesigned financial incentives with an automated digital activity coaching. Subsequently, a second empirical field study is conducted to get a detailed understanding of the effects of the revised intervention:

RQ 4: Which aspects of an incentive-based digital coaching app help users to increase daily physical activity?

This eight-week study ($N = 274$) uses baseline randomization and micro-randomization to evaluate the effects of different components of the newly developed physical activity app. This study helps to identify which components drive the app's overall effect and to assess whether the overall effect could meet the defined target for scalable physical activity interventions. Further, two surveys assess mediators of behavior change and participant's perception of the intervention.

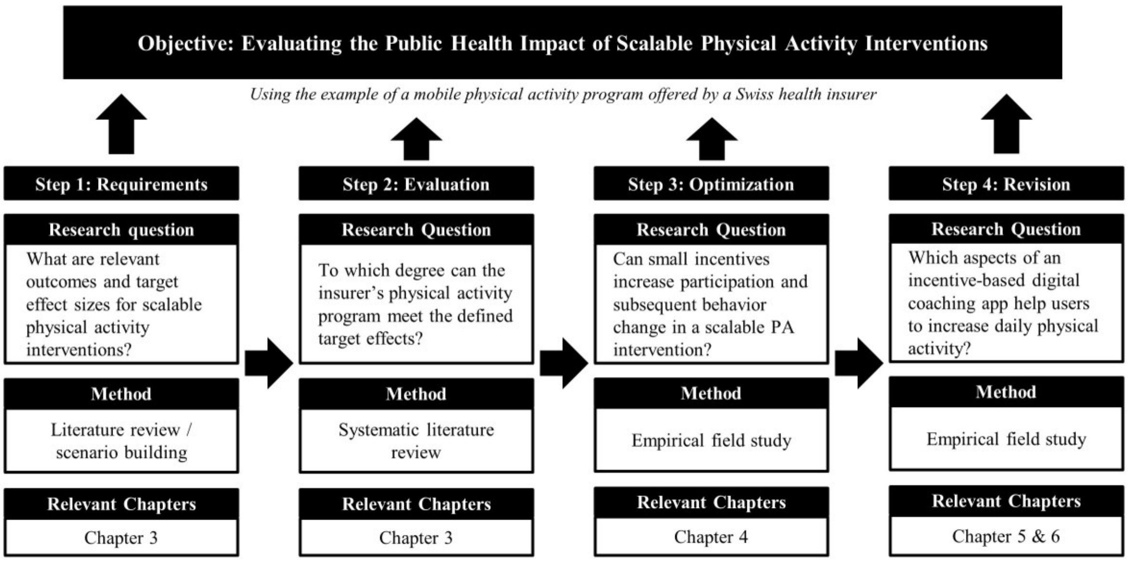


Figure 1-1. Organizing research framework and methodological approach.

Finally, the great amount of empirical evidence, that has accumulated throughout the literature reviews and field studies, is integrated to judge the public health of scalable physical activity interventions, i.e. their ability to contribute to the effective prevention of physical activity-related NCDs. This evaluation of scalable physical activity interventions and the accompanying insights are of great relevance for intervention developers, researchers and policy makers.

1.3 Structure

In line with Figure 1-1, the remainder of this dissertation is structured as follows: Before answering the proposed research questions, chapter 2 introduces relevant concepts that are necessary for a comprehensive understanding of the dissertation. Subsequently, chapter 3 reports different effect scenarios and systematic literature reviews to answer the research questions RQ 1 and RQ 2 respectively. Next, chapter 4 describes the first field study to answer research question RQ 3. Chapter 5 outlines the revision of the insurer's physical activity program and the rationale behind the novel intervention approach and chapter 6 describes the second field study to answer research question RQ 4. Chapter 7 concludes this dissertation with summarizing the key findings, evaluating the public health impact of scalable physical activity interventions, and discussing implications for intervention developers, researchers and policy makers as well as limitations and opportunities for future work.

Chapter 2

Background

This chapter introduces different concepts that are necessary for a comprehensive understanding of the research that is outlined in the subsequent chapters of this dissertation. These concepts include the MOST framework (section 2.1) that guides the intervention development approach and the accompanying empirical studies that are presented in chapter 4 and chapter 6 respectively. Further, the concept of public health impact is described (section 2.2). This concept is necessary to define relevant outcomes for scalable physical activity interventions. Lastly, section 2.3 gives an overview of health behavior change theories that are necessary to understand the mechanism behind different interventions, their potential, and their single components. Understanding the theory behind health behavior change allows for deriving the processes via which the insurer's physical activity program may affect its target outcomes.

2.1 The Multiphase Optimization Strategy

Researchers in the field of public health interventions in general (Ammerman, Smith, & Calancie, 2014), and those who study physical activity interventions in particular (Koorts et al., 2018; Reis et al., 2016), have noted that many interventions developed by researchers are not translated into practice. More specifically, it has been estimated that it takes approximately 17 years for 14% of research findings to be implemented in practice (Green, 2008). While the reasons for this research-to-practice gap are complex and manifold, the focus on RCTs in the process of intervention development likely plays a crucial role (Ammerman et al., 2014). This focus originated in the efficacy-effectiveness paradigm (Flay, 1986) that has been widely adopted by researchers, academic journals and funding agencies. Briefly, this paradigm refers to the general assumption that an intervention should demonstrate its ability to achieve positive effects on the target outcome under optimal conditions (tested in an efficacy trial, typically an RCT), before its impact in real-world settings is evaluated (in an effectiveness trial, ideally an RCT, but practical settings may require alternative designs). Why bother with the implementation of an intervention that is unable to produce a meaningful effect?

However, because the ultimate goal of the RCT is to obtain a precise estimate of the causal effect of the intervention, it prioritizes internal validity (i.e. the absence of bias in the results) over external validity (i.e. the generalizability of results to other settings or populations). To maximize the probability of proceeding to an effectiveness trial and to successfully publishing results, researchers therefore seek to demonstrate a large and unbiased intervention effect in the efficacy RCT. That is, researchers often test intensive interventions in a single setting with a highly motivated, self-selected sample, with measures in place to promote adherence and prevent attrition, and under close supervision by (also self-selected) researchers and trained staff (Glasgow, Lichtenstein, & Marcus, 2003). However, this situation differs substantially from real-world settings, where interventions may need to be brief and resource-saving, where samples are broad and heterogeneous, and where the intervention needs to be maintained not only by participants but also by the implementing organization (Glasgow et al., 2003). As a consequence, interventions that were successfully tested in efficacy RCTs often have a low probability of being successful in real-world conditions. To bridge this research-to-practice gap, researchers have called for more practice-based and practice-embedded research with a stronger focus on external validity (Green, 2008; J. Ma, Lewis, & Smyth, 2018; Reis et al., 2016) and for interventions that are "designed for dissemination" (Ammerman et al., 2014, p. 48).

The MOST framework (Collins, 2018) is an intervention development framework that was motivated by the low translation of research-based interventions into practice. In contrast to the efficacy-effectiveness paradigm (Flay, 1986), the MOST framework (Collins, 2018) does not primarily focus on intervention effectiveness during the intervention development process. Rather, the goal of MOST is to develop interventions that are effective (able to change an outcome in the desired direction), efficient (does not waste resources), economical (does not exceed budgetary constraints), and scalable (widely implementable without the need for ad hoc adjustments), before they are tested in an RCT. To accomplish this goal, the MOST framework divides the intervention development process into three different phases that differ in terms of the research objectives and the type of empirical research that is conducted (Figure 2-1): preparation, optimization and evaluation. In the preparation phase, a set of candidate intervention components is selected. In the optimization phase, the components are evaluated and the combination of components that results in the optimal intervention effect while considering important constraints for implementation (i.e. regarding efficiency, costs or scalability) is selected. In the evaluation phase, the complete optimized intervention is evaluated in a conventional RCT. Each phase is explained in greater detail below.

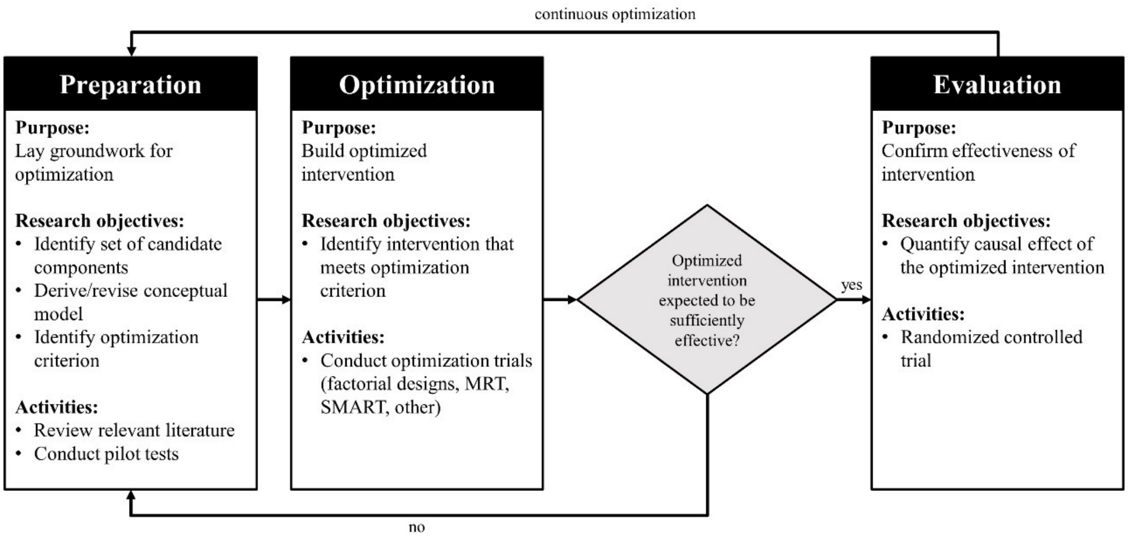


Figure 2-1. The different phases of the MOST framework.Adapted from Collins (2018, p. 24).

2.1.1 The Preparation Phase

The overall purpose of the preparation phase is to build the conceptual basis for the intervention and to inform the research to be conducted in the subsequent optimization phase. More specifically, the preparation phase has three key objectives: (1) to identify a set of candidate intervention components, (2) to derive the conceptual model of the intervention, and (3) to select an optimization criterion. To identify candidate intervention components, the researcher first defines the target outcomes of the intervention as well as the causal determinants that affect these outcomes. Based on this understanding of outcomes and their determinants, the researcher can subsequently define intervention components that change the target behavior via the hypothesized causal mechanism of action (i.e. the determinant). The combination of candidate intervention components, mechanism of action and target outcomes constitutes the conceptual model of the intervention (Collins, 2018). Thus, the conceptual model is the "engine that drives the intervention" (Collins, 2018, p. 36). It describes how the intervention is supposed to work, i.e. via which mechanisms the intervention components are hypothesized to achieve the intervention's target outcomes. Conceptual models are typically illustrated in graphical form (Figure 2-2).

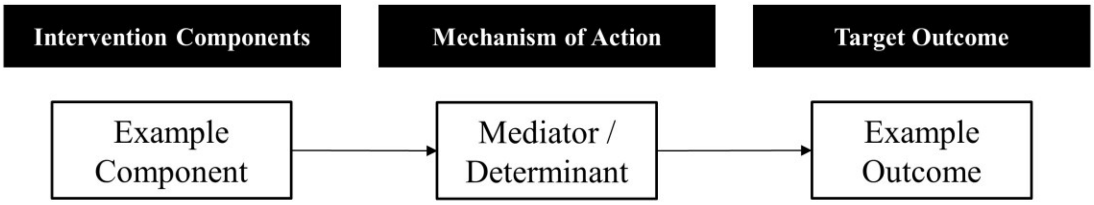


Figure 2-2. Schematic illustration of a conceptual model.

Candidate intervention components and their mechanisms of action are ideally based on behavior change theory and available empirical evidence so that a careful review of the relevant literature is advisable. If the empirical evidence on one or more intervention components is insufficient, their effects need to be evaluated in the subsequent optimization phase. The third objective of the preparation phase, selecting an optimization criterion to be used in the subsequent optimization phase, is necessary to account for implementation constraints in the development process. The optimization criterion is used to select intervention components for inclusion in the finalized intervention during the optimization phase. It defines the best expected (primary) intervention outcome that is obtainable within key constraints for implementation, such as staff time, participant time, and/or burden or intervention complexity.

Note that the setting of this dissertation differs from the preparation phase in MOST because the intervention components are developed by the insurer (presumably without an explicit conceptual model). It therefore seems reasonable to assume that the program does not exceed implementation constraints. In the present setting, the optimization criterion is thus of lesser importance. Nevertheless, the first two research questions of this dissertation are directly related to the preparation phase in MOST. Answering the first research question will determine the outcomes that the insurer's program needs to target and that are part of the program's conceptual model. Answering the second research question includes completing the conceptual model by linking the predefined program components to health behavior change theory and summarizing the empirical evidence on the component's effects. The first two research questions therefore help to derive the conceptual model of the insurer's physical activity program and inform research in the subsequent optimization phase.

2.1.2 The Optimization Phase

The optimization phase is the central element of the MOST framework. As the name implies, the purpose of the optimization phase is to optimize the intervention. Optimizing an intervention is defined as "the process of identifying an intervention that provides the best expected outcome obtainable within key constraints imposed by the need for efficiency, economy, and/or scalability" (Collins, 2018, p. 12). In other words, optimizing means selecting the set of candidate intervention components identified in the preparation phase that maximizes the effect of the intervention and can be implemented given the available resources.

To decide which candidate components to include and which to exclude, data is needed on the effects of the single intervention components on the primary intervention outcome. This data is either obtained from previous research or by conducting one or more empirical studies, so-called optimization trials. Optimization trials typically use factorial designs or fractional factorial designs to evaluate the effect of single intervention components. In contrast to a traditional RCT that contains only one experimental manipulation, and consequently only one randomization (intervention versus control), factorial experiments contain multiple experimental manipulations, called factors, thus allowing for the estimation of the effects of multiple intervention components (and their interactions) simultaneously. A detailed discussion of factorial designs is beyond the scope of this dissertation, but Collins (2018, chapters 3–5), Shadish, Cook, & Campbell (2002, chapter 8), and R. E. Kirk (2013) provide useful resources.

After information on the effects of all intervention components has been collected, MOST proposes a multi-step decision process to identify the intervention components that are to be included in the optimized intervention. As a first step, the effects of all intervention components are compared to an effect level that the researcher has defined as important. Only intervention components that have demonstrated an important effect are considered for further inclusion. Importance of effects can, but does not necessarily need to, refer to statistical significance. It may also refer to an effect level in relation to some overall target effect size that the intervention aims to achieve. The decision about which components to consider further is initially based on the main effects of the components but may be reconsidered in the light of important interactions. In a second step, the final (versions of) intervention components are selected while taking into account the constraints defined in the optimization criterion. The optimized intervention then contains the combination of components that maximizes the intervention outcome within the available resource constraints. However, as mentioned above, implementation constraints are secondary with regard to the insurer's physical activity program and the empirical studies in this dissertation focus primarily on estimating the effects of single intervention components.

More often than not, it may be necessary to return to the preparation phase after optimizing the intervention. This is the case, for example, when the intervention components are not expected to lead to significant or meaningful changes in the primary outcome, or this change can only be achieved by exceeding the available resources. In this case, proceeding to test the optimized intervention in an RCT in the evaluation phase makes little sense. Instead, the conceptual model may need to be revised and new intervention components may need to be developed (Collins, 2018).

2.1.3 Evaluation Phase

If the optimized intervention is believed to demonstrate a sufficient effect, it is evaluated in an RCT in the evaluation phase. The RCT compares the complete optimized intervention to an appropriate control condition to obtain an estimate of its causal effect on the primary outcome. The type of control (e.g. minimal intervention, standard of care or waitlist control) must be determined based on the research question and/or ethical considerations (Collins, 2018).

2.1.4 Summary

The MOST framework (Collins, 2018) provides guidance for the development of effective, efficient, economical and scalable behavior change interventions, and thereby attempts to close the research-to-practice gap of behavioral interventions. Interventions that have been developed according to the MOST cycle are assumed to be effective because all included intervention components have demonstrated important effects. They are efficient because intervention components that have demonstrated missing or unimportant effects are not included in the optimized version. They are economical because the optimization criterion guarantees that resource constraints are not exceeded.

As a result, the optimized intervention can be implemented without the need for further adjustments.

The research described in this dissertation primarily adopts two ideas that are very central to the MOST framework and that help to support the development and implementation of the insurer's physical activity program. The first idea is the conceptual model and the underlying assumption that a comprehensive understanding of the causal processes via which the intervention components affect relevant outcomes is essential for interventions to be successful. In fact, the first two research questions help to derive the program's conceptual model. The second idea is the process of optimization, which involves iterative testing of effects of single intervention components with the goal of including an optimal set of components in a finalized version of the intervention. Indeed, the two field studies reported in chapter 4 and chapter 6, respectively, can be conceptualized as optimization trials.

As mentioned above, one central aspect of the MOST framework is the conceptual model that describes via which processes the components of an intervention affect relevant intervention outcomes. In the case of the insurer's program, physical activity might seem to be the relevant intervention outcome. Although this certainly is correct, the overall objective of this dissertation adds further complexity. This is because the ability of an intervention to prevent disease (i.e. its public health impact) depends on additional factors beyond the intervention's ability to change behavior. The following chapter therefore introduces the concept of public health impact, which helps to define additional target outcomes of the insurer's physical activity program.

2.2 Public Health Impact of Interventions

Given that preventing disease and improving health are the overall objectives of mobile physical activity interventions, the target outcome of the insurer's physical activity program could simply be defined as the impact the program has on the health of its target population, e.g. measured as reductions in mortality or NCD incidence rates. Obviously, measuring these outcomes requires studies that span several years and is not feasible during the development process of the program. Therefore, subsection 2.2.1 introduces the RE-AIM framework, which allows for a more accessible way of quantifying an intervention's public health impact. The RE-AIM framework will help to define the target outcomes for scalable physical activity interventions (RQ 1). Beyond outcomes specified in the RE-AIM framework, different approaches to prevention that implicitly or explicitly underlie an intervention can also affect the public health impact of that intervention. The effects of these prevention strategies are therefore outlined in subsection 2.2.2. It is necessary to understand the implications of different prevention strategies in order to recognize strengths and limitations of the insurer's physical activity program and mobile physical activity interventions in general. Finally, an intervention's public health impact depends not only on its outcomes (e.g. behavior change) and its prevention strategy, but also on the relationship between its behavioral target (e.g. physical activity) and disease. Epidemiologists have derived methods to estimate the ability of interventions to prevent disease based on their effects on important risk factors, such as physical inactivity. One such method, the potential impact fraction, is explained in subsection 2.2.3. The potential impact fraction is later used to determine plausible target effect sizes of the insurer's physical activity program, i.e. to answer research question 2.

2.2.1 The RE-AIM Framework

Similar to MOST, the starting point for the development of the RE-AIM framework was criticism of both the strong focus on RCTs and intervention efficacy as the primary outcome of health interventions (Glasgow et al., 1999). The RE-AIM framework realizes that the overall impact of interventions is defined by factors beyond efficacy, such as the intervention's ability to be implemented and sustained in practice. Therefore, the RE-AIM framework proposes to complement efficacy with additional outcomes that collectively capture the public health impact of an intervention. These are reach, effectiveness, adoption, implementation and maintenance (Table 2-1). The RE-AIM outcomes refer to one of two different levels: the individual level that captures the impact that the intervention has on the target population if it is implemented, and the setting or organizational level that measures the extent to which the intervention is implemented in the real world. Both levels are necessary to determine the public health impact of an intervention. Although the RE-AIM framework does not provide a direct measure of an intervention's public health impact, it is based on the assumption that the intervention's impact is maximized by maximizing effects on the five RE-AIM dimensions.

Intervention reach refers to the percentage and characteristics of people who receive an intervention. It is measured on the individual level by comparing numbers and characteristics between non-participants and participants in the intervention. Reach also captures the degree to which intervention participants are representative of the target population. Calculating the participation rate for an intervention requires knowledge about the denominator, i.e. the number of people in the target population that would be eligible to participate. Like in the MOST framework (section 2.1), efficacy is defined as the intervention's ability to change important outcomes in the desired direction and is thus measured on the individual level. Important outcomes may include the intervention's primary outcome but should ideally also consider broader measures, such as satisfaction, quality of life, or attrition to adequately reflect public health impact (Gaglio, Shoup, & Glasgow, 2013; Glasgow et al., 1999). Because an intervention's benefits should outweigh its potential harms, the RE-AIM framework further recommends adjusting measures of efficacy for negative effects of the intervention.

Adoption is an outcome that is similar to reach but is measured on the organizational or setting level. It refers to the proportion and representativeness of the setting or organization that adopts the intervention and is typically measured by observation or surveys. Implementation captures the extent to which the intervention is implemented as intended by the intervention protocol and is measured on the setting level. Specifically, it refers to the degree to which the intervention is delivered as planned by the organization or institution that provides the intervention. Lastly, maintenance refers to the extent to which the intervention has been institutionalized or has become part of the provider's routine and everyday practice. On the individual level, maintenance refers to long-term effects of the intervention, defined as effects that are maintained longer than six months after the last intervention contact.

Table 2-1. Summary of the RE-AIM framework (adopted from Glasgow et al., 1999).

Outcome	Definition	Level
Reach	Proportion of the target population that participated in the intervention	Individual
Efficacy	Effect on positive outcomes minus effect on negative outcomes	Individual
Adoption	Proportion of settings or organizations that will adopt this intervention	Organizational
Implementation	Extent to which the program is implemented as intended in the real world	Organizational
Maintenance	Extent to which the intervention and its effects are sustained over time	Individual and organizational

No firm guidelines exist regarding the operationalization of all dimensions or the combination of different dimensions into an overall RE-AIM score. Glasgow, Klesges, Dzewaltowski, Estabrooks, and Vogt (2006) recommend calculating separate scores for the individual-level and setting-level impact using the product of the respective dimensions scores. Thus, the individual-level score is calculated as the product of reach and efficacy $(R * E)$[4], whereas the setting-level score is calculated as the product of adoption and implementation $(A * I)$. Similar to what has been noted in the previous section, issues regarding adoption and implementation of the program are assumed to be taken into account by the health insurer when designing the program. The focus of this dissertation is therefore on the program's individual-level impact.

Beyond reach, efficacy, and sustainable implementation in practice, the public health impact of an intervention with a prevention target can also depend on its underlying prevention strategy. Specifically, different prevention strategies can have different implications for the importance of an intervention's reach and efficacy. The following subsection outlines the existing prevention strategies and their implications.

[4] If data on the maintenance of effects is available, these data are used for evaluating the efficacy of the intervention.

2.2.2 Prevention Strategies

For interventions that ultimately aim to prevent disease, e.g. physical activity interventions, Rose (2001) distinguished two different prevention strategies: population strategies and high-risk strategies. Both strategies differ with regard to their overall goal and the type of determinants that are targeted. The approach of Rose (2001) was later extended to further differentiate between high- and low-agency strategies (J. Adams, Mytton, White, & Monsivais, 2016). Both approaches have implications for an intervention's reach and efficacy and are briefly explained below.

Population-level versus High-risk Strategies

Population-level strategies seek to lower the incidence rate of a disease in the whole population. To do so, they target factors that affect mean levels of risk factors in the population (so-called causes of incidence) so that the whole distribution of risk exposure within the population is shifted in a favorable direction (Rose, 2001). Ideally, this requires the intervention to reach the complete population. Immunizations, obligatory seatbelts in cars, lifestyle recommendations or sugar taxes are examples of population-level strategies. The compelling advantage of population-level strategies is that they target the underlying causes of disease incidence. As a result, these interventions can have great effects on the population level and may contribute to preventing or even eradicating the target disease. On the other hand, population-level strategies offer little benefit for the individual because the majority of individuals is at low risk, at least for some amount of time (Rose, 2001). For example, insurance data from the US indicate that the average driver is involved in a car accident once every 18 years (Toups, 2011). Thus, wearing a seatbelt has no benefit for most drivers most of the time. In addition, population-level strategies can be difficult to implement because they sometimes require concerted effort from researchers, government agencies and healthcare systems (e.g. immunizations), lack public acceptance (e.g. smoking bans), or face backlash from industry (e.g. sugar taxes; Adams, Mytton, White, & Monsivais, 2016).

High-risk strategies, on the other hand, seek to identify and protect individuals who are at high risk of developing a disease. Interventions applying a high-risk strategy screen the population for high-risk individuals and target known determinants of individual disease risk (so-called causes of cases) in these high-risk individuals, ideally resulting in a truncation of the risk distribution within the population (Rose, 2001). An advantage of high-risk strategies is that they reach those individuals who most benefit from the intervention and are therefore a more cost-effective use of resources. This is especially important for interventions with possible adverse effects, which must be offset by greater benefits. However, high-risk strategies are ultimately focused on risk management and do not attack the factors causing the disease (Rose, 2001). While these strategies may successfully prevent high-risk individuals from becoming sick, they do not prevent individuals from developing high disease risk. This can limit the public health impact of high-risk strategies, especially when identification of high-risk individuals is imprecise and the prevalence of the risk factor is low. For example, in the Framingham Heart Study (Shurtleff, 1970), men with clinically high levels of serum cholesterol (> 310mg/100ml) were almost twice as likely to die from coronary heart disease than men with a low serum cholesterol concentration during a ten year period. However, the high-risk group was only responsible for 9% of absolute deaths attributable to serum cholesterol, because only very few men showed a high level of cholesterol (Rose, 1981). Thus, an intervention targeted at the high-risk group, even if very powerful, would have little effect on the population level.

From an intervention development perspective, the two strategies can have different implications. The population-level strategy essentially relies on large reach to generate public health impact. In fact, efficacy may be secondary because even a small reduction of disease risk may result in a large number of prevented cases if it is applied to the whole population (Rose, 2001). Developers of population-level intervention strategies might therefore focus on increasing reach rather than increasing efficacy of the intervention. In contrast, the impact of high-risk strategies crucially depends on their protective power. Given the smaller target group of high-risk individuals, intervention developers may prioritize the intervention's efficacy over its reach.

High-agency versus Low-agency Strategies

Prevention strategies or interventions can also be differentiated according to their degree of individual agency, i.e. the extent to which participants have to be motivated and able to engage with the intervention content to benefit from the intervention (J. Adams et al., 2016). This is referred to as the degree of individual agency of the intervention. Examples of interventions with a high degree of agency include lifestyle recommendations, front of pack nutrition scores or commercial weight loss programs. The success of these interventions relies heavily on the respective target group's uptake of and continuous engagement with the intervention, e.g. adhering to lifestyle recommendations, understanding nutrition scores and acting accordingly, or participating in and reaching behavioral targets in weight loss programs. By contrast, examples of interventions with a low degree of individual agency include taxes on unhealthy consumer goods, reduction of salt in processed foods or the fluoridation of tap water. These interventions typically target the context in which behavior occurs, i.e. the environmental determinants of behavior, and thus require little to no effort on the part of the target group to be effective. Figure 2-3 further illustrates the abovementioned prevention strategies using different intervention examples of possible combinations of the targeting and individual agency dimensions.

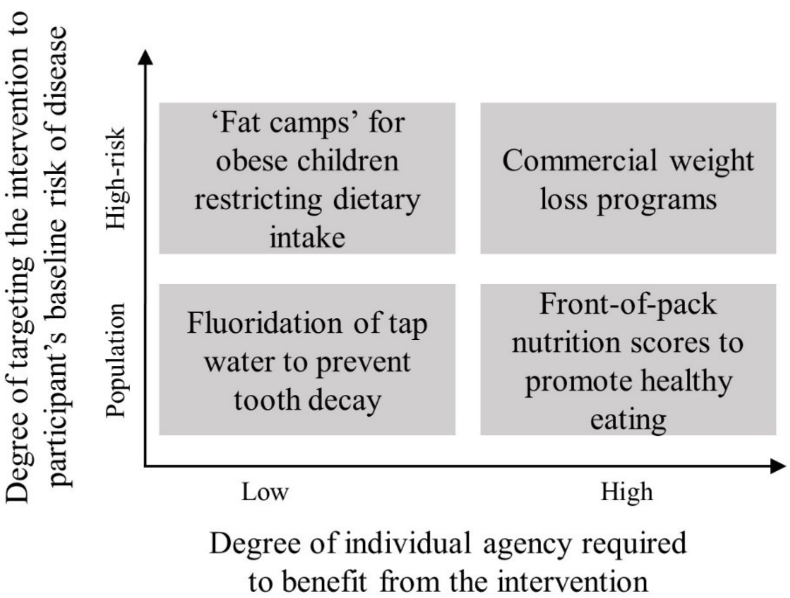


Figure 2-3. Targeting and agency dimensions of health interventions, adopted from Adams et al. (2016).

Population-level intervention strategies make the implicit assumption that the intervention produces a similar effect on all individuals irrespective of their baseline risk. However, this is unlikely to be the case for population-level interventions with a high degree of agency, where individuals must invest a significant amount of resources to participate and/or benefit from the intervention (J. Adams et al., 2016). It is well documented, for example, that individuals who are at low risk for disease, for example those who are healthier, better educated and of higher socio-economic status, are more likely to engage in health promotion activities (Glasgow, McCaul, & Fisher, 1993; Patel et al., 2017). Consequently, a high degree of agency in population-level interventions may restrict its overall reach (thus undermining its public health impact) and may widen health inequalities because high-risk groups are less likely to benefit (McLaren, McIntyre, & Kirkpatrick, 2009). Unfortunately, many interventions using mobile technologies can be characterized as population-level interventions with a high degree of individual agency. These interventions typically attempt to leverage the scalability of mobile technology and are offered to the general population without screening participants for their disease risk. Simultaneously, these interventions require the acquisition and sustained use of mobile technologies or smartphone apps.

Another factor that influences the public health impact of an intervention is the relationship between its behavioral target (e.g. physical activity) and disease, e.g. the incidence of a specific NCD. Specifically, the intervention's public health impact is greater the more the intervention changes the behavioral target and the stronger the relationship is between the target behavior and disease incidence. The following subsection introduces a way of quantifying these relationships and the resulting public health impact.

2.2.3 Quantifying Public Health Impact

The previous two subsections have illustrated the characteristics of an intervention (e.g. reach and efficacy) that determine its public health impact. They have not, however, provided a way to quantify public health impact, e.g. in terms of the number of incident cases that can be prevented by an intervention given certain assumptions regarding its reach and efficacy. One way to do so is to estimate the intervention's potential impact fraction (IF, Morgenstern & Bursic, 1982). The IF estimates the reduction of incidence of a disease resulting from a specified change in the distribution of an ordinal risk factor, e.g. due to an intervention. Note that the IF is applicable only to ordinal risk factors, i.e. risk factors with multiple categories and ascending or descending disease risk across the categories. Examples include physical activity (inactive, active, very active) or a categorization of the body mass index (normal weight, overweight, obese). For example, the IF could be used to estimate the proportion of new incident cases of type-2 diabetes mellitus (T2DM) that could be prevented by a physical activity intervention that changes the distribution of physical inactivity (i.e. the proportion of active vs. inactive individuals) in the population in some hypothesized way. As such, the IF can be expressed as:

$$IF = \frac{R - R'}{R}$$

where R is the pre-intervention risk of disease in the population and R' is the post-intervention risk of disease in the population.

Three different types of information are required to estimate the IF. First, estimating the IF requires knowledge of the distribution of the risk factor and its i categories (f_i) in the target population. This information can typically be obtained from representative surveys. Second, information on the magnitude of the relationship between the risk factor and the disease to be prevented is required. These can be disease risk estimates for each category of the risk factor (R_i), but more often risk ratios for each category of the risk factor will be available (RR_i). These can often be obtained, for example, from dose-response meta-analyses of the relationship between the risk factor and the disease. Finally, information on the change of the physical activity distribution as a result of the intervention is needed. This corresponds to the efficacy of the intervention, i.e. the effect that the intervention has on its target outcome. This effect has to be specified as the fraction of individuals in each activity category (g_i) that is shifted to the next lower risk category. Values of g_i can be based on plausible assumptions or previous studies. With these three pieces of information, the IF can be calculated using the following formula (Morgenstern & Bursic, 1982):

$$IF = \frac{\sum_{i=0}^{k}(f_i g_i - f_{i-1} g_{i-1}) RR_i}{\sum_{i=0}^{k} f_i RR_i}$$

More detailed information on the IF, including the derivation of this formula, is available in Appendix A. In chapter 3, the IF will be used to estimate the public health impact of scalable physical activity interventions for different assumptions regarding reach and efficacy. This will help to answer the second research question, that is, to determine plausible target effect sizes for outcomes of mobile physical activity interventions.

2.2.4 Summary

From a public health perspective, the overall goal of the insurer's physical activity program is to improve health among its target population by preventing adverse health outcomes, primarily the incidence of non-communicable diseases. This section outlined three perspectives that are important for understanding the public health impact of an intervention. The RE-AIM framework (Glasgow et al., 1999) illustrates that the public health impact of an intervention depends on five different intervention characteristics. Of these, reach (i.e. how many and who is reached by the intervention?) and efficacy (i.e. can the intervention change health-related outcomes?) are most relevant for the insurer's physical activity program. Beyond the outcomes of the RE-AIM framework, interventions can apply different prevention strategies that also affect their public health impact. Physical activity interventions built around mobile technologies often attempt to leverage the scalability of these technologies to reach as many individuals as possible. At the same time, they require individuals to invest effort and motivation in order to benefit from the intervention. This combination is known to lead to selection effects that can limit the intervention's public health impact. Finally, an intervention's public health impact depends on the relationship between its behavioral target (e.g. a risk factor) and disease incidence. Based on information on this relationship, the potential impact fraction offers a way to quantify an intervention's public health impact.

After having elaborated on public health impact and the resulting target outcomes of mobile physical activity interventions (cf. RQ 1), it is important to understand whether and how the program can achieve those outcomes to be able to judge its potential public health impact (cf. RQ 2). Therefore, the following section provides an introduction to and overview of health behavior change theories. These theories identify important factors that determine whether people do or do not change their health behavior and are later used to identify the processes via which the insurer's physical activity program could affect its target outcomes.

2.3 Health Behavior Change Theories

Theories are generally defined as "a systematic view of events or situations by specifying relations among variables, in order to explain and predict the events or situations" (Glanz, Rimer & Viswanath, 2008, p. 26). Accordingly, theories of health behavior change are a specific set of theories that identify key variables that explain (the change in) behaviors that influence health, such as physical activity, smoking or dietary behavior. These theories can both explain behavior and also guide the development of health behavior change interventions by specifying potential intervention targets and by explaining how a change in these targets causes a change in the respective behavior. In fact, it is considered best practice to base the development of health interventions on available evidence and appropriate theories (P. Craig et al., 2008), because there is a moral imperative to only invest resources in the development of interventions that target mechanisms with a realistic chance of changing behavior (Moore & Evans, 2017).

The following subsections provide a more detailed overview of the existing health behavior change theories and their distinct characteristics, and introduce the Health Action Process Approach (HAPA; Schwarzer & Luszczynska, 2008), a theory of particular relevance for mobile physical activity interventions. The last subsection briefly outlines the work centered around the taxonomy of Behavior Change Techniques (BCTs; Michie et al., 2013), which has tried to address some of the shortcomings of traditional health behavior change theories.

2.3.1 Overview of Theories

A review of the literature from 1960 to 2012 identified 82 different theories of behavior or behavior change (Davis, Campbell, Hildon, Hobbs, & Michie, 2015), illustrating the wide range of theories available. Yet, a few theories stand out because they have been more frequently cited in the literature. These are the Health Belief Model (Hochbaum, 1958; Rosenstock, 1960, 1974), the Transtheoretical Model of Change (Prochaska & DiClemente, 1982), the Theory of Planned Behavior (Ajzen, 1985), the Social Cognitive Theory (Bandura, 1986), and the Information-Motivation-Behavioral-Skills Model (Fisher & Fisher, 1992). Other influential theories include Control Theory (Carver & Scheier, 1982), Goal Setting Theory (Locke & Latham, 1990), the Fogg Behavior Model (Fogg, 2009), Self-Determination Theory (E. L. Deci & Ryan, 1985), and the Health Action Process Approach (Schwarzer & Luszczynska, 2008). Although many of these theories have substantial overlaps, they also differ on several dimensions.

Multiple Levels of Influence

Theories may focus on behavioral determinants on different levels, typically referred to as the individual level, the interpersonal level, and the environmental level. Determinants on the individual level include a person's beliefs, motivation and abilities. The Health Belief Model (Hochbaum, 1958; Rosenstock, 1960, 1974) is an example of an individual-level theory. The Health Belief Model assumes that disease prevention efforts (e.g. getting vaccinated) primarily depend on subjective assessments of the disease (perceived susceptibility and perceived severity of the disease) and of the target behavior (perceived benefits and barriers). On the interpersonal level, an individual's relationships with family members, friends, coworkers or health professionals can affect health behavior, for example through behavioral models (good or bad), social influence or social support. Social Cognitive Theory (SCT; Bandura, 1986) is an example of a theory that specifies how factors on the interpersonal level can impact health behaviors.

Specifically, SCT assumes that social models are an important source of self-efficacy, i.e. beliefs about one's personal ability to perform a behavior, which in turn is a powerful determinant of behavior. People are more confident to perform a health behavior, e.g. to stop smoking, if they observe others similar to themselves successfully perform the behavior. Ecological models of health behavior explicitly take multiple levels of influence into account and specify behavioral determinants on the environmental level. An example that focuses on physical activity is the Ecological Model of Four Domains of Active Living (Sallis et al., 2006). This model assumes that physical activity is, among other determinants, influenced by access to recreational activity and parks, traffic safety, public transport systems, parking regulations and neighborhood walkability, i.e. densely connected street networks with homes close to commercial and institutional facilities.

Although studies on the relative importance of determinants at each level are scarce, those that do exist conclude that individual-level determinants are most important for overall physical activity (Giles-Corti & Donovan, 2002), while all levels are of roughly equal importance for walking (Giles-Corti & Donovan, 2003), a specific type of physical activity. Similarly, researchers have argued that interventions targeting determinants on multiple levels have the highest chance of success, because providing people with sufficient motivation and skills to change behaviors is unlikely to lead to sustainable behavior change if environments and policies that aggravate change exist (Sallis, Owen, & Fisher, 2008). Yet, the majority of health behavior change theories focuses on determinants on the individual level.

Continuum Models and Stage Models

Theories make different assumptions about the nature of behavior change. Continuum models assume that behavior change is a linear process that is influenced by a set of key variables and that individuals have a certain probability of change depending on their value on these variables (Lippke & Ziegelmann, 2008). This has implications for intervention design: interventions should ideally target all determinants of behavior change identified by the theory in order to maximize the probability of change. The Theory of Planned Behavior (TPB; Ajzen, 1985) is an often cited continuum model. The TPB has identified intentions and perceived behavioral control (i.e. self-efficacy) as predictors of behavior, and assumes that intentions can be predicted accurately from attitudes toward the behavior, subjective norms (i.e. perceived social pressure), and, again, perceived behavioral control. On the contrary, stage models assume that behavior change takes place in discrete stages and that, depending on the stage, certain variables are more important than others (Lippke & Ziegelmann, 2008). Within each stage, these variables do not necessarily predict behavior but often predict progression to the next stage of change. The Transtheoretical Model of Change (Prochaska & DiClemente, 1982), for example, defines five different stages of behavior change (precontemplation, contemplation, preparation, action, and maintenance) and ten processes that help individuals to progress through the different stages. According to stage models, behavior change interventions should be matched to the individual's stage of change.

Motivation and Volition

Individual-level determinants of behavior that have been identified by health behavior change theories can be broadly categorized as either motivational determinants, i.e. predictors of an individual's intention to perform a behavior, or volitional determinants, i.e. factors that promote the effective translation of intentions into actions. Motivational determinants are in general assumed to be cognitive and include beliefs about a behavior-related disease (e.g. perceived threat, perceived susceptibility) and beliefs about the behavior (e.g. outcome expectancies, attitude, norms, self-efficacy), whereas volitional determinants focus on strategies and capabilities related to self-regulation (e.g. goal setting, progress monitoring, feedback, if-then plans; Sheeran, Klein, & Rothman, 2017). Earlier theories of health behavior typically do not specify volitional factors but rather focus only on predictors of intentions. This has been noted as a weakness of those theories because intentions alone are often insufficient to accurately predict health behavior (intention-behavior gap; Webb & Sheeran, 2006).

Criticism

Health behavior change theories have been criticized for a number of reasons that mainly relate to the theories' function of informing the development of health behavior change interventions (Sheeran et al., 2017). For example, a researcher designing an intervention is typically interested in identifying the most important determinants of a specific behavior (e.g. alcohol consumption) in a specific target population (e.g. adolescents) in a specific setting (e.g. school setting). Likewise, intervention designers need to know how exactly the key behavioral determinants (and subsequently the target behavior) can be changed most effectively. Unfortunately, most health behavior change theories provide little guidance to answer these questions. Another related criticism is the predominantly observational study designs that are used to validate theories. Because observational studies do not permit causal inference, these studies may confirm a theory's predictive validity but not its ability to identify appropriate intervention targets. Indeed, when comparing relationships of theory-based predictors with health behaviors from observational and experimental studies, Sheeran et al. (2017) noticed substantial differences for some predictors (e.g. $d_{\text{experimental}} = .38$ and $d_{\text{correlational}} = .70$ for attitudes and $d_{\text{experimental}} = .36$ and $d_{\text{correlational}} = .87$ for intentions).

In addition, existing theories have a strong focus on individual-level determinants, mostly related to reflective cognitive processes (Moore & Evans, 2017). While these processes are of great importance, this exclusive focus can lead to the neglect of interpersonal and structural determinants, automatic processes, and behavioral routines that can have equal or even greater leverage for behavior change (Brewer & Rimer, 2008; Marteau, Hollands, & Fletcher, 2012; Moore & Evans, 2017; Sheeran et al., 2017).

New Developments

More recent developments in the field of behavior change science have attempted to address some of the shortcomings of existing theories. Two research initiatives, the Human Behavior Change Project (HBCP) in the UK and the Science of Behavior Change (SOBC) project in the US, aim to identify which behavioral interventions work for which behaviors, for whom, how well, for how long, and in what setting (Sumner et al., 2018). While the HBCP approach develops a knowledge system based on annotated study reports using the taxonomy of behavior change techniques (cf. subsection 2.3.3 in this chapter), the SOBC initiative promotes a mechanism-focused experimental medicine approach to developing and evaluating behavior change interventions. Further, novel theoretical approaches highlight the role of time in the behavior change process (Hall & Fong, 2007; Spruijt-Metz & Nilsen, 2014), a topic that has particular relevance for interventions incorporating mobile technologies as these are able to operate on much smaller time-scales and with faster feedback cycles as compared to traditional behavior change interventions (Scholz, 2019).

Further, theoretical approaches to habit formation (W. Wood & Neal, 2007) and concepts that focus on automatic processes, such as the nudge framework (Thaler & Sunstein, 2009), point to less conscious and reflective processes involved in behavior change and may help to design interventions that are better at maintaining behavior change over time (Marteau et al., 2012). In fact, self-reported data on the relation of habit and behavior indicate that up to 60% of physical activity may be considered habitual, that is, occurring as a learned response to (environmental) stimuli without any awareness of the intentions or goals of the activity (Gardner, de Bruijn, & Lally, 2011). Another theory that has received attention in recent years, due to its focus on volitional factors, is the Health Action Process Approach (Schwarzer & Luszczynska, 2008). As illustrated below, this theory is of great relevance for the development of mobile health interventions and is therefore discussed in more detail in the next section.

2.3.2 The Health Action Process Approach

The HAPA model (Schwarzer & Luszczynska, 2008) is a theory of health behavior change that integrates elements from several earlier theories, most notably from Social Cognitive Theory (Bandura, 1986) and Heckhausen's volition theory (Heckhausen & Heckhausen, 1991). The HAPA (Figure 2-2) understands behavior change as the result of two consecutive phases: a motivational phase that leads to a behavioral intention and a volitional phase that subsequently translates this intention into action. Each phase is characterized by different social-cognitive predictors. The detailed specification of volitional processes in the HAPA makes this model attractive for guiding the development of mobile health behavior change interventions. In line with what is to be expected for high-agency interventions (cf. subsection 2.2.2), survey results indicate that more than 80% of users of mobile physical activity and weight loss apps report intentions to change behavior (Carroll et al., 2017) and may thus benefit particularly from strategies that support the translation of their good intentions into actions.

The Motivational Phase

The formation of a behavioral intention is preceded by different cognitive beliefs. Risk-perception (e.g. "insufficient exercise will increase my cardiovascular disease risk") is seen as the trigger of an intention formation process and may lead to more elaborated thoughts about the consequences of the target behavior and personal abilities to perform the behavior. Specifically, outcome expectancies are the result of balancing the pros and cons of anticipated behavioral outcomes (e.g. "regular exercise will help me to prevent cardiovascular disease and feel better"), and self-efficacy is the personal belief of being capable of performing the behavior even when facing difficulties (e.g. "I will be able to exercise every day even if I am tired").

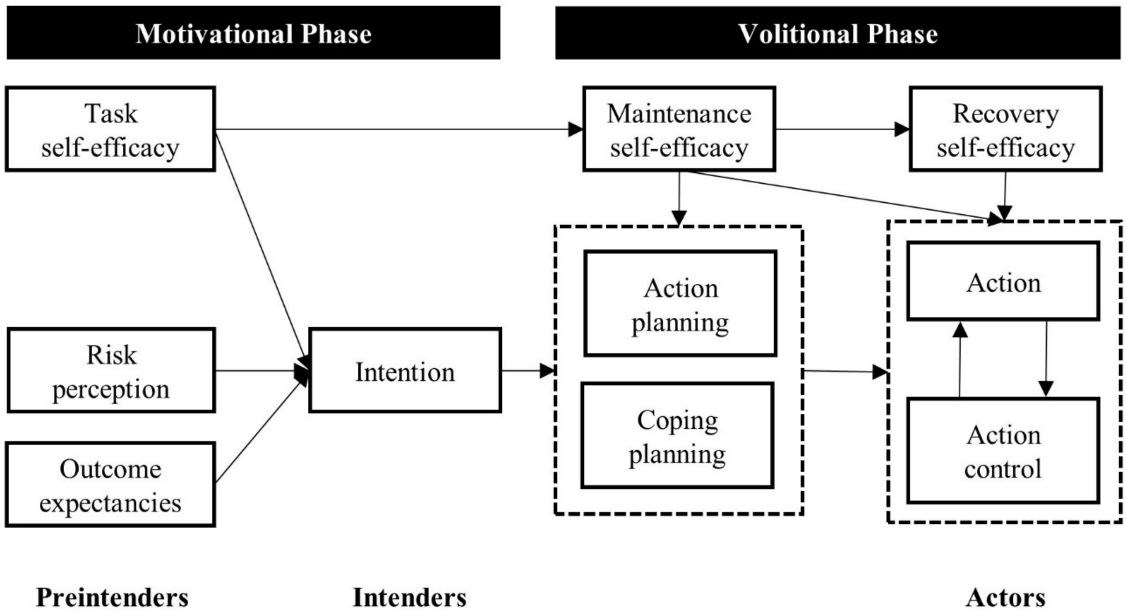


Figure 2-4. The Health Action Process Approach.

The Volitional Phase

After the intention to change has been built, it must be translated into action. Good intentions are more likely translated into action if people prepare strategies that ensure success and overcome potential difficulties. The HAPA proposes that two different types of planning, action planning and coping planning, mediate the translation from intention to action. Action planning refers to specifying when, where and how the behavior will be performed (e.g. "every day after work, I will run for 30 minutes in the park close to home"). The idea of action planning is closely related to the concept of implementation intentions (Gollwitzer, 1999), i.e. formulating if-then plans to facilitate the translation from intention to action (e.g. "if I come home from work, then I will run for 30 minutes in the park nearby"). Both action planning and implementation intentions rely on supplementing the initial intention (e.g. "I want to exercise regularly") with sufficient detail and situational cues to increase the likelihood of action. Action planning primarily promotes the initiation of behavior change and protects its execution from tempting behavioral alternatives. Coping planning, on the other hand, refers to planning effective strategies when faced with barriers or difficulties to executing the target behavior (e.g. "when I want to exercise but I feel tired, I will think of how good I feel after exercising as a way to motivate myself"). Consequently, coping planning complements action planning and protects the plan against foreseeable barriers.

Action control is the most proximal predictor of behavior in the HAPA model. Action control is a multi-faceted process that ensures the alignment of behavior with set behavioral goals. More specifically, once a behavioral goal has been set, it has to be quantified in terms of a measurable reference value (e.g. running five times per week for 30 minutes). This reference value must be remembered and be retrievable from memory over the complete course of action. During execution, goal progress is monitored and constantly evaluated against the reference value. Subsequently, the behavior is regulated based on the result of the evaluation, i.e. effort is maintained or increased if the current performance is below the reference value. For example, after running 25 minutes, effort is maintained for another five minutes to reach the goal of running 30 minutes. In summary, action control can thus be described by three sub-processes: awareness of standards, self-monitoring, and self-regulatory effort. It is conceptually equivalent to the notion of the negative feedback loop in Control Theory (Carver & Scheier, 1982).

As in the motivational phase, self-efficacy is also assumed to play an important role in the volitional phase. In contrast to the motivational phase, where self-efficacy refers to beliefs about being able to perform the behavior in general (task self-efficacy), self-efficacy in the volitional phase refers to the ability to maintain the behavior change (maintenance self-efficacy) and recover from potential setbacks (recovery self-efficacy). That is, after a behavior change goal has been adopted, it may be difficult to adhere to it. People high in maintenance self-efficacy are more confident about overcoming barriers and will invest more effort and develop more and better strategies to do so. The notion of recovery self-efficacy was adopted from addiction research (Marlatt, Baer, & Quigley, 1995), where efforts to abstain from addictive substances can be completely undermined by an initial lapse (abstinence violation effect). This effect is caused by attributing the lapse to internal and stable factors, such as lack of willpower. People high in recovery self-efficacy, however, will attribute the initial lapse to external factors (e.g. to a high-risk situation or to positive outcome expectancies) and quickly return to adhering to their initial goal.

Implications for Intervention Design

The two phases of the HAPA imply different health behavior change interventions for individuals who are at different stages of the behavior change process. Individuals in the motivational phase ("non-intenders") will benefit most from interventions that target predictors of intention, such as information that illustrates the risks and benefits of performing or not performing the behavior, or from behavioral models as a source of self-efficacy. Individuals in the volitional phase can be further divided into “intenders” and “actors” depending on whether or not they already perform the target behavior. Having already developed an intention to change, intenders are not likely to benefit from interventions targeted at non-intenders. Instead, they will benefit from formulating detailed action and coping plans to support the translation of their intentions into actions. Actors do not require an intervention per se, unless they need to be prepared for high-risk situations were lapses are anticipated (Schwarzer & Luszczynska, 2008).

Empirical Evidence

Several empirical studies have provided support for the assumptions of the HAPA model. A recent meta-analysis of 95 studies (Zhang, Zhang, Schwarzer, & Hagger, 2019) summarized the empirical evidence on the validity of the HAPA model. Using meta-analytical structural equation modelling, the authors found that across different behavioral domains, and in line with the HAPA, risk perceptions, outcome expectancies, and task self-efficacy significantly predicted intentions, with the effects of outcome expectancies ($\beta = .27$) and task self-efficacy ($\beta = .33$) being substantially larger than the effect of risk perceptions ($\beta = .07$). In turn, intentions and maintenance self-efficacy had direct and indirect effects on health behavior. The indirect effects were mediated through action planning and coping planning, although there were relatively small ($\beta = .02$ - $.04$) in comparison to the direct effects ($\beta = .14$ - $.18$). In addition, both action planning ($\beta = .09$) and coping planning ($\beta = .10$) had direct effects on behavior. The analysis could not confirm the proposed effect of recovery self-efficacy on health behavior. Overall, the HAPA model explained 17.5% of the variance in behavior. However, the authors were not able to include measures of action control in their analysis.

The meta-analysis by Zhang et al. (2019) focuses on the prediction of behavior rather than on behavior change. However, accounting for past behavior in the analysis did not substantially alter the conclusions (Zhang et al., 2019). Furthermore, the analysis is based on correlational data and the results need to be treated with caution when informing the design of interventions. Yet, evidence from experimental studies has confirmed the efficacy of planning interventions for physical activity (Bélanger-Gravel, Godin, & Amireault, 2013; Carraro & Gaudreau, 2013) and healthy eating (Adriaanse, Vinkers, De Ridder, Hox, & De Wit, 2011), and has demonstrated the superiority of HAPA-based stage-matched versus stage-mismatched interventions (Lippke, Schwarzer, Ziegelmann, Scholz, & Schüz, 2010). In addition, the HAPA has been used to guide the design of successful behavior change interventions for various health behaviors, including physical activity (Storm et al., 2016), fruit and vegetable consumption (Duan, Wienert, Hu, Si, & Lippke, 2017), sleep hygiene (Lin et al., 2018) and therapy adherence (Deng, Wang, Sun, & Chen, 2013).

Although the empirical evidence generally supports the relationships proposed by the HAPA model, like other existing theories, the HAPA model provides limited guidance on which determinants are most important for which behaviors, target groups or settings. One approach to overcoming this limitation of existing theories is to analyze the behavior change techniques (BCTs) that are used by existing interventions. The next section illustrates how analyses of BCTs can reveal which aspects of an intervention are associated with the intervention's effectiveness for different behaviors, target groups or settings.

2.3.3 Behavior Change Techniques

In an attempt to standardize the reporting of intervention content and to identify the aspects of interventions that contribute to their effectiveness, Abraham and Michie (2008) developed the first taxonomy of clearly defined behavior change techniques. BCTs are "observable, replicable, and irreducible component[s] of an intervention designed to alter or redirect causal processes that regulate behavior" (Michie et al., 2013, p. 82). In other words, BCTs are the smallest elements of an intervention that are thought to produce behavior change, such as goal setting, social comparison or self-monitoring. The first taxonomy included 26 BCTs and was later revised and extended twice, to first include 40 BCTs (Michie et al., 2011) and then 93 BCTs (Michie et al., 2013) that can be clustered into 16 groups. The taxonomy of BCTs established a standardized vocabulary for researchers with which to describe intervention content and the taxonomy is widely adopted in the scientific literature. Subsequently, the BCTs could be linked to one or more theory-based mechanisms of action that explain how each technique actually brings about behavior change (Carey et al., 2018; Johnston et al., 2018). This linkage of BCTs and mechanisms of action is based on intervention reports and expert consensus (Johnston et al., 2018).

Coding of existing interventions that have been tested in (randomized) controlled trials according to the BCT taxonomy also enables meta-regression analyses that identify which BCTs are associated with intervention effectiveness. Several of these analyses have been conducted for physical activity interventions. Michie et al. (2009) reported that physical activity interventions in adults that use the BCT “self-monitoring” in combination with at least one other BCT derived from Control Theory (Carver & Scheier, 1982), namely “prompting intention formation”, “prompting specific goal setting”, “providing feedback on performance”, and “prompting review of behavioral goals”, lead to greater changes in behavior as compared to interventions that do not include those BCTs (effect sizes of $d = 0.38$ and $d = 0.27$ respectively). Similarly, Murray et al. (2017) demonstrate that in non-clinical adult populations the BCTs “prompt self-monitoring of behavioral outcome” and “use of follow-up prompts” lead to interventions having greater effects on physical activity at more than six months post-baseline. In adults aged 55–70 years, the BCT “feedback” has been found to be associated with larger effects of physical activity interventions 12 months post-randomization, while the BCTs “information on when and where to perform the behavior” and “information on consequences of the behavior” were associated with smaller intervention effects (O’Brien et al., 2015). In another meta-analysis of physical activity and healthy eating interventions for overweight and obese adults, the BCTs “self-monitoring” and “goal setting” predicted short-term effects on behavior change (≤ six months), and the BCTs “self-monitoring”, “goal setting”, “feedback on behavioral outcome”, and “adding objects to the environment” predicted long-term effects (≥ 12 months).

Other meta-analyses report less consistent results. In a meta-analysis of physical activity and healthy eating interventions, Olander et al. (2013) found 21 BCTs associated with intervention effectiveness. However, the authors do not differentiate between physical activity and healthy eating interventions, and include both experimental and observational studies in their analysis. McDermott, Oliver, Iverson, and Sharma (2016) found that the BCT “providing information on consequences of behavior” was associated with positive changes in intention but no BCT was associated with positive changes in behavior. Instead, the BCT “provide feedback on performance” was associated with smaller changes in behavior.

The exploratory nature of the analyses, which do not allow causal inference, is a limiting factor of the meta-regressions of BCTs on intervention effectiveness. Thus, results may be subject to confounding by other intervention characteristics or systematic co-occurrence of BCTs, which has been noted in one study (O’Brien et al., 2015). Nevertheless, the research on BCTs suggests that behavior change strategies that target self-regulation, such as self-monitoring, goal setting and feedback, are associated with larger effect sizes of physical activity interventions. This highlights the importance of volitional processes for behavior change and illustrates the potential of mobile technologies to promote physical activity. Due to their inherent capability to passively collect physical activity data and feed these data back to the user in the form of numbers or visualizations, mobile technologies include characteristics and features that have been related to physical activity behavior change in empirical studies.

2.3.4 Summary

Behavior change theories enable a thorough understanding of the processes that lead to health behavior change and thus provide the basis for deriving and evaluating the conceptual model of the insurer's physical activity program. The existing theories highlight the fact that health behavior change is a complex process: a vast number of theories exist that differ on many dimensions, such as the level of influence of their determinants, the assumptions they make about the nature of behavior change, or the attention that is directed to volitional processes. One example of a health behavior change theory is the HAPA model (Schwarzer & Luszczynska, 2008). The HAPA model explicitly specifies volitional determinants of behavior change and is thus of great relevance for the insurer's physical activity program. More specifically, users of the insurer's program are likely to be motivated to change their behavior and may thus particularly benefit from volitional intervention strategies that help them to translate their motivation into actual behavior change.

Despite the undisputable importance of behavior change theories, they do not specify which behavioral determinants are most important for which behavior, which target group, or in which setting. Research on BCTs and associated mechanisms of action represents one approach to answering these questions that are essential in the process of intervention development. So far, existing studies of BCTs in physical activity interventions have highlighted the important role of intervention strategies that target behavioral regulation, such as self-monitoring, goal setting, and feedback, and thus have confirmed the significance of volitional processes for behavior change.

Collectively, this chapter has described the MOST framework (Collins, 2018), which guides the research presented in the following chapters, has introduced outcomes that capture an intervention's public health impact, (i.e. the RE-AIM framework, Glasgow, 1999), and has outlined the basics of behavior change theories. In the following chapter, this information is applied to the insurer's physical activity program to define outcomes and target effect sizes of the program (RQ 1) and to assess whether the program is capable of meeting these targets (RQ 2), e.g. by deriving its conceptual model.

Chapter 3

Requirements for Mobile Physical Activity Interventions

The main objective of this chapter is to answer the first two research questions of this dissertation. After defining target outcomes of mobile physical activity interventions based on the RE-AIM framework (Glasgow et al., 1999) and reviewing plausible effect sizes reported in the literature, the hypothetical potential impact fraction of mobile physical activity interventions is modelled in different scenarios, This helps to determine plausible minimum effect sizes for mobile physical activity interventions and to answer the first research question (section 3.1). As a next step, the potential of the insurer's program to meet the defined target effects is evaluated. First, the rationale of the program is briefly outlined from a health insurance perspective, before the program's components are described in detail. Subsequently, the program's conceptual model can be derived to investigate whether the program is targeting important behavior change determinants as proposed by behavior change theories (section 3.2). The conceptual model is complemented by two systematic literature reviews on mobile technologies (section 3.3) and financial incentives (section 3.4), the two main components of the program. These reviews help to quantify the effects that can be expected from the components of the program and, together with the conceptual model, help to answer the second research question. The chapter concludes with a brief summary (section 3.5).

3.1 Target Outcomes and Effect Sizes

In this section, relevant target outcomes and target effect sizes for mobile physical activity interventions are defined and the first research question of this dissertation is answered.

3.1.1 Target Outcomes

The first research question of this dissertation requires a definition of outcomes for scalable physical activity interventions that measure the intervention's public health impact. Recall that the RE-AIM framework (Glasgow et al., 1999) defines five outcomes on both, the individual and organizational levels that capture an intervention's public health impact (reach, efficacy, adoption, implementation, and maintenance). In the present setting, because the organization responsible for developing the program coincides with the organization responsible for its adoption and implementation, it seems reasonable to focus on the individual-level outcomes reach, efficacy and maintenance. For the purpose of this dissertation, the impact of the insurer's program (I) is therefore quantified by the product of its reach (R) and efficacy (E):

$$I = R * E$$

In line with the definitions of reach and efficacy (section 2.2), the participation rate (the proportion of participants among all those potentially eligible) is used to quantify the reach of the program. Given the high degree of agency inherent in the program, the characteristics of participants, such as age, health status, and income, can provide additional valuable information about possible selection effects. Measures of behavior change can be used to quantify the intervention's efficacy (and maintenance). Due to the program's focus on step counts, it seems evident to use changes in steps per day as the outcome to evaluate the program's efficacy. Steps per day are widely regarded as a suitable target outcome for large-scale physical activity interventions. Steps can be accurately measured by commercially available activity trackers under laboratory conditions (Case et al., 2015) and under free-living conditions for healthy adults (Feehan et al., 2018). Changes in steps per day primarily reflect walking, a subtype of physical activity that can be performed easily by the majority of the population, irrespective of age and without the need for skills, training, or equipment (J. N. Morris & Hardman, 1997). Meta-analyses of randomized studies have linked walking to reductions in all-cause mortality (Kelly et al., 2014), and type-2 diabetes risk (Aune, Norat, Leitzmann, Tonstad, & Vatten, 2015), and to increases in fitness (Murphy, Nevill, Murtagh, & Holder, 2007), especially for previously sedentary adults. Observational studies have linked increases in step counts to lower body mass index, greater insulin sensitivity, and greater stability of blood glucose values (Dwyer et al., 2011; Ponsonby et al., 2011).

After having defined reach and efficacy as the relevant outcomes for mobile physical activity interventions, target effect sizes on both outcomes need to be determined to fully answer the first research question. The next subsection tries to determine these target effects and the resulting public health impact of mobile physical activity interventions.

3.1.2 Target Effect Size

To determine target effect sizes for mobile physical activity interventions, this section first reviews effect sizes for reach and efficacy found in related work. These effects illustrate the range of plausible effects that can be expected from mobile physical activity interventions. As a next step, different scenarios of an intervention's potential impact fraction for varying degrees of reach and efficacy are created based on the previously identified plausible effects (pessimistic, realistic, and optimistic scenarios). These scenarios reveal whether plausible effects of mobile physical activity interventions will have a sufficient public health impact, or whether considerably greater effects are needed. Finally, based on the information presented in this subsection, the first research question is answered.

Effects Sizes Reported in Related Work

As mentioned in section 2.2.2, the reach of mobile physical activity interventions may be limited due to the high degree of agency required from participants. This assumption is supported by the existing literature, although the overall evidence is scarce. For example, a survey among US employers offering workplace wellness programs revealed that the majority of programs report participation rates between 0% and 20% in the comparable categories weight management, disease management and health coaching (Nyce, 2010). In a program utilizing activity trackers and rewards, which was offered by a health insurer in the US, the reported participation rate was only 1.2% during the first 1.5 years of the program (Patel et al., 2017). Thus, under real world conditions, mobile physical activity interventions can reasonably be expected to reach between 1% and 20% of their target population. However, reach appears to be substantially higher in research settings: a review of RE-AIM outcomes in mobile physical activity interventions reported a median reach of 51% (Blackman et al., 2013).

Regarding efficacy of mobile physical activity interventions, a wide range of effect sizes is reported in literature. Meta-analyses have reported effects ranging from around 400 steps (Direito, Carraça, Rawstorn, Whittaker, & Maddison, 2016) to around 2,500 steps (Bravata et al., 2007). In addition, an often reported target effect for mobile physical activity interventions is an increase of 3,000 steps per day, because this effect is required for a completely inactive participant to meet the recommended level of 150 minutes of moderate-intensity physical activity per week (Tudor-Locke et al., 2011). This approximation assumes that the 3,000 steps taken are over and above habitual activity with a cadence of at least 100 steps per minute to reasonably qualify as moderate-intensity physical activity. Given a commonly reported standard deviation of $SD = 2{,}500$ steps (Bravata et al., 2007), an increase of 3,000 steps per day corresponds to a large effect of $d = 1.2$. However, recall that the relationship between physical activity and health outcomes is generally curvilinear and substantial health benefits can be obtained at levels below the recommended level (Warburton & Bredin, 2017). Therefore, effects below 3,000 steps may be sufficient for interventions to create a substantial public health impact. Thus, related work has illustrated that an increase of 3,000 steps is an often-reported target effect for mobile physical activity interventions and some studies have reported effect sizes that come close to this effect. Nevertheless, smaller effects are likely more feasible to achieve and may confer a similar health benefit.

Effect Scenarios for the Mobile Physical Activity Interventions

To illustrate the potential public health impact of mobile physical activity interventions, the intervention's potential impact fraction (Morgenstern & Bursic, 1982) was modelled in three scenarios (pessimistic, realistic, and optimistic) with varying degrees of reach (between 1%, 10%, and 20%) and efficacy (1,500 steps and 3,000 steps). Specifically, the potential impact fraction (IF) was calculated for two physical activity-related NCDs, CHD and T2DM. As outlined in subsection 2.2.3, estimating the IF for mobile physical activity interventions requires knowledge of the distribution of physical activity (f_i) in the target population. This information was obtained from the Swiss Health Survey 2017 (Bundesamt für Statistik, 2018a). The Swiss Health Survey reports physical activity prevalence according to four ordinal categories: inactive (< 30 minutes of moderate activity per week), low active (30 – 149 minutes of moderate activity or vigorous activity once per week), sufficiently active (> 150 minutes of moderate activity or vigorous activity twice per week), and trained (vigorous activity three times or more per week). Further, information on the disease risk (risk ratio) for each category of the physical activity distribution (RR_i) is needed. This was obtained from dose-response meta-analyses of the relationship between physical activity and CHD (Sattelmair et al., 2011) and the relationship between physical activity and T2DM (Aune et al., 2015). RRs from these meta-analyses were mapped to the different activity categories as follows: for the low active category, the RR for meeting half the current physical activity recommendations was extracted; for the sufficiently active category, the RR for meeting the current physical activity recommendations was extracted; and for the trained category, the RR for meeting twice the current physical activity recommendations was extracted. Finally, information on the change of the physical activity distribution as a result of the intervention is needed, i.e. the fraction of individuals in each activity category (g_i) who are shifted to the next lower risk category. In line with the reasoning above, this fraction was set to 1 for a hypothetical intervention effect of 3,000 steps, meaning that all individuals reached by the intervention are shifted up one physical activity category. For an effect of 1,500 steps, this fraction was approximated by the ratio of disease risk reduction of 1,500 steps compared to 3,000 steps, which was obtained from the corresponding meta-analysis. For example, for T2DM an increase of 3,000 steps is associated with a risk reduction of 19% and an increase of 1,500 steps is associated with a risk reduction of 14%. To reflect this effect proportion, the fraction g_i for an effect of 1,500 steps was set to $0.14/0.19 = 0.74$.

To account for the varying degrees of reach in the three scenarios, the IF was applied only to the number of disease incidents that are expected among program participants, assuming 1% (pessimistic scenario), 10% (realistic scenario), and 20% (optimistic scenario) participation respectively for some hypothetical target population. The number of incident cases were estimated using the T2DM incidence rate reported by Huber, Schwenkglenks, Rapold, and Reich (2014) and the CHD incidence rate reported by Schäfer, Lorenz, Priess, and Bitzer (2016). Figure 3-1 depicts the results of the different scenarios for the incidence of T2DM (A) and CHD (B).

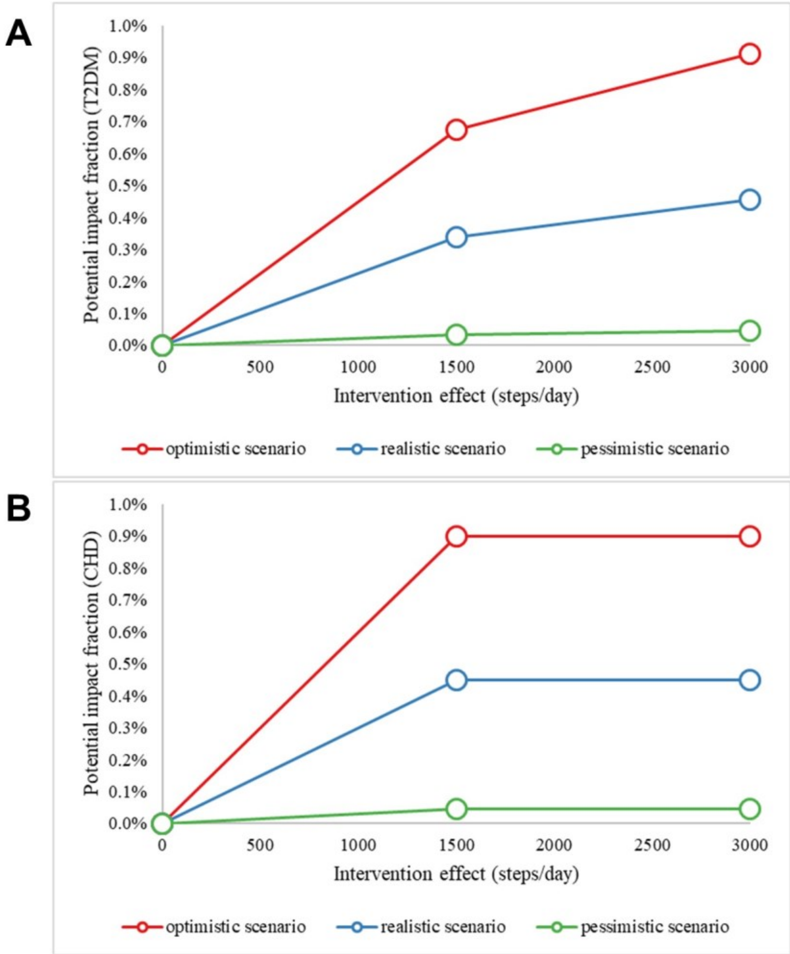


Figure 3-1. Potential impact fraction for T2DM (A) and CHD (B) for different intervention scenarios.

Three insights can be derived from Figure 3-1. First, within all three scenarios, the rise in IF is steeper between 0 and 1,500 steps than it is between 1,500 and 3,000 steps. This reflects the curvilinear relationship between physical activity and T2DM (Aune et al., 2015) and between physical activity and CHD (Sattelmair et al., 2011) and indicates that an intervention effect of 1,500 steps may confer a large proportion of the risk reduction of physical activity. In fact, for CHD, physical activity increases above 1,500 steps do not increase the IF. This is because in the meta-analysis by Sattelmair et al. (2011), the risk ratios corresponding to physical activity increases of 1,500 steps and 3,000 steps are equal (0.86). Second, the relative differences between the three scenarios remain fairly stable for different efficacy values, indicating that no effect threshold exists for the reach of the intervention. Rather, the overall impact of the program seems to increase linearly with reach, meaning that more reach consistently leads to a greater public health impact. Third, in line with the RE-AIM framework, there is a clear interaction between reach and efficacy. For example, given only 1% participation in the pessimistic scenario, even large increases of 3,000 steps do not translate to a meaningful public health impact. For the purpose of this dissertation, it therefore seems reasonable to expect a minimum reach of 10% for mobile physical activity interventions because a lower participation will likely limit the public health impact irrespective of the intervention's efficacy.

Because the IFs estimated in Figure 3-1 are relative numbers, the absolute public health impact of the three scenarios is difficult to interpret. One way to approach this problem is to compare the IFs above to the maximum effect that can be expected by increases in physical activity on the population level. If in the optimistic scenario above, an intervention effect of 3,000 steps were to reach the complete population, the resulting IF for CHD would be 4.5% indicating the upper threshold for the effect of mobile physical activity interventions. In the case of a target population of $N = 150{,}000$, which corresponds to the target population of the insurer's program (cf. subsection 3.2.2), this upper threshold would translate to an estimated 81 cases prevented annually. As a comparison, the identified target effects of 1,500 steps increase and 10% participation would translate to an estimated 8 cases prevented annually, highlighting that these are effect thresholds below which the contribution to prevention is negligible.

Some points need to be noted regarding the calculation of the IF above. Due to the number of assumptions involved in the calculations, the results are subject to uncertainty and may serve as a rule of thumb at best. For example, the models ignore intervention drop-outs, as well as failed maintenance of effects, and assume that the health benefits of physical activity take effect immediately. A further assumption is that only physical inactivity is affected by the intervention while distributions of other risk factors remain stable. Clearly, these assumptions greatly simplify reality. Another shortcoming of the scenarios above is that this uncertainty in the estimates was not quantified, e.g. by calculating 95% confidence intervals. Yet, despite these limitations, the estimates above align reasonably well with other estimates reported in the literature. For example, Lee et al. (2012) estimate that 6% of CHD cases globally and 7% of T2DM cases could be prevented if physical inactivity were to be eliminated. These proportions are somewhat above the corresponding proportions for a population-level intervention effect of 3,000 steps (i.e. 4.5%, see above), which is to be expected given that physical inactivity is less prevalent in Switzerland than in most other countries.

Answer to Research Question 1

After the relevant outcomes for the insurer's physical activity program have been defined (subsection 3.1.1) and information on target effect sizes has been gathered (subsection 3.1.2) the first research question of this dissertation can be answered:

RQ1: What are relevant outcomes and target effect sizes for scalable physical activity interventions?

Reach, i.e. participation rate, and efficacy, i.e. behavior change, are main determinants of an intervention's public health impact. Scalable physical activity interventions should increase physical activity by at least 1,500 steps per day on average and maximize uptake among the target population. To have a substantial impact on public health, the participation rate should be well above 10%.

As a next step, the insurer's physical activity program is evaluated with regard to its capability to meet the target effects defined above (RQ 2). Therefore, the insurer's program is introduced and its conceptual model is derived in the following subsection. The conceptual model illustrates via which processes the main components of the program could, in principle, affect the identified target outcomes reach and efficacy. Subsequently, the conceptual model is complemented with two systematic reviews of the literature that quantify the effects of the program's main components (sections 3.3 and 3.4). The conceptual model and results from the two reviews then help to answer the second research question.

3.2 The Insurer's Physical Activity Program

This section introduces the insurer's physical activity program. First, the rationale of the program is briefly outlined from a health insurance perspective, before the main components and the conceptual model of the program are described.

3.2.1 Background

Rising costs and tough competition have been identified as two key challenges for health insurers in Switzerland. Healthcare costs have increased by 165%, from 26.9 billion Swiss Francs (CHF) in 1990 to CHF 71.2 billion in 2014 (Ernst & Young, 2017).

Likewise, insurance premiums are expected to rise from CHF 396 per month on average in 2014 to CHF 826 per month on average in 2030 (Ernst & Young, 2017). Analyses reveal that the rise of NCD prevalence is the main driver of rising healthcare costs in Switzerland (Wieser et al., 2014). In fact, NCDs cause 80% of all direct and indirect healthcare costs in Switzerland today (Wieser et al., 2014). This development is threatening the businesses of health insurers. Motivated by continuously rising insurance premiums, the Swiss population has already voted four times on whether to restructure the health insurance market and introduce a uniform, government-controlled health insurance (1994, 2003, 2007, and 2014). Although the idea was rejected each time, acceptance in the population grew from 23% in 1994 to 38.5% in 2014 (Santesuisse, 2016). In addition, the rising healthcare costs have led to fierce competition that has resulted in a consolidation of the health insurance market. Of the 555 health insurers that were active in Switzerland in 1980 (Hefti & Frey, 2008), only 57 are still active today (Ernst & Young, 2017). Of those, the ten largest insurers hold almost 90% market share in the insurance market (Ernst & Young, 2017). As a consequence, health insurers face great pressure to find new ways to differentiate and to save costs to ensure competitiveness.

The development of innovative prevention programs, such as mobile physical activity programs, has been suggested as a strategic approach for health insurers to address their two main challenges: rising healthcare costs and strong competition (FSG, 2017; Kinder, Steingröver, & Neuhaus, 2017). In line with the concept of shared value opportunities (M. R. Kramer & Porter, 2011), this strategy is based on the fact that a health insurance company's profitability is dependent on the health outcomes of its insurees. In fact, actively engaging in prevention (and thus creating value on a societal level), has two distinct advantages for health insurers. First and foremost, successful prevention can reduce healthcare costs for insurees and thus counter the trend of rising costs and insurance premiums. Second, prevention programs offer the opportunity for health insurers to create interaction opportunities with insurees. Interaction has been identified as a key driver of customer loyalty for health insurers (Kinder et al., 2017), whose traditional activities, like claims management and cost coverage, require little to no interaction with insurees. Thus, prevention programs provide opportunities for health insurers to overcome the main challenges of the health insurance market: rising costs and fierce competition.

One such prevention program was developed by the partnering health insurance company and serves as the exemplary mobile physical activity intervention for this research. The next subsection briefly describes the main components of this program.

3.2.2 Program Components

In light of the challenges for Swiss health insurance companies outlined in the previous subsection, the health insurer supporting the research of this dissertation developed and launched a reward-based mobile physical activity promotion program on the complementary insurance market in 2015[5]. The program is part of a wide range of preventive activities, such as gym memberships, health checks, and smoking cessation programs, which are subsidized by the insurer when an insuree takes out two complementary insurance policies. The combined amount of subsidy for all activities ranges from CHF 300 to CHF 700 per year, according to the level (and premium) of complementary insurance chosen by the insuree (economy, balance, or premium), and is financed by profits made from the corresponding complementary insurance policies (M. Stäheli, personal communication, July 8, 2019). Further, to encourage uptake of a diverse set of preventive activities, the insuree must split the subsidy between different activity categories. For example, on the economy level, the insuree may spend a maximum of CHF 150 (from a total possible subsidy of CHF 300) on preventive activities related to physical activity, thus determining the maximum level of rewards that can be paid out to economy-level insurees. This amount increases up to CHF 500 on the premium level. At the time of writing, 150,000 insurees are eligible for participation in the program (M. Stäheli, personal communication, August 5, 2019).

Insurees eligible for the physical activity program are able to connect wearable devices from some of the best-known wearable manufacturers –Fitbit, Garmin and Jawbone – to their personal accounts on the insurer's online platform to share the number of steps walked per day with the insurance company. Insurees who do not own a suitable device are eligible to purchase one with a discount of 20%. Once registered, the insuree agrees that their number of steps per day is shared on a daily basis between the wearable manufacturer and the insurer via an API. No other data are recorded by the wearable

[5] Note that health insurance coverage in Switzerland is divided into mandatory and complementary insurance policies that are subject to different legislations and regulatory bodies. Due to strict regulations on the mandatory insurance market, the mobile physical activity program can only be offered on the complementary insurance market. However, 75% of the Swiss population own both mandatory and complementary insurance policies (Eisler & Lüber, 2006)

device are shared with the insurer. Participants who achieve a monthly average of at least 7,500 but less than 10,000 steps per day receive a monthly reward of CHF 5. Participants who walk 10,000 steps or more on average per day are entitled to receive a reward of CHF 10 each month. This reward is paid out to participants by the health insurer. The two levels of step goals, 7,500 and 10,000 steps respectively, were chosen based on a review article that identified these amounts as the lower and upper limits of approximate translations of the recommended 150 minutes per week of moderate aerobic physical activity into steps per day (Tudor-Locke et al., 2011). Overviews of step counts and earned rewards can be accessed by participants through a dashboard (Figure 3-2) that is available on the insurer's online platform. Rewards are also communicated monthly via email to program participants, although participants can opt out of receiving emails.

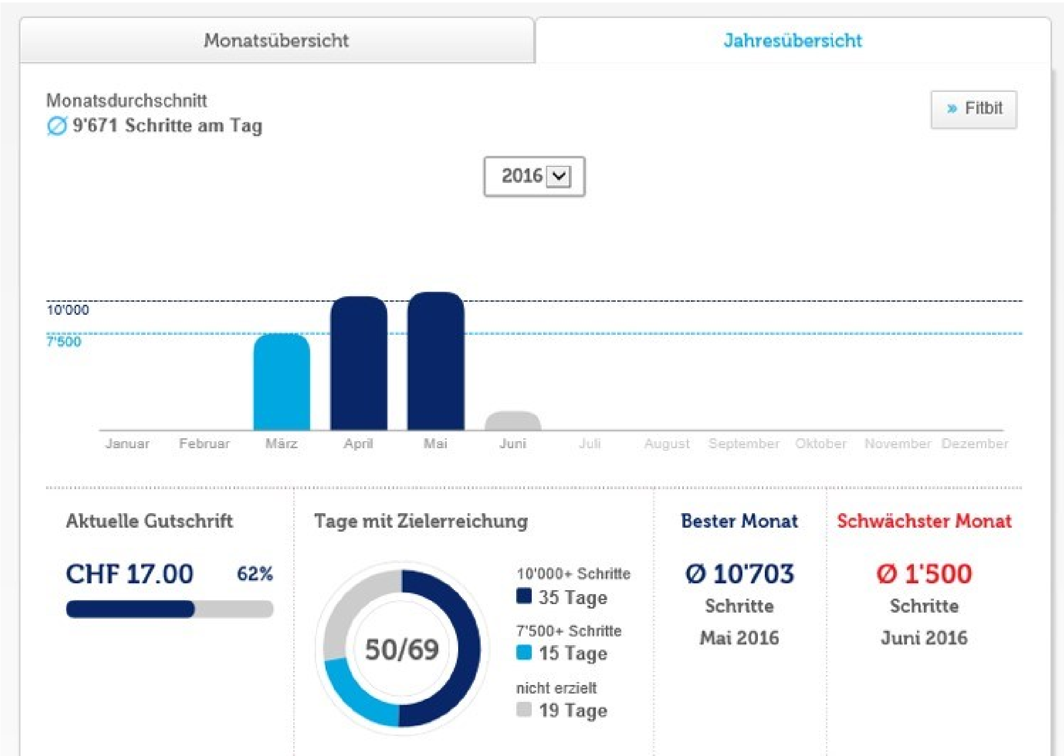


Figure 3-2. Online dashboard of the mobile physical activity program.

As mentioned briefly in subsection 2.2.2, the prevention approach that underlies the insurer's program can be classified as a population-level strategy with high individual agency. The program targets the entire population of eligible insurees because there are no health-related restrictions to participation and there is no screening prior to participation to identify at-risk individuals. Likewise, the program requires a high degree of individual agency for eligible insurees to benefit from the program: insurees are required to own or buy an activity tracker, connect the device to their account on the insurer's online platform, and sustainably engage in physical activity to meet the program's monthly step goals.

To sum up, the insurer's physical activity program contains three main components: activity trackers, financial incentives, and an online dashboard. The next subsection explains how these components could affect the program's target outcomes reach and efficacy.

3.2.3 Conceptual Model

The description of program components (subsection 3.2.2) together with information on the program's target outcomes (subsection 3.1.1) and on behavior change theories and BCTs (section 2.3) is sufficient to derive the conceptual model that explains how the insurer's physical activity program might affect its target outcomes. The program contains three major components – financial incentives, activity trackers and the online dashboard – that may contain one or multiple BCTs that are hypothesized to affect the program's reach and efficacy via two distinct mechanisms of action (Figure 3-3.).

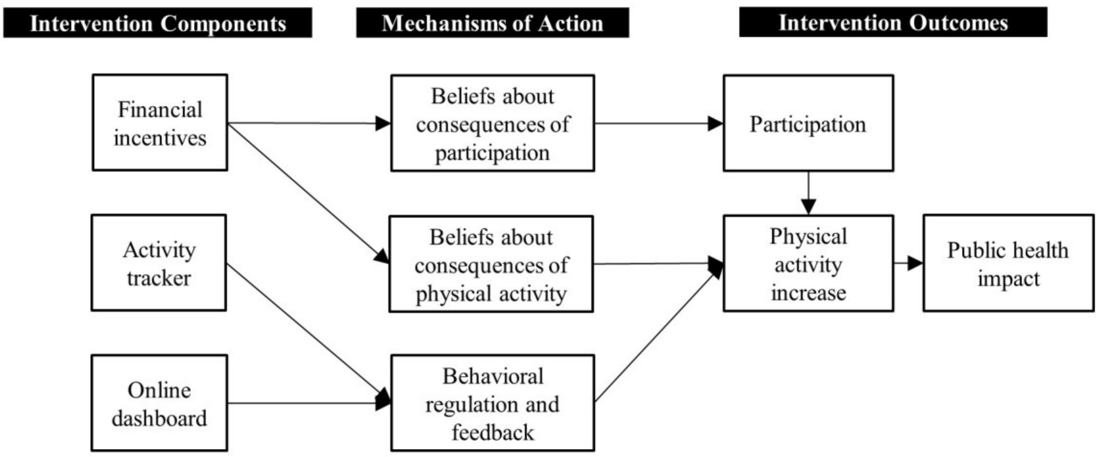


Figure 3-3. Conceptual model of the insurer's physical activity program. All effects shown are assumed to be positive.

The financial incentives (BCT 10.1: material incentive) used in the insurer's physical activity program could affect participation and behavior change by changing participants' beliefs about the consequences of participating (e.g. "if I participate in the program, I will be able to earn financial rewards") and encouraging engagement in physical activity (e.g. "if I walk more than 10,000 steps per day on average, I will earn CHF 10"). Most behavior change theories realize that evaluations of pros and cons of a behavior are a central determinant of motivation (Williams, Anderson, & Winett, 2005). When forming such outcome expectancies, individuals can be subject to the so-called present bias (Loewenstein, Asch, & Volpp, 2013), i.e. the tendency to overvalue immediate consequences and undervalue distant consequences. This suggests that more immediate incentives should be more effective, especially because the benefits inherent in health behaviors (e.g. improved health or appearance) typically lie in the distant future (Hall & Fong, 2007). The activity tracker and the online dashboard measure and visualize participants' physical activity data (BCTs 2.2: feedback on behavior and 2.3: self-monitoring of behavior) and, together with the associated step goals (BCT 1.1: goal setting), support participants' action control, i.e. self-regulatory processes that direct behavior towards an increase and maintenance of physical activity (Schwarzer & Luszczynska, 2008). As briefly outlined in sections 2.3.2 and 2.3.3, action control is assumed to be a central volitional determinant of behavior, and research on BCTs has confirmed that physical activity interventions that focus on action control processes produce greater effects. Although this increases confidence in the program's ability to produce the desired change of increased physical activity, the conceptual model also illustrates potential shortcomings of the program. For example, the program's components target (some) behavior change processes only on the individual level. Section 2.3 illustrated that physical activity is a complex behavior with important determinants on the interpersonal and environmental levels as well. This calls into question whether the program's effects will be sufficient and whether they can be sustained long enough to create the required impact on the health of its target population. For example, one might doubt whether the motivation to be more active that is induced by financial incentives can be upheld within social settings that do not value physical activity or within environments that constantly create incentives for alternative sedentary activities.

Finally, some additional points need to be kept in mind regarding the program's conceptual model. First, the insurer did not design the program's components based on knowledge of important physical activity determinants. Instead, the conceptual model is built on a retrospective assessment of the program components in the light of behavior change theory. Thus, rather than providing a rationale for the design of the program, the conceptual model in Figure 3-3 primarily explains how the intervention could, in principle, affect relevant outcomes. Second, the conceptual model is inherently subjective and different mechanisms of action are also plausible. For example, activity trackers could additionally act as an environmental cue that prompts physical activity during participants' everyday lives, a change process proposed by the Transtheoretical Model of Change (Prochaska & DiClemente, 1982). Likewise, the financial incentives are tied to graded step goals and their effect may partly rely on affecting self-efficacy, i.e. participants' beliefs about their ability to perform the target behavior (Bandura, 1986). In addition, learning theories and positive reinforcement represent a complementary conceptualization of the mechanism of action of financial incentives (Domjan, 1998). Lastly, two additional components, monthly emails and the discount on activity trackers, are left out of the model to avoid greater complexity. While these components do not necessarily apply to all program participants, they may still affect the program's reach and efficacy.

3.2.4 Summary

The insurer's physical activity program is offered on the complementary insurance market and consists of three main components, financial incentives, an activity tracker and an online dashboard. To have a substantial impact on public health, the program should increase the physical activity levels of its participants by a minimum of 1,500 steps per day and maximize uptake among the target population. At the very minimum, at least 10% of the program's target population should be reached. The program's conceptual model illustrates that its components can be linked to both motivational and volitional determinants as well as corresponding mechanisms of action on the individual level that have been identified by behavior change theories (cf. section 2.3). These theories, and the accompanying empirical evidence, support the assumption that components of the insurer's mobile physical activity program can, in principle, positively affect the program's reach and efficacy, although the size and sustainability of these potential effects are unclear. To complement the conceptual model of the insurer's physical activity program, and to answer the second research question of this dissertation, a review of the empirical evidence on the program's components is necessary. The following sections systematically review the empirical evidence on two of the program's components – activity trackers and financial incentives – and quantify their effects in steps per day.

3.3 A Review of Mobile Physical Activity Interventions

3.3.1 Introduction

Due to its great potential, there are extensive scientific publications that have examined the effect of mobile technologies, such as activity trackers, smartphones and pedometers, on physical activity. Further, there have been multiple attempts to review the existing literature and quantify the effect of mobile physical activity interventions. For example, a simple hand search on Google Scholar using the search terms "physical activity", "mobile", and "review" revealed at least eight potentially relevant reviews among the first 20 hits, which summarize the evidence on mobile physical activity interventions. To consolidate knowledge around mobile physical activity interventions, an overview of systematic reviews was conducted. The process of preparing the overview followed the recommendations for overviews of reviews in the Cochrane Handbook for Systematic Reviews of Interventions (Becker & Oxman, 2008). The overview was first conducted in 2016 in an unsystematic way and revised and updated in 2019 to reflect the current level of empirical evidence.

3.3.2 Methods

Selection Criteria

To qualify for inclusion in the overview, reviews must summarize the results of RCTs that evaluate the effects of mobile interventions on physical activity in adult populations, and must report a summary effect on steps per day. Mobile physical activity interventions are defined as interventions with at least one mobile technology component, such as smartphones, smartwatches, pedometers or activity trackers, that aim to increase physical activity. No selection criteria were specified with regard to comparison groups. The included reviews must be published in a peer-reviewed journal in English.

Search methods

To identify candidate reviews, the PubMed database was searched with the following search string "((Physical Activity[Title] OR Walking[Title] OR Exercise[Title]) AND (Review[Title] OR Systematic Review[Title] OR Meta-Analysis[Title] OR Meta Analysis[Title]) AND (Mobile[Title/Abstract] OR Smartphone[Title/Abstract] OR mHealth*[Title/Abstract] OR App*[Title/Abstract] OR Wearable*[Title/Abstract] OR

Pedometer*[Title/Abstract] OR Activity Tracker*[Title/Abstract]) AND (Randomized Controlled Trial*[Title/Abstract] OR RCT*[Title/Abstract]) NOT (Child*[Title] OR Adolescent*[Title] OR Protocol[Title]))". Titles and abstracts of all resulting articles were screened and checked against the inclusion criteria to exclude irrelevant articles. Subsequently, full texts of the remaining potentially relevant reviews were screened to make a final inclusion or exclusion decision. Reference lists of included reviews were screened and a hand search on Google Scholar was performed to identify other relevant reviews that may have been missed during the search process. The search was performed by a single reviewer.

Data Collection and Analysis

The following data were extracted from each included review: author, title, year of publication, mobile technologies used in primary studies, intervention characteristics, types of comparison groups, average intervention duration, average sample size, number of comparisons, and the summary effect. For each summary effect, the point estimates and 95% confidence intervals were extracted, along with the number of studies and total sample size. Data extraction was performed by a single reviewer.

Search results were compared to the selection criteria using a checklist. Subsequently, extracted data from included reviews were recorded into a spreadsheet. The review authors were contacted to resolve uncertainties or if data was missing. To assess the quality of included reviews, the tool for the Assessment of Multiple Systematic Reviews (AMSTAR; Shea, Grimshaw, et al. (2007); Appendix B) was used. AMSTAR is a reliable and valid 11-item quality measurement tool for systematic reviews (Shea, Bouter, et al., 2007). Further, the quality of evidence for each included review was evaluated using the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework (Guyatt et al., 2008). The GRADE framework provides a systematic approach to rating the quality of a body of evidence according to four certainty categories:

- High: We are very confident that the true effect lies close to that of the estimate of the effect.
- Moderate: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- Low: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.

- Very low: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

According to GRADE, reviews of RCTs initially start with a high certainty rating that is subsequently adjusted based on additional criteria (Balshem et al., 2011). More specifically, and in line with the GRADE recommendations, the quality of evidence of each included review is rated down one level if one the following criteria is met:

- Risk of bias: most information that contributes to the summary effect comes from studies that are at least at moderate risk of bias; that is, there is at least one crucial limitation for one bias category or some limitations for multiple categories, which are sufficient to lower the confidence in the overall estimate (Guyatt, Oxman, Vist, et al., 2011).
- Publication bias: the evidence consists of a number of small studies or publication bias is suggested by funnel plot asymmetry or statistical tests (Guyatt, Oxman, Montori, et al., 2011).
- Imprecision: the combined sample size of all studies that inform the summary effect does not meet the optimal information size (OIS) criterion or both potential harms and benefits are included in the limits of the summary effect's confidence interval (Guyatt, Oxman, Kunz, Brozek, et al., 2011). The OIS criterion refers to the overall sample size of a review that is necessary to detect a meaningful intervention effect (see below).
- Inconsistency: there is substantial variation of point estimates and minimal overlap between CIs of studies that contribute to the review's summary effect. Statistical tests of heterogeneity are significant at $\alpha < .05$ and the I^2-statistic exceeds 50% (Guyatt, Oxman, Kunz, Woodcock, Brozek, Helfand, Alonso-Coello, Glasziou, et al., 2011). The I^2-statistic reflects the proportion of variance of the effects in primary studies that is due to true differences of effects and not to chance (Borenstein, Hedges, Higgins, & Rothstein, 2011).
- Indirectness: the majority of interventions and populations in the review are substantially different from the insurer's physical activity program and its targeted populations, and these differences are likely to lead to a difference in outcomes (Guyatt, Oxman, Kunz, Woodcock, Brozek, Helfand, Alonso-Coello, Falck-Ytter, et al., 2011). Specifically, the evidence of included reviews was rated down for indirectness if the majority of the primary studies reported effects of interventions that included face-to-face components (e.g. counselling) in

patient populations where physical activity interventions are known to be more effective (Conn, Valentine, & Cooper, 2002; Ebrahim et al., 2011).

Data across included reviews were synthesized using comparisons of included summary effects in tabular and graphical forms.

A Note on Bias in RCTs

To assess risk of bias in included studies, review authors typically evaluate the included studies with regard to multiple categories of bias that can occur in RCTs. These categories are, for example, defined in chapter 8 of the Cochrane Handbook of Systematic Reviews of Interventions (Higgins & Altman, 2008). They include: selection bias (systematic differences in baseline characteristics between study groups, e.g. due to lack of randomization or failure to conceal the allocation sequence from the person who is responsible for deciding whether to include participants in the trial); performance bias (systematic differences between study groups with regard to efforts made by study participants or personnel, e.g. due to lack of blinding to group allocation); detection bias (systematic differences in outcome assessments between the study groups, e.g. due to lack of blinding); attrition bias (differences in outcomes between study groups due to systematic differences in participant withdrawal); and reporting bias (systematic differences between reported and unreported findings). In the present overview of reviews, less importance was given to bias that originates from lack of blinding, because evidence suggests that bias associated with lack of blinding is negligible in trials where outcomes are objectively assessed (L. Wood et al., 2008).

A Note on the OIS criterion

The OIS criterion was suggested by the authors of the GRADE framework to evaluate imprecision of a summary effect (Guyatt, Oxman, Kunz, Brozek, et al., 2011). Similar to a power analysis of a regular RCT, the OIS defines the overall sample size that is necessary for a summary effect to make adequate conclusions about the effect of an intervention. Because the goal of a summary effect is to quantify the true effect of an intervention, the total sample size of included studies should yield sufficient power to detect the minimum intervention effect that seems feasible to detect. For the present overview of reviews, the OIS was therefore set to a small standardized effect of $d = 0.2$, which corresponds to an unstandardized intervention effect of 500 steps, given a standard deviation of step counts of $SD = 2{,}500$ steps (cf. section 3.1.1). This effect is very similar to the effect of non-mobile physical activity interventions in healthy adults (Conn et al., 2011). The resulting OIS to detect such an effect is $N = 800$ assuming an alpha level of $\alpha = .05$ and a power of $1\text{-}\beta = .80$ (Faul, Erdfelder, Lang, & Buchner, 2007). The OIS is preferred to alternative methods of evaluating imprecision, such as judging the width and borders of confidence intervals of included studies, because confidence intervals are known to be fragile in small samples and can thus be potentially misleading (Guyatt, Oxman, Kunz, Brozek, et al., 2011).

3.3.3 Results

Search Process

The search of the PubMed database identified 30 potentially relevant reviews, of which four were included in the overview. Five reviews were additionally identified from screening reference lists of included reviews and conducting an additional search on Google Scholar, bringing the total number of included reviews to nine (Figure 3-4).

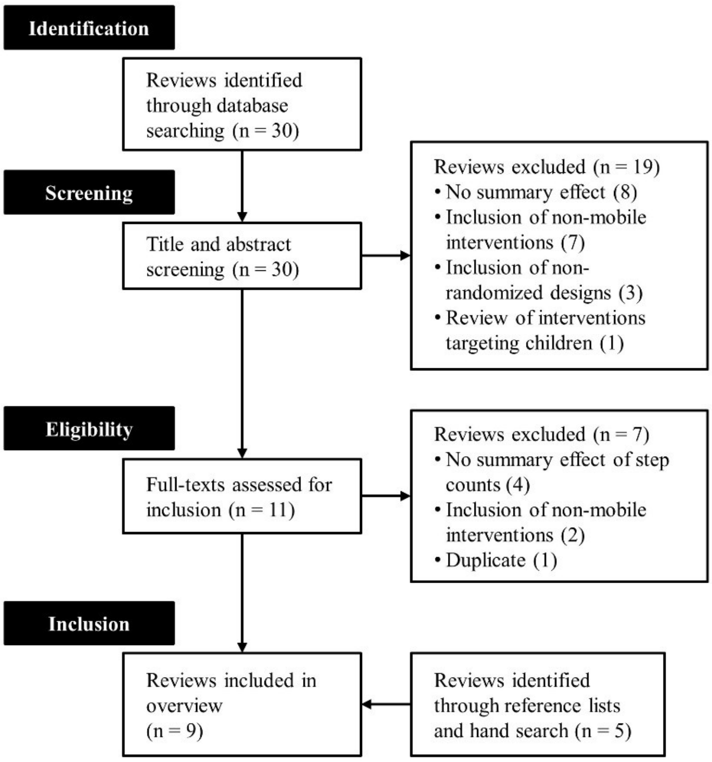


Figure 3-4. Selection process of reviews of physical activity interventions.

The main reasons reviews were excluded was related to reviews that did not perform a meta-analysis, and consequently did not report a summary effect of reviewed interventions, and to reviews that included interventions that did not utilize mobile technologies.

Characteristics of Included Reviews

Table 3-1 summarizes the characteristics of included reviews. All reviews were published between 2007 and 2019 and mostly included studies in at-risk populations (e.g. elderly or overweight) and patient populations. No review explicitly focused on the effect of mobile technology in the general population. Reviews differed somewhat with regard to the technology used, with some reviews (3/9, 33%) including different mobile technologies (e.g. pedometers, wearables and smartphone apps) and others focusing on single technologies, mostly pedometers (6/9, 66%). Beyond mobile technologies, most interventions included in the reviews utilized combinations of goal setting, in-person or telephone coaching, and the provision of educational or informational material. Yet, implementation varied greatly between studies, which ultimately led to a great variety of interventions included in each review. For example, interventions included in the reviews ranged from the use of pedometers and daily step goals (e.g. Araiza, Hewes, Gashetewa, Vella, & Burge, 2006) to wearables in combination with educative print material, weekly group sessions and monthly telephone coaching (e.g. Shuger et al., 2011). However, reviews that explicitly focused on patient populations tended to include more intensive interventions (e.g. mobile technology in combination with in-person counselling), while reviews in at-risk populations tended to include interventions with mobile technology as the core component.

Although rarely specified explicitly in the study selection criteria, most primary studies compared the interventions to an active control group that also received components to increase physical activity. Examples include alternative interventions, interventions that excluded the mobile technology component, or standard of care interventions in patient populations. However, the reviews of Brickwood et al. (2019) and de Vries et al. (2016) included mostly passive control groups that received no intervention, such as waitlist control groups or blinded pedometer controls. Most reviews included studies with small sample sizes and short interventions. Average sample sizes of reviews ranged from N = 35 (Bravata et al., 2007) to N = 199 (Gal et al., 2018), with only three reviews reporting an average sample size greater than 150 (Brickwood et al., 2019; Gal et al., 2018; Romeo et al., 2019). Duration of included interventions ranged from 2 weeks to 52 weeks but was roughly comparable between reviews with an average intervention duration of around 16 weeks (± 3 weeks). Only one review of interventions in type 2 diabetes patients (Qiu et al., 2014) included substantially longer interventions and reported an average duration length of 28 weeks.

Table 3-1. Characteristics of mobile physical activity intervention reviews.

Review	Population	Mobile Technology	Interventions	Comp	NOC	Ø Length	Ø *N*	AMSTAR
Bravata et al. (2007)	At-risk and patients	Pedometers	Mostly pedometer and goal setting; some include supervised exercise sessions	active	8	13.4 (9.2)	35	5
Brickwood et al. (2019)	At-risk	Commercially available wearables	Mostly wearable, goal setting and education; some include rewards	passive	13	13.4 (9.2)	165	9
de Vries et al. (2016)	Overweight and obese	Pedometers	Pedometer, goal setting and coaching	passive	5	16.9 (13.2)	83	7
Direito et al. (2016)	At-risk	Wearables, pedometers, apps	Mobile technology, goal setting and education	active	6	11.6 (3.6)	67	8
Gal et al. (2018)	At-risk and patients	Wearables, pedometers, apps	Mobile technology and goal setting; some include education or coaching	active	7	19.7 (14.2)	199	9
M. A. Kirk et al. (2019)	Cardio-metabolic patients	Pedometers and accelerometers	Mobile technology, goal setting and coaching; some include education	active	19	16.1 (16.7)	77	6
Qiu et al. (2014)	T2DM patients	Pedometers	Pedometer, coaching and activity diary	active	7	28.2 (20.3)	123	6
Romeo et al. (2019)	At-risk and patients	Smartphone apps	Standalone smartphone apps that include goal setting and other features	active	6	17.0 (17.3)	196	9
Vaes et al. (2013)	T2DM patients	Pedometers	Pedometer and coaching; some include an activity diary	active	7	23.4 (15.4)	90	7

Note. Comp: comparator, NOC: number of comparisons that contribute to the summary effect, T2DM: type 2 diabetes mellitus. Average length is given in weeks with standard deviation in brackets. Characteristics of interventions and comparators are main characteristics of primary studies that contributed to the reported summary effect. At-risk populations are populations with known low activity levels, such as sedentary, elderly, or overweight populations.

The methodological quality of reviews ranged from 5 to 9 (out of a maximum of 11) on the AMSTAR scale (mean: 7.3). The most common methodological shortcomings were not providing a list of excluded studies (9/9 reviews, 100%), not reviewing unpublished and grey literature (8/9, 89%), not providing an a-priori specified design of the review (5/9, 56%), and not considering the quality of included studies appropriately when formulating conclusions and recommendations (5/9, 56%). The full AMSTAR assessment of each review is available in 0.

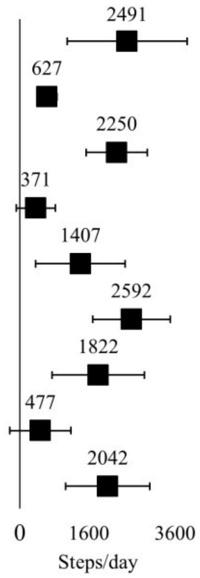
Summary Effects of Mobile Physical Activity Interventions

The summary effects of mobile physical activity interventions on steps per day that were reported in the included reviews are illustrated in Table 3-2. All but two reviews (Direito et al., 2016; Romeo et al., 2019) reported statistically significant effects in favor of mobile physical activity interventions. However, the size of reported summary effects varied greatly between reviews and ranged from a statistically non-significant effect of 371 steps (Direito et al., 2016) to a highly significant effect of more than 2,500 steps (M. A. Kirk et al., 2019). The effects of mobile physical activity interventions seemed to be smaller in at-risk populations (Brickwood et al., 2019; Direito et al., 2016; Romeo et al., 2019) than in patient populations (M. A. Kirk et al., 2019; Qiu et al., 2014; Vaes et al., 2013). Similar to the great variation of effects between reviews, the confidence intervals in Table 3-2 suggest great variation between the primary studies in all but two reviews (Brickwood et al., 2019; Direito et al., 2016). The large I^2 values (i.e. $I^2 > 50\%$) indicate that this variation is not random but largely due to true effect differences of primary studies. None of the reviews reported on the reach of the interventions investigated in primary studies.

The quality of evidence of the included reviews was rated as low or very low according to GRADE in all but two reviews (Brickwood et al., 2019; Direito et al., 2016). The main reasons for rating down the quality of evidence referred to potential publication bias that was strongly suspected in 4/9 (44%) reviews, imprecision of effects (4/9, 44%), and inconsistency of effects (6/9, 67%). The two reviews with moderate and high-quality evidence reported small effects of 371 steps and 627 steps respectively with no heterogeneity. The complete GRADE rating of all reviews is available in Appendix D.

Table 3-2. Summary effects reported in included reviews.

Review	I^2	GRADE	Summary Effect
Bravata et al. (2007)	91%	very low	
Brickwood et al. (2019)	3%	high	
de Vries et al. (2016)	54%	low	
Direito et al. (2016)	0%	moderate	
Gal et al. (2018)	90%	low	
M.A. Kirk et al. (2019)	91%	very low	
Qiu et al. (2014)	86%	low	
Romeo et al. (2019)	72%	low	
Vaes et al. (2013)	73%	very low	



Note. All summary effects favor the intervention group.

3.3.4 Discussion

Principal Findings

Most reviews on mobile physical activity interventions focus on effects in at-risk and patient populations, are of moderate methodological quality, and produce low quality evidence of intervention effects. However, a few higher quality reviews exist. These reviews indicate that interventions that combine mobile technologies with goal setting and/or educational components can increase physical activity by up to 600 steps per day in at-risk populations. This effect is similar to the increase of 496 steps found in a meta-analysis of non-mobile physical activity interventions in healthy adults (Conn et al., 2011). Most studies were of short duration so that no conclusions can be drawn regarding the sustainability of effects. However, one recent trial from the UK (N = 1,023) reported a sustained effect of similar size for a 12-week, pedometer-based intervention including a physical activity handbook and a walking plan for older adults (45–75 years) at one year (642 steps, 95%CI [329, 955]) and three year (627 steps, 95%CI [198, 1056]) follow-ups (Harris et al., 2017; Wahlich et al., 2017).

Overall, there is great variation between the summary effects of reviews of mobile physical activity interventions. Explaining these differences is challenging due to the wide variety of populations, intervention components and control groups that were included in the primary studies. However, the intervention's target group may be an important factor. Effects of mobile physical activity interventions appear to be substantially greater in patient populations than in at-risk populations, a pattern that has also been observed for non-mobile physical activity interventions (Conn et al., 2002). For example, summary effects of reviews that focus exclusively on the effects of mobile physical activity interventions in patient populations vary between 1,800 and 2,600 steps per day. Several factors could explain this difference in effect sizes. First, in line with the assumptions of the Health Belief Model (Rosenstock, 1974), patients may experience a greater sense of urgency and thus greater motivation for behavior change. This may be especially true for patients who are willing to participate in clinical studies. Second, physical activity levels in patients are typically low (Morrato, Hill, Wyatt, Ghushchyan, & Sullivan, 2007; Sisson et al., 2010) and increasing physical activity may be easier for this population. Third, mobile physical activity interventions in patient populations often include additional intervention components, such as in-person counselling. Thus, in the included reviews, patient populations may have received more powerful interventions that in turn had stronger effects.

Implications

Mobile physical activity interventions can increase physical activity by 600 steps at best in at-risk populations, and potentially by a much larger volume in diseased populations. Thus, based on the existing literature, an effect of 600 steps at best can be expected from the activity tracker component in the insurer's physical activity program, although some interventions included in the reviews leveraged additional components. This effect alone is insufficient to produce a meaningful reduction of disease incidence in the target population of the insurer's program. As outlined in section 3.1.2, an effect of 1,500 steps would be a reasonable target effect size for mobile physical activity interventions. Thus, the insurer's physical activity program, as well as mobile prevention programs in general, need to combine mobile technologies with additional intervention components to produce a sufficient overall effect on behavior change.

Limitations and Future Work

Some limitations of the present overview of reviews have to be noted. All reviews of interventions in patient populations reported great heterogeneity, which limits the informative value of summary effects and indicates the presence of moderating variables. Beyond the intervention's target group, the great variety of intervention components utilized in the included studies may be one important source of heterogeneity. In addition, most RCTs included in the reviews were small-scale pilot studies. As a consequence, many reported summary effects lack precision and may be subject to publication bias. Thus, results of this overview, especially with regard to patient populations, need to be interpreted with caution. Further, assessing the characteristics of intervention and control groups was aggravated by the lack of detailed reporting in both reviews and primary study reports, and was therefore restricted to rather broad categories of intervention components. In addition, the systematic literature search was restricted to one database and the selection and coding of reviews was performed by a single reviewer. This may limit the representativeness of reviews and the objectivity of results.

Large-scale randomized controlled trials with longer follow-up periods are needed to confirm the findings of this overview. In addition, and as illustrated by the MOST framework (Collins, 2018), optimization trials need to investigate which intervention components of mobile physical activity interventions work in which populations. In particular, surprisingly little is known about the isolated effect of mobile technologies in the general population, a research gap that is becoming more and more important as the adoption of wearable technology continues to increase (Statista, 2017).

Conclusion

The results of mobile physical activity interventions seem to mirror those of non-mobile interventions. Keeping the limitations of this overview in mind, the best available estimate of the effect of the mobile technology component in the insurer’s program on physical activity levels of its participants is around 600 steps per day. Thus, despite increasing adoption and the potential for scalable and population-level interventions, mobile technologies alone might only make a small contribution to public health and prevention. However, mobile technologies may be a valuable addition to chronic disease management programs where substantially larger effects can be expected.

3.4 A Review of Financial Incentives

3.4.1 Introduction

Similar to mobile physical activity interventions, the literature on financial incentives and physical activity is abundant. One reason for the great interest in the effects of financial incentives on health behaviors is regulations that have been introduced in various countries in recent years that allow insurers or employers to incentivize health behaviors. For example, the Affordable Care Act introduced in 2010 in the US allowed American insurance companies to tie premiums to the achievement of prevention-related health goals under certain conditions (Madison, Schmidt, & Volpp, 2013). Similarly, since 2004, health insurers in Germany have been able to provide incentives for participation in quality assured prevention programs as long as the incentives are financed by savings in healthcare costs that are attributable to the prevention programs (Stock et al., 2010).

The numerous empirical investigations on incentives and health behavior have utilized many different forms of financial incentives. For example, mobile technologies enable insurers to offer financial incentives (for physical activity) that are contingent on participants engaging in the target behavior and are obtained immediately after they reach their behavioral goals (e.g. a daily step goal). Traditional incentives, on the other hand, often target intermediate outcomes, such as participation in exercise sessions. J. Adams, Giles, McColl, and Sniehotta (2014) have identified a total of nine domains that describe the differences between financial incentives used in the scientific literature (Table 3-3).

Table 3-3. Design dimensions of financial incentives, adopted from J. Adams et al. (2014).

Dimension	**Explanation**
Direction	Whether the incentive is a positive gain or the avoidance of a negative loss
Form	The type of incentive, e.g. cash or voucher
Magnitude	Total monetary value of the incentive
Certainty	The degree of certainty that the incentive will be received
Target	Whether the incentive targets a process (e.g. participation in exercise classes), an intermediate (e.g. physical activity), or an outcome (e.g. weight loss)
Frequency	Whether all or only some of the occurrences of the behavior are incentivized
Immediacy	The delay between the behavior and the incentive
Schedule	Whether the magnitude of the incentive is fixed or varies over time
Recipient	Whether the recipient of the incentive is the individual performing the behavior, a group, a significant other, the physician etc.

Multiple attempts have been made to summarize the empirical evidence on the effects of financial incentives and on one or several of the design dimensions listed in Table 3-3. These reviews suggest that financial incentives can improve preventive behaviors (e.g. immunizations or cancer screenings; Kane, Johnson, Town, & Butler, 2004), smoking cessation (Giles, Robalino, McColl, Sniehotta, & Adams, 2014), dietary behavior (Purnell, Gernes, Stein, Sherraden, & Knoblock-Hahn, 2014), and exercise session attendance (Mitchell et al., 2013; Strohacker, Galarraga, & Williams, 2014), for as long as the incentives are in place and, in particular, when they are of larger value, tied to specific behavioral goals, and are received with certainty upon goal achievement.

To consolidate the existing knowledge on financial incentives for the promotion of physical activity, this section describes the conduction and results of an overview of reviews. Again, the process of preparing the overview followed the recommendations for overviews of reviews in the Cochrane Handbook for Systematic Reviews of Interventions (Becker & Oxman, 2008). The overview was first conducted in 2016 in an unsystematic way and revised and updated in 2019 to reflect the current level of empirical evidence.

3.4.2 Methods

Selection Criteria

To qualify for inclusion in the overview, reviews must summarize the results of RCTs that evaluate the effects of financial incentives on physical activity in adult populations, and must report a summary effect on steps per day. Financial incentives were defined as any reward with a clearly quantifiable monetary value, such as cash or vouchers. No selection criteria were specified with regard to comparison groups. The included reviews must be published in a peer-reviewed journal in English.

Search Methods

To identify candidate reviews, the PubMed database was searched with the following search string: "((review[Title] OR meta-analysis[Title]) AND (incentive*[Title] OR reward*[Title] OR penalt*[Title] OR bonus[Title] OR benefit[Title]) AND (walk*[Title] OR exercise*[Title] OR physical activity[Title] OR activity[Title] OR lifestyle[Title]))". Title and abstract of all resulting articles were screened and checked against the inclusion criteria to exclude irrelevant articles. Subsequently, full texts of the remaining potentially relevant reviews were screened to make a final inclusion or exclusion decision. Reference lists of included reviews were screened, and a hand search on Google Scholar was performed to identify other relevant reviews that might have been missed during the search process. The search was performed by a single reviewer.

Data Collection and Analysis

For this overview, the same data were extracted and analyzed as in the overview of mobile physical activity interventions (section 3.3). Likewise, the GRADE criteria for rating the quality of evidence of included reviews remained unchanged, with the exception of rating the indirectness of an included summary effect. Specifically, the quality of evidence of a summary effect was rated down for indirectness if the majority of primary studies were conducted in patient populations and with financial incentives substantially greater than those of the insurer's physical activity program (cf. section 3.2). Data extraction was performed by a single reviewer.

3.4.3 Results

Search Process

Searching the PubMed database revealed 26 reviews of which six were considered potentially relevant. The full texts of these six reviews were inspected in detail to make a final inclusion or exclusion decision. Two reviews (Gong, Trentadue, Shrestha, Losina, & Collins, 2018; Mitchell et al., 2019) provided a summary effect for incentives measured in steps per day and were consequently included in the overview (Figure 3-5). Two reviews focused on exercise session attendance and were excluded. Two further reviews included various measures related to physical activity but did not perform meta-analysis.

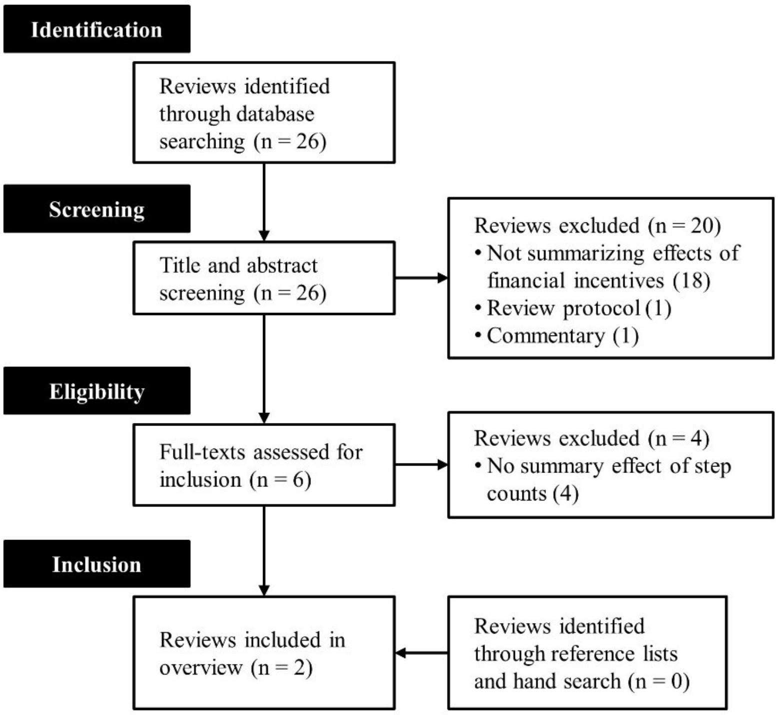


Figure 3-5. Selection process of reviews of financial incentives.

Further screening of reference lists and a hand search conducted in Google Scholar did not lead to the inclusion of additional reviews.

Characteristics of Included Reviews

Presumably reflecting the growing adoption of and interest in mobile technology over the past years, the two existing reviews that reported summary effects of financial incentives on physical activity measured in steps per day were published very recently

(Table 3-4). The two included reviews were of acceptable methodological quality (AMSTAR rating of 7 and 8 respectively, detailed rating available in Appendix E). However, neither of the two reviews searched unpublished literature, provided a list of studies that were excluded, and did consider limitations of primary studies appropriately when formulating conclusions and recommendations.

The review by Gong et al. (2018) primarily focused on incentive effects on weight loss in patients with chronic diseases, but summarized incentive effects on steps per day that were reported in two studies. In contrast, Mitchell et al. (2019) identified 12 RCTs of financial incentives in adults that reported 23 relevant comparisons. The two RCTs identified in Gong et al. (2018) were also included in the meta-analysis by Mitchell et al. (2019), which explains the similarities of the reviews with regard to average incentive value, intervention length and sample size of included primary studies. Therefore, the remainder of this result section will focus mostly on the review by Mitchell et al. (2019).

Table 3-4. Characteristics of reviews of financial incentives.

ation	Incentives	Ø value	NOC	Ø Length	Ø *N*	AMSTAR
patients	Certain outcome and progress incentives (cash) with high immediacy	US$ 69.90 / month	2	15.9 (4.0)	116	7
ients,	Mostly outcome incentives (vouchers, cash, and donations) with high immediacy and varying certainty (1%-100%)	US$ 70.50 / month	23	20.0 (5.7)	219	8

NOC: number of comparisons that contribute to the summary effect. Average length is given in weeks with ckets. At-risk populations are populations with known low activity levels such as elderly or overweight

The included primary studies were moderately sized (average *N*: 219, range: 45 – 800) and primarily evaluated short-term effects (intervention length: 12 – 24 weeks) of immediate incentives tied to daily or weekly step goals. Some studies additionally evaluated effects after incentive removal with follow-up periods between three and six months. The average incentive value in the included reviews (ca. US$ 70 per month, range: US$ 30 – US$ 167) was seven times as large as the incentive value of the insurer's physical activity program

(CHF 10 per month, roughly equal to US$ 10.3). To measure steps per day, both incentive and control groups received either a commercially available wearable device, a pedometer or a smartphone.

Summary Effects of Incentives

Gong et al. (2018) reported a statistically significant summary effect of incentives on physical activity of 940 steps, 95%CI [306, 1574]. The authors did not report heterogeneity of this effect, which was calculated from the effects of the two included studies and found to be nonexistent ($I^2 = 0\%$, $p = .64$). Similarly, Mitchell et al. (2019) report a statistically significant summary effect of 607 steps, 95%CI [422, 792], with substantial and statistically significant heterogeneity ($I^2 = 81\%$, $p < .001$). The smallest incentive amount that led to a significant increase in steps per day was around US$ 30 per person per month, or US$ 1 per person per day, in several included studies (M. A. Adams et al., 2017; Patel et al., 2016; Patel et al., 2018), while the average incentive value in both reviews was around US$ 70 (Table 3-4). Pooling data from nine studies reporting 19 relevant comparisons, the authors found that this effect decreased marginally and remained statistically significant at three to six months after the incentives were withdrawn (514 steps, 95%CI [313, 715], $I^2 = 85\%$, $p < .001$). In further subgroup analyses, Mitchell et al. (2019) found that incentive effects were larger in studies with greater incentive values and low baseline physical activity levels.

None of the studies included in the review by Mitchell et al. (2019) reported the effects of incentives on participation in the intervention or the reach of the intervention. However, one study noted that incentives increased the proportion of participants wearing the activity tracker (Finkelstein et al., 2016). Drop-out rates between incentive and control conditions were comparable in all studies. Applying the GRADE criteria, the quality of evidence in both reviews was rated as low (Appendix F). Reasons for rating down the quality of evidence include the small overall sample size on which the summary effect is based (Gong et al., 2018), as well as evidence for publication bias and great inconsistency of effects of primary studies (Mitchell et al., 2019).

3.4.4 Discussion

Principal Findings

So far, two systematic reviews have reported summary effects of financial incentives on objectively measured physical activity. These reviews have produced empirical evidence of low quality. The existing evidence suggests that immediate financial incentives tied to physical activity with an average value of US$ 70 per month can lead to a small increase in physical activity of 600 steps per day on average. In several studies, statistically significant effects were observed for incentives as small as US$ 30 per month. If the incentives are removed, effects can be sustained for up to three to six months after incentive removal. The included studies do not allow for any conclusions regarding the effects of incentives on participation.

The reported effects of incentives on physical activity are partly in line with those of other systematic reviews. Barte and Wendel-Vos (2017) reviewed randomized studies that tested the effect of financial incentives on a variety of physical activity behaviors. They found that incentives tied to exercise session attendance or objectively measured physical activity increased the respective target behavior. Likewise, the systematic reviews by Strohacker et al. (2014) and Mitchell et al. (2013) reported positive effects of incentives on exercise session attendance. However, the empirical evidence regarding the sustainability of incentive effects is less clear. Of the four studies with follow-up periods after incentive removal in the review by Strohacker et al. (2014), two reported sustained effects, whereas the other two reported a regression of behavior to baseline levels. In a systematic review of the effects of incentives on various health behaviors, incentive effects were not sustained beyond three months after removal for smoking cessation, healthy eating and physical activity (Mantzari et al., 2015). A potential explanation for sustained effects in the review by Mitchell et al. (2019) may be the use of mobile technologies in primary studies. The continuous self-monitoring of physical activity levels by study participants could have facilitated effect maintenance after incentive removal or, at least, delayed the decay of incentive effects.

However, incentivizing health behaviors with financial incentives, especially mere cash incentives, has also been criticized for various reasons inside and outside of academia (Honegger, 2018; Promberger & Marteau, 2013; Schmidt, Voigt, & Wikler, 2010). One prevailing argument against financial incentives is a potential undermining effect of incentives on intrinsic motivation, i.e. engagement in a behavior for its own sake rather than for any other reward or outcome. Intrinsic motivation has been identified as an important predictor of behavior change maintenance (Teixeira, Carraça, Markland, Silva, & Ryan, 2012). The argument is based on a vast amount of psychological research that demonstrates that adding a reward to a previously unrewarded and intrinsically motivated activity reduces intrinsic motivation operationalized as the time spent performing the activity after the reward has been removed (Edward L Deci, Koestner,

& Ryan, 1999). However, Promberger and Marteau (2013) have reviewed the scientific literature and found no evidence of an undermining effect in the context of health behaviors. Specifically, they point out that health behaviors are seldomly intrinsically motivated and therefore systematically differ from behaviors where an undermining effect has been found. The results of the meta-analysis by Mitchell et al. (2019) also do not support the presence of an undermining effect for physical activity, because the undermining effect predicts that physical activity levels of previously incentivized participants should be lower than those of control participants at post-incentive follow-up. This is evidently not the case. Yet, no study has so far investigated the effects of incentives on direct measures of motivation and while the incentive is in place. Other points of criticism relate to the ethics of incentives, especially when offered by health insurance companies. Schmidt et al. (2010) argue, for example, that incentivizing health behaviors can lead to cost shifting, i.e. a general raise in insurance premiums, so that individuals who obtain incentives merely pay the pre-incentive premium. Further, they point out that financial incentives can be unfair if not all people have the same chance of reaching the target behavior, as well as the fact that incentives do not differ between those who try but fail and those who do not try. Others fear an undermining of the solidarity principle (Honegger, 2018), i.e. the imperative of equal insurance premiums for healthy and sick individuals, that characterizes healthcare systems in Switzerland and many other countries. Some of these points may be mitigated by the design of the incentive, e.g. by using other incentive forms rather than mere cash payments or by incentivizing relative behavior change instead of absolute levels of behavior. Nevertheless, the arguments against financial incentives may need to be considered carefully when implementing incentives as part of a health behavior change intervention to avoid resistance among members of the intervention's target group.

Implications

Mobile physical activity promotion programs can increase their overall effect by 600 steps per day on average when offering goal-contingent financial incentives with high immediacy and as long as incentives are in place. At minimum, the incentive value should equal US$30 per month but greater values tend to lead to larger effects. Although incentive effects may be sustained for shorter time periods, long-term maintenance is unclear. Despite these benefits, financial incentives often face substantial criticism and need to be designed and implemented with care.

Limitations

Some limitations of this overview of reviews have to be noted. First, the two reviews in this overview produced empirical evidence of low quality. Specifically, the review by Mitchell et al. (2019) reports great variation in effects of primary studies, substantial heterogeneity and potential publication bias. Thus, the confidence in the found effect is limited and it is possible that the effects of incentives in the insurer's physical activity program are substantially different. Further, almost all primary studies included in the two reviews used pedometers or smartphones to measure physical activity. These devices do not allow for differentiating increases in physical activity from increases in wear time. Thus, it is possible that some of the physical activity differences between incentive and control groups can be explained simply by increases in device wear time. However, one larger study that used accelerometers for measuring physical activity found no differences between effect estimates that were adjusted and estimates that were unadjusted for wear time (Finkelstein et al., 2016). Finally, the high value of the investigated incentives calls into question the external validity of the results of the reviews. As illustrated by the insurer's physical activity program (section 3.2), real-world programs operate with incentive values far below the average incentive values investigated in empirical studies. In fact, the average incentive values in both reviews exceed the incentive value in the insurer's physical activity program by a factor of seven, and even the minimum incentive value included in the primary studies is three times as large as the incentives in the insurer's program. Given the evidence for a dose-response relationship between incentive value and physical activity, this calls into question whether the effects found in empirical studies translate to the insurer's physical activity program. Similar to the overview of reviews of mobile physical activity interventions, the systematic literature search for this overview was restricted to one database and the selection and coding of reviews was performed by a single reviewer. Again, this may limit the representativeness of reviews and the objectivity of results.

Future research could focus on alternative incentive designs that rely less on mere cash payments, as well as investigating incentive characteristics that have received less attention in previous studies, such as incentive plans with smaller magnitude and varying frequency. Future studies also need to overcome the limitations of the existing studies by employing longer incentive and follow-up periods, and by using measurement devices, such as accelerometers, that allow distinguishing increases in physical activity from increases in wear time. Cost-effectiveness analyses that complement the results of RCTs could further facilitate the adoption of incentives in real-world settings.

Conclusion

Enabled by mobile technologies, financial incentives for physical activity can be tied to behavioral goals and paid out frequently and immediately. Evidence suggests that incentives succesfully increase physical activity by 600 steps per day on average and that these effects may be sustained after incentive removal – at least for some time. However, incentive values in real-world settings are often much smaller than those investigated so far. It is therefore unclear whether the found effects can be applied to the incentives used in the insurer's physical activity program.

3.5 Summary

In this chapter, the insurer's mobile physical activity program has been thoroughly described and analyzed to evaluate its potential for a public health impact. In a first step, the overall target effect size of the program was determined. Models of potential impact fractions on T2DM and CHD revealed that the insurer's program should lead to an average increase in physical activity of 1,500 steps and should maximize uptake among its target population to create a substantial public health impact. Next, two overviews of systematic reviews that focused on the program's key components – commercially available activity trackers and financial incentives – shed light on the program's potential to reach the abovementioned targets. For activity trackers, high-quality evidence suggests that this component can increase participants' physical activity by 600 steps per day. For financial incentives, the empirical evidence suggests an effect of a similar size for considerably larger incentives. Thus, in principle, the combination of mobile technology and financial incentives could add to an overall effect of 1,200 steps (assuming no interaction between the two components). Effects of similar size were found in a randomized trial evaluating the combined effects of Fitbit activity trackers and financial incentives (1050 steps, 95%CI [600, 1490]; Finkelstein et al., 2016) and in a subgroup analysis in the review by Mitchell et al. (2019) that specifically investigated financial incentives in combination with wearable devices (1242 steps, 95%CI [745, 1739]). However, the evidence on financial incentives is of low quality and the incentive values investigated in the reviews exceeded the incentive value of the insurer's program by a factor of seven on average. Given a possible dose-response relationship between incentive value and physical activity, it is unclear whether the small financial incentives of the insurer's program can produce similar effects. Further, neither overview of reviews allowed any conclusions regarding the reach of the reviewed interventions. Collectively, the second research question can therefore be answered as follows:

RQ 2: To what degree can the insurer's physical activity program meet the defined target effects?

The degree to which the insurer's physical activity program can meet the defined target effects on reach and efficacy are unclear. Too few studies have reported intervention reach. Randomized studies on mobile technologies and financial incentives suggest that the insurer's program may increase physical activity levels by up to 1,200 steps but the investigated incentive values were several times larger than the incentives in the insurer's program.

Given that neither the reach nor the efficacy of the insurer's physical activity program can be reliably approximated based on data reported in previous studies, the potential public health impact of the program cannot be evaluated. Therefore, the insurer's program was investigated in an empirical study. The goal of this study was to explore the program's reach and evaluate the effects of the financial incentive component of the program. In addition to financial incentives, a different incentive design was evaluated that could avoid some of the ethical problems associated with financial incentives. In line with the MOST framework, the results of this study help to decide whether the public health impact of the program can be expected to suffice or whether a revision of the program and its conceptual model is advisable. The following chapter describes the conduction and the results of the study and discusses the study results and their implications.

Chapter 4

Study I: Small Incentives to Promote Physical Activity[6]

Based on the conclusions of chapter 3, this chapter presents the first empirical study of this dissertation, which explores the reach of the insurer's physical activity program and evaluates the effects of incentives on the program's reach and efficacy. The empirical study was conducted as part of a pilot test of the insurer's program. For this pilot test, a smaller group of insurees was invited to try out the program for a six-month period.

4.1 Introduction

The overview of reviews on financial incentives (section 3.4) revealed that previous studies report the average effect of incentives on physical activity to be an increase of 600 steps per day, but these reviews typically use incentive values several times larger than those used in the insurer's physical activity program. It is unclear whether similar effects can be obtained with incentives as small as those in the insurer's physical activity program, i.e. CHF 10 per month. Further, no study has yet investigated how incentives can affect the reach of an intervention, making it difficult to judge the public health impact of the insurer's physical activity program. The absence of studies investigating these factors instigated the empirical study described in this chapter.

The overview of reviews of financial incentives also pointed to possible ethical problems, especially when financial incentives are provided by health insurers, i.e. cost-shifting and undermining of the solidarity principle. Alternative incentive designs might

[6] Parts of this chapter, relating in particular descriptions of methods, results and discussion of the reported study, are published in the context of the following academic publications: J.-N. Kramer and Kowatsch (2017), J.-N. Kramer, Tinschert, Scholz, Fleisch, and Kowatsch (2019).

avoid some of these problems. For example, incentives in the form of donations to charity (charity incentives) are a promising incentive design that has also been investigated by some of the studies included in the review by Mitchell et al. (2019). Because charity incentives are paid out to charity organizations and not to the participant, they eliminate the problem of cost-shifting and do not undermine the solidarity principle. What is more, donating to charity activates an additional neural reward system in participants that is not activated when they merely receive cash incentives (Moll et al., 2006), and this philanthropy has been related to happiness in observational and experimental studies (Dunn, Aknin, & Norton, 2008). This affective component, sometimes described as the "warm glow of giving" (Andreoni, 1990), has been successfully used in marketing strategies to motivate purchase behavior (Varadarajan & Menon, 1988). Thus, charity incentives could potentially lead to greater changes in behavior than mere cash incentives.

However, previous studies on the effects of charity incentives have found mixed results. In a large RCT in Singapore, weekly cash incentives ($30 for walking 70,000 steps per week) increased full-time workers' daily steps during the study's six-months incentive period, but similar charity incentives did not have the same positive effect (Finkelstein et al., 2016). In a smaller RCT with older adults in Philadelphia, both weekly cash and charity incentives ($20 for meeting a personalized step goal on at least five out of seven days) were successful in promoting step-goal achievement over the 16-week study period – but not during the 4-week follow-up period (Harkins, Kullgren, Bellamy, Karlawish, & Glanz, 2017). In addition to mere cash incentives, the study reported in this chapter therefore examined the effects of charity incentives of comparable value on the reach and efficacy of the insurer's mobile physical activity promotion program. Further, the study allowed for the gathering of first feedback on how the program is perceived and accepted by its target group.

4.2 Methods

4.2.1 Recruitment

In June 2015, eligible insurees were randomly assigned to a type of incentive (cash incentive, charity incentive, or no incentive), and invited via email to participate in the mobile physical activity program. Invitation emails contained information about the promotion program, the incentive condition of the email recipient and a link to the insurance's online platform where insurees could log in and register for the program (Appendix G). Additionally, the invitation email informed insurees about the opportunity to buy a compatible activity tracker at a reduced price if they did not own one already. Insurees who were not interested in the program were asked to complete a brief survey via a separate link at the bottom of the invitation email that asked them to indicate their agreement with potential predefined reasons not to participate in the program.

On the insurance provider's online platform, insurees received detailed information about the program, data protection policies and eligibility criteria and could provide consent to participate. To provide consent, participants had to confirm that they do not have a medical condition that prohibits increased levels of physical activity by ticking a checkbox. Insurees were advised to consult a physician if they were in doubt. Insurees did not receive any information about the existence of different incentive groups. To complete registration, insurees linked their wearable manufacturer's customer account to the insurance provider's online platform so that their steps would be synchronized daily via an application programming interface. To facilitate automatic synchronization of step counts, compatible activity trackers were limited to devices developed by the major wearable manufacturers on the Swiss market Garmin (Olathe, KS), Jawbone (formerly San Francisco, CA) and Fitbit (San Francisco, CA). Alternatively, insurees could use the Fitbit smartphone application to track daily steps. Once registered, participants were able to add family members who also met the eligibility criteria to the pilot program.

At the beginning and at the end of the program, participants were asked to complete a web-based survey to collect data on demographic variables and covariates of physical activity. Participants received 10 Swiss Francs (CHF) for the completion of both surveys. The insurance company provided data on age, gender, nationality and federal state (canton) of all invited insurees as well as step data of participants. The institutional review board of the University of St. Gallen approved the study. Analyses of primary and secondary outcomes were completed in 2018.

4.2.2 Sample

Due to legal regulations in Switzerland, the mobile physical activity promotion program was offered to insurees with complementary insurance plans only. To facilitate enrolment, only insurees that met the predefined eligibility criteria were invited to participate: insurees had to be at least 18 years old, German-speaking, enrolled in a complementary insurance plan and registered on the insurer’s online platform. There was no racial or gender bias in the selection of invited insurees. Naturally, invited insurees resided primarily in the German-speaking parts of Switzerland. No eligibility criteria were defined on the canton level.

4.2.3 Incentive Schedules

In the cash incentive condition, participants received CHF 10 for each month they walked more than 10,000 steps per day on average and CHF 5 for each month they walked less than 10,000 but over 7,500 steps per day (cf. section 3.2.2). Charity financial incentives coincided with cash incentives, with the exception that participants could donate a chosen proportion of their earned reward to one of three charity organizations (a foundation supporting the rights and needs of Swiss children and adolescents, a foundation for health promotion of Swiss adolescents and an organization committed to preserving the Swiss hiking track network). Participants in the control group were informed that participation can enhance health and well-being but did not receive any incentives during the first three months of the program. However, from the fourth month onwards until the end of the program, the control group was entitled to personal financial incentives of CHF 20 for walking 10,000 steps on average and CHF 10 for walking 7,500 steps on average per month. The option to choose a proportion of the earned reward to be donated in the charity group and the addition of incentives in the control group gave all participants the opportunity to earn a maximum of CHF 60 during the six-month pilot test. This design aspect was of great importance to the health insurer.

4.2.4 Randomization

An important study design consideration was minimizing the risk of spillover effects between study groups, especially between the incentivized groups and the control group, to prevent frustration and drop-out among insurees. Insurees were therefore randomized using a blocked cluster-randomization based on their’ canton of residence ($N = 20$) with a block size of five and an allocation ratio of 2:2:1 with fewer insurees allocated to the control condition. Each block consisted of two pairs of neighboring cantons and one single canton. An additional consideration in the study’s randomization scheme was to account for differences in activity preferences between urban and rural areas in

Switzerland (Lamprecht, Fischer, & Stamm, 2014). The blocks were therefore matched for population density. Next, canton pairs within each block were randomized to the incentive conditions using the toss of a coin and the remaining canton was allocated to the control group.

4.2.5 Outcomes

Primary outcome was the participation rate in the three different groups. Insurees were considered as participants if they provided consent to participate and shared their step counts at least once during the first three months of the study. Daily step counts and the proportion of participant days with more than 10,000 steps during the first three months of the program were analyzed as secondary outcomes. Commercially available activity trackers have been shown to accurately measure step counts under free living conditions in healthy adults (Feehan et al., 2018). Participants' non-usage attrition (Eysenbach, 2005) was used as a measure of engagement with the program and explored as an additional outcome. A participant was coded as "non-usage attrition observed", when she or he stopped synchronizing step counts with the insurer. Additionally, participants' perceptions and acceptance of the program including improvement suggestions were assessed in the follow-up survey.

4.2.6 Statistical Analysis

Gao and colleagues' approach for non-aggregate cluster-randomized controlled trials with binary outcomes (Gao, Earnest, Matchar, Campbell, & Machin, 2015) was used to determine the minimum number of insurees to invite to the pilot program. Accordingly, a sample size of $N = 15{,}822$ invited insurees is necessary to detect a 5% difference in participation rates between control and incentive groups assuming an α-level of .05, a power of .80, an intra-cluster correlation of $\rho = .01$ (Donner & Klar, 2000) and a mean cluster size of 879 (SD = 1,326; based on data from the health insurer).

Linear mixed models and generalized linear mixed models (Raudenbush & Bryk, 2002) were fit to the data to analyze group differences of primary and secondary outcomes. The model of participation rate included a fixed effect for incentive condition and a random intercept for canton. Models of step counts and participant days with more than 10,000 steps included fixed effects for time, self-reported physical activity measures at baseline, known covariates of physical activity (Bauman et al., 2012), incentive condition, the incentive condition by time interaction and a random intercept for participants. The time variable was mean-centered before entering the model. In addition, all models of secondary outcomes were adjusted for group differences at baseline. To be able to adjust models for covariates and group differences at baseline, only participants that completed the baseline survey were included in the analyses of secondary outcomes. Several sensitivity analyses were performed to assess the robustness of the results. Differences in participation rates between groups were further adjusted using fixed effects for age, gender and nationality of participants and cantonal population density. For the secondary outcomes, steps and participant days with more than 10,000 steps, a nested random effect for canton was added to the model to account for potential clustering on the canton level. Additionally, sensitivity analyses with regard to missing data were performed for all outcomes (Appendix H). Because the incentive structure changed for the control group in the fourth month, all outcomes and analyses refer to the first three months of the pilot test. Cox proportional hazard regression models were fit to the data to analyze participants' non-usage attrition. This analysis was exploratory and included effects for age, gender, self-reported activity at baseline, purchasing an activity tracker to participate, and incentive condition.

The reported results on primary and secondary outcomes are pooled over separate analyses of 20 imputed datasets created using 50 imputation iterations for each dataset. Data was imputed in wide format using all variables in the statistical models plus additional auxiliary variables to set up the imputation models. Subsequently, convergence of the imputation algorithm was examined by visual inspection of plots of mean and standard deviation of imputations across all iterations. Based on data from a Swiss representative sample (Sequeira, Rickenbach, Wietlisbach, Tullen, & Schutz, 1995), daily step counts exceeding 25,000 or below 1,000 steps were set to missing before starting the imputation. Data were analyzed in R (version 3.4.2, R Core Team, Vienna, Austria) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2014) for fitting (generalized) linear mixed models and the mice package (van Buuren & Groothuis-Oudshoorn, 2011) for multiple imputation.

4.2.7 Protocol deviations

Due to an error in the randomization process, six out of all 26 Swiss cantons were non-randomly allocated to the three study groups. These cantons contained 831 (3.09%) of invited insurees and were excluded from all analyses.

4.3 Results

4.3.1 Recruitment and Sample

In total, $N = 26{,}091$ insurees (mean age 45.48 years, SD = 14.97 years, 38.52% female) met the predefined eligibility criteria and were randomized. Of those, 1,338 (5.13%) participated in the program.

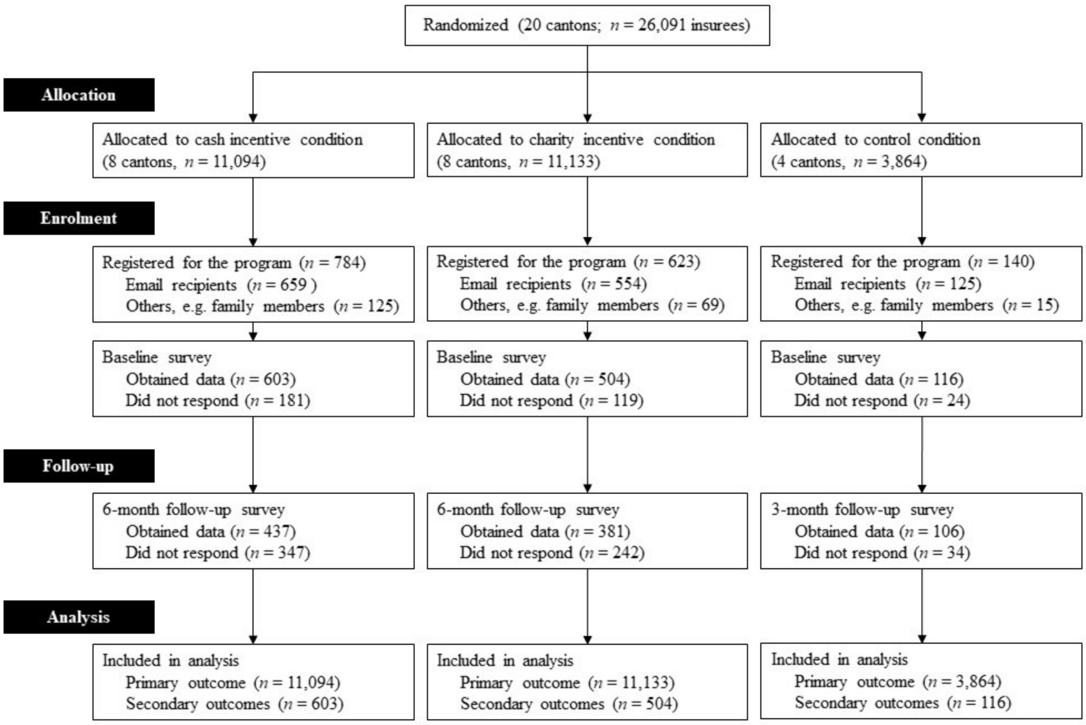


Figure 4-1. Flow of participants through the study.

Additionally, 209 family members of invited insurees participated, bringing the total number of participants to n = 1,547. Loss to follow-up was 20.9% (324/1,547) for the baseline survey and 40.3% (623/1,547) for the follow-up survey. The proportion of women among participants was higher than among all invited insurees (47.83% vs. 38.52%), indicating that women were more likely to participate. Participants were on average 42.65 (SD = 13.03) years old and mostly Swiss (90.20%).

Table 4-1. Characteristics of invited insurees (N = 26,091).

	CSI (*n* = 11,094)	**CHI (*n* = 11,133)**	**CG (*n* = 3,864)**	**p-value[a]**	**CSI vs CG[b]**	**CHI vs CG[b]**	**CSI vs CHI[b]**
Canton level							
Number of cantons	8	8	4				
Cluster size	1,353.38 (2,080.96)	1,362.75 (1,439.82)	934.00 (1,263.82)	0.907	0.25	0.32	0.01
Population density[c] (residents/km^2)	619.32 (244.93)	505.92 (1,721.03)	270.15 (826.41)	0.732	0.65	0.19	0.11
Individual level							
Age	44.81 (14.73)	45.87 (15.12)	46.30 (15.15)	**<0.001**	0.10	0.03	0.07
Female	4441 (40.03)	4261 (38.27)	1348 (34.89)	**<0.001**	0.11	0.07	0.04
Nationality				**<0.001**	0.16	0.07	0.13
Swiss	9,491 (85.55)	9,973 (89.58)	3,486 (90.22)				
German	582 (5.25)	384 (3.45)	111 (2.87)				
Other	654 (5.90)	476 (4.28)	191 (4.94)				
NA	367 (3.31)	300 (2.69)	76 (1.97)				

Note. Boldface indicates statistical significance ($p < 0.05$). SMD, absolute standardized difference; CSI, cash incentives; CHI, charity incentives; CG, control group; IQR, interquartile range; table displays mean (standard deviation) for continuous and *n* (percentage) for categorical variables unless stated otherwise.
[a] Based on one-way ANOVA for normal, Kruskal-Wallis test for non-normal, and χ^2-test of independence for categorical variables.
[b] Absolute standardized difference (SMD). Values greater than 0.20 are defined as small effect size. Non-normal variables were log-transformed before calculating SMD.
[c] Reported values are weighted by cluster size.

Based on baseline survey data (*n* = 1223), 43.58% of participants had a university degree and only 3.84% reported poor or fair health conditions. There were statistically significant small group differences with regard to participants' residential environments, self-reported health status, and minutes sitting per week (Table 4-2), with the greatest differences observed between the cash incentive and the control group. Compared to the Swiss population, program participants were more likely to hold a university degree and earned higher wages (Bundesamt für Statistik, 2016, 2018b). In the charity incentive condition, the mean proportion of donated rewards was 20.29% (SD = 26.81). However, 310 of 623 participants (49.76%) chose not to donate any proportion of their reward to charity.

Table 4-2. Characteristics of program participants (N = 1,547).

	CSI (*n* = 784)	CHI (*n* = 623)	CG (*n* = 140)	p-value[a]	CSI vs CG[b]	CHI vs CG[b]	CSI vs CHI[b]
Age	42.71 (12.99)	42.14 (12.95)	44.63 (12.53)	0.126	0.15	0.18	0.04
Female	382 (48.72)	299 (47.99)	59 (42.14)	0.424	0.13	0.12	0.01
Educational attainment				0.473	0.25	0.19	0.09
Secondary school	12 (1.99)	8 (1.59)	2 (1.72)				
Vocational school	181 (30.02)	163 (32.34)	35 (30.17)				
High school	92 (15.26)	85 (16.87)	27 (23.28)				
University	278 (46.10)	213 (42.26)	42 (36.21)				
NA	40 (6.63)	35 (6.94)	10 (8.62)				
Residential environment				**<0.001**	0.29	0.15	0.32
Town	85 (14.10)	42 (8.33)	11 (9.48)				
Outskirts of town	173 (28.69)	102 (20.24)	24 (20.69)				
Village	253 (41.96)	260 (51.59)	64 (55.17)				
Countryside	73 (12.11)	88 (17.46)	14 (12.07)				
NA	19 (3.15)	12 (2.38)	3 (2.59)				
Monthly net income				0.343	0.26	0.24	0.17
< CHF 2500	27 (4.48)	30 (5.95)	3 (2.59)				
CHF 2501 –5000	83 (13.76)	80 (15.87)	18 (15.52)				
CHF 5001 –7500	190 (31.51)	147 (29.17)	33 (28.45)				
CHF 7501 –10000	101 (16.75)	78 (15.48)	26 (22.41)				
> CHF 10000	73 (12.11)	46 (9.13)	8 (6.90)				
No answer	110 (18.24)	111 (22.02)	25 (21.55)				
NA	19 (3.15)	12 (2.38)	3 (2.59)				
Health status				**0.032**	0.23	0.21	0.18
Poor	2 (0.33)	1 (0.20)	0 (0.00)				
Fair	15 (2.49)	20 (3.97)	9 (7.76)				
Good	230 (38.14)	225 (44.64)	43 (37.07)				
Very good	266 (44.11)	203 (40.28)	48 (41.38)				
Excellent	68 (11.28)	40 (7.94)	12 (10.34)				
NA	22 (3.65)	15 (2.98)	4 (3.45)				
Activity tracker brand				0.520	0.13	0.17	0.04
Fitbit	511 (84.74)	432 (85.71)	94 (81.03)				
Garmin	67 (11.11)	55 (10.91)	14 (12.07)				
Jawbone	25 (4.15)	17 (3.37)	8 (6.90)				
Bought an activity tracker to participate	320 (53.07)	303 (60.12)	62 (53.45)	**0.048**	0.01	0.15	0.14
Sitting (min./week)[d]	2,435.80 (1,378.83)	2,263.40 (1,303.22)	2,165.52 (1,247.31)	**0.039**	0.21	0.08	0.13
Moderate activities and walking (MET-min./week), median (IQR)[d]	2,628.00 (3,306)	2,745.75 (3,327)	2,079.00 (3,720)	0.243	0.09	0.04	0.05

Note. Boldface indicates statistical significance ($p < 0.05$). CSI, cash incentives; CHI, charity incentives; CG, control group; IQR, interquartile range; table displays mean (standard deviation) for continuous and *n* (percentage) for categorical variables unless stated otherwise.

[a] Based on one-way ANOVA for normal, Kruskal-Wallis test for non-normal, and χ^2-test of independence for categorical variables.

[b] Absolute standardized difference (SMD). Values greater than 0.20 are defined as small effect size. Non-normal variables were log-transformed before calculating SMD.

[c] Reported values are weighted by cluster size.

[d] Assessed using the International Physical Activity Questionnaire (C. L. Craig et al., 2003).

4.3.2 Participation

Among invited insurees, 5.94% participated in the cash incentive group and 4.98% participated in the charity incentive group, compared to 3.23% in the control group. Differences between incentive groups and control group were statistically significant (cash incentives: odds ratio [OR] = 1.96, 95%CI [1.55, 2.49], p < .001; charity incentives: OR = 1.59; 95%CI [1.25, 2.01], p < .001). Contrast analysis revealed that participation rates also differed significantly between insurees in the cash incentive and the charity incentive group (OR = 1.24, 95% CI: [1.06, 1.44], p = .006).

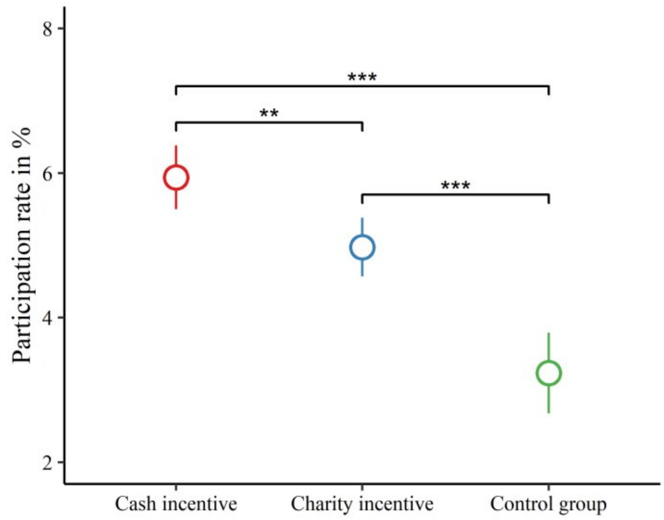


*Figure 4-2. Participation rate by incentive group. Error bars are 95% confidence intervals. ***: p < .001, **: p < .01.*

In total, N = 972 invited insurees indicated which of the potential reasons best reflected their decision not to participate in the program (Table 4-3). The requirement of spending money on an activity tracker emerged as the most important reason for non-participation, selected by 41.05% of survey respondents. Several other reasons, such as incompatibility of existing devices (17.90%), lack of interest (15.47%) and the small size of the rewards (14.99%), were selected by some respondents, whereas the requirement to share data with the health insurer was selected only by a few respondents (9.47%).

Table 4-3. Reasons for refusing to participate (N = 972).

Response	***n* (%)**
I do not want to spend money on an activity tracker	399 (41.05)
I already have an activity tracker, but it's not compatible	174 (17.90)
I have generally no interest in tracking my daily steps	153 (15.74)
The expected reward does not correspond to my effort[a]	119 (14.99)
I believe that I am not achieving the given step goals	142 (14.61)
I see no personal benefit	134 (13.79)
The technical effort is too high	119 (12.24)
I do not want to share my step data with my health insurance	92 (9.47)
I would like to buy an activity tracker from another brand	4 (0.41)

[a] This answer option was given only to insurees in the personal and charity financial incentive group (n = 794)

4.3.3 Behavior Change

On average, participants walked 10,709 (SD = 4,555) steps per day and reached the 10,000 steps goal on 54.24% of all days during the first three months of the program. Participants' step counts showed a small but statistically significant positive correlation with age, $r(1541) = .19$, $p < .001$. To investigate group differences, Table 4-4 compares steps per day and the probability of walking more than 10,000 steps per day between the groups at the beginning of the study, in the middle of the study, and at the end of the first three months. Although participants in the charity incentive group consistently accumulated a higher number of steps than participants in both the cash incentive and control groups (Figure 4-3), these step count differences were not statistically significant at any time point. The difference between the incentive groups and the control group diminished over time, but this trend was also not statistically significant (change of the charity incentive effect over time: -3 steps/day, 95%CI [-6, 1], $p = .16$; change of the cash incentive effect over time: -3 steps/day, 95%CI [-6, 1], $p = .15$).

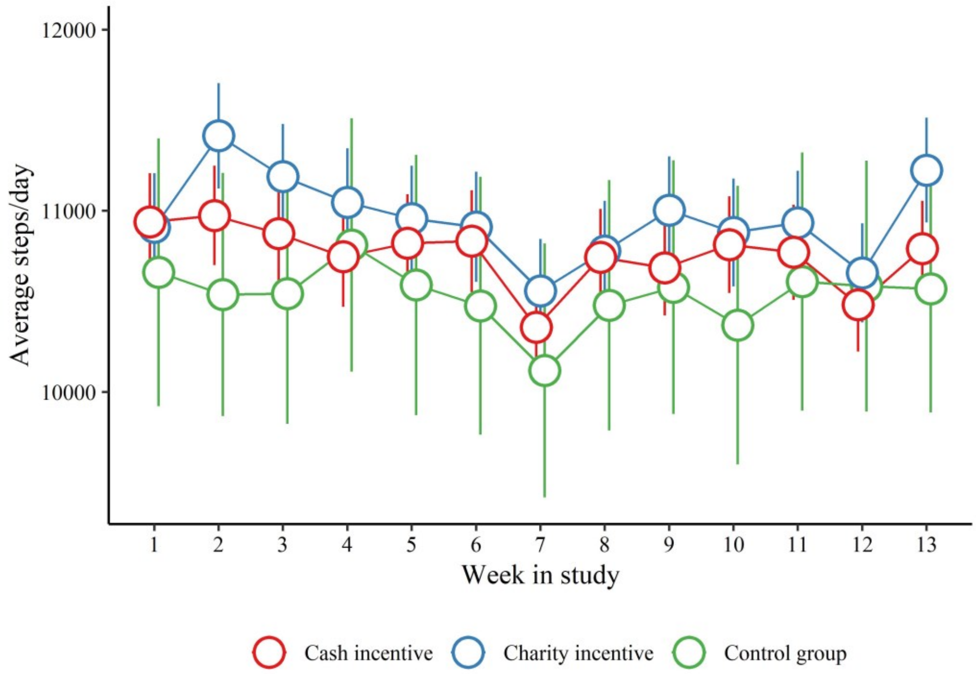


Figure 4-3. Unadjusted weekly mean number of steps by incentive condition. Error bars are 95% confidence intervals.

At the beginning of the program, participants receiving charity incentives had a 12% higher chance (OR = 1.68; 95% CI [1.23, 2.30], p = .004) of walking at least 10,000 steps per day compared to the control group. This difference diminished significantly over time (change of the charity incentive effect over time: -0.003, 95% CI [-0.01, -0.001], p = .003) and was no longer significant three months after the start of the program (Table 4-4). After adjusting p-values for multiple comparisons across time, the probability of walking at least 10,000 steps per day did not differ significantly between the cash incentive group and the control group at any time point. Likewise, the probability of walking at least 10,000 steps per day did not differ significantly between the cash and the charity incentive groups. Interestingly, the adjusted models reveal that participants who bought an activity tracker in order to participate walked 1004 steps more, 95%CI [708, 1301], p < .001, and had a significantly higher chance of walking at least 10,000 steps per day, OR = 1.75, 95%CI [1.48, 2.10], p < .001, compared to participants who already owned a tracker prior to the invitation to participate in the program.

Table 4-4. Group comparisons for behavior change outcomes.

	Day 1	Day 46	Day 92
Difference in steps/day (95% CI, *p*-value[a])			
CSI vs CG	206 [-322, 735]; $p = 1.00$	92 [-416, 600]; $p = 1.00$	-23 [-557, 512]; $p = 0.934$
CHI vs CG	507 [-29, 1044]; $p = 0.192$	384 [-128, 896]; $p = 0.284$	260 [-281, 802]; $p = 0.346$
CSI vs CHI	-301 [-620, 19]; $p = 0.13$	-292 [-597, 14]; $p = 0.184$	-283 [-604, 38] $p = 0.084$
Odds ratio walking > 10,000 steps/day (95% CI, *p*-value[a])			
CSI vs CG	1.41 [1.03, 1.92]; $p = 0.093$	1.20 [0.89, 1.61]; $p = 0.477$	1.02 [0.74, 1.39]; $p = 0.919$
CHI vs CG	**1.68 [1.23, 2.30]; *p* = 0.004**	**1.45 [1.07, 1.96]; *p* = 0.031**	1.25 [0.91, 1.71]; $p = 0.168$
CSI vs CHI	0.84 [0.69, 1.01]; $p = 0.063$	0.83 [0.69, 0.99]; $p = 0.069$	0.81 [0.68, 0.98]; $p = 0.092$

Note: Boldface indicates statistical significance (p<0.05). Table depicts point estimates with 95% confidence intervals in brackets. CSI, cash incentives; CHI, charity incentives; CG, control group.
[a] *P*-values are adjusted for multiple testing across time points using the Holm-Bonferroni method. 95% confidence intervals are not adjusted, because Holm-Bonferroni-adjusted confidence intervals are non-informative (Holm, 1979).

4.3.4 Engagement

At six months, 25.6% of participants had stopped sharing their step data with the health insurer (Figure 4-4). During the first three months, non-usage attrition was slightly lower in the cash incentive group (9.6%) than in the charity incentive group (13.3%) as well as in the control group (12.1%). However, differences in attrition rates between the incentive groups and the control group during the first three months of the study were not statistically significant (hazard ratio [HR] = 0.75, 95%CI [0.45, 1.26], p = .28 for the cash incentive group; HR = 1.08, 95%CI: [0.65, 1.79], p = .77 for the charity incentive group). In the adjusted model, older insurees (one year HR = 0.96, 95%CI [0.94, 0.98], p < .001) and insurees who purchased an activity tracker in order to participate (HR = 0.64, 95%CI [0.44, 0.92], p = .02) were at significantly lower risk for non-usage attrition (Appendix I).

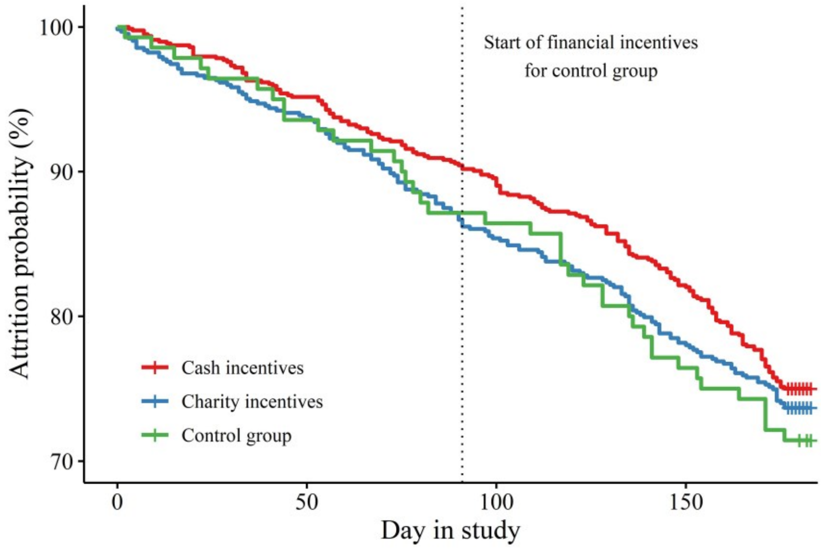


Figure 4-4. Kaplan-Meier attrition curves by incentive group.

4.3.5 Participants' perceptions

In total, *n* = 924 participants responded to the follow-up survey and answered questions about the mobile physical activity program. The majority of survey respondents enjoyed participating in the program and indicated that it was motivating and helped them to increase their daily physical activity (Figure 4-5). In addition, 86.1% of respondents were highly motivated to continue actively participating in the program. Of all survey respondents, *n* = 97 gave suggestions for improving the program. The suggestions related mostly to technical issues (29%), e.g. improving the synchronization between the activity tracker and the insurer's online platform, the integration of additional physical activities, such as bicycling and swimming (36%), and additional support functions, such as reminders or social comparisons (16%).

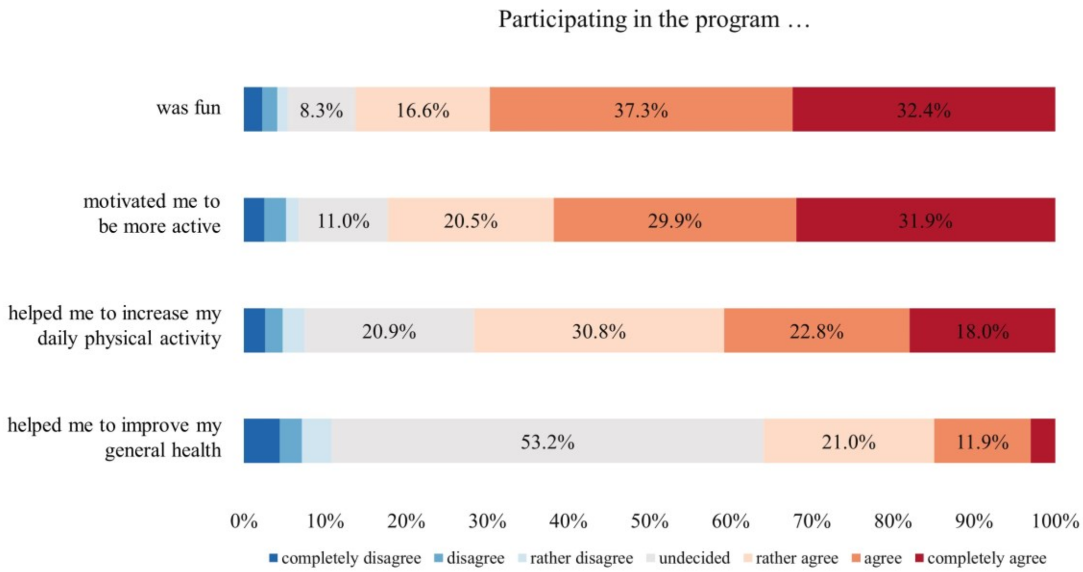


Figure 4-5. Participants' perceptions of the physical activity program (n = 924).

4.4 Discussion

4.4.1 Principal Findings

This three-arm cluster-randomized controlled trial explored the reach of the insurer's physical activity program and investigated whether small cash and charity incentives can promote participation and behavior change. Cash and charity incentives with a value of CHF 10 per person per month significantly increased participation, with cash incentives being more effective than charity incentives. With cash incentives in place, the participation rate of the program was just under six percent. This reflects a considerably higher uptake compared to an almost identical program offered by a US health insurer (Patel et al., 2017). On the one hand, this is a remarkable result, especially because invitations to participate were only sent out via email and further measures to support uptake of the program (e.g. marketing campaigns) were not implemented in order to preserve the study's experimental manipulation. On the other hand, a higher participation rate is necessary for the program to have a significant impact on public health (e.g. considerably more than 10%, cf. section 3.1.1). Although uptake of the program may increase over time, the results revealed that some barriers for participation are inherent in the program's design. Most notably, the need to purchase an activity tracker emerged as a major barrier for participation. Despite the limited participation, the program was perceived very positively among those who participated. Additionally, and similar to other studies (Glasgow et al., 1993; Patel et al., 2017), comparisons of participants' baseline characteristics with representative data from the Swiss population suggest that individuals at lower risk of disease are more likely to participate in the program, which could reduce the program's public health impact and contribute to health inequalities. The positive correlation that was observed between age and step counts of participants provides further support for such a selection effect. This correlation is somewhat counterintuitive because there is compelling empirical evidence for a negative association between age and physical activity in general (Sallis, 2000) and step counts in particular (Bassett Jr, Wyatt, Thompson, Peters, & Hill, 2010; Sequeira et al., 1995). One possible explanation for the positive correlation found in the present study is that the program might appeal especially to active elderly individuals, for whom walking is the preferred type of activity. Indeed, higher step counts for elderly participants were also reported by a similar physical activity promotion program in the US (Patel et al., 2017).

The effects of the program's small cash and charity incentives on behavior are likely to be limited. The data in this study suggest that charity incentives only increase the likelihood of walking 10,000 steps per day in the short term, with effects dissipating after three months. Cash incentives did not lead to significant changes in physical activity. These results deviate from previous studies that typically find a stable effect of incentives as long as they are in place. As indicated in the review by Mitchell et al. (2019), these previous studies used larger incentive values as well as more frequent and immediate reward schedules. Research in the field of behavioral economics suggests that people tend to place more importance on immediate rewards and discount future rewards (Volpp, Asch, Galvin, & Loewenstein, 2011), a phenomenon known as present bias (also cf. section 3.2.3). The monthly reward schedule in the present study may have caused participants to further attenuate the subjective value of the already small incentives. Both factors, the small incentive value and the monthly incentive schedule, might therefore explain the limited effects of incentives on behavior change found in the present study.

In sum, the third research question of this dissertation can be answered as follows:

> **RQ 3: Can small incentives increase participation and subsequent behavior change in a scalable physical activity intervention?**
>
> *Small incentives, especially cash incentives, can increase participation in a scalable physical activity intervention. The effects of incentives on subsequent behavior change are short-lived at best and are not sustained beyond a time period of three months.*

Participants who received charity incentives recorded slightly more steps than participants who received cash incentives, but this difference did not reach statistical significance. Thus, the data in this study do not support the hypothesis that charity incentives can lead to greater behavior change than mere cash incentives. Nevertheless, due to their inherent characteristics, charity incentives remain an attractive incentive design, especially for health insurance companies, and are worth investigating in future research.

Irrespective of incentives, attrition of the program was substantial. After six months, more than a quarter of participants had abandoned the program. This number is similar to market research data on wearable devices that suggests that 30% of new owners of wearable devices no longer user their device within six months of purchase (Ledger & McCaffrey, 2014). Although some attrition is to be expected in large-scale physical activity programs, considerable attrition can significantly attenuate the overall health impact of the program. Participants' underlying motives for taking part in the program may point to reasons for the large attrition of the program. Apart from improving health (named by 31% of survey respondents), curiosity was named as a main motivation to participate by 30% of respondents to the follow-up survey. For those participants, continuing to participate in the program may have been of little value after they learned about their daily step counts and related activity patterns. Similarly, younger participants were at higher risk of dropping out of the program presumably because quantifying activity as steps per day may appeal more to elderly participants for whom walking is likely to be the preferred type of physical activity.

A less anticipated observation was the higher physical activity levels and lower attrition rates among insurees who bought an activity tracker in order to participate in the program. It's possible that the financial investment in the activity tracker increased commitment to the physical activity program and its goals, thereby leveraging the fact that people have the desire to appear consistent with their prior commitments and behaviors (Cialdini & Goldstein, 2004). This "consistency principle" has been applied successfully in previous research as a persuasion technique (the so-called foot-in-the-door technique; Freedman & Fraser, 1966) to change peoples' attitudes, e.g. towards organ donation (Girandola, 2002), and to increase compliance with (typically) one-time behavioral requests, such as donations to charity (Bell, Cholerton, Fraczek, Rohlfs, & Smith, 1994) or having a drink (Guéguen, Marchand, Pascual, & Lourel, 2008). The present study suggests that this strategy also holds potential value for health behavior change interventions. Further exploration of this potential may be a promising avenue for future research.

4.4.2 Implications

In general, physical activity promotion programs can use small cash or charity incentives (CHF 10 per month) to increase participation among the target population, but not to influence behavior change in participants. Small cash incentives are more effective in promoting participation than charity incentives. However, programs that require participants to own or purchase an activity tracker are likely to limit uptake in the target population, probably in particular among those who would most benefit from participating. Cash incentives aimed at increasing physical activity should be larger and more immediate than the incentives used in the present study. To prevent attrition, mobile physical activity programs may need to create value for participants that goes beyond the mere tracking of activity.

Moreover, the present study has important implications for the design and further development of the insurer's physical activity program. The data from this pilot test suggest that the public health impact of the program is limited due mainly to the low overall reach, large attrition and insufficient effects of incentives on behavior change. Consequently, and in line with the MOST framework, a revision of the program is recommended. The present study points to several aspects that could improve the program's impact. In line with the implications mentioned above, enabling insurees to participate without requiring an activity tracker (e.g. by allowing physical activity to be tracked with a smartphone) could substantially improve the program's reach. Additionally, greater and more immediate incentives are needed to affect behavior change. Lastly, survey results revealed participants' need for integration of further physical activities and additional behavioral support. The revision of the insurer's physical activity program is described in detail in the next chapter.

4.4.3 Limitations and Future Work

The present study has some limitations to consider. First, the results on the effects of incentives on physical activity have to be interpreted with caution. Because all insurees were randomized prior to participation, differential participation between the study groups could potentially confound the effects of incentives on physical activity. Indeed, small but statistically significant differences were observed between the incentive groups at baseline. Although the statistical models controlled for those differences, it is still possible for unobserved group differences to affect the results. For example, the study design did not allow for collecting baseline measurements on physical activity outcomes (e.g. steps per day). Similarly, non-significant effects on steps per day may not be taken as evidence of the absence of meaningful effects because the study was not powered for secondary outcomes. In fact, because participants could respond to the baseline survey after the program had started, it is possible that some self-reported baseline information was affected by the incentive strategies. This could potentially lead to overly conservative estimates of group differences in adjusted models. Similar to most studies of mobile physical activity interventions, physical activity increases in the present study cannot be separated from increases in wear time. Thus, it cannot be excluded that the reported short-term effects of incentives mainly reflect increases in the time the activity tracker was worn by participants. Further, participants in the charity incentive group had the opportunity to chose the proportion of their reward that they wished to donate to charity. Thus, the results of this study are not directly comparable to those of previous studies, which did not offer this option. Lastly, the comparison of participants' baseline characteristics with representative data from the Swiss population suggested a selection effect of the program. However, detailed data on characteristics of non-participants are necessary to rigorously evaluate selection effects and whether these are amplified or attenuated by the presence of incentives.

Future work needs to complement the present research by investigating alternative incentive designs with small incentive values that can be implemented and sustained in practice. As financial incentives gain popularity as a behavior change tool in practice, there is a clear need to investigate incentive designs with a reasonable chance of implementation. Further, this study revealed the need to research and identify strategies that successfully promote reach and prevent attrition within mobile physical activity interventions. The identification of such strategies addresses the most urgent shortcomings of those interventions and is likely to have considerable impact on their success.

4.4.4 Conclusion

Small monthly cash and charity incentives increased participation in the insurer's physical activity program. However, the short-term effects of these incentives on physical activity and attrition limits the utility of small incentives in the context of health promotion programs. Consequently, a revision of the insurer's physical activity program is required. Barriers to participation need to be removed, incentives need to be modified, and additional measures to prevent attrition are needed in order to increase the program's public health impact.

Chapter 5

Revision of the Insurer's Program

The study presented in the previous chapter concluded that the public health impact of the insurer's physical activity program is limited, due mainly to limited reach, selection effects, short-term effects of incentives, and attrition. To address these limitations, and in line with the MOST framework (Collins, 2018), a revision of the program is recommended. The revision of the insurer's physical activity program is described in this chapter.

5.1 Introduction

In order to address the shortcomings of the insurer's physical activity program, several measures were taken; these are briefly described in the following paragraphs and summarized in Table 5-1. The study in the previous chapter revealed that the reach of the insurer's program was limited mainly by the requirement that participants own an activity tracker. A way of surmounting this barrier and of facilitating uptake of the program is to use a smartphone, rather than an activity tracker, as the mobile device for monitoring physical activity. Similar to activity trackers, smartphones are capable of tracking a user's daily step count with acceptable accuracy in controlled conditions (Case et al., 2015), although smartphones may underestimate step counts in free-living conditions if they are not carried continuously (Duncan, Wunderlich, Zhao, & Faulkner, 2017). The key benefit of smartphone-based interventions is that they have the potential to reach a far larger proportion of the target population than activity trackers. At the time of writing, it is estimated that 77% of the Swiss population own a smartphone, placing Switzerland among the top ten countries worldwide in terms of smartphone adoption (eMarketer, 2017). As a comparison, the adoption rate of wearables and activity trackers in Switzerland is only around 10% (BVDW, 2016). What is more, smartphone ownership is high, or at least rising, among less educated and elderly individuals worldwide (Pew Research Center, 2019), a subgroup of the population at higher disease risk. Thus, smartphone-based interventions may also be less susceptible to selection effects than interventions built around wearable devices. To leverage the advantages of smartphone-based interventions, researchers at the Center for Digital Health Interventions developed a smartphone app for measuring and promoting daily physical activity: the Assistant to Lift your Level of Activity (Ally). The Ally app serves as the central component in a research prototype of a revised version of the insurer’s physical activity program.

High attrition was identified as another limitation of the insurer’s physical activity program. Within the Ally app, a digital coach was added as a tool to prevent attrition and to design interventions in an interactive and engaging way. Broadly, digital coaches are computer programs that mimic real-life conversations (e.g. chatbots), thereby enabling elements from traditional face-to-face coaching to be scaled-up via digital and mobile interfaces, primarily a smartphone. Digital coaches and automated conversational agents are increasingly being used as a way to deliver health interventions (Laranjo et al., 2018), and existing intervention platforms allow digital coaches to be easily integrated into smartphone apps (Barata, Kowatsch, Tinschert, & Filler, 2016). One potential advantage of digital coaches is the capability to build lasting, positive relationships with the user (so-called working alliances or therapeutic alliances), thereby increasing user adherence and preventing user attrition from digital health interventions (Kowatsch et al., 2018). The Ally app therefore includes a digital physical activity coach, also called “Ally”, which supports users as they increase their daily activity. Additional intervention components were designed to be delivered specifically via the digital coach within the Ally app.

Finally, the financial incentives in the insurer’s physical activity program had limited effects on behavior change. The existing literature on financial incentives and physical activity (cf. section 3.4) suggests that greater and more immediate incentives are needed to effectively increase physical activity. Thus, the incentives were redesigned accordingly during the development of the Ally app. The incentives included in the Ally app were daily financial incentives with a value of CHF 1 per day, which corresponds to the minimum incentive value that had produced significant effects on behavior change in previous studies (Mitchell et al., 2019).

Table 5-1. Comparison of the insurer's program and the Ally app.

Limitation of the insurer's physical activity program	**Addressed in the Ally app as follows**
Limited reach and selection effects	Use of a smartphone app to track physical activity
Limited effects of financial incentives on behavior change	Increasing the value and immediacy of incentives, plus additional intervention components
Substantial attrition	Use of an automated digital physical activity coach to build a positive relationship with the user
Focus on step counts as the only measure of physical activity	Not addressed

Table 5-1 summarizes the measures that were taken to overcome the limitations of the insurer's mobile physical activity program. The exclusive focus on step counts could not be addressed because smartphone-based, wholistic physical activity monitoring requires the user to self-report different physical activities. However, tying financial incentives to self-reported physical activity was considered inappropriate due to the likely entry of biased activity data. While an objective and wholistic monitoring of physical activity is possible through the inclusion of additional sensors (e.g. accelerometers), it was not possible to integrate these sensors in the limited time available for app development. Additionally, inclusion of these sensors gives rise to a concern about limited intervention reach, which the smartphone app is meant to address, as they require participants to procure and use additional features.

The following sections elaborate in greater detail on some noteworthy aspects of the revision process and the Ally app. The next section elaborates on the concept of digital coaching, a central element of the Ally app. Section 5.3 briefly introduces the MobileCoach platform, the software that was used to implement the digital coaching component of the Ally app. Finally, section 5.4 provides an overview of the Ally app and its intervention components, and section 5.5 outlines the conceptual model of the app.

5.2 Digital Coaching

A central new element in the revised version of the insurer's physical activity program is the addition of a digital physical activity coach. The following subsections explain the concept of digital coaching and its potential advantages and discuss the existing empirical evidence.

5.2.1 Definition and Examples

The term digital coaching is not yet clearly defined and is often used broadly as a synonym for digital health interventions. For the purpose of this dissertation, and in line with Kowatsch et al. (2019), digital coaching is understood more narrowly as the use of conversational agents (e.g. chatbots) to apply strategies, tools, and techniques that promote desirable and sustainable health behavior change. Within this definition, an important class of digital coaches are text-based healthcare chatbots (Kowatsch et al., 2018) that use either rule-based approaches or more advanced natural language processing capabilities to enable text-based communication similar to communication via mobile messaging platforms like iMessage or WhatsApp. Examples of applications of text-based healthcare chatbots include: Florence, for blood pressure management (Cottrell, Chambers, & O'Connell, 2012); Anna and Lukas, to support obesity treatment in children (Kowatsch, Nißen, et al., 2017); Woebot, for improving symptoms of depression (Fitzpatrick, Darcy, & Vierhile, 2017); and Lark, a digital weight-loss coach (Stein & Brooks, 2017). Another class of digital coaches that has been empirically investigated is embodied conversational agents (Bickmore & Cassell, 2005), i.e. digital coaches that are connected to a human-like character or avatar. In addition to engaging in mere text-based communication, these embodied conversational agents are capable of mimicking human gestures and non-verbal behavior.

It is no coincidence that many digital coaches, including the examples above, have been developed very recently and concurrent with a rising interest in digital health interventions. The following sections explain the advantages that digital coaches can have, especially when they are added to digital and mobile health interventions.

5.2.2 Rationale

Researchers in the field of human-computer interaction discovered early on that relationships between humans and computers are fundamentally social (Nass, Steuer, & Tauber, 1994). In a series of laboratory experiments, Nass and colleagues demonstrated that phenomena of human social interaction (e.g. politeness norms and gender stereotypes) can be replicated in interactions with computers, and that humans perceive different computer programs as different social actors (Nass et al., 1994). In a follow-up experiment, Nass and colleagues demonstrated that humans assign human personality traits to computers based on the way the computers present information (Nass, Moon, Fogg, Reeves, & Dryer, 1995). These findings suggest that humans and computers can build relationships that might, at least in part, resemble social relationships among humans.

The implications of these findings for digital and mobile health interventions have been summarized in the so-called Talk and Tools paradigm (Beun et al., 2017). Beun and colleagues noticed that the user interfaces of digital health interventions were almost exclusively designed for direct manipulation (e.g. clicking or touching, a “Tool”) and lacked communication interfaces (e.g. chats and dialogues, “Talk”) that could facilitate the establishment of positive human-computer relationships. Beun et al. (2017) therefore suggested augmenting digital health interventions with communication interfaces, such as digital coaches. There are two potential benefits of adding digital coaching to a mobile health intervention. First, building and maintaining social-emotional relationships between the digital coach and users can serve to maintain engagement over time and thus prevent intervention attrition (Beun et al., 2017; Bickmore, Schulman, & Yin, 2010). Second, these positive relationships can, in turn, promote adherence to interventions and thus increase intervention success. This has been consistently observed in face-to-face interventions, especially in psychotherapy (Di Blasi, Harkness, Ernst, Georgiou, & Kleijnen, 2001; Flückiger, Del Re, Wampold, & Horvath, 2018; Martin, Garske, & Davis, 2000). Therapeutic alliance, a measure of the client-therapist relationship quality, for example, is responsible for a moderate treatment effect of d = .58 (Flückiger et al., 2018). Although reverse causality could also explain the relationship between therapeutic alliance and treatment outcomes (i.e. more effective therapy leads to a better client-therapist relationship), longitudinal data support the role of therapeutic alliance as a causal factor for treatment success (Flückiger et al., 2018). In sum, using a digital coach to establish and maintain positive relationships with users could support long-term engagement with the intervention and promote adherence to the intervention's behavioral goals.

Some studies have investigated which characteristics of digital coaches contribute most to building a positive relationship with the user. These studies indicate that the use of empathy, humor, and self-disclosure are promising strategies for establishing a positive relationship (Bickmore, Gruber, & Picard, 2005; Bickmore & Picard, 2005). Those strategies were applied when scripting the digital physical activity coach Ally. For example, Ally uses informal greetings and farewells, regularly inquires about the user's current situation, and provides support and comfort in cases when low well-being is reported. Further, Ally makes witty remarks every now and then and, over time, reveals more about "herself" and what's on her mind[7].

Although there is clearly great potential for the application of digital coaching on a conceptual level, its benefits also need to be demonstrated empirically. The following section therefore reviews the empirical evidence regarding digital coaching for health and health behavior change interventions.

5.2.3 Empirical Evidence

Although research on digital coaching is still in its infancy, recently published review articles have made first attempts at summarizing the current state of research. Provoost, Lau, Ruwaard, and Riper (2017) reviewed 49 studies on embodied conversational agents used in mental health applications. The majority of reviewed conversational agents focused on social skills training and cognitive behavioral therapy, mostly in patients with autism spectrum disorder, depression, or anxiety (Provoost et al., 2017). Reported acceptance and user satisfaction was high, and the reviewed studies indicated that embodied conversational agents can increase patients' involvement in therapy and their therapy adherence. However, most studies were small-scale development and pilot studies. Another review of conversational agents used in the treatment of depression, anxiety, schizophrenia and substance abuse disorders reported similar results (Vaidyam, Wisniewski, Halamka, Kashavan, & Torous, 2019).

Laranjo et al. (2018) reviewed 17 articles about conversational agents in healthcare in general, although many of the identified studies targeted mental health issues. Most

[7] Assuming that users would connect the name Ally to a female digital coach, a female gender identity was given to Ally. The gender of the digital coach Ally is suggested to users by the avatar icon included in the chat (cf. Figure 5-2) but never revealed or discussed explicitly.

reviewed conversational agents were either supporting the patient (e.g. providing health education) or the clinician (e.g. supporting diagnosis). Satisfaction with the conversational agents in the review by Laranjo et al. (2018) was high, although some studies reported user experience problems and moderate levels of ease of use. One study included in the review evaluated the effect of a conversational agent in a randomized controlled trial, and reported a positive and statistically significant effect on depression symptoms (Fitzpatrick et al., 2017).

Examples of interventions using digital coaching also exist outside of mental health. Kowatsch, Nißen, et al. (2017) reported a study with obese children who were supported by a digital coach for six months in addition to their regular therapy. Preliminary results of $N = 15$ children suggest that, although the interaction frequency with the digital coach was high, the achievement rate of behavioral goals dropped from 80% in the beginning to below 40% at four months. In another study, the addition of a digital coach to a smoking cessation app more than doubled user engagement with the app and possibly increased quit success, although the evidence supporting the latter conclusion was of low quality (Perski, Crane, Beard, & Brown, 2019). In an observational study of a digital coaching app for weight loss, overweight and obese participants voluntarily interacted with the digital coach for 15 weeks on average before abandoning the app and lost an average of 2.38% of body weight during that time period (Stein & Brooks, 2017). The number of conversations with the digital coach was associated with weight loss, i.e. the more conversations participants had with the coach, the more weight they lost. With regard to physical activity, one RCT of a web-based digital coaching intervention found significant differences regarding self-reported activity when comparing the intervention to a no-intervention control group, but not in comparison to a content-identical intervention without the digital coach (Friederichs, Bolman, Oenema, Guyaux, & Lechner, 2014). In another RCT, adding a digital coach to an online pedometer-based intervention prevented a decline in step counts that was observed in the control group over the three-month study period (Watson, Bickmore, Cange, Kulshreshtha, & Kvedar, 2012).

Finally, a meta-analysis of randomized controlled trials of virtual humans in health interventions demonstrated that interventions using digital coaches can produce statistically significant effects on non-clinical and clinical outcomes, SMD = 0.49, 95%CI [0.27, 0.72], $I^2 = 75\%$ (T. Ma, Sharifi, & Chattopadhyay, 2019). "Virtual humans" is a broad concept used by the authors that includes digital coaches but also other digital representations of humans whose function may not relate primarily to communication, such as digital avatars in video games. Given this broad definition of interventions and the great heterogeneity of outcomes (e.g. hallucinations, exercise, depression, social skills, substance abuse, well-being), the meaningfulness of this summary effect may well be questioned. In addition, the authors did not report a risk of bias assessment of primary studies and did not assess publication bias. Because primary studies were mostly small (average $N = 80$), publication bias may be suspected.

In sum, a wide variety of digital coaches has been described and evaluated in the literature. So far, the primary area of application for digital coaching interventions appears to be in mental health, although some examples on lifestyle behaviors, such as physical activity, do exist. As of today, small development studies and pilot studies suggest that acceptance and user satisfaction with digital coaches are high. Some examples have demonstrated that interventions that include a digital coach can successfully increase engagement and change relevant intervention outcomes. However, this needs to be confirmed in large-scale RCTs with long-term follow-up measurements. It is unclear whether delivering an intervention via a digital coach has any benefits compared to other modes of delivery, e.g. due to increased adherence.

5.3 The MobileCoach Intervention Platform

To implement a digital coach, specialized software is necessary. This section briefly introduces the MobileCoach intervention platform (www.mobile-coach.eu), an open-source behavioral intervention platform that can be used to develop and implement behavior change interventions that are delivered via digital coaches. This platform was used to develop the Ally app, including its digital coaching component.

5.3.1 History

The MobileCoach system was initially developed as an SMS-based intervention platform (Filler et al., 2015) with the first effective applications developed for smoking cessation and curbing heavy drinking in adolescents (Haug et al., 2017; Haug, Schaub, Venzin, Meyer, & John, 2013). Subsequently, the MobileCoach platform was advanced to a more complex dialogue system (Kowatsch, Volland, et al., 2017) that allowed longer and more dynamic automated chat interactions via dedicated smartphone apps similar in style to popular messaging apps like iMessage or WhatsApp. First applications included a digital coach for obesity treatment in children (Kowatsch, Nißen, et al., 2017) and a digital coach to support data collection in asthmatics (Tinschert et al., 2019). MobileCoach-based smartphone apps include a chat window that displays the conversation between the user and the digital coach and a dashboard that summarizes relevant behavioral outcomes (e.g. steps per day, goal achievement). One important aspect in developing digital coaching applications is the dialogue management, that is, the rules and components that connect user input to responses from the digital coach and thus handle the flow of the conversation. The next subsection briefly explains the dialogue management of the MobileCoach system.

5.3.2 Dialogue Management

The MobileCoach version used in the development of the Ally smartphone app uses Extensible Markup Language (XML) scripts for dialogue management (Figure 5-1). Higher-order scripts determine rules that specify which conversation script is delivered to which user under which conditions (i.e. the intervention logic). These rules are based on if-statements and operate on variables that are specified in the MobileCoach system. Figure 5-1 A illustrates an example of a higher-order script of the Ally app. If by the end of the day someone using the Ally app recorded fewer than 1,000 steps that day, the conversation “invalid-steps” is called. If the user’s step count is above 1,000 steps but below the daily step goal, then either the conversation “goal-not-achieved” or the conversation “close-to-goal” is called depending on whether the difference between the user’s step count and the goal is less than or greater than 1,000 steps. If the user achieved their goal, then the conversation “goal-achieved” is called.

A

```
<select-case>
  <case variable="$day" in="20|23|35|37|45">
    <if variable="$current_steps" smallerThan="1000">
      <call-conversation ref="invalid-steps"/>
    </if>
    <if variable="$current_steps" greaterOrEqual="1000">
      <if variable="$current_steps" smallerThan="$current_goal">
        <if variable="$differenceStepsToGoal" greaterOrEqual="1000">
          <call-conversation ref="goal-not-achieved1"/>
        </if>
        <if variable="$differenceStepsToGoal" smallerThan="1000">
          <call-conversation ref="close-to-goal1"/>
        </if>
      </if>
      <if variable="$current_steps" greaterOrEqual="$current_goal">
        <call-conversation ref="goal-achieved1"/>
      </if>
    </if>
  </case>
```

B

```
<conversation id="goal-achieved1">
  <question text="Lass uns zum Abschluss des Tages schauen, wieviele Schritte du geschafft hast." type="select_one_horizontal">
    <answer text="Auf geht's"/>
  </question>
  <message>Du hast dein Ziel erreicht und bist insgesamt $current_steps Schritte gelaufen!</message>

  <if variable="$cashGroup" equals="cs">
    <question text="Dafür bekommst du einen Franken gutgeschrieben!" type="select_one_horizontal">
      <answer text="Top"/>
    </question>
  </if>
  <if variable="$cashGroup" equals="ch">
    <question text="Mit deiner Leistung spendest du einen Franken an: $organization!" type="select_one_horizontal">
      <answer text="Top"/>
    </question>
  </if>
```

Figure 5-1. Examples of dialogue management in MobileCoach. A: Different conversational scripts are called based on variables specified in the MobileCoach system (referenced by a $ symbol). B: Conversations consist of repeated sequences of question and answer tags and can be personalized by variable references and if-statements.

Within each conversational script, the dialogues are made up of repeated sequences of questions (messages from the digital coach to the user) and answers (predefined responses that can be selected by the user; Figure 5-1 B). By saving and referencing variables in the conversational script, conversations can be made more dynamic since messages can be tailored and personalized. Variables can be specified a priori (e.g. to define the user's group membership and implement different randomization schemes) or dynamically within the conversation based on recorded sensor data (e.g. the current step count) or user input (e.g. the user's name). The script-based dialogue management implies that all conversations are initiated by the digital coach and have to be defined a priori by the intervention author. This makes the MobileCoach platform a feasible tool for time-limited interventions but less feasible for long-term interventions. A key strength of the MobileCoach system is the integration of sensor data that allows dynamic personalization of conversations.

Having introduced the concept of digital coaching and the MobileCoach intervention platform, the next section describes how the digital coaching was implemented in the Ally app. Beyond digital coaching, the Ally app contains additional intervention components that are outlined as well below.

5.4 App Description & Intervention Components

In this section, the Ally app is described. This section starts with a general description of the app before its intervention components are explained in greater detail.

5.4.1 Description

The Ally app supports users in increasing and maintaining daily step counts by combining the physical activity monitoring capabilities of a smartphone with financial incentives and digital coaching. Specifically, the app sets daily step goals, tracks the user's daily step counts, rewards goal achievement, and provides additional digital coaching to help the user reach their daily goal. Generally, the Ally app consists of two different modules: the dashboard that displays physical activity-related data and the chat window that displays the conversations between the user and the digital coach (Figure 5-2). The Ally app runs on the common operating systems Android and iPhone operating system (iOS). On Android smartphones, Ally obtains all physical-activity-related information from GoogleFit, a health-tracking platform developed by Google. On iOS smartphones, the same information is obtained from the HealthKit, an application programming interface for health apps provided by Apple.

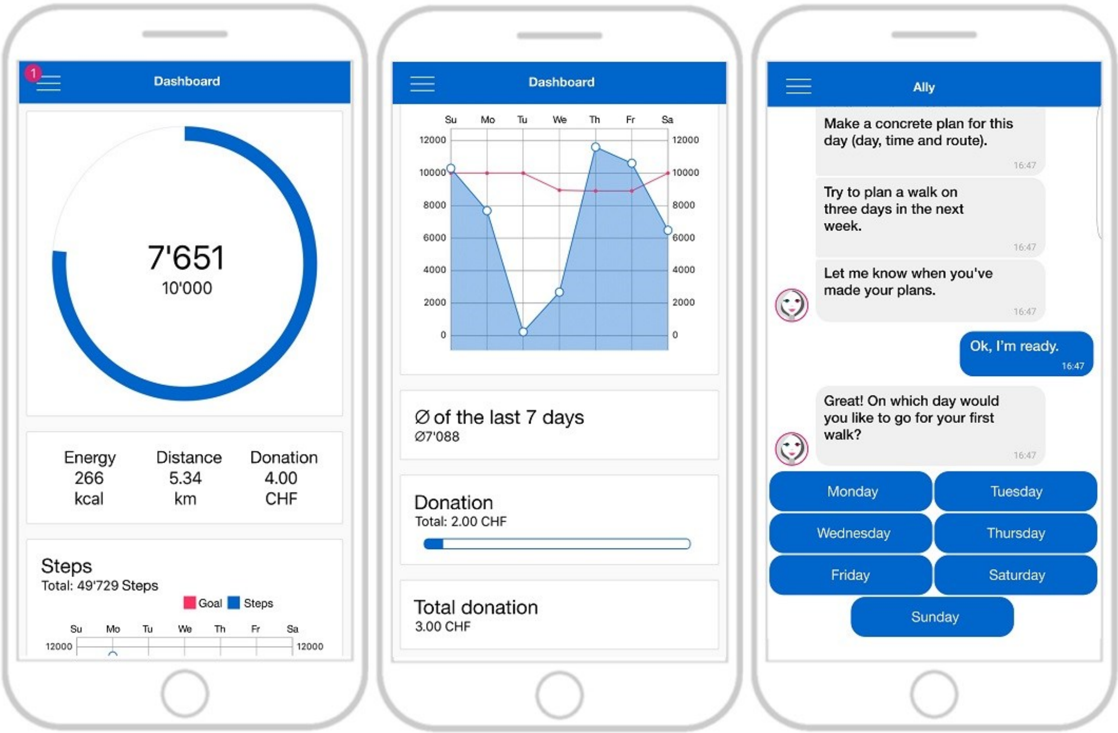


Figure 5-2. The Ally app. Dashboard with daily (left) and weekly overviews (middle), and interactions with the digital coach (right).

Due to their strong focus on self-monitoring behavior, physical activity apps likely appeal more to individuals who are motivated to become more active. Indeed, as noted earlier, users of mobile physical activity apps report greater intentions to change their physical activity than non-users (Carroll et al., 2017). The target population of the insurer’s physical activity program is no exception as indicated by possible selection effects (cf. section 4.3). When developing intervention components of the Ally app, especially the digital coaching, the focus was therefore on volitional behavior change processes that support motivated individuals in translating their intentions into actions and reaching behavioral goals (Schwarzer & Luszczynska, 2008). Beyond the dashboard and financial incentives, intervention components of the Ally app include personalized step goals, prompts to monitor and increase activity, and planning exercises. The latter two components were delivered via the digital coach. All intervention components are described in greater detail in the next subsection.

5.4.2 Intervention Components

Personalized Step Goals

Similar to the insurer's physical activity program, the Ally app provided step goals whose achievement was rewarded with financial incentives. In contrast to the static monthly goals of 7,500 and 10,000 steps, however, the Ally app set daily step goals (BCT 1.1: goal setting) that were personalized for each user based on the user's activity over the preceding 9 days, employing the moving-window percentile-rank algorithm described by Adams and colleagues (M. A. Adams et al., 2017). This adaptive goal-setting algorithm sets the daily step goal to the sixtieth percentile of the participant's step count distribution of the preceding 9 days, meaning that the participant reaches their step goal 40% of the time when maintaining their recent activity level. Previous studies have demonstrated that this adaptive goal setting outperforms static step goals (M. A. Adams et al., 2017; M. A. Adams et al., 2013). To facilitate maintenance of behavior change, adaptive step goals are capped at 10,000 steps per day, which approximates the amount of activity needed to reach the recommended 150 minutes of moderate-to-vigorous physical activity per week (Tudor-Locke et al., 2011). Ally communicates the personalized step goal to each user at the start of the day. Goals are a fundamental part of behavioral action control because they provide the standard against which one's own behavior is evaluated (Carver & Scheier, 1982) and thus control direction, effort and persistence of behavior (Latham & Locke, 1991). For goals to regulate behavior effectively, commitment to the goal is essential. Because the daily step goals of the Ally app are personalized, they may result in higher commitment over time compared to the static step goals in the insurer's physical activity program.

Dashboard

Similar to the online dashboard in the insurer's physical activity program, the Ally app's dashboard visualizes recorded physical activity data (Figure 5-2) and thus provides a convenient way for the user to self-monitor physical activity (BCT 2.3: self-monitoring of behavior) and obtain information on goal achievement (BCT 2.2: feedback on behavior). In line with the HAPA model of behavior change (Schwarzer & Luszczynska, 2008) and in addition to behavioral goals, these processes are central to action control, the most proximal determinant of behavior.

Financial Incentives

Based on the suggested explanations for the limited effects of the monthly incentives in the insurer’s physical activity program on behavior change (section 4.4), the Ally app rewards the achievement of the daily step goal with a payment of CHF 1 (BCT 10.1: material incentive). Thus, financial incentives of the Ally app are greater in value and more immediate and thereby more strongly leverage the user’s present bias (Loewenstein et al., 2013). Yet, the incentive value corresponds to the lowest financial incentive that has so far produced positive effects on behavior change in empirical studies (Mitchell et al., 2019).

Physical Activity Prompts

The daily step goals and the app’s dashboard will unlikely have an effect on behavior for users who do not engage with the Ally app on a regular basis, e.g. to check their current physical activity level. Physical activity prompts were designed to engage the user with the Ally app and increase the user’s awareness of the daily step goal and their current physical activity level. The prompts include short conversations with the digital coach that remind users of their daily step goal (BCT 1.1: goal setting), compare the user’s current step count to their daily goal (BCT 1.6: discrepancy between current behavior and goal, BCT 2.2: feedback on behavior), and provide an estimate of walking minutes necessary to reach the goal together with an actionable tip on how to increase physical activity (BCT 4.1: instruction on how to perform the behavior). Similar to the dashboard, physical activity prompts thus support subprocesses of action control, albeit in a much more active way since this intervention is pushed to the user via a notification in the smartphone’s status bar.

Tips on how to increase physical activity, such as “integrate a detour into your daily walking routines (e.g. walking to work) to collect additional steps” were gathered from websites and from reports of physical activity intervention studies. In total, 31 different tips were collected. In a pilot test, these tips were evaluated by a convenience sample (N = 29, M_{Age} = 31.0 years [SD = 10.2 years], 55% female) with regard to their ability to increase step counts on the same day the tip was given. Answers were given on a four-point Likert scale ranging from “not at all helpful” (1) to “very helpful” (4). Tips that were judged as difficult to implement were excluded or improved based on comments from participants. In total, 18 tips were deemed satisfactory for inclusion in the Ally app (Appendix J).

Planning Exercises

Even if users of the Ally app are highly motivated to increase their activity, previous studies show that, on average, 47% of people fail to act upon their good intentions (Sheeran, 2002). In addition, lack of time is often reported as a barrier to physical activity (Ashton, Hutchesson, Rollo, Morgan, & Collins, 2017; Cerin, Leslie, Sugiyama, & Owen, 2010) suggesting that competing intentions and short-term desires may prevent physical activity intentions from being implemented. In line with the HAPA model (cf. section 2.3.2), forming specific plans about when and how to act increases the likelihood of performing the intended behavior (Bélanger-Gravel et al., 2013; Scholz, Schüz, Ziegelmann, Lippke, & Schwarzer, 2008) and helps to bridge the abovementioned intention behavior gap. Planning can be further divided into action planning (specifying when, where, and how to act) and coping planning (specifying behavioral responses for barriers and difficult situations; cf. section 2.3.2). Plans that are articulated in an if-then format (e.g. "if I am tired at work, I will go for a brief walk to get new energy") are typically referred to as implementation intentions (Gollwitzer, 1999).

The Ally app included both an action planning (BCT 1.4: action planning) and a coping planning exercise (BCT 1.2: problem solving). In the action planning exercise, Ally asks the user to plan at least one and up to three walks for the upcoming week. To plan a single walk, the user needs to specify the day of the week, the time, and the route that they intend to walk. To create flexible plans, and thus increase the likelihood of adherence, Ally advises the user to choose event-related times (e.g. after work) instead of actual times. Action planning directly supports behavioral regulation because it helps people to identify and seize opportunities for action (Sheeran, Milne, Webb, & Gollwitzer, 2005). An important aspect is linking the initiation of the target behavior to a situational or environmental cue (e.g. the end of a workday) and committing to the initiation of the target behavior. This essentially transfers the behavioral control from the self to the situational cue (Sheeran et al., 2005) thereby facilitating the initiation of action once the cue is encountered.

In the coping planning exercise, Ally asks the user to identify barriers to physical activity by reflecting on the 2 least active days from the previous week. The user is then prompted to develop counterstrategies for each barrier using the if-then format of implementation intentions. Ally guides this process using examples for common barriers to physical activity that have been identified in previous studies (Cerin et al., 2010; Reichert, Barros, Domingues, & Hallal, 2007; Zunft et al., 1999), for example: "If I want to go for a walk but I lack motivation, I will think of the health benefits of walking to motivate myself." Finally, the user has the option to anticipate days of the upcoming week where the barrier may arise again. The mechanisms by which coping planning affects behavior change are thought to be similar to the mechanisms by which action planning works (Scholz et al., 2008), i.e. by committing to situational cues that elicit behavioral responses. In addition, coping planning may affect the user's self-efficacy, i.e. their beliefs about their capability to increase physical activity even when faced with difficulties. As illustrated in section 2.3, self-efficacy is a powerful individual-level predictor of behavior. Both action planning and coping planning exercises include reminders for the user on days when either a walk or a coping reaction was scheduled.

5.5 Conceptual Model

Based on the intervention components and their potential mechanisms of action outlined in section 5.4, the conceptual model of the Ally app can be specified as follows:

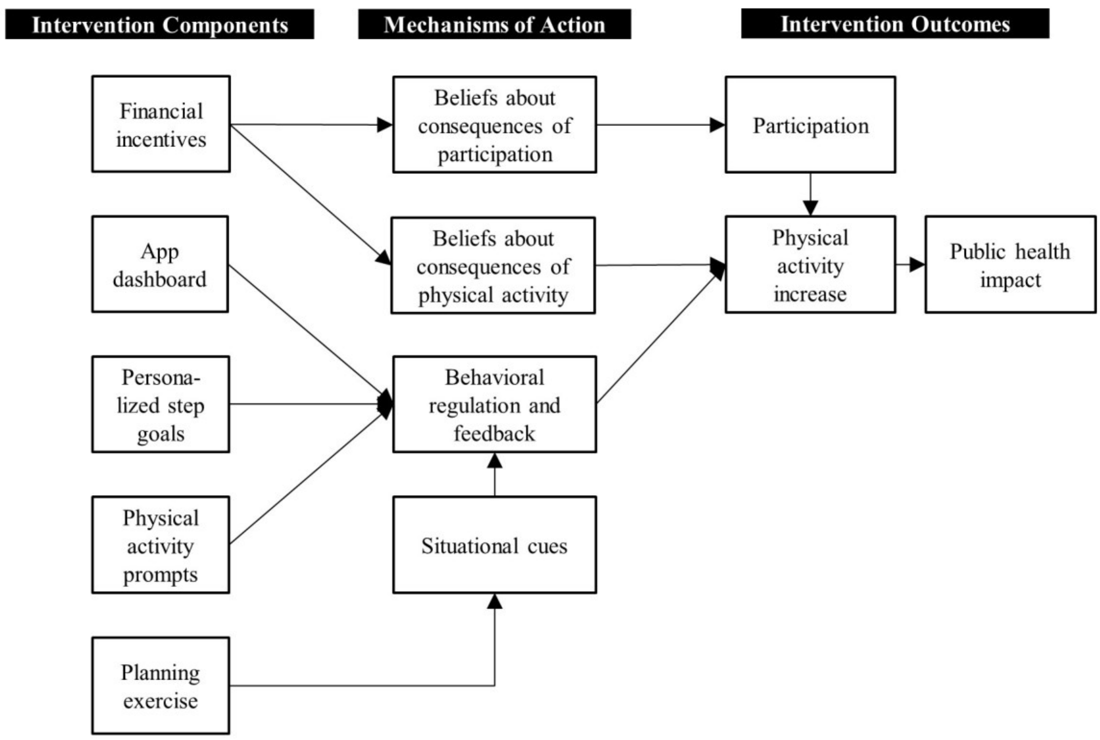


Figure 5-3. Conceptual model of the Ally app. All effects shown are assumed to be positive.

Note that, again (cf. subsection 3.2.3), mechanisms of action other than those proposed are possible but have been left out of the conceptual model to reduce complexity. For example, as mentioned in section 5.4, effects of personalized step goals and coping planning may be mediated through increased commitment or self-efficacy respectively.

Compared to the insurer's physical activity program, the mechanisms of action of the Ally app focus more on the volitional determinants of behavior thereby taking into account the characteristics of its target group. Yet, it is unclear whether and which of the novel intervention components of the Ally app can indeed change behavior, and whether the resulting overall effect of the app is sufficient to create substantial public health impact. Therefore, and in line with the MOST framework (Collins, 2018), a further optimization trial was conducted to assess the effects of single components of the Ally app. This optimization trial is reported in the following chapter.

Chapter 6

Study II: An Optimization Trial of the Ally App[8]

This chapter presents this dissertation's second empirical study, which investigates the effects of different components of the Ally app on users' physical activity. The results of this study help to answer the fourth research question of this dissertation and have informed optimization of the Ally app.

6.1 Introduction

The study described in this chapter was motivated by the concept of optimization as described in the MOST framework (cf. section 2.1). Recall that optimization trials evaluate the effects of single intervention components in order to identify which components do (and which do not) make an important contribution to the overall effect of the intervention. As such, optimization trials evaluate the effect of single intervention components on the primary intervention outcome. When optimizing the Ally app, it is not feasible to evaluate all intervention components listed in Figure 5-3. Some components, for example the app's dashboard and (personalized) step goals, are not only intervention components but can also be considered as very basic functionalities of the app itself. As such, they have to be included in the app and are therefore not subject to the process of optimization. Consequently, the primary objective of this study was to quantify the main effects of the remaining components: financial incentives, physical

[8] Parts of this chapter, relating in particular to the methods, results and discussion of the reported study, are published in the context of the following academic publications: J.-N. Kramer, Künzler, Tinschert, and Kowatsch (2019), J.-N. Kramer et al. (2020).

activity prompts, and planning exercises. Beyond effects on physical activity, the impact of financial incentives on participants' intrinsic and extrinsic motivation was examined because the potential undermining of intrinsic motivation constitutes a prevailing argument against the use of financial incentives to promote health behaviors (cf. section 3.4). A secondary objective was to explore participant engagement with and acceptance of the Ally app, which can provide additional relevant insights for optimization. As part of a related research project, this study also collected data to explore participants' states of receptivity (Nahum-Shani et al., 2016), i.e. moments in time where participants were particularly receptive to engaging with interventions delivered via smartphone push notifications. As a result, intervention-related push notifications were delivered to participants at random time-points within pre-specified time windows that were considered appropriate for the delivery of interventions. Note that due to the focus on optimization and the evaluation of single intervention components, the present study did not evaluate the reach of the Ally app.

It was hypothesized that all intervention components encourage participants to engage in physical activity, i.e. to walk. Further, based on the distinction between motivational and volitional determinants of behavior (cf. section 2.3), an interaction between incentives and physical activity prompts and planning was expected. Specifically, the effects of planning and self-monitoring prompts were expected to be greater if they are accompanied by incentives.

6.2 Methods

The optimization trial of the Ally app was conducted from October to December 2017. Again, this study was conducted in collaboration with the partnering health insurance company and data were collected in the German-speaking part of Switzerland. Informed consent was obtained from all individual participants included in the study. The ethical review board of ETH Zurich approved all study procedures. The study is registered on ClinicalTrials.gov (NCT03384550).

6.2.1 Study Design

Optimization trials typically use factorial experiments to evaluate the effect of single intervention components. Recently, Klasnja et al. (2015) proposed the micro-randomized trial (MRT), an optimization trial design related to the factorial experiment and specifically developed for optimizing mobile health interventions. MRTs use repeated randomization (micro-randomization) of participants to different versions and/or the presence and absence of individual intervention components over the course of the intervention. Thus, micro-randomization can be thought of as the repeated conduction of short-term factorial experiments. Like a regular factorial experiment, this enables the estimation of the intervention components' main effects and their interactions on (short-term) outcomes. Further, micro-randomization enables estimating time-varying causal effects, which can be highly beneficial in the development of adaptive mobile health interventions. The present study uses both, baseline randomization (as used in classic factorial experiments) and micro-randomization to estimate marginal and time-varying causal effects of financial incentives, physical activity prompts, and planning exercises. While marginal effects represent the average effect of the intervention over the complete study period and are of primary interest for the purpose of optimization, time-varying effects can add important insights regarding a possible deterioration of effects.

The study consisted of a 2-week run-in and baseline period and a 6-week intervention period. Data was collected from participants' smartphones via the Ally app and from two online questionnaires at the beginning and at the end of the study. Participants received CHF 10 (equal to $10) for participation in the study and completing both questionnaires. We invited 30,000 insurees of the health insurance company to participate in the study via an email invitation. Interested insurees could click on a link in the invitation email to be forwarded to an online survey platform where they were screened for eligibility. Eligibility criteria were: 1) German-speaking, 2) aged 18 years or older, 3) enrolled in a complementary insurance program, 4) being free of any medical condition that prohibits increased levels of physical activity, 5) not actively using an activity tracker or a comparable smartphone app, and 6) not working night shifts. Eligible insurees could subsequently obtain detailed information about the study goals and procedures, provide consent to participate, and enroll in the study. After enrollment, participants completed the first online questionnaire and received a 6-digit code, together with instructions on how to download and install the Ally app. Participants had to enter the code once upon first opening the Ally app to connect survey data and app data, and to ensure that only study participants were using the app. The baseline period started once participants had installed the app. During this period, Ally counted and displayed steps per day and sent occasional messages that were unrelated to physical activity to foster participants' interest in the study. However, the app's dashboard did not display any information related to financial incentives and no intervention-related messages were sent. Two weeks after sending out the invitation emails the baseline period ended and the six-week intervention period started for all participants. During the intervention period, the Ally app set daily step goals and delivered interventions to support step goal achievement.

6.2.2 Intervention Components and Randomization

Because the MobileCoach version used in this study requires dissemination time points for dialogues to be known a priori (cf. section 5.3), randomization for all intervention components (including sequences of micro-randomized component delivery) was performed in advance, upon enrollment of participants in the study.

Incentives

Similar to the first study of the insurer's physical activity program, cash incentives were again compared to charity incentives and to a no incentive control group. At the beginning of the study, participants were randomized to one of the three groups for the duration of the study with a randomization probability of .33 for each group. Participants in the cash incentive group received CHF 1 for each day they reached their personalized step goal. Participants in the charity incentive group earned the same amount, which was donated automatically to a charity organization. Participants were given four preselected charities (the Swiss Red Cross, Pro Natura [an organization committed to nature conservation], Rega [Swiss air rescue organization], and the foundation for children's cancer research) to choose from plus the option to name a charity of choice. In contrast to the charity incentives used in the first study of the insurer's physical activity program, participants did not have the opportunity to keep a proportion of the reward to themselves.

Planning

Planning interventions were randomized on a weekly basis. Every Sunday, participants received an action planning intervention, a coping planning intervention, or no planning intervention. Planning interventions included reminders on days when a brisk walk was planned or if a barrier for physical activity was anticipated (cf. subsection 5.4.2). Planning interventions were sent out according to a uniform and strongly balanced intervention schedule (Table 6-1) that controlled for time and carry-over effects during the six-week intervention period. At the beginning of the study, participants were randomized to one of the nine different sequences of the intervention schedule that determined the order of planning and control conditions during the study. To guarantee balance between the sequences, we used blocked randomization with a block size of nine and a randomization probability of .11. Planning interventions were delivered Sundays at a random point in time between 10am and 6pm.

Table 6-1. Intervention delivery schedule for planning interventions.

Sequence	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
S_1	AP	AP	CP	CC	CC	CP
S_2	CP	CP	CC	AP	AP	CC
S_3	CC	CC	AP	CP	CP	AP
S_4	AP	CP	CP	AP	CC	CC
S_5	CP	CC	CC	CP	AP	AP
S_6	CC	AP	AP	CC	CP	CP
S_7	AP	CC	CP	CP	CC	AP
S_8	CP	AP	CC	CC	AP	CP
S_9	CC	CP	AP	AP	CP	CC

Note. AP: action planning, CP: coping planning, CC: control condition (no planning)

Physical Activity Prompts

Participants were randomized daily from Monday through Saturday with a probability of .50 to either receive or not receive a physical activity prompt. Physical activity prompts were delivered at a random point in time between 10am and 6pm. To avoid interference between physical activity prompts and planning interventions, participants were not randomized to receiving or not receiving physical activity prompts on Sundays.

6.2.3 Outcomes

The proportion of participant days that daily step goals were achieved and step counts obtained from participants' smartphones were pre-specified as the primary and secondary outcome respectively. Due to the greater relevance of steps for estimating the resulting public health impact, the results section focuses mainly on the effects of intervention components on step counts. Postintervention differences in intrinsic and extrinsic motivation and differences in app engagement during the intervention period are evaluated as additional secondary outcomes. Dimensions of intrinsic and extrinsic motivation are measured using the Behavioral Regulation for Exercise Questionnaire-2

(BREQ-2; Markland & Tobin, 2004). As the external regulation subscale in the BREQ-2 exclusively relates to external regulation by other people, it is substituted by the more generally worded external regulation subscale of the Situational Motivation Scale (SIMS; Guay, Vallerand, & Blanchard, 2000). Subscales of both instruments have shown good reliability (Cronbach alpha=.73-.86 for the BREQ-2; Markland & Tobin, 2004; Cronbach alpha=.86 for the SIMS external regulation subscale; Guay, Vallerand, & Blanchard, 2000). Validity has been confirmed by factor analysis for the BREQ-2 (Markland & Tobin, 2004) and by correlational analysis for the SIMS (Guay et al., 2000). Similar to the first study, non-usage attrition, i.e. whether participants stopped using the Ally app, was analyzed as a measure of participant engagement. Non-usage attrition was operationalized via the daily number of app launch sessions, an objective measures of participants' app usage. An app launch session was defined as any interaction of the participant with the Ally app, separated by 5 minutes between events. If a participant left the app open and did not take action for 5 minutes or more, then the next interaction with the app counts as a new session. A participant was coded as "non-usage attrition observed" when she/he stopped using the Ally app at least 7 days before the end of the study.

Additionally, self-reported health status was measured with the SF-12 (Ware Jr, Kosinski, & Keller, 1996), and potential mediators of behavior change, i.e. self-efficacy, action control, action planning and coping planning, were assessed at baseline and at postintervention follow-up on a five-point Likert scale using adaptions of the measures of Scholz, Keller, and Perren (2009), Schwarzer, Lippke, and Luszczynska (2011), Sniehotta, Nagy, Scholz, and Schwarzer (2006), and Sniehotta, Schwarzer, Scholz, and Schüz (2005). Participant's perceptions of the Ally app, of intervention components, and of the chatbot as well as predictors of technology acceptance were measured using items from Venkatesh, Thong, and Xu (2012) at postintervention follow-up.

6.2.4 Statistical Analyses

To evaluate the effect of each intervention component and the interactions of interest, outcomes were aggregated to the time-scale of randomization of the respective intervention component. That is, to estimate the effect of incentives, randomized once at baseline, participants' average steps per day (calculated over the complete intervention period) were compared between the incentive groups using a linear regression model with incentive group membership represented by dummy-coded variables. For planning interventions, randomized weekly, participants' weekly step count averages (calculated separately for each of the six weeks of the intervention period) were compared between the different planning conditions. For physical activity prompts, randomized daily, steps per day were compared between days when physical activity prompts were either present or absent. To estimate the treatment effects of the micro-randomized components, planning exercises and physical activity prompts, we followed the analysis approach by Boruvka et al. (Boruvka, Almirall, Witkiewitz, & Murphy, 2017) for data from micro-randomized trials. This method produces unbiased causal treatment effects in situations where treatments are repeatedly randomized and covariates are time-varying. In the case of the present study, this method simplifies to an analysis using generalized estimating equations (GEE; Zeger, Liang, & Albert, 1988). Similar to multi-level modelling, GEE models account for the nested structure of longitudinal data. Intervention effects on step goal achievement are estimated using the same analysis approach. Analysis of variance was used to investigate differences between incentive groups with regard to intrinsic and extrinsic motivation at post-intervention follow up. Cox proportional hazard regression models were fit to the data to analyze participants' non-usage attrition. Similar to the first study, the cox regression model included effects for age, gender, baseline step count, smartphone operating system, intention to change physical activity, and incentive condition. Further details on the statistical analysis are available in Appendix K.

All effects of intervention components were estimated in a complete case analysis using available data only. Sensitivity analyses were conducted for missing data (intention-to-treat analysis) and for adjustment of covariates of physical activity. Covariates included in all adjusted models were age, gender, baseline step count, smartphone operating system, and employment. In addition, longitudinal models were further adjusted for linear time trends and a binary indicator for weekend days. To account for missing observations, missing data were assumed to be missing at random and multiple imputation was used to create ten complete datasets. Models were then fitted to each complete dataset separately and results were pooled over all datasets using Rubin's rules (Rubin, 2004).

A priori power analyses were conducted using a simulation-based approach that assumed a proportion of step goals achieved of 50% without interventions, intervention main effects on step goal achievement of 15%, and interaction effects of 5%. Based on these assumptions, we require a sample size of 220 to detect interaction effects with a power of 1-β = .80, assuming a type-1 error rate of 5%. All analyses were conducted in R, version 3.5.1 (R Core Team, 2014).

6.3 Results

6.3.1 Recruitment and Sample

Of all 30,000 invited insurees, 749 were screened for eligibility, of which 382 were classified as eligible and provided their consent to participate. Of those, N = 274 completed the baseline survey, installed the Ally app, and were randomized. Due to technical errors, six participants did not always receive the interventions they were randomized to receive. These participants were analyzed according to their randomized intervention schedules. After completion of the study, n = 181 insurees participated in the follow-up survey.

Comparisons of participants' baseline step counts with large-scale step count data from physical activity app users in Switzerland (Althoff et al., 2017), and of SF-12 component summary scores with the German 12-item Short Form norm sample (Morfeld, Kirchberger, & Bullinger, 2011), indicate that on average participants in our study were healthier and more active than the general population. Table 6-2 illustrates baseline characteristics of participants.

Table 6-2. Baseline characteristics of participants (N = 274).

Variable	Value
Age	41.73 (13.54)
Sex	
Female	158 (57.66)
Male	111 (40.51)
Missing	5 (1.82)
Education	
No university degree	100 (36.50)
University degree	164 (59.85)
Missing	10 (3.65)
Employment	
Full-time	152 (55.47)
Part-time	76 (27.74)
Not working	38 (13.87)
Missing	8 (2.92)
Smartphone operating system	
iOS	186 (67.88)
Android	88 (32.12)
Intention to increase physical activity	
Yes	223 (81.39)
No	48 (17.52)
Missing	3 (1.09)
Baseline step count	6,336 (2701)
Intrinsic motivation	3.96 (0.88)
Extrinsic motivation	2.93 (0.75)
Sitting (hours/day)[a]	7.00 [4.00, 9.00]
MVPA[b] (hours/day)[a]	1.75 [1.17, 3.00]
BMI	24.44 (4.15)
SF-12 physical component summary	53.32 (4.58)
SF-12 mental component summary	51.17 (8.11)

Note. Reported numbers are mean (standard deviation) for continuous variables and *n* (%) for categorical variables unless indicated otherwise.
[a] Reported numbers are median (interquartile range) due to non-normality
[b] Moderate-to-vigorous physical activity

6.3.2 Physical Activity

During the baseline period, participants walked 6,336 steps per day on average (SD = 2,701). During the intervention period, participants' mean step counts increased significantly to 6,774 steps per day (SD = 2,996), $t(200) = 3.0$, $p = .005$. A graphical illustration of participants' step counts over time suggests a curvilinear increase of step counts that starts during the baseline period (Figure 6-1). Specifically, step counts increased from around 5,500 steps per day at the beginning of the baseline period to just below 7,000 steps per day roughly two weeks into the intervention period. Subsequently, participants' step counts remained at this level throughout the rest of the study.

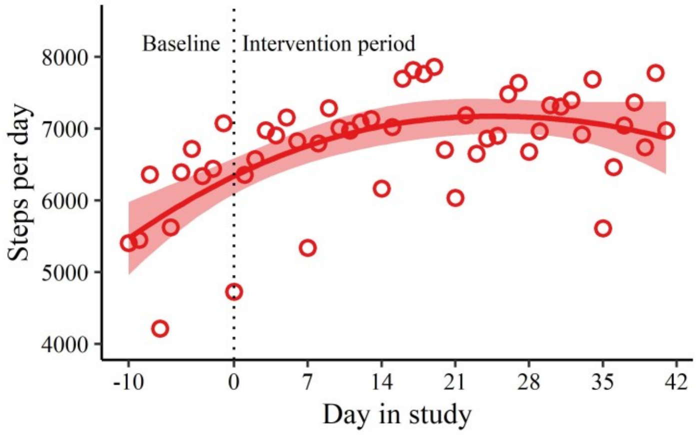


Figure 6-1. Daily step counts during baseline and intervention period. The solid line represents a LOESS smoothed curve with 95% confidence band.

There was a positive but statistically not significant correlation between participants' age and baseline step count ($r(193) = .09$, $p = .19$) and a positive and statistically significant correlation between age and step counts during the intervention period ($r(222) = .15$, $p = .02$).

6.3.3 Intervention Components

Incentives

Both cash and charity incentives led to a substantial but statistically not significant increase in step counts. During the study, participants in the cash incentive group walked on average 783 steps more per day, 95% CI [-135, 1701], $p = .10$, and participants in the charity incentive group walked 602 steps more per day, 95% CI [-305, 1509], $p = .19$, compared to the no incentive control group (Figure 6-2). However, cash incentives significantly increased the proportion of days that step goals were achieved. Participants in the cash incentive group had an 8.1% greater probability of reaching their daily step goals, 95%CI [2.1, 14.1], $p = .01$, than control group participants. Charity incentives were associated with a 6.9% greater probability of goal attainment, 95% CI [1.0, 12.8],

p = .02, compared to the no incentive control group. In the sensitivity analyses, only the effect of cash incentives remained statistically significant (Appendix Table L-1).

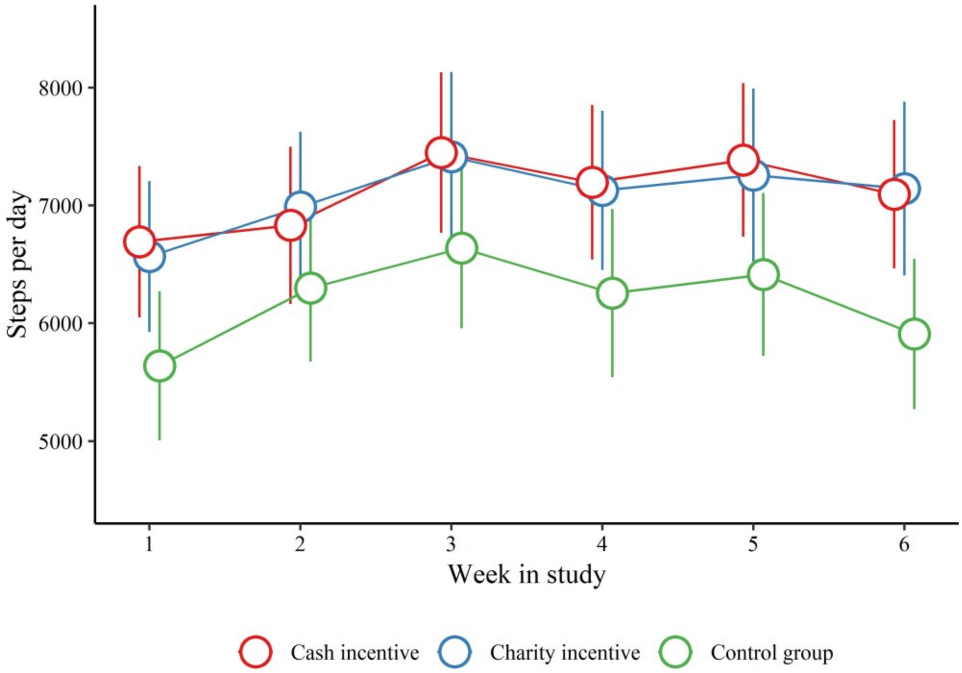


Figure 6-2. Unadjusted daily steps by incentive condition and study week. Error bars are 95% confidence intervals.

At baseline, participants' levels of intrinsic motivation were higher compared to their levels of extrinsic motivation (Table 6-2), and values remained stable at post-intervention follow-up. After the end of the intervention period, the incentive groups had similar levels of intrinsic motivation (M_{cash} = 3.92, SD_{cash} = 0.88; $M_{charity}$ = 3.90, $SD_{charity}$ = 0.93; $M_{control}$ = 3.97, $SD_{control}$ = 0.75) and group differences were not statistically significant (F[2, 178] = 0.11, p = .89). Likewise, levels of extrinsic motivation were similar between incentive groups (M_{cash} = 2.83, SD_{cash} = 0.77; $M_{charity}$ = 2.78, $SD_{charity}$ = 0.70; $M_{control}$ = 2.80, $SD_{control}$ = 0.80) and differences were not statistically significant (F[2, 178] = 0.08, p = .92).

Physical Activity Prompts

Averaged over incentive conditions, the main effect of physical activity prompts was small and not statistically significant for both steps per day (43 steps, 95%CI [-114, 200], p = .59) and daily step goal achievement (1.1%, 95%CI [-1.1, 3.2], p = .33). When adding the interaction effect between incentive conditions and physical activity prompts to the model, there was a somewhat larger but not statistically significant effect of physical activity prompts on step counts in the no incentive control group (108 steps, 95%CI [-191, 406], *p* = .48;). The effect of physical activity prompts was slightly smaller in the cash and charity incentive groups but this difference was not statistically significant (interaction effect cash incentives: -67 steps, 95%CI [-485, 325], *p* = .74; interaction effect charity incentives: -135 steps, 95%CI [-538, 167], *p* = .51). Likewise, no meaningful or statistically significant interaction effects were observed for step goal achievement (Appendix Table L-3).

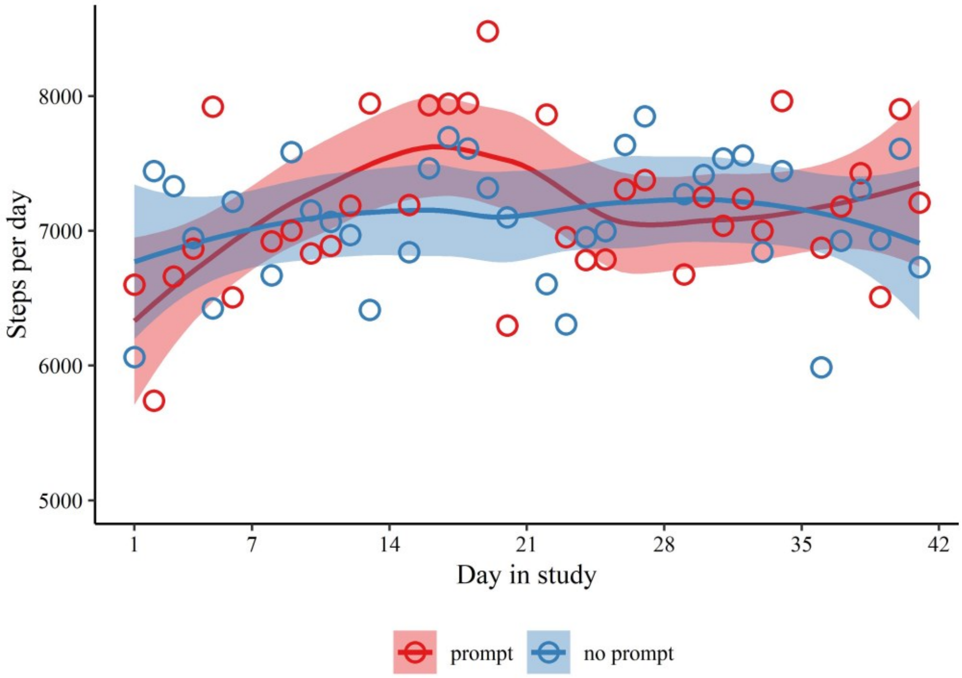


Figure 6-3. Unadjusted daily steps with and without prompts. Solid lines represent LOESS smoothed curves with 95% confidence bands.

The time-varying effect was investigated by adding the interaction between physical activity prompts and day of study to the model (Figure 6-3). On the first day of the study, the effect of physical activity prompts was negative but statistically not significant (steps per day: -43 steps, 95%CI [-407,321], *p* = .81); step goal achievement: -3.5%, 95%CI [-8.0%, 0.7%], *p* = .10). For step goal achievement, there was a statistically significant linear change in the effect over time (Appendix Table L-5), leading to a positive effect of physical activity prompts that became statistically significant around four weeks into the study. However, this time-varying effect was not robust to sensitivity analyses. No statistically significant time-varying effect was found in the models of steps per day (Appendix Table L-6).

In the post-intervention follow-up survey, the majority of participants indicated that physical activity prompts were useful (63.2%) and the included tips were easy to implement during everyday life (53.8%). Nevertheless, some participants reported that they lost interest in the content of the prompts after some time (35.6%). *Planning Exercises*Out of three possible plans, participants articulated 0.6 action plans and 0.4 coping plans per week on average. Neither action planning nor coping planning significantly affected participants' step counts (action planning: 101 steps, 95%CI [-163, 366], p = .45; coping planning: -113 steps, 95%CI [-351, 125], p = .35) or weekly step goal achievement (action planning: 1.1%, 95%CI [-2.0%; 4.1%], p = .49; coping planning: -0.3%, 95%CI [-3.1%, 2.5%], p = .84).

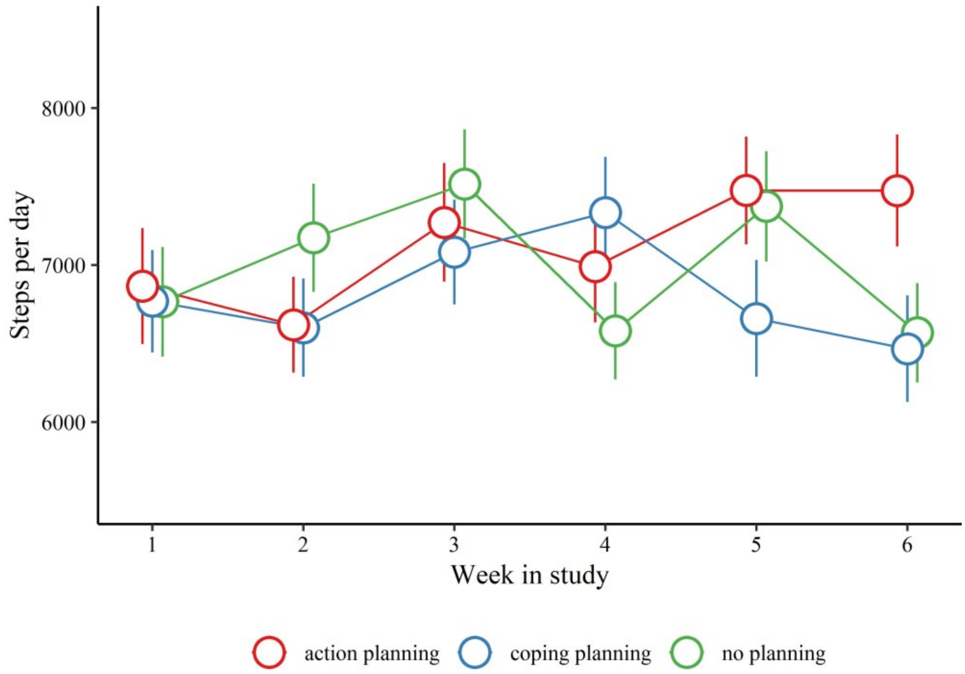


Figure 6-4. Unadjusted daily steps by planning condition and study week. Error bars are 95% confidence intervals.

Adding the interaction between planning exercises and incentive types revealed larger effects from both planning exercises in the no incentive control group and decreased effects in the cash and charity incentive groups, although neither simple effects nor interactions were statistically significant (Appendix Table L-7, Appendix Table L-8). However, when adjusting the analysis for missing data, there was a statistically significant simple effect of action planning on step goal achievement in the no incentive control group (5.8%, 95%CI [1.2, 10.4], $p = .01$) that decreased significantly in the cash incentive group (interaction effect: -7.1%, 95%CI [-14.0, -0.1], $p = .047$) and not significantly in the charity incentive group (interaction effect: -4.9%, 95%CI [-11.8, 2.0], $p = .16$). While the pattern of results was similar in models of steps per day, no effect was statistically significant (Appendix Table L-8). Again, time-varying effects were investigated by adding the interaction of action planning and coping planning with week in study to the model (Figure 6-4). Time-varying effects were small and statistically not significant for both step counts and step goal achievement (Appendix Table L-9 and Appendix Table L-10).

Participants' evaluations of planning exercises in the follow-up survey were mixed. While some participants agreed with the statement that the planning exercises were easy to complete (44.4% for action planning and 32.9% for coping planning), others disagreed (36.4% for action planning and 35.6% for coping planning). Some participants also reported that they did not always adhere to their plans (25.8% for action planning and 27.9% for coping planning) and that they would prefer to plan on a daily instead of a weekly basis (54.5% for action planning and 53.0% for coping planning).

6.3.4 Exploratory Analyses

The effects of planning interventions and physical activity prompts likely depend on whether participants engage with the respective intervention content. Yet the number of action and coping plans made by participants was low, and response rates to intervention conversations varied between 40.6% for coping planning and 55.4% for physical activity prompts. This suggests that the overall engagement with physical activity prompts and planning interventions was too low for the interventions to produce an effect. Therefore, an exploratory analysis of intervention components' main effects was conducted, which included recoded treatment indicators to differentiate whether a participant engaged with the intervention content or not (see Appendix K for details). Engagement with the intervention was defined as responding to the first message of an intervention-related conversation with the digital coach. To adjust for possible confounding in this analysis, we added known covariates of physical activity to the model.

Participants who engaged with physical activity prompt conversations recorded on average 405 steps more per day, 95%CI [189, 621], $p < .001$, compared to participants who did not receive a prompt. Conversely, participants who did not engage with the prompt decreased their daily step count on average by -745 steps, 95%CI [-1077, -431], $p < .001$. Likewise, participants engaging in action planning and coping planning interventions recorded on average 421 steps, 95%CI [127, 715], $p = .005$, and 475 steps, 95%CI [128, 822], $p = .007$, respectively, per day more in the following week compared to participants who did not receive planning interventions. Participants not engaging in the planning exercises recorded fewer steps than participants that did not receive a planning intervention. This difference was statistically significant for coping planning (-579 steps, 95%CI [-942, -216], $p = .002$), but not for action planning (-250 steps, 95%CI [-701, 201], $p = .28$). Detailed model results of the exploratory analyses are available in Appendix Table L-11 and Appendix Table L-12.

6.3.5 Mediators of Behavior Change

Examination of changes in self-reported mediators of behavior change, from baseline to post-intervention follow-up, might point to mechanisms by which the Ally app promotes physical activity. Participants reported a statistically significant medium-sized increase in action control ($t(179) = 7.17$, $p < .001$, $d_z = 0.54$, Figure 6-5 A) and a statistically significant small increase in coping planning ($t(179) = 3.02$, $p = .002$, $d_z = 0.22$, Figure 6-5 C). Participants reported no significant differences for action planning and maintenance self-efficacy, two mediators with comparatively higher baseline values (Figure 6-5 B and E). In addition, participants reported a small and statistically significant decrease in task self-efficacy ($t(179) = -4.86$, $p < .001$, $d_z = -0.37$, Figure 6-5 D).

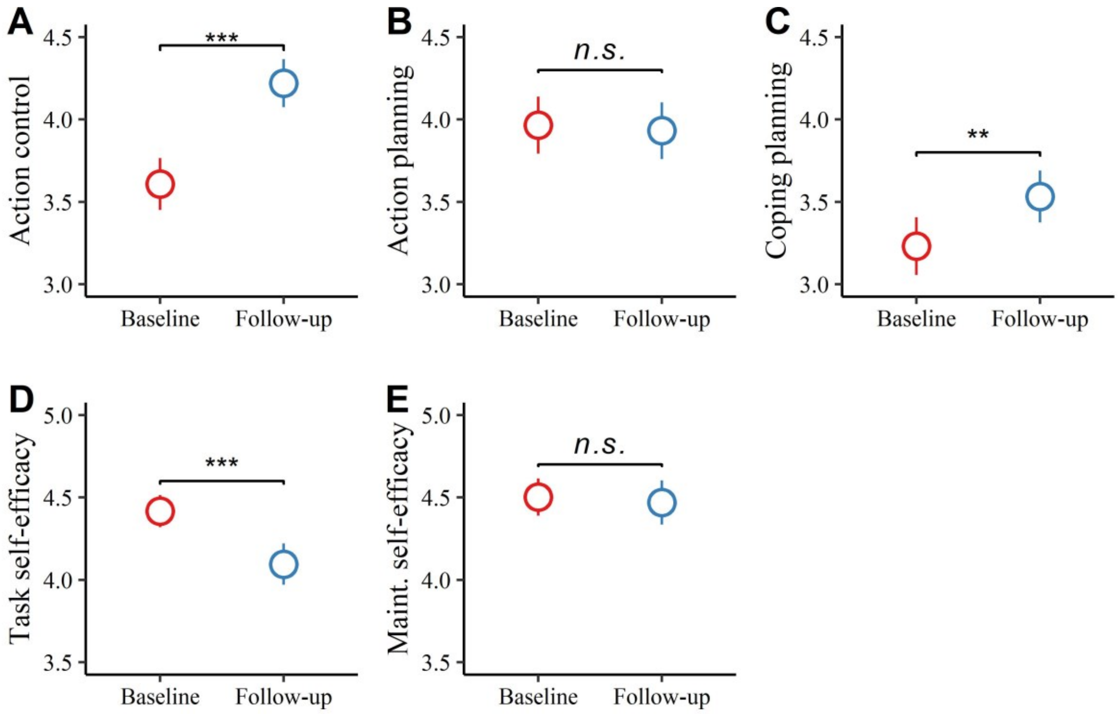


*Figure 6-5. Pre-post comparisons of behavior change mediators. Error bars are 95% confidence intervals. ***: p < .001, **: p < .01, n.s.: not significant. Answers were given on a five-point Likert scale.*

6.3.6 Engagement

On average, participants launched the Ally app 3.4 (SD = 3.5) times per day. The proportion of participants using the Ally app declined over the course of the study and, at the end of the study, 83 of 274 participants (30.3%) had stopped using the Ally app. As illustrated in Figure 6-6, attrition was lowest in the cash incentive group (25.8%), followed by the control group (28.4%), and the charity incentive group (36.1%). The differences in attrition rates were not significantly different between the control group and the cash incentive group (HR = 0.64, 95%CI [0.28, 1.42], p = .27) and the charity incentive group (HR = 1.58, 95%CI [0.84, 3.00], p = .16). In the adjusted model, none of the included variables significantly predicted non-usage attrition (Appendix Table L-13). Yet, participants who stopped using the app differed significantly from participants who continuously used the app with regard to steps per day recorded during the intervention period (4441 steps [SD = 2653] vs. 6979 steps [SD = 2909]) but not with regard to steps per day at baseline (5916 steps [SD = 2544] vs. 6408 steps [SD = 2727], see also Appendix Figure M-1).

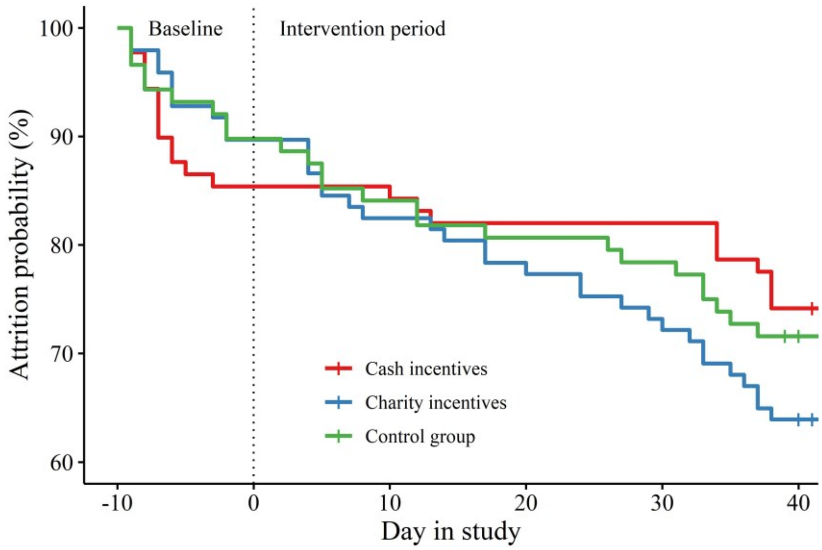


Figure 6-6. Kaplan-Meier attrition curves by incentive group.

6.3.7 Participants' Perceptions

At post-intervention follow-up, participants reported moderate to high overall satisfaction with the Ally app, with 60.2% of participants indicating that they were satisfied or very satisfied with the App. Participants rated the app positively on a 7-point response scale regarding predictors of technology acceptance. Specifically, participants agreed that the Ally app was easy to use ($M = 5.94$, $SD = 1.27$) and fun to use ($M = 4.78$, $SD = 1.70$), and that using the app became a habit ($M = 5.18$, $SD = 1.68$) and helped increase everyday physical activity ($M = 4.89$, $SD = 1.72$). Participants developed neither positive nor negative attitudes towards the digital coach. The mean values were close to the neutral scale midpoint for the following items, which participants were asked to rate: their intention to continue working with the digital coach; the closeness of their relationship with the digital coach; their preference of the digital over a human coach; and their adherence to the advice of the digital coach (Appendix Table N-1).

In open-ended questions regarding positive and negative aspects of the app, two different negative and two different positive themes emerged. Participants criticized technical errors (e.g. crashes, delayed responses from the digital coach, or incorrect displays of step counts) and dialogue management issues (e.g. repetitive dialogues or inadequate response options). Further, participants mentioned the restrictive focus on steps per day and expressed the desire to track other types of physical activity. On the positive side, participants appreciated the design of the app (e.g. ease and simplicity of navigation, and visualizations) and the conversations with the digital coach (e.g. humorous and motivating character, constant friendly reminders to reach step goals).

6.4 Discussion

6.4.1 Principal Findings

This optimization trial quantified the main effects and interactions of three components of the Ally app: financial incentives, physical activity prompts, and planning interventions. Additional analyses explored behavior change mediators, intervention effects conditional on engagement with the intervention, and attrition. Over the course of the study, participants increased their physical activity by almost 1,500 steps. With the caveat that the observed rise in step counts relies on observational data only, this increase seemed to be driven by participants who used the app over the complete study period, while step counts of participants who stopped using the app declined over time. Notably, the observed increase in physical activity started already during the baseline period, indicating that the intervention components that are not subject to experimental manipulation (e.g. the app's dashboard and personalized step goals) may contribute substantially to the overall effect of the app.

Among the intervention components that were randomized throughout the study, cash incentives supported participants' increase in physical activity by significantly increasing the achievement of daily step goals. The effect on step goal achievement translated to an increase of 780 steps per day, which is comparable to incentive effects found in previous studies (Mitchell et al., 2019). However, this effect was not statistically significant. Lack of statistical power may provide a likely explanation for the non-significant effect of incentives on step counts because this study was powered for its primary outcome step goal achievement. Indeed, a post-hoc power analysis revealed a power of only $1-\beta = .52$ for the found effect size of 780 steps. Charity incentives did not promote step goal achievement or steps per day after adjusting the analysis for missing data. Taking into account the results of the first study reported in chapter 4, there is no support for the use of charity incentives to increase physical activity when donating to charity is compulsory. Benefits of mere charity incentives may be restricted to certain populations, such as the elderly, who might be more responsive to social and emotionally rewarding goals (Harkins et al., 2017).

The remaining intervention components – physical activity prompts and planning exercises – did not contribute to the app's overall effect. Thus, these components do not qualify for inclusion in an optimized version of the Ally app and a revision of both components is necessary. The exploratory analyses illustrated that engagement with both components is crucial. For example, the results of the exploratory analyses are in line with a potential positive effect of both components when participants engage with the intervention, although other explanations for this effect are possible (see below). Further, the exploratory analyses also illustrate that the interventions are ignored by a considerable proportion of participants. Participants who ignored the interventions tended to record fewer steps than those that did not receive any intervention presumably because they were not carrying their smartphone and thus could neither be reached by the intervention nor record steps. This highlights the role of accurately timing the delivery of intervention-related push notifications to moments in time where participants are able to engage with the intervention. It further suggests that, similar to the negative effects that were observed when participants ignored the interventions, potential positive effects of intervention components could, either fully or in part, reflect increases in the time the smartphone was carried by participants. This may also be true for the effects of some mobile physical activity interventions reported in previous studies, that did not control their analyses for wear time of mobile devices (Glynn et al., 2014; Shcherbina et al., 2019).

Though the timing of delivery is one important aspect that needs to be optimized when revising physical activity prompts and planning exercises, feedback from participants revealed further improvement opportunities. For instance, some participants stated that they lost interest in the physical activity prompts over time, indicating that the relevance of the information conveyed by the prompts decreased over the course of the study, possibly because participants started to learn about their activity patterns. Decreasing the prompt frequency and including additional insights regarding participants' physical activity could ensure that physical activity prompts remain relevant over time. The planning exercises, on the other hand, were too great a burden for participants as illustrated by the low number of plans that were made by participants via the Ally app. Consequently, the planning exercises need to be simplified, for example, by planning on a daily basis or by planning outside the chat via a dedicated user interface.

Contrary to what was hypothesized, the effects of physical activity prompts and planning exercises were not enhanced by the presence of incentives. While this could likely be a consequence of missing main effects, the result pattern points to another potential explanation. Given the high activity levels of participants at baseline, ceiling effects may prevent positive interactions of intervention components. For example, the action planning exercise produced a significant positive effect on step goal achievement for participants in the no incentive control group, but negligible and statistically non-significant effects for participants receiving cash or charity incentives. Thus, it may be difficult for active individuals to raise their daily activity level beyond increases attributable to the app's basic components (i.e. the dashboard and personalized step goals) and incentives.

Summarizing the results above, the fourth research question of this dissertation can therefore be answered as follows:

RQ 4: Which aspects of an incentive-based digital coaching app help users to increase daily physical activity?

Daily incentives, physical activity prompts and planning exercises were evaluated. Of these, only daily cash incentives significantly increased physical activity. Physical activity prompts and planning may require sufficient participant engagement to be effective. Although not explicitly tested, the remaining components of the app, i.e. the dashboard and personalized step goals, also appear to contribute to the app's overall effect.

Beyond the main effects of intervention components, this optimization trial also revealed selection effects and attrition that can considerably limit the overall public health impact of the Ally app. First, there seemed to be selective participation in the study among all invited insurees. Similar to what has been observed for the insurer's physical activity program (cf. chapter 4) and for health promotion programs in general (Glasgow et al., 1993), baseline characteristics illustrated that study participants were healthier and more active than the general population. Likewise, the missing negative correlation between participants' age and baseline step counts and the positive correlation between age and step counts during the intervention period suggest that specifically elderly participants were more active and perhaps more motivated than their age-group average. This demonstrates that relying on smartphones for monitoring physical activity (rather than wearable devices) does not necessarily prevent selection effects. Further, although a digital coach was integrated into the Ally app to maintain engagement and prevent attrition, the analysis of app usage revealed that 30% of participants stopped using the Ally app over the course of the study. Those who did use the app continuously were more active during the intervention, but not during the baseline period, suggesting that participants who did use the app seemed to benefit. Nevertheless, the digital coach failed to establish a positive relationship with participants, suggesting that the strategies applied by the digital coach Ally to maintain engagement were insufficient. In fact, the attrition rate in the shorter present study was greater than the attrition rate in the insurer's physical activity program observed in the first study (cf. chapter 4). While technical errors may have played a role, this attrition rate can potentially be explained by the relatively low barriers to deleting a smartphone app as compared to abandoning a wearable device that was possibly purchased just recently, perhaps in order to participate in the study. Additionally, recent real-world data illustrate great attrition and low engagement for the majority of mobile health apps (Baumel, Muench, Edan, & Kane, 2019; Dorsey, McConnell, Shaw, Trister, & Friend, 2017), and earlier research demonstrated similar challenges for behavioral open access websites (Eysenbach, 2005). Thus, attrition and lack of engagement appear to be challenges for mobile health apps in general. Collectively, these results highlight that the benefits of the Ally app are restricted to a subset of participants, thereby limiting the app's potential impact on public health.

Interestingly, and despite an observed increase in physical activity, participants reported a decrease in task self-efficacy over the course of the intervention. Two different processes may explain this effect. First, the decrease in self-efficacy may simply be the consequence of the observed average increase in physical activity by participants over the course of the study. Participants who increased their physical activity are likely to be less optimistic regarding further positive increases in the future. Second, for participants who failed to increase their activity, this decrease may be the consequence of optimism bias, i.e. the tendency of individuals to overestimate the likelihood of positive outcomes when thinking about the future (Sharot, 2011). Participants may have been overly optimistic in estimating their ability to increase physical activity at baseline and subsequently slightly corrected their perceptions after six weeks of trying. Indeed, motivating health-promoting behaviors, such as physical activity, has been suggested as a possible explanation for the adaptivity of holding optimism bias (Giltay, Geleijnse, Zitman, Buijsse, & Kromhout, 2007; Sharot, 2011).

6.4.2 Implications

Combining physical activity smartphone apps with cash incentives may be a promising approach for large-scale physical activity promotion, but selection effects, attrition, and lack of engagement are major challenges. Physical activity programs built around smartphone apps therefore need to actively promote (long-term) engagement and prevent selection effects. For example, targeted marketing and endorsement of apps via healthcare providers may be one way to counter selection effects and reach those most in need of physical activity. This is currently being facilitated in some countries through the development of databases that certify trusted and evidence-based health apps, such as the NHS apps library in the UK[9] or the digimedia library in Germany[10]. Within physical activity apps, notification dependent interventions may require additional engagement strategies in order to be effective. The present study illustrated that simply relying on participants' willingness to engage with the interventions is insufficient.

In addition, the results of this study also have implications for the further use and development of the Ally app. Empirical results and participant feedback demonstrated that the Ally app requires substantial improvements with regard to technical quality, digital coaching and engagement strategies if it is to produce sustained effects that are sufficient to result in public health impact. Planning exercises and physical activity prompts in particular need to be revised according to participant feedback. Further, while the incentives used in the Ally app produced significant effects, they are substantially larger than the incentives of the insurer's program in the first study. In fact, the maximum attainable incentive amount of CHF 365 per year in the present study exceeds the subsidy of economy and balance-level insurees for physical activity, which was set to CHF 150 and CHF 250 respectively (cf. section 3.2). Thus, it would be necessary to adjust the subsidy, e.g. by allowing for a more flexible split of the complete subsidy between different activity categories, to incorporate incentives with a reasonable likelihood of changing behavior into the insurer's physical activity program.

6.4.3 Limitations and Future Work

Several limitations of this study merit consideration. To begin with, the selective sample of participants may limit the generalizability of the reported intervention effects. For

[9] https://www.nhs.uk/apps-library/

[10] https://digimeda.de/

example, physical activity intervention effects are known to be greater in patient populations (Conn et al., 2002) and incentive effects are known to be greater in more deprived populations (Mantzari et al., 2015). Second, the present study does not allow to separate increases in physical activity from increases in the amount of time the smartphone was carried by participants. The exploratory analyses suggest that carrying the smartphone could be a driver of physical activity levels in the present study. Third, the reported effects are limited to the six-week intervention period only. It is unclear, for example, whether the effects of incentives are maintained over longer periods of time. Similarly, the MobileCoach version used in this study limits the digital coaching of the Ally app to intervention dialogues that had to be initiated by the digital coach and had to be manually prespecified before the start of the study. It is therefore not feasible to extend the digital coaching of the Ally app to time periods substantially longer than the duration of the study, which could limit long-term effects of digital coaching components. Fourth, technical errors were mentioned by a substantial proportion of participants in the follow-up survey. This may affect both the internal and external validity of the results. For example, technical errors can impact participants' use of the Ally app and subsequently the effectiveness of intervention components. Lastly, although all participants indicated upon enrollment that they were not using any comparable apps or devices for tracking physical activity, it is nevertheless possible that such apps or devices were used or that participants primarily used the Apple Health or Google Fit applications that were required for the Ally app to obtain step counts. Like technical errors, additional apps or devices could potentially affect the use of the Ally app and thus the effectiveness of interventions.

Future research needs to investigate the individual, interpersonal and contextual determinants of uptake and engagement with mobile physical activity apps. One promising strategy to increase engagement is, for example, the integration of human support into mobile health apps (Schueller, Tomasino, & Mohr, 2017). In previous research, human support increased efficacy of web-based interventions for depression (Andersson & Cuijpers, 2009) and health apps that include social support had higher engagement and lower attrition than other health apps (Baumel et al., 2019). Another strategy to increase engagement with interventions delivered via push notifications, such as those investigated in the present study, is to use intelligent notification management algorithms. These algorithms utilize smartphone sensor data to predict opportune moments for intervention delivery, i.e. moments when participants will most likely react to a smartphone push notification. Research in the field of interruptibility has revealed that the application of such algorithms can increase response rates to smartphone notifications (Künzler, Kramer, & Kowatsch, 2017). Moreover, research is needed on how to use digital conversational agents to build and maintain positive relationships with users and on whether these positive relationships translate to higher engagement, adherence, and, ultimately, higher intervention effects. The mixed perceptions of participants in this study indicate that building these positive relationships is complex and a one-fits-all solution may not exist. Likewise, more advanced dialogue management software can enable more dynamic and complex digital coaching interventions, such as user-initiated dialogues or event-related interventions (e.g. triggered by the achievement of activity goals or user states) that are potentially of greater relevance compared to interventions that are pushed to users at predefined or random times. Finally, future work needs to advance mobile physical activity monitoring with the goal of enabling large-scale and objective measurements of total physical activity as the basis for interventions.

6.4.4 Conclusion

The Ally app could potentially increase physical activity in the short term. While financial incentives substantially contributed to the app's overall effect, notification-dependent intervention components did not. Selection effects, attrition, and lack of engagement emerged as challenges that prevented all participants from benefiting from the app. This study pointed to several improvements that need to be implemented before the Ally app can be tested in a larger randomized controlled trial.

Chapter 7

General Discussion & Implications

This chapter summarizes the key findings of this dissertation and highlights implications for the public health impact of mobile physical activity interventions. The first subsection of this chapter recaps the context and motivation of this dissertation and briefly summarizes key findings and answers to the four research questions that were presented in chapters 3 to 6. Next, implications of the results of this dissertation for the public health impact of mobile physical activity interventions are discussed and recommendations for the design of these interventions are derived. Subsequently, limitations of this dissertation and opportunities for future work are outlined. This chapter concludes with a brief conclusion of the dissertation.

7.1 Summary

7.1.1 Motivation and Key Findings

This dissertation was motivated by three distinct patterns observed in the literature around physical activity and physical activity interventions over the last years. The first is the tremendous health benefits of physical activity. To date, a vast amount of empirical evidence has accumulated detailing the health benefits of physical activity. In particular, physical activity has emerged as an important preventive factor for non-communicable diseases, one of the major health challenges of the 21st century (WHO, 2014). In fact, large epidemiological studies have illustrated that physical inactivity is an NCD risk factor of equal importance to established health risks, such as smoking and obesity (Lee et al., 2012). Further, large dose-response meta-analyses (e.g. Arem et al., 2015) have illustrated the type and amount of physical activity needed for effective disease prevention. The second pattern is the strong limitations of existing physical activity interventions. Traditional, non-mobile physical activity interventions suffer from high resource demands and low effect sizes (Conn et al., 2002; Foster et al., 2005) making them unsuitable for large-scale adoption in practice and thus effective prevention at the population level (Ammerman et al., 2014). And third is the potential of novel interventions designed around mobile technologies, such as smartphones and wearables, to overcome these limitations. This potential is primarily dependent on the wide adoption of mobile technologies, their monitoring and feedback capabilities, and the opportunity to provide support independent of time and location and in a personalized and adaptive manner. Given the potential of mobile technologies for the promotion of physical activity, the objective of this thesis was to determine whether interventions designed around mobile technologies can be an effective tool for the prevention of physical-activity-related NCDs, such as cardiovascular diseases and diabetes. To investigate the potential public health impact of these interventions, a novel physical activity program from a Swiss health insurance company that was based on mobile technologies and financial rewards was used as an example.

In a first step, necessary requirements that mobile physical activity interventions need to fulfill in order to serve as effective prevention tools were identified. Reach and efficacy were determined as key outcomes using the RE-AIM framework (Glasgow et al., 1999), and target effect sizes on both outcomes were derived from modelling the potential impact fraction of mobile physical activity interventions under different scenarios. The scenarios revealed that mobile physical activity interventions need to reach at least 10% of the target population and increase physical activity by at least 1,500 steps per day in order to effectively prevent incidence of the selected NCDs. In a next step, two extensive literature reviews were conducted to approximate the effects that are to be expected from the insurer’s physical activity program. These reviews revealed that the program’s main components can, in principle, increase physical activity by around 1,200 steps per day. However, the reach of the program is unclear, and the program’s financial incentives are several times smaller than those investigated in the literature, calling into question whether they can produce effects similar to those reported in previous studies. Subsequently, a first field study was conducted to estimate the program’s potential reach and examine the effects of its small financial incentives. In this study, a large dataset was collected, containing six months of real-world behavioral data from a pilot phase of the insurer’s physical activity promotion program. The study found that the program’s financial incentives significantly increased participation from around 3% to almost 6%, but did not affect participants’ physical activity. The program's requirement that participants own an activity tracker limited overall participation considerably. Participants' characteristics and 25% attrition during the six-month study suggested that the program does not reach those most in need of physical activity.

To address the limitations identified in the first study, the insurer's program was comprehensively revised. A smartphone app for monitoring physical activity (Ally) was developed in order to increase the program's reach and counteract selection effects. Additionally, the program's incentives were redesigned according to findings from previous research, and a scalable digital physical activity coaching was integrated into the app to support behavior change and prevent attrition. In line with the concept of optimization in the MOST framework (Collins, 2018), a second eight-week field study investigated both main and time-varying effects of incentives, as well as two components of the digital activity coaching: physical activity prompts and planning exercises. An impressive high-resolution dataset of behavioral and app usage data was collected as part of the second study. This study found that participants increased their physical activity by roughly 1,500 steps per day after starting to use the app. Of the investigated interventions, only financial incentives contributed to this overall effect. Lack of participant engagement emerged as a key barrier to digital coaching interventions and, again, selection effects were observed, as was over 30% attrition.

7.1.2 Contributions

This dissertation makes important contributions to different streams of the scientific literature. By reporting two empirical field studies which stand out due to limited participant exclusion criteria, minimal contact between study participants and the research team, and collection of real-world data, this dissertation answers earlier (Ammerman et al., 2014; Glasgow, Klesges, Dzewaltowski, Bull, & Estabrooks, 2004) and more recent (Collins, 2018; Koorts et al., 2018; J. Ma et al., 2018) calls for practice-embedded research and higher external validity of research on behavioral interventions. Importantly, by illustrating the limitations of scalable physical activity interventions in real-world settings, this dissertation damps the enthusiasm surrounding mobile health interventions and paves the way for more focused and relevant research that can drive the development and implementation of interventions with greater public health impact.

Further, the individual studies reported in this dissertation each make important contributions to the scientific literature around financial incentives for the promotion of physical activity. The study reported in chapter 4 was the first to investigate financial incentives that are small enough to be paid out continuously in practice thereby addressing a crucial research gap because incentive effects are typically not sustained after the incentive is withdrawn. Contrary to previous research on financial incentives, this study illustrated that the effects of small incentives are not sustained over time even when the incentives are in place. This suggests that an incentive threshold exists, i.e. a certain amount above which incentives start to produce sustained effects. Indeed, the revised incentives in the second study produced stable effects over the (short) study period, suggesting that this threshold likely is around an incentive value of CHF 1 per day in industrialized western countries. This is supported by previous studies on financial incentives and physical activity. In sum, the two studies provide an important evidence base that can guide other researchers and practitioners in designing and implementing financial incentives for the promotion of physical activity.

In the field of mobile health, the findings from the second study contribute to the understanding of engagement and attrition and complement research around just-in-time adaptive interventions (Nahum-Shani et al., 2016). Specifically, this study was the first to quantify the proportion of participants that effectively react to intervention prompts in a mobile health study and thereby highlighted the importance of just-in-time delivery of intervention-related push notifications. In fact, the study illustrated that delivering interventions at random points in time can undermine their effectiveness, because participants ignored the intervention prompts around 50% of the time. To the degree that these response rates can be generalized to other mobile health interventions, this finding illustrates that interventions delivered via push notifications may benefit only a subgroup of participants and that overall effect sizes of the corresponding interventions need to be interpreted with caution. Similarly, a related finding challenged the interpretation of intervention effects that are commonly reported in mobile health intervention studies, especially those relying on smartphone-based interventions. The pattern of results observed in the exploratory analyses suggested that overall effects of mobile physical activity interventions could reflect increases in wear time of the smartphone instead of increases in physical activity. Few studies in the field of mobile health, that evaluate physical activity apps, do appropriately control effect estimates for possible increases in wear time, for example by additionally measuring physical activity using an accelerometer.

Finally, this dissertation reported studies that were among the first to evaluate single components of mobile physical activity interventions thereby facilitating the understanding of how interventions produce behavior change. The detailed and specific insights obtained from both studies, such as effect sizes, interaction effects, attrition rates, motives for participation, response rates, participant feedback on intervention repetitiveness and burden interventions, is of great value for researchers and practitioners seeking to develop mobile physical activity interventions. On a more general level, the focus in this dissertation on the effects of single intervention components, instead of effects of complete interventions, contributes to understanding which intervention components work for which behaviors and which target groups and thereby helps to build a cumulative science of behavior change (Collins et al., 2014; Sumner et al., 2018).

7.2 Public Health Impact and Implications

7.2.1 Public Health Impact

Even though some positive effects were observed when evaluating mobile physical activity interventions, this dissertation illustrates that, at present, these interventions contribute little to the prevention of NCDs. While mobile physical activity interventions may, at least in the short term, increase physical activity to an extent that is sufficient to substantially reduce NCD risk, the two field studies identified three major barriers that limit the public health impact of mobile physical activity interventions considerably. These are:

1) Limited reach:

Although scalability is one of the key advantages of mobile physical activity interventions, the first study illustrated that uptake among the target population is insufficient for the effective prevention of NCDs. This is especially true for interventions designed around wearable devices because the requirement that participants own and continuously use a dedicated device for monitoring physical activity prevents many people from participating. Because mobile physical activity interventions rely on active participation to be effective, low intervention reach results in low public health impact.

2) Selection effects:

Those people who are reached by mobile physical activity interventions are not necessarily the ones who are at risk of developing NCDs. Participants in the two field studies were systematically better educated, healthier, better paid and more active than the general population. These factors are directly or indirectly related to a lower NCD risk. Because increases in physical activity have greater effects on NCD risk for inactive and at-risk individuals (cf. section 1.1), voluntary selection into health interventions is accompanied by reduced public health impact.

3) Attrition:

In both studies, a substantial proportion of participants dropped out and stopped monitoring their physical activity within a relatively short period of time. This effect appeared to be greater for interventions centered around smartphone apps than for wearable-based interventions, although attrition in both cases was substantial. Usage data from real-world settings demonstrated similar attrition rates for wearable devices (Ledger & McCaffrey, 2014) and even higher attrition rates for mobile health apps (Baumel et al., 2019) as compared to the attrition reported in this dissertation. As a consequence, mobile physical activity interventions only benefit a subset of participants over a longer period of time, which in turn limits their public health impact.

The identified barriers exemplify the conflict inherent in population-level prevention strategies with a high degree of individual agency (J. Adams et al., 2016). The scalability of mobile physical activity interventions is undermined by of the fact that they require active and sustained participation and engagement (cf. subsection 2.2.2). Although these barriers limit the contribution of mobile physical activity interventions to NCD prevention, the results of this dissertation have numerous implications for different stakeholders involved in the development and implementation of physical activity interventions. In fact, learnings from the two studies may contribute to overcoming some of the abovementioned barriers.

7.2.2 Implications

The results of this dissertation have important implications for intervention developers, practitioners, and researchers. In the following, each implication is stated and briefly elaborated in a separate paragraph.

Build multi-component interventions. Intervention developers need to have a clear plan of how their intervention produces sufficient effects on the dimensions reach and efficacy. As the conceptual models of this dissertation illustrate, this almost always requires multiple intervention components that affect outcomes via evidence-based mechanism of action. Specifically, simply relying on mobile technologies is insufficient in the general population (cf. systematic review of mobile physical activity interventions in section 3.3).

Optimize the intervention before conducting an RCT. One of the most important insights of this dissertation is the illustration of the value of intervention optimization. Evaluating either the insurer's physical activity program or the Ally app in an RCT would not only have resulted in more resource intensive empirical studies and results that are more difficult to interpret, but also in significantly fewer insights. For example, effect estimates and participant feedback regarding single components of the Ally app are of unparalleled value for their further development.

Remove barriers to participation. Researchers developing interventions tend to focus on the intervention's efficacy, but ensuring sufficient intervention reach is of equal importance. In fact, the way in which the target group is reached needs to be an explicit part of the intervention and its conceptual model. In particular, the intervention's characteristics must not limit its reach. For example, if wearable devices are part of the intervention these need to be made available at minimal or no cost. If a mobile app is part of the intervention, strategies need to be in place to guarantee awareness and motivation within the target group about downloading the app.

Develop a powerful value proposition. The results of this dissertation highlight that certain subgroups of participants are more attracted to and engage for longer and more frequently with mobile physical activity interventions. Essentially, these were subgroups for whom regular monitoring of step counts is of greater value and relevance, e.g. health-conscious, motivated and elderly individuals. This implies that, to overcome the limitations of mobile physical activity interventions, it is worth thinking about how the intervention provides (or does not provide) value for different target groups, a question that has been studied intensively in the business and marketing disciplines and is inherent in the concept of the value proposition. The value proposition, i.e. the promise of benefits that a customer will receive from a product, is a core element of the product development process, and is thought to be a key driver of adoption, user engagement and customer loyalty (Osterwalder, Pigneur, Bernarda, & Smith, 2014). Products with a strong value proposition help customers to achieve desired outcomes and reduce barriers that aggravate achievement of everyday goals (Osterwalder et al., 2014). Applied to mobile physical activity interventions in this dissertation, the value proposition is a parsimonious model that can explain participants' adoption, engagement, and attrition. An in-depth understanding of the target group is necessary to understand their desired outcomes, goals and barriers, and subsequently to derive value propositions and intervention components. A user-centered approach to intervention development is therefore recommended (Yardley, Morrison, Bradbury, & Muller, 2015). Developing a value proposition can help intervention developers get a better understanding of the target group, the target behavior, the context in which the behavior is (mostly) performed, and barriers that aggravate performing the behavior. It becomes apparent that, for mobile physical activity interventions, additional services and functionalities beyond monitoring and digital coaching might be necessary to create value for other target groups, e.g. high-risk individuals. This is currently mirrored by recent developments on the wearable device market, where successful companies have abandoned their hardware-reliant business models and have shifted to hybrid models that include additional services for consumers and enterprises (such as offering wholistic health coaching and corporate wellness programs; Muoio, 2019).

Target at-risk or diseased populations. Developers of mobile physical activity interventions can, at least in part, avoid the inherent tension of population-level and high agency interventions by adopting a high-risk strategy, or even by moving to the management of NCDs, where physical activity also plays a major role (Warburton, Nicol, & Bredin, 2006). While adopting a high-risk strategy has its own drawbacks (cf. section 2.2.2), it avoids the problems of selection effects, and limited reach is of lesser importance than in population-level strategies. Considering the identified limitations, mobile physical activity interventions are therefore more likely to produce a substantial public health impact in high-risk strategies. The situation is similar for interventions focusing on disease management although, admittedly, these interventions cannot contribute to primary prevention of NCDs. A compelling argument for the use of mobile technologies in disease management is the large effect sizes reported for mobile physical activity interventions in patient populations (cf. section 3.3).

Focus on temporary interventions whose effects are easy to maintain. An alternative strategy for intervention developers to approaching the identified limitations of mobile physical activity interventions is to develop intervention components that do not require continuous engagement and whose effects are comparatively easy for participants to maintain. Habit formation and skills training are two potential examples. One option for further improving the Ally app, for example, would be to offer action and coping planning exercises only for a limited time, with a focus on teaching participants how to plan efficiently, rather than guiding participants through the planning process in detail.

Use immediate, goal-contingent financial incentives with a value of at least CHF 1 per day. Financial incentives are an attractive intervention component because they can affect both the reach and the efficacy of the intervention. Based on the currently available evidence, including the studies within this dissertation, the minimum value for financial incentives to significantly change physical activity is around CHF 1 per day. Incentives with a lower value are unlikely to produce changes in physical activity, although they could possibly affect other outcomes, e.g. participation. The systematic reviews in section 3.4 demonstrated that the effects of financial incentives are maintained at three months after withdrawal but maintenance of effects for longer follow-up periods is unclear. Therefore, incentives should be offered without withdrawal if possible.

Understand intervention development as a software development project. The Ally study revealed that technical errors can limit the internal and external validity of study findings. In retrospect, key reasons for technical errors were lack of time and personnel, in addition to the high complexity and unpredictability inherent in software development. Using systematic methods to manage this high degree of complexity could avoid errors while allowing for efficient work that still meets the requirements of all stakeholders. Behavioral scientists need to familiarize themselves with the different roles and frameworks that are applied in software development (Schwaber & Beedle, 2002) to understand how best to collaborate and communicate with software developers. In addition, the necessary infrastructure, i.e. a powerful development team, needs to be in place. For this purpose, behavioral scientists can collaborate with computer scientists at their universities, approach IT service providers, or cooperate with businesses.

Target interpersonal and environmental determinants of behavior. By monitoring activity and providing feedback, mobile technologies are inherently focused on individual-level determinants of physical activity. Though important, initial changes in physical activity may not be sustained in settings that do not facilitate or even aggravate change (cf. section 2.3). For example, mobile technologies can communicate social norms or facilitate social support, e.g. via online communities or communication with health coaches and physicians. It is important that mobile physical activity interventions explicitly consider and leverage those determinants in order to produce sustainable effects.

7.3 Limitations and Future Work

Every scientific work comes with limitations and this dissertation is no exception. This section outlines the most important limitations of the present work. Subsequently, opportunities for future work are discussed, some of which directly relate to the implications and limitations presented in this chapter.

7.3.1 Limitations

The following paragraphs discuss the more general limitations of this dissertation, which primarily relate to its overall objective. Specific limitations pertaining to the literature reviews that are part of this dissertation have been discussed in the subsections 3.3.4 and 3.4.4, and specific limitations of the two field studies have been discussed in sections 4.4 and 6.4, respectively. To avoid redundancies, these limitations are not repeated here. The general limitations of this dissertation can be classified into three main categories: limitations regarding quantification of public health impact, limitations inherent in the intervention development process, and the similarity of the investigated mobile physical activity interventions.

First, although the overall objective of this dissertation was to evaluate the potential public health impact of mobile physical activity interventions, it was not possible to obtain a direct measure of public health impact, such as reductions in disease incidence resulting from an exposure to a mobile physical activity intervention. Clearly, obtaining such direct measures of public health impact was not considered feasible with the available resources. Instead, the public health impact of mobile physical activity interventions was approximated in this dissertation using the dimensions reach and efficacy of the RE-AIM framework (Glasgow et al., 1999) and by defining lower effect thresholds that rely on calculating the potential impact fraction (cf. subsection 2.2.3) resulting from an intervention's reach and efficacy. While this approximation of public health impact is a plausible alternative, some limitations have to be noted. Most importantly, perhaps, the lower effect thresholds applied in this dissertation assume long-term maintenance of intervention effects which was not investigated in this dissertation. They also rely on physical inactivity prevalence estimates from Switzerland

and average estimates of the strength of the relationship between physical activity and disease obtained from meta-analyses. Both parameters likely vary between different countries, for example due to differences in urbanization, employment, age, health care quality and health care utilization. On the one hand, this illustrates that, in addition to the factors considered in this dissertation, the public health impact of mobile physical activity interventions depends on the setting where the intervention is implemented. On the other hand, this demonstrates the number of assumptions involved in the calculation of the lower effect thresholds that provided the basis for evaluating the mobile physical activity interventions in this dissertation. As a consequence, it is possible that one draws different conclusions when evaluating mobile physical activity interventions with direct and objective measures of public health impact.

Second, the process of intervention development that motivated the two field studies in this dissertation broadly followed the MOST framework (Collins, 2018). As such, this dissertation shares some of its limitations. For example, the MOST framework begins the intervention development process by establishing the conceptual model of the intervention's target behavior based on theory and available empirical evidence. As briefly discussed in section 2.3, information from these sources, and thus the conceptual models, tends to be rather general and might only apply to a limited extent to the situation and problem at hand. As a consequence, the two studied interventions make some implicit assumptions that may not necessarily be correct, e.g. that walking is an appropriate target behavior for most participants, and that self-regulation and outcome expectancies are core behavioral determinants in the target group. An elaborate engagement with the target group to identify its most prominent strengths, characteristics, and problems is missing in the MOST framework. In line with the implications mentioned above, explicitly interviewing or surveying members of the target group could help to develop interventions that create more value and are better accepted. Other intervention development approaches, such as the intervention mapping protocol (Bartholomew, Parcel, Kok, & Gottlieb, 2006), more explicitly highlight this so-called need assessment phase in intervention development. Therefore, it is possible that some of the limitations identified in this dissertation and attributed to mobile technologies are, at least in part, affected by the intervention development process.

Third, although external validity has been mentioned as a key strength of this dissertation, one aspect of external validity may actually limit the conclusions that can be drawn from the reported studies. This aspect refers to the generalizability of the results to different types of mobile physical activity interventions. Although the two field studies investigated different mobile physical activity interventions that leveraged different components and different mobile technologies, both interventions were similar in nature and both partly targeted the same mechanisms of action. As a consequence, the results of this dissertation may not generalize well to mobile physical activity interventions with substantially different components or characteristics. Similarly, one could argue that some important benefits of mobile technologies, such as the adaptive and context-aware delivery of interventions, have not been realized in the interventions studied in this dissertation. It is possible that interventions fully realizing the potential of mobile technologies are not subject to the same limitations as the interventions investigated in the two field studies. Nevertheless, the limited engagement and high attrition found in the interventions in this study confirm a pattern that has begun to emerge in the broader mobile health literature. This suggests that, at least today, the identified limitations pertain to mobile interventions in general irrespective of components and other characteristics.

Finally, it must be noted that this dissertation examined mobile physical activity interventions from a public health perspective. Specifically, it focused on the potential public health impact of mobile physical activity interventions. It is important to mention that this is not a wholistic perspective and public health impact is not sufficient to determine the overall value of mobile physical activity interventions. For example, mobile physical activity interventions could have substantial effects on well-being (Karapanos, Gouveia, Hassenzahl, & Forlizzi, 2016) or on selected clinical outcomes (Hou, Carter, Hewitt, Francisa, & Mayor, 2016) that may not translate into public health impact. This dissertation merely assessed the value that mobile physical activity interventions hold for the effective prevention of NCDs.

7.3.2 Future Work

This dissertation revealed several opportunities for future research to advance science around mobile physical activity interventions and in turn to contribute to maximizing their public health impact. First and foremost, research is urgently needed to facilitate a comprehensive understanding of reach, engagement, and attrition in mobile health interventions. To overcome the current limitations of mobile health interventions, this research needs to provide answers to the following research questions: What are viable strategies for increasing the uptake of mobile physical activity interventions? What prevents individuals at higher disease risk from accessing mobile physical activity interventions and how can these barriers be overcome? How can participants in mobile physical activity interventions be engaged in the long-term? Recently, some researchers have questioned whether maximizing engagement in digital interventions is actually desirable and have instead coined the phrase “effective engagement” (Yardley et al., 2016), i.e. the level of engagement that is sufficient to produce lasting behavior change. In line with this reasoning, it is crucial to investigate what level of engagement constitutes effective engagement for which intervention component.

Another important point to address in future work refers to the monitoring of physical activity. While current mobile technologies, and smartphones in particular, are capable of monitoring steps per day, a novel way of ubiquitously monitoring physical activity of any kind using a simple and understandable metric is highly desirable. First, participants in both field studies expressed the need for such a metric, reflecting the fact that steps per day only account for some subtypes (e.g. walking, jogging) of physical activity. Second, recent research has questioned the value of measuring step counts for health promotion. Although earlier research has demonstrated substantial health benefits of walking (Murphy et al., 2007) independent of total physical activity levels (Kelly et al., 2014), a recent large randomized study suggests that increases in step counts do not necessarily translate into health benefits (Finkelstein et al., 2016). A possible explanation for this finding is that physical activity has to be performed at a certain intensity (i.e. moderate-to-vigorous physical activity or more than 3 METs) to affect health. Steps, however, are counted regardless of the intensity of walking. Thus, a digital biomarker for exercise intensity would enable interventions that promote more meaningful behavior changes. A promising example is the activity score reported by Nes and colleagues who derived a personalized and intensity-sensitive activity score based on individuals’ heart rate data. The authors subsequently demonstrated that achieving a score above 100 was associated with a 17% to 23% reduction in cardiovascular disease mortality risk (Nes, Gutvik, Lavie, Nauman, & Wisløff, 2017).

Although research in the area of mobile physical activity interventions is growing, the field of mobile health is still in its infancy. This is illustrated, for example, by the predominantly small-scale and pilot studies that have been included in the existing reviews of mobile physical activity interventions (section 3.3). While larger randomized studies with longer follow-up periods are clearly needed, the field of mobile health should carefully avoid the mistakes made in the intervention and behavioral sciences over the past years, i.e. focusing too much on intervention efficacy and RCTs. As illustrated in this dissertation, it is well worth complementing RCTs with building conceptual models and optimization trials and focusing on practice-embedded research that evaluates all aspects of an intervention that can determine its public health impact. In addition to this emphasis on understanding the effects of an intervention and its components, future research needs to investigate for whom mobile physical activity interventions are most effective – and for whom they are not. For example, researchers have pointed out that constant monitoring (as in mobile physical activity interventions) draws attention to the output of behavior and can make performing the behavior feel less enjoyable (Etkin, 2016). In fact, some people tend to actively avoid monitoring behavior to protect their positive self-view in the case of failure (Kangovi & Asch, 2018), a phenomenon known as the “ostrich problem” (Webb, Chang, & Benn, 2013).

Finally, and in line with the points above, future works needs to refine the theories of health behavior change. Although the general limitations of existing health behavior change theories have already been discussed (section 2.3), it is worth pointing out that novel theories are required to specifically take into account the questions that arise during the process of mobile health intervention development. Beyond specifying behavioral determinants and their effects on the individual, interpersonal and environmental levels, this includes hypotheses about time-varying states (e.g. fluctuations in cognitive load or motivation), time-invariant characteristics (e.g. age, socio-economic status), and contextual factors (e.g. location, weather) that moderate relationships between determinants and outcomes (Hekler et al., 2016; Riley et al., 2011). The present work has illustrated, for example, that the success of mobile health interventions can ultimately depend on whether participants are willing and able to engage with the intervention content. More elaborate theoretical models could inform the development of adaptive mobile interventions that leverage information on effect moderators to deliver interventions when they have the greatest probability of actually changing behavior (Nahum-Shani et al., 2016). Similarly, health behavior change theories have to become more explicit about the time-scales of the included variables and relationships (Scholz, 2019; Spruijt-Metz & Nilsen, 2014), because mobile interventions often target outcomes on smaller time-scales than traditional interventions, and different behavioral determinants may operate on different time-scales. For example, while absolute values of aggregated step counts are driven by biological factors and health indicators (such as age or body mass index), daily step count distributions show peaks in the morning, during lunch time and in the evening, suggesting that on smaller time-scales physical activity is driven greatly by daily routines (Althoff et al., 2017).

7.4 Conclusion

Mobile technologies, such as smartphones, smart watches and wearable devices, have had a profound impact on many aspects of our lives, from the way we communicate to the way we conduct business. This dissertation was intended to investigate the potential that mobile technologies can have for health, undeniably one of the most important aspects of our lives. In particular, this dissertation looked at the potential that large-scale mobile physical activity interventions have for the effective prevention of NCDs, one of the most pressing challenges in healthcare today.

This dissertation's findings illustrate that work is still needed for mobile physical activity interventions to become powerful, population-level prevention strategies. To date, limited reach, selection effects, and attrition prevent mobile physical activity interventions from having substantial public health impact. These barriers exemplify the tension between these interventions' goals of reaching the majority of their target population and the simultaneous requirement for members of this target population to actively participate and continuously engage with the intervention. The consequence of this tension is that only a small subgroup within the target population actually benefits. Although, of course, the benefits for this subgroup are favorable, the observed pattern of participation, engagement and attrition undermines the interventions' overall public health impact. The reasons for the existence of the abovementioned barriers are manifold and relate not only to the way these interventions are offered (e.g. requiring investment in mobile devices) but also to the way these interventions are designed (e.g. with a strong focus on self-regulation and feedback).

Despite these limitations, the importance of mobile technology for public health and prevention cannot be overstated. In comparison to traditional prevention approaches, mobile technologies have fundamentally changed the way interventions are designed, have greatly simplified the way they can be accessed, and have offered new opportunities to apply potentially more powerful behavior change strategies. But most importantly, perhaps, mobile physical activity interventions are increasingly adopted outside of research settings. Although many of these existing physical activity programs are not (yet) evidence-based, their existence illustrates that mobile technologies can bridge the evidence-practice gap, one of the greatest challenges for traditional physical activity interventions. There is no doubt, therefore, that mobile physical activity interventions are here to stay. Whether they will contribute to disease prevention will depend on a number of factors. Specifically, mobile physical activity interventions have to be easily accessible for high-risk populations and must leverage multiple evidence-based components that target determinants on different levels, to produce changes in physical activity that can actually be maintained. Further, just like other smartphone apps and digital products, they have to provide clear value to those populations in order to be able to facilitate uptake and engagement. For individuals at high risk of disease this may go beyond monitoring and feedback on physical activity.

References

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health Psychology, 27*(3), 379-387.

Adams, J., Giles, E. L., McColl, E., & Sniehotta, F. F. (2014). Carrots, sticks and health behaviours: a framework for documenting the complexity of financial incentive interventions to change health behaviours. *Health Psychology Review, 8*(3), 286-295.

Adams, J., Mytton, O., White, M., & Monsivais, P. (2016). Why are some population interventions for diet and obesity more equitable and effective than others? The role of individual agency. *PLoS Medicine, 13*(4), e1001990.

Adams, M. A., Hurley, J. C., Todd, M., Bhuiyan, N., Jarrett, C. L., Tucker, W. J., . . . Angadi, S. S. (2017). Adaptive goal setting and financial incentives: a 2×2 factorial randomized controlled trial to increase adults' physical activity. *BMC Public Health, 17*(1), 286.

Adams, M. A., Sallis, J. F., Norman, G. J., Hovell, M. F., Hekler, E. B., & Perata, E. (2013). An adaptive physical activity intervention for overweight adults: a randomized controlled trial. *PloS One, 8*(12), e82901.

Adriaanse, M. A., Vinkers, C. D., De Ridder, D. T., Hox, J. J., & De Wit, J. B. (2011). Do implementation intentions help to eat a healthy diet? A systematic review and meta-analysis of the empirical evidence. *Appetite, 56*(1), 183-193.

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett Jr, D. R., Tudor-Locke, C., . . . Leon, A. S. (2011). 2011 Compendium of Physical Activities: a second update of codes and MET values. *Medicine and Science in Sports and Exercise, 43*(8), 1575-1581.

Aitken, M., Clancy, B. & Nass, D. (2017). *The Growing Value of Digital Health*. Retrieved from Parsippany, NJ: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control* (pp. 11-39). Heidelberg: Springer.

Althoff, T., Sosič, R., Hicks, J. L., King, A. C., Delp, S. L., & Leskovec, J. (2017). Large-scale physical activity data reveal worldwide activity inequality. *Nature, 547*, 336-339. Retrieved from https://doi.org/10.1038/nature23018

Ammerman, A., Smith, T. W., & Calancie, L. (2014). Practice-based evidence in public health: improving reach, relevance, and results. *Annual Review of Public Health, 35*, 47-63.

Andersson, G., & Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cognitive Behavior Therapy, 38*(4), 196-205.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal, 100*(401), 464-477.

Araiza, P., Hewes, H., Gashetewa, C., Vella, C. A., & Burge, M. R. (2006). Efficacy of a pedometer-based physical activity program on parameters of diabetes control in type 2 diabetes mellitus. *Metabolism: Clinical and Experimental, 55*(10), 1382-1387. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-33748455432&doi=10.1016%2fj.metabol.2006.06.009&partnerID=40&md5=bc5196f2bc5b8aaf1ab263c82902562e

Ashton, L. M., Hutchesson, M. J., Rollo, M. E., Morgan, P. J., & Collins, C. E. (2017). Motivators and barriers to engaging in healthy eating and physical activity: A cross-sectional survey in young adult men. *American journal of men's health, 11*(2), 330-343.

Aune, D., Norat, T., Leitzmann, M., Tonstad, S., & Vatten, L. J. (2015). Physical activity and the risk of type 2 diabetes: a systematic review and dose–response meta-analysis. *European Journal of Epidemiology, 30*(7), 529-542. Retrieved from https://doi.org/10.1007/s10654-015-0056-z

Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., . . . Norris, S. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology, 64*(4), 401-406.

Bandura, A. (1986). *Social Foundations of Thought and Action*. Inglewood Cliffs, NJ: Prentice Hall.

Barata, F., Kowatsch, T., Tinschert, P., & Filler, A. (2016). *Personal MobileCoach: tailoring behavioral interventions to the needs of individual participants.* Paper presented at the ACM International Joint Conference on Pervasive and Ubiquitous Computing.

Barte, J. C., & Wendel-Vos, G. W. (2017). A systematic review of financial incentives for physical activity: the effects on physical activity and related outcomes. *Behavioral Medicine, 43*(2), 79-90.

Bartholomew, L. K., Parcel, G. S., Kok, G., & Gottlieb, N. H. (2006). *Planning health promotion programs: An intervention mapping approach*. San Francisco, CA: John Wiley & Sons.

Bassett Jr, D. R., Wyatt, H. R., Thompson, H., Peters, J. C., & Hill, J. O. (2010). Pedometer-measured physical activity and health behaviors in United States adults. *Medicine and Science in Sports and Exercise, 42*(10), 1819-1825.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bauman, A. E., Reis, R. S., Sallis, J. F., Wells, J. C., Loos, R. J., Martin, B. W., & Group, L. P. A. S. W. (2012). Correlates of physical activity: why are some people physically active and others not? *The Lancet, 380*(9838), 258-271.

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. *Journal of Medical Internet Research, 21*(9), e14567. Retrieved from http://www.jmir.org/2019/9/e14567/

Becker, L. A., & Oxman, A. D. (2008). Overviews of reviews. In J. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 607-631). Chichester, UK: John Wiley & Sons.

Beets, M. W., Beighle, A., Erwin, H. E., & Huberty, J. L. (2009). After-school program impact on physical activity and fitness: a meta-analysis. *American Journal of Preventive Medicine, 36*(6), 527-537.

Bélanger-Gravel, A., Godin, G., & Amireault, S. (2013). A meta-analytic review of the effect of implementation intentions on physical activity. *Health Psychology Review, 7*(1), 23-54.

Bell, R. A., Cholerton, M., Fraczek, K. E., Rohlfs, G. S., & Smith, B. A. (1994). Encouraging donations to charity: A field study of competing and complementary factors in tactic sequencing. *Western Journal of Communication, 58*(2), 98-115.

Beun, R. J., Fitrianie, S., Griffioen-Both, F., Spruit, S., Horsch, C., Lancee, J., & Brinkman, W.-P. (2017). Talk and Tools: the best of both worlds in mobile user interfaces for E-coaching. *Personal and ubiquitous computing, 21*(4), 661-674.

Bickmore, T., & Cassell, J. (2005). Social Dialongue with Embodied Conversational Agents. In J. C. J. van Kuppevelt, L. Dybkjær, & N. O. Bernsen (Eds.), *Advances in Natural Multimodal Dialogue Systems* (pp. 23-54). Dordrecht, NL: Springer.

Bickmore, T., Gruber, A., & Picard, R. (2005). Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Education and Counseling, 59*(1), 21-30.

Bickmore, T., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(2), 293-327.

Bickmore, T., Schulman, D., & Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence, 24*(6), 648-666.

Blackman, K. C., Zoellner, J., Berrey, L. M., Alexander, R., Fanning, J., Hill, J. L., & Estabrooks, P. A. (2013). Assessing the Internal and External Validity of Mobile Health Physical Activity Promotion Interventions: A Systematic Literature Review Using the RE-AIM Framework. *Journal of Medical Internet Research, 15*(10), e224. Retrieved from http://www.jmir.org/2013/10/e224/

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons.

Boruvka, A., Almirall, D., Witkiewitz, K., & Murphy, S. A. (2017). Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association, 113*, 1112-1121.

Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., . . . Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health: a systematic review. *JAMA, 298*(19), 2296-2304.

Brewer, N. T., & Rimer, B. K. (2008). Perspectives on Health Behavior Theories that Focus on Individuals. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health Behavior and Health Education* (pp. 149-165). San Francisco, CA: John Wiley & Sons.

Brickwood, K.-J., Watson, G., O'Brien, J., & Williams, A. D. (2019). Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. *JMIR mHealth and uHealth, 7*(4), e11819.

Brown, W. J., Eakin, E., Mummery, K., & Trost, S. G. (2003). 10,000 Steps Rockhampton: Establishing a multi-strategy physical activity promotion project in a community. *Health Promotion Journal of Australia, 14*(2), 95-100.

Bundesamt für Statistik. (2016). Häufigkeitsverteilung (monatlicher Nettolohn), Voll- und Teilzeitbeschäftigte nach Lohnhöhenklassen - Privater und öffentlicher Sektor zusammen - Schweiz. Retrieved from https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/tabellen.assetdetail.327851.html

Bundesamt für Statistik. (2018a). Schweizerische Gesundheitsbefragung 2017.

Bundesamt für Statistik. (2018b). Ständige Wohnbevölkerung ab 15 Jahren nach höchster abgeschlossener Ausbildung, 2016. Retrieved from https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/tabellen.assetdetail.4242913.html

Burke, L. E., Ma, J., Azar, K. M., Bennett, G. G., Peterson, E. D., Zheng, Y., . . . Suffoletto, B. (2015). Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. *Circulation, 132*(12), 1157-1213.

BVDW. (2016). *DACH-Studie 2016 - Digitale Mediennutzung in der DACH-Region 2016*. Retrieved from https://www.bvdw.org/fileadmin/bvdw/upload/publikationen/mobile/DACH-Studie_2016.pdf

Carey, R. N., Connell, L. E., Johnston, M., Rothman, A. J., de Bruin, M., Kelly, M. P., & Michie, S. (2018). Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. *Annals of Behavioral Medicine, 53*(8), 693-707.

Carraro, N., & Gaudreau, P. (2013). Spontaneous and experimentally induced action planning and coping planning for physical activity: A meta-analysis. *Psychology of Sport and Exercise, 14*(2), 228-248.

Carroll, J. K., Moorhead, A., Bond, R., LeBlanc, W. G., Petrella, R. J., & Fiscella, K. (2017). Who uses mobile phone health apps and does use matter? A secondary data analytics approach. *Journal of Medical Internet Research, 19*(4), e125.

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin, 92*(1), 111-135.

Case, M. A., Burwick, H. A., Volpp, K. G., & Patel, M. S. (2015). Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA, 313*(6), 625-626.

Cerin, E., Leslie, E., Sugiyama, T., & Owen, N. (2010). Perceived barriers to leisure-time physical activity in adults: an ecological perspective. *Journal of Physical Activity and Health, 7*(4), 451-459.

Choi, K. W., Chen, C.-Y., Stein, M. B., Klimentidis, Y. C., Wang, M.-J., Koenen, K. C., & Smoller, J. W. (2019). Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample mendelian randomization study. *JAMA psychiatry, 76*(4), 399-408.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591-621.

Cohen, J. (1988). *Statistical power analysis for the social sciences* (2nd ed.). Hillsdale, N.J.: Erlbaum.

Collins, L. M. (2018). *Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST)*. New York, NY: Springer.

Collins, L. M., Trail, J. B., Kugler, K. C., Baker, T. B., Piper, M. E., & Mermelstein, R. J. (2014). Evaluating individual intervention components: making decisions based on the results of a factorial screening experiment. *Translational Behavioral Medicine, 4*(3), 238-251.

Comstock, J. (2014). Humana: Engaged Vitality users had lower costs, fewer absences. Retrieved from https://www.mobihealthnews.com/32786/humana-engaged-vitality-users-had-lower-costs-fewer-absences

Conn, V. S., Hafdahl, A. R., & Mehr, D. R. (2011). Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. *American Journal of Public Health, 101*(4), 751-758.

Conn, V. S., Valentine, J. C., & Cooper, H. M. (2002). Interventions to increase physical activity among aging adults: a meta-analysis. *Annals of Behavioral Medicine, 24*(3), 190-200.

Cottrell, E., Chambers, R., & O'Connell, P. (2012). Using simple telehealth in primary care to reduce blood pressure: a service evaluation. *BMJ Open, 2*(6), e001391. Retrieved from https://bmjopen.bmj.com/content/bmjopen/2/6/e001391.full.pdf

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Pratt, M., . . . Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise, 35*, 1381-1395.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ, 337*, a1655.

Davis, R., Campbell, R., Hildon, Z., Hobbs, L., & Michie, S. (2015). Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health Psychology Review, 9*(3), 323-344.

De Cocker, K. A., De Bourdeaudhuij, I. M., Brown, W. J., & Cardon, G. M. (2007). Effects of "10,000 steps Ghent": a whole-community intervention. *American Journal of Preventive Medicine, 33*(6), 455-463.

de Vries, H. J., Kooiman, T. J., van Ittersum, M. W., van Brussel, M., & de Groot, M. (2016). Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. *Obesity, 24*(10), 2078-2091.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627-668.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York, NY: Plenum.

Deng, T., Wang, Y., Sun, M., & Chen, B. (2013). Stage-matched intervention for adherence to CPAP in patients with obstructive sleep apnea: a randomized controlled trial. *Sleep and Breathing, 17*(2), 791-801.

Di Blasi, Z., Harkness, E., Ernst, E., Georgiou, A., & Kleijnen, J. (2001). Influence of context effects on health outcomes: a systematic review. *The Lancet, 357*(9258), 757-762.

Dimensions. (2019). Re-imagining discovery and access to research. Retrieved from https://www.dimensions.ai/

Ding, D., Lawson, K. D., Kolbe-Alexander, T. L., Finkelstein, E. A., Katzmarzyk, P. T., van Mechelen, W., . . . Committee, L. P. A. S. E. (2016). The economic burden of physical inactivity: a global analysis of major non-communicable diseases. *Lancet, 388*(10051), 1311-1324.

Direito, A., Carraça, E., Rawstorn, J., Whittaker, R., & Maddison, R. (2016). mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. *Annals of Behavioral Medicine, 51*(2), 226-239.

Dmitrienko, A., & D'Agostino Sr, R. (2013). Traditional multiplicity adjustment methods in clinical trials. *Statistics in Medicine, 32*(29), 5172-5218.

Domjan, M. (1998). *The principles of learning and behavior*. Belmont, CA: Thomson Brooks/Cole Publishing Co.

Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. London, UK: Arnold.

Dorsey, E. R., McConnell, M. V., Shaw, S. Y., Trister, A. D., & Friend, S. H. (2017). The use of smartphones for health research. *Academic Medicine, 92*(2), 157-160.

Duan, Y. P., Wienert, J., Hu, C., Si, G. Y., & Lippke, S. (2017). Web-based intervention for physical activity and fruit and vegetable intake among Chinese university students: a randomized controlled trial. *Journal of Medical Internet Research, 19*(4), e106.

Duncan, M. J., Wunderlich, K., Zhao, Y., & Faulkner, G. (2017). Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. *Journal of Sports Sciences, 36*, 1695-1704.

Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008). Spending money on others promotes happiness. *Science, 319*(5870), 1687-1688.

Dwyer, T., Ponsonby, A.-L., Ukoumunne, O. C., Pezic, A., Venn, A., Dunstan, D., . . . Zimmet, P. (2011). Association of change in daily step count over five years with insulin sensitivity and adiposity: population based cohort study. *BMJ, 342*, c7249.

Ebrahim, S., Taylor, F., Ward, K., Beswick, A., Burke, M., & Smith, G. D. (2011). Multiple risk factor interventions for primary prevention of coronary heart disease. *Cochrane database of systematic Reviews*, CD001561.

Eisler, R., & Lüber, A. (2006). Wie wichtig ist den Schweizern eine Spitalzusatzversicherung?

eMarketer. (2017). Denmark 'has highest smartphone penetration rate in the world'. Retrieved from http://www.netimperative.com/2017/12/denmark-highest-smartphone-penetration-rate-world/

Ernst & Young. (2017). Kasse für Kranke? Oder Partner für Gesundheit? Eine strategische Analyse des Schweizer Krankenversicherungsmarkts.

Etkin, J. (2016). The hidden cost of personal quantification. *Journal of Consumer Research, 42*(6), 967-984.

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

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Feehan, L. M., Geldman, J., Sayre, E. C., Park, C., Ezzat, A. M., Yoo, J. Y., . . . Li, L. C. (2018). Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. *JMIR Mhealth Uhealth, 6*(8), e10527. Retrieved from http://mhealth.jmir.org/2018/8/e10527/

Filler, A., Kowatsch, T., Haug, S., Wahle, F., Staake, T., & Fleisch, E. (2015). *MobileCoach: A Novel Open Source Platform for the Design of Evidence-based, Scalable and Low-Cost Behavioral Health Interventions - Overview and Preliminary Evaluation in the Public Health Context*. Paper presented at the 14th annual Wireless Telecommunications Symposium (WTS 2015), New York, USA.

Finkelstein, E. A., Haaland, B. A., Bilger, M., Sahasranaman, A., Sloan, R. A., Nang, E. E. K., & Evenson, K. R. (2016). Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. *Lancet Diabetes and Endocrinology, 4*(12), 983-995.

Fisher, J. D., & Fisher, W. A. (1992). Changing AIDS-risk behavior. *Psychological Bulletin, 111*(3), 455-474.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health, 4*(2), e19.

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15*(5), 451-474.

Flückiger, C., Del Re, A., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55*(4), 316-340.

Fogg, B. J. (2009). *A behavior model for persuasive design.* Paper presented at the 4th international Conference on Persuasive Technology, Claremont, CA.

Foster, C., Hillsdon, M., Thorogood, M., Kaur, A., & Wedatilake, T. (2005). Interventions for promoting physical activity. *Cochrane database of systematic Reviews, 1*, Art. No. CD003180.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of Personality and Social Psychology, 4*(2), 195-202.

Friederichs, S., Bolman, C., Oenema, A., Guyaux, J., & Lechner, L. (2014). Motivational Interviewing in a Web-Based Physical Activity Intervention With an Avatar: Randomized Controlled Trial.

Journal of Medical Internet Research, 16(2), e48. Retrieved from http://www.jmir.org/2014/2/e48/

FSG. (2017). Insuring Shared Value.

Gaglio, B., Shoup, J. A., & Glasgow, R. E. (2013). The RE-AIM framework: a systematic review of use over time. *American Journal of Public Health, 103*(6), e38-e46.

Gal, R., May, A. M., van Overmeeren, E. J., Simons, M., & Monninkhof, E. M. (2018). The Effect of Physical Activity Interventions Comprising Wearables and Smartphone Applications on Physical Activity: a Systematic Review and Meta-analysis. *Sports medicine-open, 4*(1), 42.

Gao, F., Earnest, A., Matchar, D. B., Campbell, M. J., & Machin, D. (2015). Sample size calculations for the design of cluster randomized trials: A summary of methodology. *Contemporary Clinical Trials, 42*, 41-50.

Gardner, B., de Bruijn, G.-J., & Lally, P. (2011). A systematic review and meta-analysis of applications of the self-report habit index to nutrition and physical activity behaviours. *Annals of Behavioral Medicine, 42*(2), 174-187.

Giles-Corti, B., & Donovan, R. J. (2002). The relative influence of individual, social and physical environment determinants of physical activity. *Social Science and Medicine, 54*(12), 1793-1812.

Giles-Corti, B., & Donovan, R. J. (2003). Relative influences of individual, social environmental, and physical environmental correlates of walking. *American Journal of Public Health, 93*(9), 1583-1589.

Giles, E. L., Robalino, S., McColl, E., Sniehotta, F. F., & Adams, J. (2014). The effectiveness of financial incentives for health behaviour change: systematic review and meta-analysis. *PloS One, 9*(3), e90347.

Giltay, E. J., Geleijnse, J. M., Zitman, F. G., Buijsse, B., & Kromhout, D. (2007). Lifestyle and dietary correlates of dispositional optimism in men: The Zutphen Elderly Study. *Journal of Psychosomatic Research, 63*(5), 483-490.

Girandola, F. (2002). Sequential Requests and Organ Donation. *The Journal of Social Psychology, 142*(2), 171-178. Retrieved from https://doi.org/10.1080/00224540209603893

Glanz, K., Rimer, B. K., & Viswanath, K. (2008). Theory, Research, and Practice in Health Behavior and Health Education. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health Behavior and Health Education* (pp. 23-40). San Francisco, CA: John Wiley & Sons.

Glasgow, R. E., Klesges, L. M., Dzewaltowski, D. A., Bull, S. S., & Estabrooks, P. (2004). The future of health behavior change research: what is needed to improve translation of research into health promotion practice? *Annals of Behavioral Medicine, 27*(1), 3-12.

Glasgow, R. E., Klesges, L. M., Dzewaltowski, D. A., Estabrooks, P. A., & Vogt, T. M. (2006). Evaluating the impact of health promotion programs: using the RE-AIM framework to form summary measures for decision making involving complex issues. *Health Education Research, 21*(5), 688-694.

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health, 93*(8), 1261-1267.

Glasgow, R. E., McCaul, K. D., & Fisher, K. J. (1993). Participation in worksite health promotion: a critique of the literature and recommendations for future practice. *Health Education Quarterly, 20*(3), 391-408.

Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *American Journal of Public Health, 89*(9), 1322-1327.

Glynn, L. G., Hayes, P. S., Casey, M., Glynn, F., Alvarez-Iglesias, A., Newell, J., . . . Murphy, A. W. (2014). Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. *British Journal of General Practice, 64*(624), e384-e391.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54*(7), 493-503.

Gong, Y., Trentadue, T. P., Shrestha, S., Losina, E., & Collins, J. E. (2018). Financial incentives for objectively-measured physical activity or weight loss in adults with chronic health conditions: A meta-analysis. *PloS One, 13*(9), e0203939. Retrieved from https://doi.org/10.1371/journal.pone.0203939

Google. (2013). Android KitKat. Retrieved from https://developer.android.com/about/versions/kitkat

Gourlan, M., Trouilloud, D., & Sarrazin, P. (2011). Interventions promoting physical activity among obese populations: a meta-analysis considering global effect, long-term maintenance, physical activity indicators and dose characteristics. *Obesity Reviews, 12*(7), e633-e645.

Green, L. W. (2008). Making research relevant: if it is an evidence-based practice, where's the practice-based evidence? *Family Practice, 25*(suppl_1), i20-i24.

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and emotion, 24*(3), 175-213.

Guéguen, N., Marchand, M., Pascual, A., & Lourel, M. (2008). Foot-in-the-Door Technique Using a Courtship Request: A Field Experiment. *Psychological Reports, 103*(2), 529-534. Retrieved from https://doi.org/10.2466/pr0.103.2.529-534

Guyatt, G. H., Oxman, A. D., Kunz, R., Brozek, J., Alonso-Coello, P., Rind, D., . . . Vist, G. (2011). GRADE guidelines 6. Rating the quality of evidence—imprecision. *Journal of Clinical Epidemiology, 64*(12), 1283-1293.

Guyatt, G. H., Oxman, A. D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., . . . Vist, G. (2011). GRADE guidelines: 8. Rating the quality of evidence—indirectness. *Journal of Clinical Epidemiology, 64*(12), 1303-1310.

Guyatt, G. H., Oxman, A. D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., . . . Akl, E. A. (2011). GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *Journal of Clinical Epidemiology, 64*(12), 1294-1302.

Guyatt, G. H., Oxman, A. D., Montori, V., Vist, G., Kunz, R., Brozek, J., . . . Falck-Ytter, Y. (2011). GRADE guidelines: 5. Rating the quality of evidence—publication bias. *Journal of Clinical Epidemiology, 64*(12), 1277-1282.

Guyatt, G. H., Oxman, A. D., Vist, G., Kunz, R., Brozek, J., Alonso-Coello, P., . . . Falck-Ytter, Y. (2011). GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *Journal of Clinical Epidemiology, 64*(4), 407-415.

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ, 336*(7650), 924-926.

Hall, P. A., & Fong, G. T. (2007). Temporal self-regulation theory: A model for individual health behavior. *Health Psychology Review, 1*(1), 6-52.

Harkin, B., Webb, T. L., Chang, B. P., Prestwich, A., Conner, M., Kellar, I., . . . Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin, 14*(2), 198-229.

Harkins, K. A., Kullgren, J. T., Bellamy, S. L., Karlawish, J., & Glanz, K. (2017). A Trial of Financial and Social Incentives to Increase Older Adults' Walking. *American Journal of Preventive Medicine, 52*(5), e123-e130.

Harris, T., Kerry, S. M., Limb, E. S., Victor, C. R., Iliffe, S., Ussher, M., . . . Furness, C. (2017). Effect of a primary care walking intervention with and without nurse support on physical activity levels in 45-to 75-year-olds: The Pedometer And Consultation Evaluation (PACE-UP) cluster randomised clinical trial. *PLoS Medicine, 14*(1), e1002210.

Haug, S., Paz Castro, R., Kowatsch, T., Filler, A., Dey, M., & Schaub, M. P. (2017). Efficacy of a web- and text messaging-based intervention to reduce problem drinking in adolescents: Results of a cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology, 85*(2), 147-159.

Haug, S., Schaub, M. P., Venzin, V., Meyer, C., & John, U. (2013). Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial. *Journal of Medical Internet Research, 15*(8), e171.

Heckhausen, J., & Heckhausen, H. (1991). *Motivation and Action*. Berlin, Germany: Springer.

Hefti, C., & Frey, M. (2008). Die Entwicklung der Versicherungslandschaft in der Krankenversicherung 1996-2006: Schlussbericht.

Hekler, E. B., Michie, S., Pavel, M., Rivera, D. E., Collins, L. M., Jimison, H. B., . . . Spruijt-Metz, D. (2016). Advancing models and theories for digital behavior change interventions. *American Journal of Preventive Medicine, 51*(5), 825-832.

Higgins, J. P., & Altman, D. G. (2008). Assessing risk of bias in included studies. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 187-241). Chichester, UK: John Wiley & Sons.

Hochbaum, G. M. (1958). *Public participation in medical screening programs: A socio-psychological study*. Retrieved from

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.

Honegger, L. (2018). Weitere Krankenkassen planen umstrittene Belohnungs-App für fitte Kunden. Retrieved from https://www.watson.ch/wirtschaft/digital/412199381-weitere-krankenkassen-planen-umstrittene-belohnungs-app-fuer-fitte-kunden

Hou, C., Carter, B., Hewitt, J., Francisa, T., & Mayor, S. (2016). Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. *Diabetes Care, 39*(11), 2089-2095.

Huber, C. A., Schwenkglenks, M., Rapold, R., & Reich, O. (2014). Epidemiology and costs of diabetes mellitus in Switzerland: an analysis of health care claims data, 2006 and 2011. *BMC Endocrine Disorders, 14*(1), 44.

Johnston, M., Carey, R. N., Connell Bohlen, L., Johnston, D. W., Rothman, A., de Bruin, M., . . . Michie, S. (2018). Linking behavior change techniques and mechanisms of action: Triangulation of findings from literature synthesis and expert consensus.

Kane, R. L., Johnson, P. E., Town, R. J., & Butler, M. (2004). A Structured Review of the Effect of Economic Incentives on Consumers' Preventive Behavior. *American Journal of Preventive Medicine, 27*(4), 327-352.

Kangovi, S., & Asch, D. A. (2018). Behavioral phenotyping in health promotion: embracing or avoiding failure. *JAMA, 319*(20), 2075-2076.

Karapanos, E., Gouveia, R., Hassenzahl, M., & Forlizzi, J. (2016). Wellbeing in the making: peoples' experiences with wearable activity trackers. *Psychology of well-being, 6*(1), 4.

Kelly, P., Kahlmeier, S., Götschi, T., Orsini, N., Richards, J., Roberts, N., . . . Foster, C. (2014). Systematic review and meta-analysis of reduction in all-cause mortality from walking and cycling and shape of dose response relationship. *International Journal of Behavioral Nutrition and Physical Activity, 11*(1), 132.

Kinder, C., Steingröver, M., & Neuhaus, K. (2017). Kundenloyalität bei Krankenkassen: Jede Interaktion ist eine Chance.

Kirk, M. A., Amiri, M., Pirbaglou, M., & Ritvo, P. (2019). Wearable Technology and Physical Activity Behavior Change in Adults With Chronic Cardiometabolic Disease: A Systematic Review and Meta-Analysis. *American Journal of Health Promotion, 33*(5), 778-791. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/0890117118816278

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences*. Los Angeles, CA: SAGE.

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

Koorts, H., Eakin, E., Estabrooks, P., Timperio, A., Salmon, J., & Bauman, A. (2018). Implementation and scale up of population physical activity interventions for clinical and community settings: the PRACTIS guide. *International Journal of Behavioral Nutrition and Physical Activity, 15*(1), 51.

Kowatsch, T., Fischer-Taeschler, D., Putzing, F., Bürki, P., Stettler, C., Chiesa-Tanner, G., & Fleisch, E. (2019). Die digitale Pille für chronische Krankheiten. In M. A. Pfannstiel, P. Da Cruz, & H. Mehlich (Eds.), *Digitale Transformation von Dienstleistungen im Gesundheitswesen VI* (pp. 205-231). Wiesbaden, Germany: Springer.

Kowatsch, T., Nißen, M., Rüegger, D., Stieger, M., Flückiger, C., Allemand, M., & von Wangenheim, F. (2018). *The Impact of Interpersonal Closeness Cues in Text-based Healthcare Chatbots on Attachment Bond and the Desire to Continue Interacting: An Experimental Design.* Paper presented at the Twenty-Sixth European Conference on Information Systems (ECIS2018), Portsmouth, UK.

Kowatsch, T., Nißen, M., Shih, C.-H. I., Rüegger, D., Volland, D., Filler, A., . . . Büchter, D. (2017). *Text-based Healthcare Chatbots Supporting Patient and Health Professional Teams: Preliminary Results of a Randomized Controlled Trial on Childhood Obesity.* Paper presented at the Persuasive Embodied Agents for Behavior Change (PEACH2017) Workshop at the 17th International Conference on Intelligent Virtual Agents, Stockholm, Sweden.

Kowatsch, T., Volland, D., Shih, I., Rüegger, D., Künzler, F., Barata, F., . . . Heldt, K. (2017, 30 May - 1 June). *Design and Evaluation of a Mobile Chat App for the Open Source Behavioral Health Intervention Platform MobileCoach.* Paper presented at the International Conference on Design Science Research in Information Systems, Karlsruhe, Germany.

Kramer, J.-N., & Kowatsch, T. (2017). Using feedback to promote physical activity: the role of the feedback sign. *Journal of Medical Internet Research, 19*(6), e192.

Kramer, J.-N., Künzler, F., Mishra, V., Smith, S. N., Kotz, D., Scholz, U., . . . Kowatsch, T. (2020). Which Components of a Smartphone Walking App Help Users to Reach Personalized Step Goals? Results From an Optimization Trial. *Annals of Behavioral Medicine*.

Kramer, J.-N., Künzler, F., Tinschert, P., & Kowatsch, T. (2019). *Trajectories of Engagement with a Digital Physical Activity Coach: Secondary Analysis of a Micro-Randomized Trial*. Paper presented at the 10th ISRII Scientific Meeting, Auckland, New Zealand.

Kramer, J.-N., Tinschert, P., Scholz, U., Fleisch, E., & Kowatsch, T. (2019). A Cluster-Randomized Trial on Small Incentives to Promote Physical Activity. *American Journal of Preventive Medicine, 56*(2), e45-e54.

Kramer, M. R., & Porter, M. (2011). Creating shared value. *Harvard Business Review, 89*(1/2), 62-77.

Künzler, F., Kramer, J.-N., & Kowatsch, T. (2017). *Efficacy of Mobile Context-aware Notification Management Systems: A Systematic Literature Review and Meta-Analysis*. Paper presented at the Fifth International Workshop on Pervasive and Context-Aware Middleware (PerCAM'17), Rome, Italy.

Kvedar, J. C., Fogel, A. L., Elenko, E., & Zohar, D. (2016). Digital medicine's march on chronic diseases. *Nature Biotechnology, 34*(3), 239-246.

Lamprecht, M., Fischer, A., & Stamm, H. (2014). *Sport Schweiz 2014: Sportaktivität und Sportinteresse der Schweizer Bevölkerung*. Retrieved from https://www.baspo.admin.ch/de/dokumentation/publikationen/sport-schweiz-2014.html

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., . . . Lau, A. Y. (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association, 25*(9), 1248-1258.

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes, 50*(2), 212-247.

Ledger, D., & McCaffrey, D. (2014). *Inside Wearables - How the Science of Human Behavior Change Offers the Secret to Long-Term Engagement*. Retrieved from Cambridge, MA:

Lee, I.-M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., Katzmarzyk, P. T., & Group, L. P. A. S. W. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet, 380*(9838), 219-229.

Lin, C.-Y., Strong, C., Scott, A. J., Broström, A., Pakpour, A. H., & Webb, T. L. (2018). A cluster randomized controlled trial of a theory-based sleep hygiene intervention for adolescents. *Sleep, 41*(11), zsy170.

Lippke, S., Schwarzer, R., Ziegelmann, J. P., Scholz, U., & Schüz, B. (2010). Testing stage-specific effects of a stage-matched intervention: A randomized controlled trial targeting physical exercise and its predictors. *Health Education and Behavior, 37*(4), 533-546.

Lippke, S., & Ziegelmann, J. P. (2008). Theory-based health behavior change: Developing, testing, and applying theories for evidence-based interventions. *Applied Psychology, 57*(4), 698-716.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Loewenstein, G., Asch, D. A., & Volpp, K. G. (2013). Behavioral economics holds potential to deliver better results for patients, insurers, and employers. *Health Affairs, 32*(7), 1244-1250.

Ma, J., Lewis, M. A., & Smyth, J. M. (2018). Translational behavioral medicine for population and individual health: gaps, opportunities, and vision for practice-based translational behavior change research. *Translational Behavioral Medicine, 8*(5), 753-760.

Ma, T., Sharifi, H., & Chattopadhyay, D. (2019). *Virtual Humans in Health-Related Interventions: A Meta-Analysis.* Paper presented at the 2019 CHI Conference on Human Factors in Computing Systems.

Madison, K., Schmidt, H., & Volpp, K. G. (2013). Smoking, obesity, health insurance, and health incentives in the Affordable Care Act. *JAMA, 310*(2), 143-144.

Mantzari, E., Vogt, F., Shemilt, I., Wei, Y., Higgins, J. P., & Marteau, T. M. (2015). Personal financial incentives for changing habitual health-related behaviors: A systematic review and meta-analysis. *Preventive Medicine, 75*, 75-85.

Markland, D., & Tobin, V. (2004). A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. *Journal of Sport and Exercise Psychology, 26*(2), 191-196.

Marlatt, G. A., Baer, J. S., & Quigley, L. A. (1995). Self-efficacy and addictive behavior. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 289-315). New York, NY: Cambridge University Press.

Marteau, T. M., Hollands, G. J., & Fletcher, P. C. (2012). Changing human behavior to prevent disease: the importance of targeting automatic processes. *Science, 337*(6101), 1492-1495.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*(3), 438-450.

McDermott, M. S., Oliver, M., Iverson, D., & Sharma, R. (2016). Effective techniques for changing physical activity and healthy eating intentions and behaviour: A systematic review and meta-analysis. *British Journal of Health Psychology, 21*(4), 827-841.

McLaren, L., McIntyre, L., & Kirkpatrick, S. (2009). Rose's population strategy of prevention need not increase social inequalities in health. *International Journal of Epidemiology, 39*(2), 372-377.

Michie, S., Abraham, C., Whittington, C., McAteer, J., & Gupta, S. (2009). Effective techniques in healthy eating and physical activity interventions: a meta-regression. *Health Psychology, 28*(6), 690-701.

Michie, S., Ashford, S., Sniehotta, F. F., Dombrowski, S. U., Bishop, A., & French, D. P. (2011). A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. *Psychology & Health, 26*(11), 1479-1498.

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., . . . Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine, 46*(1), 81-95.

Mindell, J., Ison, E., & Joffe, M. (2003). A glossary for health impact assessment. *Journal of Epidemiology and Community Health, 57*(9), 647-651.

Mitchell, M. S., Goodman, J. M., Alter, D. A., John, L. K., Oh, P. I., Pakosh, M. T., & Faulkner, G. E. (2013). Financial incentives for exercise adherence in adults: systematic review and meta-analysis. *American Journal of Preventive Medicine, 45*(5), 658-667.

Mitchell, M. S., Orstad, S. L., Biswas, A., Oh, P. I., Jay, M., Pakosh, M. T., & Faulkner, G. (2019). Financial incentives for physical activity in adults: systematic review and meta-analysis. *British Journal of Sports Medicine*, Published Online First: 15 May 2019.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto–mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences, 103*(42), 15623-15628.

Moore, G. F., & Evans, R. E. (2017). What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation of complex population health interventions. *SSM-population health, 3*, 132-135.

Morfeld, M., Kirchberger, I., & Bullinger, M. (2011). *SF-36 Fragebogen zum Gesundheitszustand: Deutsche Version des Short Form-36 Health Survey*. Göttingen, Germany: Hogrefe.

Morgenstern, H., & Bursic, E. S. (1982). A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. *Journal of Community Health, 7*(4), 292-309.

Morrato, E. H., Hill, J. O., Wyatt, H. R., Ghushchyan, V., & Sullivan, P. W. (2007). Physical activity in US adults with diabetes and at risk for developing diabetes, 2003. *Diabetes Care, 30*(2), 203-209.

Morris, J. N., & Hardman, A. E. (1997). Walking to health. *Sports Medicine, 23*(5), 306-332.

Morris, J. N., Heady, J., Raffle, P., Roberts, C., & Parks, J. (1953). Coronary heart-disease and physical activity of work. *The Lancet, 262*(6796), 1111-1120.

Muoio, D. (2019). Fitbit unveils data-driven personal health subscription, Fitbit Premium. Retrieved from https://www.mobihealthnews.com/news/north-america/fitbit-unveils-data-driven-personal-health-subscription-fitbit-premium

Murphy, M. H., Nevill, A. M., Murtagh, E. M., & Holder, R. L. (2007). The effect of walking on fitness, fatness and resting blood pressure: a meta-analysis of randomised, controlled trials. *Preventive Medicine, 44*(5), 377-385.

Murray, J. M., Brennan, S. F., French, D. P., Patterson, C. C., Kee, F., & Hunter, R. F. (2017). Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: a systematic review and meta-analysis. *Social Science and Medicine, 192*, 125-133.

Nafus, D., & Sherman, J. (2014). This one does not go up to 11: the quantified self movement as an alternative big data practice. *International journal of communication, 8*, 1784–1794.

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2016). Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 1-17.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, C. (1995). *Can computer personalities be human personalities?* Paper presented at the Conference on Human Factors in Computing Systems.

Nass, C., Steuer, J., & Tauber, E. R. (1994, April 24-28). *Computers are social actors.* Paper presented at the Conference on Human Factors in Computing Systems, Boston, MA.

Nes, B. M., Gutvik, C. R., Lavie, C. J., Nauman, J., & Wisløff, U. (2017). Personalized activity intelligence (PAI) for prevention of cardiovascular disease and promotion of physical activity. *The American journal of medicine, 130*(3), 328-336.

Nyce, S. (2010). Boosting wellness participation without breaking the bank. Retrieved from https://www.towerswatson.com/en/Insights/Newsletters/Americas/insider/2010/boosting-wellness-participation-without-breaking-the-bank

O'Brien, N., McDonald, S., Araújo-Soares, V., Lara, J., Errington, L., Godfrey, A., . . . White, M. (2015). The features of interventions associated with long-term effectiveness of physical activity interventions in adults aged 55–70 years: a systematic review and meta-analysis. *Health Psychology Review, 9*(4), 417-433.

Olander, E. K., Fletcher, H., Williams, S., Atkinson, L., Turner, A., & French, D. P. (2013). What are the most effective techniques in changing obese individuals' physical activity self-efficacy and behaviour: a systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity, 10*(1), 1.

Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (2014). *Value proposition design: How to create products and services customers want*. Hoboken, NJ: John Wiley & Sons.

Patel, M. S., Asch, D. A., Rosin, R., Small, D. S., Bellamy, S. L., Eberbach, K., . . . Wesby, L. (2016). Individual Versus Team-Based Financial Incentives to Increase Physical Activity: A Randomized, Controlled Trial. *Journal of General Internal Medicine*, 1-9.

Patel, M. S., Foschini, L., Kurtzman, G. W., Zhu, J., Wang, W., Rareshide, C. A., & Zbikowski, S. M. (2017). Using wearable devices and smartphones to track physical activity: initial activation, sustained use, and step counts across sociodemographic characteristics in a national sample. *Annals of Internal Medicine, 167*(10), 755-757.

Patel, M. S., Volpp, K. G., Rosin, R., Bellamy, S. L., Small, D. S., Heuer, J., . . . Lee, S. M. (2018). A Randomized, Controlled Trial of Lottery-Based Financial Incentives to Increase Physical Activity Among Overweight and Obese Adults. *American Journal of Health Promotion*, 0890117118758932.

Penedo, F. J., & Dahn, J. R. (2005). Exercise and well-being: a review of mental and physical health benefits associated with physical activity. *Current opinion in psychiatry, 18*(2), 189-193.

Perski, O., Crane, D., Beard, E., & Brown, J. (2019). Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. *DIGITAL HEALTH, 5*, 2055207619880676.

Pew Research Center. (2019). Smartphone Ownership is Growing Rapidly Around the World, but Not Always Equally.

Physical Activity Guidelines Advisory Committee. (2008). Physical activity guidelines advisory committee report, 2008. *Washington, DC: US Department of Health and Human Services, 2008*, A1-H14.

Ponsonby, A.-L., Sun, C., Ukoumunne, O. C., Pezic, A., Venn, A., Shaw, J. E., . . . Cochrane, J. (2011). Objectively measured physical activity and the subsequent risk of incident dysglycemia: the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). *Diabetes Care, 34*(7), 1497-1502.

Prochaska, J. O., & DiClemente, C. C. (1982). Transtheoretical therapy: Toward a more integrative model of change. *Psychotherapy: theory, research & practice, 19*(3), 276-288.

Promberger, M., & Marteau, T. M. (2013). When do financial incentives reduce intrinsic motivation? comparing behaviors studied in psychological and economic literatures. *Health Psychology, 32*(9), 950-957.

Provoost, S., Lau, H. M., Ruwaard, J., & Riper, H. (2017). Embodied conversational agents in clinical psychology: a scoping review. *Journal of Medical Internet Research, 19*(5), e151.

Purnell, J. Q., Gernes, R., Stein, R., Sherraden, M. S., & Knoblock-Hahn, A. (2014). A systematic review of financial incentives for dietary behavior change. *Journal of the Academy of Nutrition and Dietetics, 114*(7), 1023-1035.

Qiu, S., Cai, X., Chen, X., Yang, B., & Sun, Z. (2014). Step counter use in type 2 diabetes: a meta-analysis of randomized controlled trials. *BMC Medicine, 12*(1), 36.

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: SAGE Publications.

Reichert, F. F., Barros, A. J., Domingues, M. R., & Hallal, P. C. (2007). The role of perceived personal barriers to engagement in leisure-time physical activity. *American Journal of Public Health, 97*(3), 515-519.

Reis, R. S., Salvo, D., Ogilvie, D., Lambert, E. V., Goenka, S., Brownson, R. C., & Committee, L. P. A. S. E. (2016). Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. *The Lancet, 388*(10051), 1337-1348.

Riley, W. T., Rivera, D. E., Atienza, A. A., Nilsen, W., Allison, S. M., & Mermelstein, R. (2011). Health Behavior Models in the Age of Mobile Interventions: Are Our Theories Up to the Task? *Translational Behavioral Medicine, 1*(1), 53-71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21796270

Romeo, A., Edney, S., Plotnikoff, R., Curtis, R., Ryan, J., Sanders, I., . . . Maher, C. (2019). Can Smartphone Apps Increase Physical Activity? Systematic Review and Meta-Analysis. *Journal of Medical Internet Research, 21*(3), e12053.

Rose, G. (1981). Strategy of prevention: lessons from cardiovascular disease. *British Medical Journal (Clinical Research Ed.), 282*(6282), 2135-2136.

Rose, G. (2001). Sick individuals and sick populations. *International Journal of Epidemiology, 30*(3), 427-432.

Rosenstock, I. M. (1960). What research in motivation suggests for public health. *American Journal of Public Health and the Nations Health, 50*(3 Pt 1), 295-302.

Rosenstock, I. M. (1974). The health belief model and preventive health behavior. *Health Education Monographs, 2*(4), 354-386.
Rowe, D. A., Mahar, M. T., Raedeke, T. D., & Lore, J. (2004). Measuring physical activity in children with pedometers: Reliability, reactivity, and replacement of missing data. *Pediatric Exercise Science, 16*(4), 343-354.
Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: John Wiley & Sons.
Sallis, J. F. (2000). Age-related decline in physical activity: a synthesis of human and animal studies. *Medicine and Science in Sports and Exercise, 32*(9), 1598-1600.
Sallis, J. F., Bull, F., Guthold, R., Heath, G. W., Inoue, S., Kelly, P., . . . Hallal, P. C. (2016). Progress in physical activity over the Olympic quadrennium. *The Lancet, 388*(10051), 1325-1336.
Sallis, J. F., Cervero, R. B., Ascher, W., Henderson, K. A., Kraft, M. K., & Kerr, J. (2006). An ecological approach to creating active living communities. *Annual Review of Public Health, 27*, 297-322.
Sallis, J. F., Owen, N., & Fisher, E. (2008). Ecological models of health behavior. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health Behavior and Health Education* (pp. 466-485). San Francisco, CA: John Wiley & Sons.
Santesuisse. (2016). Bevölkerung will Wettbewerb und Wahlfreiheit und kein Staatsmonopol. Retrieved from https://www.santesuisse.ch/de/politik-medien/themen/einheitskasse/
Sattelmair, J., Pertman, J., Ding, E. L., Kohl III, H. W., Haskell, W., & Lee, I.-M. (2011). Dose response between physical activity and risk of coronary heart disease: a meta-analysis. *Circulation, 124*(7), 789-795.
Schäfer, T., Lorenz, C., Priess, H.-W., & Bitzer, E. M. (2016). *Prävalenz und Inzidenz der Koronaren Herzkrankheit (KHK) in der ambulanten ärztlichen Versorgung*. Paper presented at the HEC 2016: Health – Exploring Complexity. Joint Conference of GMDS, DGEpi, IEA-EEF, EFMI, Munich. https://www.egms.de/static/de/meetings/gmds2016/16gmds083.shtml
Schmidt, H., Voigt, K., & Wikler, D. (2010). Carrots, sticks, and health care reform—problems with wellness incentives. *New England Journal of Medicine, 362*(2), e3.
Scholz, U. (2019). It's time to think about time in health psychology. *Applied Psychology: Health and Well-Being, 11*(2), 173-186.
Scholz, U., Keller, R., & Perren, S. (2009). Predicting behavioral intentions and physical exercise: a test of the health action process approach at the intrapersonal level. *Health Psychology, 28*(6), 702-708.
Scholz, U., Schüz, B., Ziegelmann, J. P., Lippke, S., & Schwarzer, R. (2008). Beyond behavioural intentions: Planning mediates between intentions and physical activity. *British Journal of Health Psychology, 13*(3), 479-494. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1348/135910707X216062
Schueller, S. M., Tomasino, K. N., & Mohr, D. C. (2017). Integrating human support into behavioral intervention technologies: the efficiency model of support. *Clinical Psychology: Science and Practice, 24*(1), 27-45.
Schwaber, K., & Beedle, M. (2002). *Agile software development with Scrum*. Upper Saddle River: Prentice Hall.
Schwarzer, R., Lippke, S., & Luszczynska, A. (2011). Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). *Rehabilitation Psychology, 56*(3), 161-170.
Schwarzer, R., & Luszczynska, A. (2008). How to overcome health-compromising behaviors: The health action process approach. *European Psychologist, 13*(2), 141-151.
Sequeira, M. M., Rickenbach, M., Wietlisbach, V., Tullen, B., & Schutz, Y. (1995). Physical activity assessment using a pedometer and its comparison with a questionnaire in a large population survey. *American Journal of Epidemiology, 142*(9), 989-999.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth, Cengage Learning.
Sharot, T. (2011). The optimism bias. *Current Biology, 21*(23), R941-R945.
Shcherbina, A., Hershman, S. G., Lazzeroni, L., King, A. C., O'Sullivan, J. W., Hekler, E., . . . Ashley, E. A. (2019). The effect of digital physical activity interventions on daily step count: a randomised controlled crossover substudy of the MyHeart Counts Cardiovascular Health Study.

The Lancet Digital Health. Retrieved from http://www.sciencedirect.com/science/article/pii/S2589750019301293

Shea, B. J., Bouter, L. M., Peterson, J., Boers, M., Andersson, N., Ortiz, Z., . . . Grimshaw, J. M. (2007). External validation of a measurement tool to assess systematic reviews (AMSTAR). *PloS One, 2*(12), e1350.

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., . . . Bouter, L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology, 7*(1), 10.

Sheeran, P. (2002). Intention—Behavior Relations: A Conceptual and Empirical Review. *European Review of Social Psychology, 12*(1), 1-36.

Sheeran, P., Klein, W. M., & Rothman, A. J. (2017). Health behavior change: Moving from observation to intervention. *Annual Review of Psychology, 68*, 573-600.

Sheeran, P., Milne, S., Webb, T. L., & Gollwitzer, P. M. (2005). Implementation intentions and health behaviour. In M. Conner & P. Norman (Eds.), *Predicting health behavior* (pp. 276–323). Buckingham, UK: Open University Press.

Shuger, S. L., Barry, V. W., Sui, X., McClain, A., Hand, G. A., Wilcox, S., . . . Blair, S. N. (2011). Electronic feedback in a diet-and physical activity-based lifestyle intervention for weight loss: a randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity, 8*(1), 41.

Shurtleff, D. (1970). Some characteristics related to the incidence of cardiovascular disease and death: Framingham Study, 16-year follow-up. In W. Kannel & G. T (Eds.), *The Framingham Study: An Epidemiological Investigation of Cardiovascular Disease, section 26.* Washington D.C.: US Government Printing Office No. 0-414-297.

Sisson, S. B., Camhi, S. M., Church, T. S., Tudor-Locke, C., Johnson, W. D., & Katzmarzyk, P. T. (2010). Accelerometer-determined steps/day and metabolic syndrome. *American Journal of Preventive Medicine, 38*(6), 575-582.

Sniehotta, F. F., Nagy, G., Scholz, U., & Schwarzer, R. (2006). The role of action control in implementing intentions during the first weeks of behaviour change. *British Journal of Social Psychology, 45*(1), 87-106.

Sniehotta, F. F., Schwarzer, R., Scholz, U., & Schüz, B. (2005). Action planning and coping planning for long-term lifestyle change: theory and assessment. *European Journal of Social Psychology, 35*(4), 565-576.

Somauroo, J. (2019). Fitbit Partners With Singapore Government To Offer Trackers To 5.6 Million People.

Spruijt-Metz, D., & Nilsen, W. (2014). Dynamic models of behavior for just-in-time adaptive interventions. *IEEE Pervasive Computing*(3), 13-17.

Statista. (2017). Fitness tracker device unit shipments worldwide from 2016 to 2022. Retrieved from https://www.statista.com/statistics/610390/wearable-healthcare-device-shipments-worldwide/

Stein, N., & Brooks, K. (2017). A Fully Automated Conversational Artificial Intelligence for Weight Loss: Longitudinal Observational Study Among Overweight and Obese Adults. *JMIR Diabetes, 2*(2), e28. Retrieved from http://diabetes.jmir.org/2017/2/e28/

Steinhubl, S. R., Muse, E. D., & Topol, E. J. (2013). Can mobile health technologies transform health care? *JAMA, 310*(22), 2395-2396. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24158428

Stock, S., Schmidt, H., Büscher, G., Gerber, A., Drabik, A., Graf, C., . . . Stollenwerk, B. (2010). Financial incentives in the German Statutory Health Insurance: new findings, new questions. *Health Policy, 96*(1), 51-56.

Storm, V., Dörenkämper, J., Reinwand, D. A., Wienert, J., De Vries, H., & Lippke, S. (2016). Effectiveness of a web-based computer-tailored multiple-lifestyle intervention for people interested in reducing their cardiovascular risk: a randomized controlled trial. *Journal of Medical Internet Research, 18*(4), e78.

Strohacker, K., Galarraga, O., & Williams, D. M. (2014). The impact of incentives on exercise behavior: a systematic review of randomized controlled trials. *Annals of Behavioral Medicine, 48*(1), 92-99.

Sumner, J. A., Carey, R. N., Michie, S., Johnston, M., Edmondson, D., & Davidson, K. W. (2018). Using rigorous methods to advance behaviour change science. *Nature Human Behaviour, 2*(11), 797-799.

Sumra, H. (2013). iPhone 5s Includes New 'M7' Motion Coprocessor for Health and Fitness Tracking. Retrieved from https://www.macrumors.com/2013/09/10/iphone-5s-includes-new-m7-motion-coprocessor-for-health-and-fitness-tracking/

Teixeira, P. J., Carraça, E. V., Markland, D., Silva, M. N., & Ryan, R. M. (2012). Exercise, physical activity, and self-determination theory: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 9*(1), 78.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. New York: Penguin.

Tinschert, P., Rassouli, F., Barata, F., Steurer-Stey, C., Fleisch, E., Puhan, M. A., . . . Kowatsch, T. (2019). Prevalence of nocturnal cough in asthma and its potential as a marker for asthma control (MAC) in combination with sleep quality: protocol of a smartphone-based, multicentre, longitudinal observational study with two stages. *BMJ Open, 9*(1), e026323.

Toups, D. (2011). How many times will you crash your car. Retrieved from https://www.forbes.com/sites/moneybuilder/2011/07/27/how-many-times-will-you-crash-your-car/#3257d524e621

Tudor-Locke, C., Burkett, L., Reis, J., Ainsworth, B., Macera, C., & Wilson, D. (2005). How many days of pedometer monitoring predict weekly physical activity in adults? *Preventive Medicine, 40*(3), 293-298.

Tudor-Locke, C., Craig, C. L., Brown, W. J., Clemes, S. A., Cocker, K. D., Giles-Corti, B., . . . Blair, S. N. (2011). How many steps/day are enough? For adults. *International Journal of Behavioral Nutrition and Physical Activity, 8*(1:79), 1-17.

Vaes, A. W., Cheung, A., Atakhorrami, M., Groenen, M. T., Amft, O., Franssen, F. M., . . . Spruit, M. A. (2013). Effect of ‘activity monitor-based’counseling on physical activity and health-related outcomes in patients with chronic diseases: a systematic review and meta-analysis. *Annals of Medicine, 45*(5-6), 397-412.

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry, 64*(7), 456-464.

Van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: Chapman and Hall/CRC.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1-67.

Varadarajan, P. R., & Menon, A. (1988). Cause-related marketing: A coalignment of marketing strategy and corporate philanthropy. *The Journal of Marketing*, 58-74.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly, 36*(1), 157-178.

Volpp, K. G., Asch, D. A., Galvin, R., & Loewenstein, G. (2011). Redesigning employee health incentives—lessons from behavioral economics. *New England Journal of Medicine, 365*(5), 388-390.

Wahlich, C., Cook, D., Kerry, S., Limb, E., Victor, C., Iliffe, S., . . . Fox-Rushby, J. (2017). Primary care pedometer-based walking intervention: mixed-methods results from 3 year follow-up of PACE-UP cluster-randomised controlled trial. *The Lancet, 390*, S11.

Warburton, D. E., & Bredin, S. S. (2017). Health benefits of physical activity: a systematic review of current systematic reviews. *Current Opinion in Cardiology, 32*(5), 541-556.

Warburton, D. E., Nicol, C. W., & Bredin, S. S. (2006). Health benefits of physical activity: the evidence. *Canadian Medical Association Journal, 174*(6), 801-809.

Ware Jr, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*(3), 220-233.

Watson, A., Bickmore, T., Cange, A., Kulshreshtha, A., & Kvedar, J. (2012). An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. *Journal of Medical Internet Research, 14*(1), e1.

Webb, T. L., Chang, B. P., & Benn, Y. (2013). 'The Ostrich Problem': Motivated avoidance or rejection of information about goal progress. *Social and Personality Psychology Compass, 7*(11), 794-807.

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin, 132*(2), 249-268.

WHO. (1999). Health impact assessment: Main concepts and suggested approach. *Gothenberg consensus paper. Geneva: WHO*.

WHO. (2010). *Global recommendations on physical activity for health*. Retrieved from Geneva, Switzerland: http://whqlibdoc.who.int/publications/2010/9789241599979_eng.pdf

WHO. (2014). Global status report on noncommunicable diseases.

Wieser, S., Tomonaga, Y., Riguzzi, M., Fischer, B., Telser, H., Pletscher, M., . . . Schwenkglenks, M. (2014). Die Kosten der nichtübertragbaren Krankheiten in der Schweiz.Schlussbereicht im Auftrag des Bundesamts für Gesundheit (BAG), Abteilung Nationale Präventionsprogramme.

Williams, D. M., Anderson, E. S., & Winett, R. A. (2005). A review of the outcome expectancy construct in physical activity research. *Annals of Behavioral Medicine, 29*(1), 70-79.

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G., . . . Sterne, J. A. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ, 336*(7644), 601-605.

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review, 114*(4), 843-863.

Yang, D., & Dalton, J. E. (2012). *A unified approach to measuring the effect size between two groups using SAS®.* Paper presented at the SAS Global Forum, Orlando, FL.

Yardley, L., Morrison, L., Bradbury, K., & Muller, I. (2015). The person-based approach to intervention development: application to digital health-related behavior change interventions. *Journal of Medical Internet Research, 17*(1), e30.

Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., . . . Blandford, A. (2016). Understanding and promoting effective engagement with digital behavior change interventions. *American Journal of Preventive Medicine, 51*(5), 833-842.

Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics, 44*, 1049-1060.

Zhang, C.-Q., Zhang, R., Schwarzer, R., & Hagger, M. S. (2019). A meta-analysis of the health action process approach. *Health Psychology, 38*(7), 623-637.

Zunft, H.-J. F., Friebe, D., Seppelt, B., Widhalm, K., de Winter, A.-M. R., de Almeida, M. D. V., . . . Gibney, M. (1999). Perceived benefits and barriers to physical activity in a nationally representative sample in the European Union. *Public Health Nutrition, 2*(1a), 153-160.

Appendix

Appendix A - Calculation of Potential Impact Fraction

This appendix describes the calculation of the potential impact fraction (IF) and derives the necessary formulas based on the work by Morgenstern and Bursic (1982). Specifically, the formulas in Morgenstern and Bursic (1982) are adapted for negative risk factors (e.g. physical activity). As mentioned in the main text, the calculation of the IF requires knowledge of the distribution of a risk factor in the population (e.g. physical inactivity), the magnitude of the relationship of the risk factor and the disease, i.e. RRs for each category of the risk factor, and the expected change in the risk factor distribution due to the intervention. Generally, the IF can be defined as the proportional reduction of the number of incindent cases of a disease over a defined time period resulting from a specific change of the distribution of a risk factor in the population, e.g. due to an intervention. Thus, assuming no changes in the population size N and constant distributions of other disease risk factors, the IF of an intervention can be expressed as:

$$IF = \frac{C - C'}{C} = \frac{R - R'}{R} \; (A1)$$

where C and C' and R and R' are the number of cases and the disease risk before and after the intervention respectively and $R = C/N$ (Morgenstern & Bursic, 1982). Further, for a negative risk factor with k ordinal categories (k = 0,1, …, i) and a known fraction f_i of the population in each category the pre-intervention risk R in the population can be expressed as:

$$R = \sum_{i=0}^{k} f_i R_i \; (A2)$$

where R_i is the risk of an individual in category i to develop the disease over a specified time period (i.e. the category-specifc risk). Assuming that for the negative risk factor the risk decreases monotonically over the risk categories i (so that $R_i \geq R_{i+1}$), we can use the IF to calculate the impact of an intervention that shifts a certain fraction of individuals in the i-th category (g_i) to the next lower-risk category (i+1) for all i < k. The post-intervention distribution of the risk factor (f'_i) can then be expressed as:

$$f'_i = f_i(1 - g_i) + f_{i-1}g_{i-1} \ (A3)$$

where $f_{0-1} = g_{0-1} = g_k = 0$. Therefore the post-intervention risk R' in the population can be expressed as:

$$R' = \sum_{i=0}^{k}(f_i(1 - g_i)R_i + f_{i-1}g_{i-1}R_i)(A4)$$

Substituting A2 and A4 in A1 gives:

$$IF = \frac{\sum_{i=0}^{k} f_i R_i - \sum_{i=0}^{k}(f_i(1 - g_i)R_i + f_{i-1}g_{i-1}R_i)}{\sum_{i=0}^{k} f_i R_i} (A5)$$

Because category-specific risk ratios are more commonly reported than category-specific risks, it is advisable to divide all terms by R_0. Rearranging then gives:

$$IF = \frac{\sum_{i=0}^{k}(f_i g_i - f_{i-1}g_{i-1})RR_i}{\sum_{i=0}^{k} f_i RR_i} (A6)$$

which equals the formula reported in the main text.

Appendix B - The AMSTAR Tool

Table B-1. The AMSTAR checklist.

Item	Answer
1. Was an 'a priori' design provided? The research question and inclusion criteria should be established before the conduct of the review.	☐ yes ☐ no ☐ can't answer ☐ not applicable
2. Was there duplicate study selection and data extraction? There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.	☐ yes ☐ no ☐ can't answer ☐ not applicable
3. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.	☐ yes ☐ no ☐ can't answer ☐ not applicable
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.	☐ yes ☐ no ☐ can't answer ☐ not applicable
5. Was a list of studies (included and excluded) provided? A list of included and excluded studies should be provided.	☐ yes ☐ no ☐ can't answer ☐ not applicable
6. Were the characteristics of the included studies provided? In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic	☐ yes ☐ no ☐ can't answer ☐ not applicable

data, disease status, duration, severity, or other diseases should be reported.

7. Was the scientific quality of the included studies assessed and documented? A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

☐ yes
☐ no
☐ can't answer
☐ not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

☐ yes
☐ no
☐ can't answer
☐ not applicable

9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

☐ yes
☐ no
☐ can't answer
☐ not applicable

10. Was the likelihood of publication bias assessed? An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

☐ yes
☐ no
☐ can't answer
☐ not applicable

11. Was the conflict of interest included? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

☐ yes
☐ no
☐ can't answer
☐ not applicable

Appendix C - AMSTAR Ratings of Reviews on Mobile Physical Activity Interventions

Table C-1. AMSTAR rating of the review of Bravata et al. (2007).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	No	Review provides no information on review protocol or pre-registration.
2. Was there duplicate study selection and data extraction?	Yes	“Two authors independently abstracted 4 categories of variables from each of the included studies:”, “We resolved discrepancies by repeated review and discussion between abstractors.” (p. 2297)
3. Was a comprehensive literature search performed?	Yes	“In collaboration with a professional librarian, we developed individualized search strategies for 7 databases […] We used search terms such as pedometer, activity monitor, and step counter. We also reviewed the bibliographies of retrieved articles and relevant conference proceedings and contacted experts in exercise physiology for additional studies.” (p. 2297)
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	No	Included studies restricted to search results and English language.
5. Was a list of studies (included and excluded) provided?	No	Only a list of included studies is provided.
6. Were the characteristics of the included studies provided?	Yes	Table 1
7. Was the scientific quality of the included studies assessed and documented?	No	The quality was assessed by the authors (“Overall, the quality of the reporting of the included studies was relatively good”, p. 2298) but no information on criteria and assessment processes is provided.
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	No	No sensitivity analysis with regard to study quality was performed. Conclusions do not consider study quality.
9. Were the methods used to combine the findings of studies appropriate?	Yes	Heterogeneity was evaluated and a random-effects model was used.
10. Was the likelihood of publication bias assessed	No	Publication bias assessed and reported only for observational studies.
11. Was the conflict of interest included?	Yes	Financial disclosures and funding reported on page 2303.

Table C-2. AMSTAR rating of the review of Brickwood et al. (2019).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	No	Review provides no information on review protocol or pre-registration.
2. Was there duplicate study selection and data extraction?	Yes	"All manuscripts identified as requiring full-text review were reviewed independently by 2 authors […]. A third reviewer (GW) resolved any conflicts. […] data extraction [was] performed by both authors […] individually and differences resolved by consensus." (p.3)
3. Was a comprehensive literature search performed?	Yes	"The following Web-based databases were searched […]: CENTRAL, MEDLINE, PubMed, Scopus, Web of Science, CINHAL, SPORTDiscus, and Health Technology Assessment. […] Reference lists of retrieved articles were checked, and citation searches were performed on key articles" (p.2)
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	Yes	"The authors of identified ongoing studies were contacted to obtain study progress and request available results for inclusion in the meta-analysis." (p.3)
5. Was a list of studies (included and excluded) provided?	No	Only a list of included studies is provided.
6. Were the characteristics of the included studies provided?	Yes	Appendix 3
7. Was the scientific quality of the included studies assessed and documented?	Yes	Scientific quality was assessed using the Cochrane Risk of Bias Tool (Figure 2) and the GRADE framework (Appendix 4).
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	Yes	e.g. "Participants who received an intervention including a consumer-based wearable activity tracker demonstrated a significant improvement in daily steps […]. The quality of evidence was low for daily steps […]" (p. 13).
9. Were the methods used to combine the findings of studies appropriate?	Yes	"Due to the variability of the included studies, random-effects meta-analyses were performed" (p.3).
10. Was the likelihood of publication bias assessed	Yes	"Publication bias was assessed for daily step count and MVPA with no bias identified" (p.6).
11. Was the conflict of interest included?	Yes	p. 14

Table C-3. AMSTAR rating of the review of De Vries et al. (2016).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	Yes	"The protocol for this systematic review and meta-analysis was based on the PRISMA-P statement (17) and registered at PROSPERO (CRD42015024086) (18)." (p. 2079)
2. Was there duplicate study selection and data extraction?	Yes	"Two independent content area experts (HJDV and TJMK) screened potentially eligible articles for inclusion […] Differences in appraisal were resolved by reaching consensus.", "Data extraction was performed by the reviewers utilizing a standard extraction form" (p. 2079)
3. Was a comprehensive literature search performed?	Yes	"MEDLINE, Embase, CINAHL, PsycINFO, CENTRAL, and PEDro were searched for eligible articles published before 1 July 2015. The employed MeSH terms and keywords included overweight, obesity, accelerometry, actigraphy, physical activity, exercise, and energy expenditure. Furthermore, a reference tracking strategy was performed by searching the reference lists and citations of included articles in Web of Science and Scopus" (p. 2079)
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	No	Included studies restricted to search results and English language.
5. Was a list of studies (included and excluded) provided?	No	Only a list of included studies is provided.
6. Were the characteristics of the included studies provided?	Yes	Table 1
7. Was the scientific quality of the included studies assessed and documented?	Yes	"The risk of bias was scored by two independent reviewers (HJDV and TJMK) using the Cochrane Collaboration's tool (19)." (p. 2079)
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	No	Conclusions are formulated without considering the relatively high risk of bias in included studies (e.g. "Behavioral physical activity interventions with an activity monitor increase physical activity in adults with overweight or obesity", p. 2090)
9. Were the methods used to combine the findings of studies appropriate?	Yes	Moderate heterogeneity is observed (I^2 = 49%, n.s.) and a random-effects model is used. Inclusion criteria are broad with regard to interventions and comparators.
10. Was the likelihood of publication bias assessed	No	No due to the small number of included studies
11. Was the conflict of interest included?	Yes	Yes, no conflicts are reported

Table C-4. AMSTAR rating of the review of Direito et al. (2017).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	Yes	“The criteria for considering studies for this review and the outcomes of interest, as well as the methods for data extraction, assessing risk of bias, and statistical analysis were prespecified” (p. 227)
2. Was there duplicate study selection and data extraction?	Yes	“Two authors (AD, JR) independently screened the titles and abstracts of the search results […] The retrieved full-text articles were then scanned by two authors […] and discrepancies were resolved by discussion until a consensus was reached. […] For each included study, reviewers (AD, EC or JR) independently extracted data” (p. 228)
3. Was a comprehensive literature search performed?	Yes	“Seven electronic databases were searched [..]Review articles and the reference lists of selected studies were searched for additional articles.” (p. 228)
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	No	“Only English language-published studies were accepted” (p. 228)
5. Was a list of studies (included and excluded) provided?	No	Only a list of included studies is provided.
6. Were the characteristics of the included studies provided?	Yes	Table 1
7. Was the scientific quality of the included studies assessed and documented?	Yes	Scientific quality of the included studies was assessed using the Cochrane risk of bias tool
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	No	High likelihood of performance bias not considered in conclusions and recommendations (e.g. “Current mHealth interventions have small effects on total PA, MVPA, walking and SB”, p. 238)
9. Were the methods used to combine the findings of studies appropriate?	Yes	Broad inclusion criteria with regard to participants, interventions and comparators. Random-effects model used but no heterogeneity observed for step counts.
10. Was the likelihood of publication bias assessed	Yes	“To assess publication bias, we examined funnel plots for asymmetry.” (p. 229)
11. Was the conflict of interest included?	Yes	p.238

Table C-5. AMSTAR rating of the review of Gal et al. (2018).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	Yes	"This review was registered in the PROSPERO register of systematic reviews" (p. 2)
2. Was there duplicate study selection and data extraction?	Yes	"First, the title and abstract of the search yield were independently screened by four authors […] the definite selection was […] also independently screened by three authors. Disagreement was resolved by consensus […] Three authors independently extracted the data from each of the included studies" (p. 2/3)
3. Was a comprehensive literature search performed?	Yes	"we searched on titles and abstracts in PubMed, EMBASE and the Cochrane Central Register of Controlled Trials (CENTRAL; 2008 to 2017) […]We additionally searched the reference list of relevant reviews/studies" (p. 2)
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	No	Included studies restricted to search results.
5. Was a list of studies (included and excluded) provided?	No	Only a list of included studies is provided.
6. Were the characteristics of the included studies provided?	Yes	Table 1
7. Was the scientific quality of the included studies assessed and documented?	Yes	Scientific quality of included studies was assessed using the Cochrane risk of bias tool (p.3)
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	Yes	Subgroup analyses of studies with low risk of bias studies are considered in conclusions (p.12)
9. Were the methods used to combine the findings of studies appropriate?	Yes	Inclusion criteria are relatively broad with regard to populations, interventions and comparators. A random-effects model was used and high heterogeneity was observed (I^2 = 90%)
10. Was the likelihood of publication bias assessed	Yes	"To investigate publication bias, we assessed funnel plots by visual inspection for asymmetry." (p.3)
11. Was the conflict of interest included?	Yes	p.14

Table C-6. AMSTAR rating of the review of Kirk et al. (2018).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	No	Review provides no information on review protocol or pre-registration.
2. Was there duplicate study selection and data extraction?	Yes	"Two reviewers (M.K. and M.A.) evaluated search results for potentially relevant trials and obtained full-text versions of included studies […] eligible study data […]were independently extracted and assessed by 2 reviewers" (p. 2/3)
3. Was a comprehensive literature search performed?	Yes	"an electronic literature search of English, peer-reviewed controlled trials was conducted using 5 databases […]an electronic literature search of English, peer-reviewed controlled trials was conducted using 5 databases" (p.2)
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	No	"Eligible RCTs were English language peer-reviewed RCTs" (p.2)
5. Was a list of studies (included and excluded) provided?	No	Only a list of included studies is provided.
6. Were the characteristics of the included studies provided?		Yes, in the supplementary material
7. Was the scientific quality of the included studies assessed and documented?	Yes	"Three review authors (M.K., M.A., and M.P.) independently assessed the overall methodological quality of the relevant articles using the Cochrane Collaboration Tool for assessing bias" (p.3)
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	Yes	Sensitivity analysis of studies with low risk of bias was performed
9. Were the methods used to combine the findings of studies appropriate?	Yes	Inclusion criteria are broad with regard to interventions and comparators and moderately broad with regard to populations. A random-effects model was used and heterogeneity was very high (I^2 = 91%)
10. Was the likelihood of publication bias assessed	No	No information on publication bias is provided.
11. Was the conflict of interest included?	Yes	p. 11

Table C-7. AMSTAR rating of the review of Romeo et al. (2019).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	No	Review provides no information on review protocol or pre-registration.
2. Was there duplicate study selection and data extraction?	Yes	"Studies were screened for eligibility in duplicate under blinded conditions by 2 independent reviewers […] 2 independent reviewers then screened the full-text studies […] Pairs of reviewers […] independently extracted data from each included study" (p.3)
3. Was a comprehensive literature search performed?	Yes	"A systematic search was conducted on January 8, 2018, and included 7 electronic databases […] In addition, 5 prominent researchers in the field were contacted with the list of identified studies and asked to recommend any additional studies that met the inclusion criteria." (p. 2)
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	No	Included studies restricted to search results and English language.
5. Was a list of studies (included and excluded) provided?	No	Only a list of included studies is provided.
6. Were the characteristics of the included studies provided?	Yes	Table 2
7. Was the scientific quality of the included studies assessed and documented?	Yes	"Risk of bias was assessed using a 25-item tool developed by Maher [18] and based on the Consolidated Standards of Reporting Trials (CONSORT) checklist […] Studies were also graded using the 2011 Centre for Evidence Based Medicine Levels of Evidence"
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	No	Results on study quality (item 7) are not considered in the discussion and in formulating the conclusions (p.10/11).
9. Were the methods used to combine the findings of studies appropriate?	Yes	Inclusion criteria are broad with regard to populations and possibly interventions and a random-effects model is used. Substantial heterogeneity is observed (I^2 = 72%)
10. Was the likelihood of publication bias assessed	No	No information on the assessment of publication bias is provided.
11. Was the conflict of interest included?	Yes	p. 12

Table C-8. AMSTAR rating of the review of Qiu et al. (2014).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	Yes	Review registered on PROSPERO platform
2. Was there duplicate study selection and data extraction?	Yes	"Two authors (SHQ and XC) selected and independently assessed the studies. Discrepancies were resolved by discussion or consensus" (p.2)
3. Was a comprehensive literature search performed?	Yes	"The following electronic databases were searched from January 1994 to June 2013: PubMed, Web of Science and Cochrane Library. […] The related references of all included articles were collected and hand-searched to make sure no suitable and relevant studies were missed."
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	No	Included studies were restricted to search results and English language. (p. 2)
5. Was a list of studies (included and excluded) provided?	No	List of included studies provided only.
6. Were the characteristics of the included studies provided?	Yes	Table 1
7. Was the scientific quality of the included studies assessed and documented?	Yes	"Quality was assessed independently by 2 authors (SHQ and XC) using the Cochrane Collaboration's 'Risk of Bias' Tool [22]" (p.2)
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	Yes	Subgroup analysis with regard to study quality was performed (Appendix 4) and revealed no differences.
9. Were the methods used to combine the findings of studies appropriate?	Yes	Broad inclusion criteria with regard to intervention and control groups. Random-effects model used. Very high heterogeneity observed (I^2 = 86%).
10. Was the likelihood of publication bias assessed	Yes	"Publication bias was detected and assessed by Begg's test and Egger's test." (p.3)
11. Was the conflict of interest included?	Yes	p. 8

Table C-9. AMSTAR rating of the review of Vaes et al. (2013).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	No	No information on review protocol or preregistration provided.
2. Was there duplicate study selection and data extraction?	Yes	"Titles and abstracts were screened against inclusion criteria by two authors (A.W.V. and A.C.), and potentially eligible articles were retrieved. All articles were independently selected for inclusion by two reviewers (A.W.V. and A.C.). Disagreements could be resolved by consulting a third reviewer (M.A.)." (p. 398)
3. Was a comprehensive literature search performed?	Yes	"A computerized literature search was performed in Medline/ PubMed, Web of Knowledge (WOK), Embase, and BIOSIS in August 2012. […]In addition two potentially relevant articles were found by checking reference lists of included trials" (p. 398)
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	No	Included studies were restricted to search results and English language (p. 398).
5. Was a list of studies (included and excluded) provided?	No	Only list of included studies is provided.
6. Were the characteristics of the included studies provided?	Yes	Table 2
7. Was the scientific quality of the included studies assessed and documented?	Yes	"The methodological quality of the included trials was scored independently using the Physiotherapy Evidence-based Database (PEDro) scale (26)."
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	No	Potential bias is acknowledged but not included in the conclusions (p. 410)
9. Were the methods used to combine the findings of studies appropriate?	Yes	Inclusion criteria are broad with regard to populations, interventions and comparators. A random-effects model is used and substantial heterogeneity is observed (I^2 = 73%).
10. Was the likelihood of publication bias assessed	Yes	"The presence of publication bias was checked by a funnel plot for change in physical activity level." (p.398)
11. Was the conflict of interest included?	Yes	p. 410

Appendix D - GRADE Ratings of Reviews on Mobile Physical Activity Interventions

Table D-1. GRADE rating of the review of Bravata et al. (2007).

Quality Dimension	Rating	Explanation
Risk of bias	NA	No systematic risk of bias assessment reported
Publication bias	Strongly suspected	Summary effect combines very small studies only (average *N*: 35, range: 21-62)
Imprecision	Serious imprecision	OIS criterion not met (total *N*: 277)
Inconsistency	Serious inconsistency	Point estimates vary (range: 395-5066 steps); significant heterogeneity
Indirectness	No serious indirectness	Populations and interventions relevant
Overall GRADE rating	Very low	

Table D-2. GRADE rating of the review of Brickwood et al. (2019).

Quality Dimension	Rating	Explanation
Risk of bias	No serious limitations	No crucial risk of bias for the majority of studies
Publication bias	Undetected	Based on author's judgement (p. 6)
Imprecision	No serious imprecision	OIS criterion met (total N = 2,144)
Inconsistency	No serious inconsistency	Point estimates vary somewhat, CIs overlap, no significant heterogeneity
Indirectness	No serious indirectness	Populations and interventions relevant
Overall GRADE rating	high	

Table D-3. GRADE rating of the review of De Vries et al. (2016).

Quality Dimension	Rating	Explanation
Risk of bias	No serious limitations	No crucial risk of bias for the majority of studies
Publication bias	Strongly suspected	Summary effect combines very small studies only (average *N*: 83, range: 47-106)
Imprecision	Serious imprecision	OIS criterion not met (total *N*: 417)
Inconsistency	No serious indirectness	Insignificant heterogeneity (I^2 = 49%) due to the study of Morgan et al. (2013); impact on summary effect judged as not crucial.
Indirectness	No serious indirectness	Interventions and populations partly relevant (Bond et al. (2014) and Tudor-Locke (2004) include multiple coaching sessions with patients); GRADE recommends rating down conservatively, therefore the quality of evidence is not further adjusted.
Overall GRADE rating	low	

Table D-4. GRADE rating of the review of Direito et al. (2017).

Quality Dimension	Rating	Explanation
Risk of bias	No serious limitations	No crucial risk of bias for the majority of studies
Publication bias	Undetected	Based on authors' judgement (p. 236) and insignificant results of included small studies
Imprecision	Serious imprecision	OIS criterion not met (total $N = 401$)
Inconsistency	No serious inconsistency	Some variation in point estimates, overlap of CIs and no observed heterogeneity ($I^2 = 0\%$)
Indirectness	No serious indirectness	Populations and interventions relevant
Overall GRADE rating	Moderate	

Table D-5. GRADE rating of the review of Gal et al. (2018).

Quality Dimension	Rating	Explanation
Risk of bias	No serious limitations	2/7 studies with non-significant results at moderate risk of bias (attrition bias); Studies reduce the summary effect to a non-crucial degree.
Publication bias	Strongly suspected	4/7 studies are small ($N < 100$) and show moderate to large effects (SMD: 0.55-1.67). Evidence among larger studies is mixed.
Imprecision	No serious imprecision	OIS criterion met (total $N = 1392$)
Inconsistency	Serious inconsistency	Strongly varying point estimates (SMD: -0.13-1.67) and significant heterogeneity ($I^2 = 90\%$)
Indirectness	No serious indirectness	Populations and interventions mostly relevant
Overall GRADE rating	Low	

Table D-6. GRADE rating of the review of Kirk et al. (2018).

Quality Dimension	Rating	Explanation
Risk of bias	No serious limitations	8/19 comparisons at moderate or high risk of bias; Conservatively judged as not crucial
Publication bias	Strongly suspected	Not assessed; strongly suspected due to the large number of small studies included.
Imprecision	No serious imprecision	OIS criterion met (total $N = 1471$)
Inconsistency	Serious inconsistency	Point estimates vary strongly; very strong and significant heterogeneity ($I^2 = 91\%$)
Indirectness	Serious indirectness	Interventions may have different effects in (voluntary) patient populations where need and motivation for behavior change is likely greater than in general or at-risk populations. Further, 7/19 comparisons with no applicable interventions (e.g. multiple in-person coaching sessions).
Overall GRADE rating	Very low	

Table D-7. GRADE rating of the review of Romeo et al. (2019).

Quality Dimension	Rating	Explanation
Risk of bias	Serious limitations	3/6 comparisons with moderate risk of bias and considerable impact on the summary effect
Publication bias	Undetected	Not assessed; small and larger studies included; small studies do not produce systematically larger effects
Imprecision	No serious imprecision	OIS criterion met (total $N = 1178$)
Inconsistency	Serious inconsistency	Point estimates of included studies vary; significant heterogeneity observed ($I^2 = 72\%$)
Indirectness	No serious indirectness	Interventions and populations mostly relevant
Overall GRADE rating	Low	

Table D-8. GRADE rating of the review of Qiu et al. (2014).

Quality Dimension	**Rating**	**Explanation**
Risk of bias	No serious limitations	3/7 comparisons with moderate risk of bias
Publication bias	Undetected	Based on author's judgement (p. 3)
Imprecision	No serious imprecision	OIS criterion met (total $N = 861$)
Inconsistency	Serious inconsistency	Point estimates vary and significant heterogeneity observed ($I^2 = 86\%$)
Indirectness	Serious indirectness	Diabetes patient populations only; some very intensive interventions
Overall GRADE rating	Low	

Table D-9. GRADE rating of the review of Vaes et al. (2013).

Quality Dimension	Rating	Explanation
Risk of bias	No serious limitations	4/7 comparisons with moderate risk of bias; impact on summary effect is judged not crucial
Publication bias	Undetected	No funnel plot asymmetry (Figure 2)
Imprecision	Serious imprecision	OIS criterion not met (total $N = 633$)
Inconsistency	Serious inconsistency	Point estimates vary strongly (SMD: 0.35-1.62) and significant heterogeneity is observed ($I^2 = 73\%$)
Indirectness	Serious indirectness	Interventions combine pedometers and face-to-face counselling for diabetes patients
Overall GRADE rating	Very low	

Appendix E - AMSTAR Ratings of Reviews on Financial Incentives

Table E-1. AMSTAR rating of the review of Gong et al. (2018).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	Yes	This study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol [35] and was registered in the PROSPERO International Prospective Register of Systematic Reviews (2017:CRD42017058063)" (p. 2)
2. Was there duplicate study selection and data extraction?	Yes	"Two reviewers (YG and TPT) screened the titles and abstracts of the studies and excluded those failing to meet inclusion criteria. […] two reviewers (YG and TPT) independently reviewed the full text of articles […] YG and TPT independently abstracted data from the eligible manuscripts and entered data in a Research Electronic Data Capture (REDCap) database" (p. 3).
3. Was a comprehensive literature search performed?	Yes	"We performed a systematic search of PubMed, Embase, and Web of Science" (p.3)
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	No	Included studies restricted to search results and English language.
5. Was a list of studies (included and excluded) provided?	No	Only a list of included studies is provided.
6. Were the characteristics of the included studies provided?	Yes	Supplement Table B
7. Was the scientific quality of the included studies assessed and documented?	Yes	"We evaluated the articles that met the inclusion criteria for quality using the Jadad assessment tool" (p. 3)
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	No	No sensitivity analysis with regard to study quality was performed. Conclusions do not consider study quality.
9. Were the methods used to combine the findings of studies appropriate?	Yes	Random random-effects model was used, which is considered appropriate given differences between incentive design and population of the two studies.
10. Was the likelihood of publication bias assessed	No	Not possible for two studies
11. Was the conflict of interest included?	Yes	Funding and competing interests reported on page 1 and 2

Table E-2. AMSTAR rating of the review of Mitchell et al. (2019).

Item	Rating	Explanation/Quote
1. Was an 'a priori' design provided?	Yes	Review updates a previous review (p.2)
2. Was there duplicate study selection and data extraction?	Yes	"Article records (titles, abstracts) were independently screened by two reviewers (SO and SK). A third reviewer (MM) was consulted where uncertainty existed […] Study-level […] and participant-level information […] was extracted by one reviewer […] A second reviewer (AB) audited all retrieved step count estimates […] Two authors (SO and MM) independently assessed the methodological quality of included studies" (p.2/3)
3. Was a comprehensive literature search performed?	Yes	"Seven electronic databases […] were searched"
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?	No	Included studies restricted to search results (i.e. published studies in English language)
5. Was a list of studies (included and excluded) provided?	No	Only list of included studies provided
6. Were the characteristics of the included studies provided?	Yes	Table 1 & Table 2
7. Was the scientific quality of the included studies assessed and documented?	Yes	"Two authors (SO and MM) independently assessed the methodological quality of included studies using the Effective Public Health Practice Project (EPHPP) Quality Assessment Tool for Quantitative Studies" (p.3)
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	No	Recommendations are based on sub-group analyses whose exploratory and non-causal nature is not acknowledged explicitly. Behavioral economics principles (Table 3) are recommended although other mechanisms may be responsible for the incentive effects.
9. Were the methods used to combine the findings of studies appropriate?	Yes	Heterogeneity was assessed and a random-effects model was used.
10. Was the likelihood of publication bias assessed	Yes	"The possibility of publication bias was examined by visually inspecting funnel plots for their skew and asymmetric shape and quantitatively by Egger's test for asymmetry"
11. Was the conflict of interest included?	Yes	Funding and competing interest disclosed on page 10.

Appendix F - GRADE Ratings of Reviews on Financial Incentives

Table F-1. GRADE rating of the review of Gong et al. (2018).

Quality Dimension	Rating	Explanation
Risk of bias	No serious limitations	No blinding in the study of Losina et al. (2017) which is judged as not serious
Publication bias	NA	No evaluation possible due to the small number of studies
Imprecision	Serious imprecision	OIS criterion not met (total *N*: 232)
Inconsistency	No serious inconsistency	Point estimates similar and substantial overlap between CIs.
Indirectness	Serious indirectness	Patient population in one study and large incentive amounts ($69.90/month)
Overall GRADE rating	Low	

Table F-2. GRADE rating of the review of Mitchell et al. (2019).

Quality Dimension	Rating	Explanation
Risk of bias	No serious limitations	No crucial risk of bias for the majority of studies
Publication bias	Strongly suspected	Funnell plot asymmetry and statistical tests suggest publication bias (p. 6)
Imprecision	No serious imprecision	OIS criterion met (total *N*: 2,628)
Inconsistency	Serious inconsistency	Point estimates vary (range: 93-3907 steps); significant heterogeneity
Indirectness	No serious indirectness	Populations relevant, but very large incentives ($70.5/month). GRADE recommends rating down conservatively, therefore the quality of evidence is not further adjusted
Overall GRADE rating	Low	

Appendix G - Invitation Mail (Example from the Cash Incentive Group)

Subject: Take the first step... out of 10'000!

Dear Ms. / Mr.

The CSS insurance takes another step towards health promotion.

Among selected CSS customers - and you are one of them - we are looking for participants for the CSS pilot project myStep, which is scientifically supported by the University of St. Gallen and the ETH Zurich. The temporary promotion will run from 1st of July to 31st of December 2015.

All you have to do is register, strap on your activity tracker and off you go!

You take 10,000 steps a day and not only benefit from more well-being and vitality - the CSS also rewards you with a credit to your complementary insurance premium (VVG).

You don't own an activity tracker yet? You have the option of buying an activity tracker at special conditions. Sign up right now!

You can learn more here: **[link to the insurer's online platform]**

We look forward to seeing you!

Best regards

You don't want to participate? Tell us why here **[link to survey]**. Your opinion is important to us. This survey is conducted by the Health-IS Lab of the University of St. Gallen and ETH Zurich as part of the accompanying research and is evaluated anonymously

Appendix H - Study I: Result Tables of Sensitivity Analyses

Table H-1. Sensitivity analyses of incentive effects on program participation.

	Dependent Variable: Participation (yes/no)		
	Model 1 (*n* = 25,348)	**Model 2 (*n* = 25,348)**	**Model 3[a] (*n* = 26,091)**
Fixed Effects			
(Intercept)	**-3.42 [-3.62, -3.23]**	**-3.47 [-3.71, -3.27]**	**-3.46 [-3.68, -3.24]**
Incentive Condition (reference group: control)			
Cash incentive	**0.71 [0.52, 0.96]**	**0.70 [0.45, 0.95]**	**0.67 [0.44, 0.91]**
Charity incentives	**0.49 [0.29, 0.71]**	**0.48 [0.26, 0.71]**	**0.46 [0.22, 0.70]**
Sex: Female		**0.17 [0.07, 0.28]**	**0.16 [0.05, 0.28]**
Age		**0.01 [0.01, 0.02]**	**0.01 [0.01, 0.02]**
Cantonal population density		< -0.001 [< -0.001, < 0.001]	< -0.001 [-0.001, 0.001]
Nationality (reference group: Swiss)			
German		0.01 [-0.29, 0.28]	0.01 [-0.25, 0.27]
Other		**-0.50 [-0.82, -0.21]**	**-0.50 [-0.80, -0.20]**
Random Effects: Standard Deviation			
Canton	0.07 [<0.0001; 0.14]	0.08 [<0.0001; 0.14]	-
Model Fit Parameters			
χ^2 - difference	-	**$\chi^2(5) = 64.43$**	**-**
AIC	10,316.50	10,262.10	-
BIC	10,349.10	10,335.40	-

Note: Boldface indicates statistical significance (p<0.05). Table depicts point estimates and associated 95% confidence intervals on the logit scale. AIC, Akaike's information criterion; BIC, Bayesian information criterion.

[a] Pooled results from fitting model 2 to 20 complete datasets. Random effects are not pooled and therefore not reported.

Table H-2. Sensitivity analyses of incentive effects on steps/day.

	Dependent Variable: Steps/Day		
	Model 1 (*n* = 1,005)	**Model 2 (*n* = 1,005)**	**Model 3[a] (*n* = 1,223)**
Fixed Effects			
(Intercept)	**4,924 [3,617, 6,459]**	**4,924 [3,377, 6,450]**	**5,339 [4,055, 6,622]**
Time	**-4 [-8, -0.2]**	**-4 [-7, -0.1]**	-3 [-6, 1]
Age	**40 [22, 56]**	**40 [24, 55]**	**35 [23, 47]**
Sex: Male	165 [-182, 460]	165 [-123, 506]	198 [-99, 495]
University degree	-157 [-516, 211]	-159 [-558, 210]	-200 [-516, 117]
Residential environment (reference group: city center)			
Outskirts	-132 [-702, 411]	-132 [-708, 411]	25 [-496, 545]
Village	**-710 [-1283, -64]**	**-704 [-1221, -123]**	**-538 [-1033, -43]**
Country side	-478 [-1217, 153]	-468 [-1098, 220]	-358 [-961, 246]
Participation of family member or friend	326 [-164, 690]	329 [-182, 778]	345 [-17, 707]
Subjective health status	68 [-148, 331]	69 [-163, 317]	92 [-121, 305]
Purchase of an activity tracker	**1,066 [733, 1,413]**	**1,068 [702, 1,455]**	**1,004 [708, 1,301]**
Leisure time physical activity	**529 [366, 692]**	**529 [367, 664]**	**483 [348, 619]**
Physical activity on the job	118 [-0.2, 194]	**118 [28, 233]**	87 [-3, 178]
Physical activity on the way to work	**550 [200, 911]**	**555 [207, 882]**	**663 [336, 990]**
Sitting minutes per week	**-0.2 [-0.3, -0.1]**	**-0.18 [-0.3, -0.1]**	**-0.2 [-0.3, -0.07]**
Moderate activities and walking	**0.08 [0.03, 0.1]**	**0.08 [0.03, 0.1]**	**0.09 [0.05, 0.1]**
Incentive condition (reference group: control)			
Cash incentives	273 [-310, 848]	231 [-365, 734]	92 [-416, 600]
Charity incentives	**612 [19, 1270]**	598 [-80, 1151]	382 [-128, 896]
Incentive condition ∗ time (reference group: control ∗ time)			
Cash incentives ∗ time	-4 [-7, 1]	-4 [-7, 0.04]	-3 [-6, 1]
Charity incentives ∗ time	-3 [-7, 1]	**-3 [-7, -0.002]**	-3 [-6, 1]
Random Effects: Standard Deviation			
Participants	**2,396 [2,263, 2,524]**	**2,447 [2,245, 2,480]**	-
Canton	-	91 [<0.0001, 251]	-
Model Fit Parameters			
χ^2 - difference	-	$\chi^2(1) = 0.05$	-
AIC	1,550,608.90	1,550,610.80	-
BIC	1,550,813.20	1,550,824.40	-

Note: Boldface indicates statistical significance (p<0.05). Table depicts point estimates and associated 95% confidence intervals in brackets. AIC, Akaike's information criterion; BIC, Bayesian information criterion.

[a] Pooled results from model 1 fitted to 20 complete datasets. Random effects are not pooled and therefore not reported.

Table H-3. Sensitivity analyses of incentive effects on step goal achievement.

	Dependent variable: daily goal achievement (yes/no)		
	Model 1 (*n* = 1,005)	**Model 2 (*n* = 1,005)**	**Model 3[a] (*n* = 1,223)**
Fixed Effects			
(Intercept)	**-2.78 [-3.56, -1.88]**	**-2.78 [-3.46, -1.95]**	**-2.48 [-3.23, -1.73]**
Time	-0.002 [-0.004, 0.001]	-0.002 [-0.004, 0.0003]	-0.0003 [-0.002, 0.002]
Age	**0.03 [0.02, 0.03]**	**0.03 [0.02, 0.03]**	**0.02 [0.02, 0.03]**
Sex: Male	-0.04 [-0.21, 0.12]	-0.04 [-0.24, 0.15]	-0.03 [-0.21, 0.14]
University degree	-0.13 [-0.32, 0.07]	-0.13 [-0.33, 0.10]	-0.13 [-0.32, 0.06]
Residential environment (reference group: city center)			
Outskirts	-0.01 [-0.34, 0.24]	-0.01 [-0.35, 0.32]	0.05 [-0.26, 0.35]
Village	**-0.34 [-0.71, 0.01]**	**-0.34 [-0.63, -0.04]**	**-0.24 [-0.53, 0.05]**
Country side	-0.12 [-0.54, 0.25]	-0.12 [-0.52, 0.34]	-0.07 [-0.42, 0.28]
Participation of family member or friend	0.24 [-0.05, 0.47]	0.24 [-0.01, 0.50]	**0.24 [0.03, 0.45]**
Subjective health status	0.01 [-0.14, 0.16]	0.01 [-0.11, 0.16]	0.02 [-0.10, 0.15]
Purchase of an activity tracker	**0.64 [0.43, 0.78]**	**0.64 [0.44, 0.78]**	**0.56 [0.39, 0.74]**
Leisure time physical activity	**0.26 [0.17, 0.35]**	**0.26 [0.17, 0.34]**	**0.24 [0.16, 0.32]**
Physical activity on the job	**0.06 [0.004, 0.11]**	**0.06 [0.02, 0.11]**	0.04 [-0.01, 0.09]
Physical activity on the way to work	**0.23 [0.05, 0.42]**	**0.23 [0.02, 0.42]**	**0.29 [0.10, 0.48]**
Sitting minutes per week	**< -0.0001 [< -0.0001, < -0.0001]**	**-0.0001 [-0.0002, < -0.0001]**	**-0.0001 [-0.0002, < -0.0001]**
Moderate activities and walking	< 0.0001 [< -0.0001, 0.0001]	< 0.0001 [< -0.0001, 0.0001	**< 0.0001 [< 0.0001, 0.0001]**
Incentive condition (reference group: control)			
Cash incentives	0.27 [-0.06, 0.58]	0.27 [-0.02, 0.58]	0.18 [-0.12, 0.48]
Charity incentives	**0.52 [0.21, 0.89]**	**0.52 [0.26, 0.80]**	**0.37 [0.07, 0.67]**
Incentive condition * time (reference group: control x time)			
Cash incentives * time	**-0.004 [-0.01, -0.001]**	**-0.004 [-0.006, -0.001]**	**-0.004 [-0.01, -0.002]**
Charity incentives * time	**-0.003 [-0.01, -0.0004]**	**-0.003 [-0.01, -0.0004]**	**-0.003 [-0.01, -0.001]**
Random Effects: Standard Deviation			
Participants	1.43 [1.34, 1.48]	1.43 [1.35, 1.49]	-
Canton	-	< 0.0001 [< 0.0001, 0.14]	-
Model Fit Parameters			
χ^2 - difference	-	$\chi^2(1) = 0.00$	-
AIC	86,541.90	86,543.90	-
BIC	86,736.90	86,748.20	-

Note: Boldface indicates statistical significance (*p<0.05). Table depicts point estimates and associated 95% confidence intervals on the logit scale. AIC, Akaike's information criterion; BIC, Bayesian information criterion.
[a] Pooled results from fitting model 1 to 20 complete datasets. Random effects are not pooled and therefore not reported.

Appendix I - Study I: Analysis of Program Engagement

Table I-1. Sensitivity analyses of incentive effects on non-usage attrition.

	Dependent variable: non-usage attrition (yes/no)	
	Model 1 (*n* = 1,547)	**Model 2 (*n* = 1,160)**
Age		**-0.04 [-0.06, -0.02]**
Sex: male		-0.12 [-0.5, 0.26]
Purchase of an activity tracker		**-0.45 [-0.82, -0.08]**
Moderate activities and walking		<-0.0001 [<-0.0001, <0.0001]
Incentive condition (reference group: control)		
Cash incentives	-0.28 [-0.79, 0.23]	-0.09 [-0.74, 0.56]
Charity incentives	0.08 [-0.43, 0.59]	0.07 [-0.58, 0.72]

Note: Boldface indicates statistical significance ($p<0.05$). Table depicts point estimates and associated 95% confidence intervals on the logit scale.

Appendix J - Physical Activity Tips Included in the Ally App

Table J-1. Physical activity tips included in the Ally app.

No.	Tip
1	Drinking water throughout the day is good way to stay healthy. Plus, going to the toilet more often will boost your step count.
2	Whenever you can, use the stairs instead of the elevator.
3	Try to collect some more steps during daily activities like listening to music, making a phone call, or brushing your teeth.
4	Incorporate a detour into your daily walking routines to make some more steps
5	Think ahead what you have planned for the rest of your day. If you have five minutes to spare, take a five-minute walk.
6	Studies show that prolonged sitting is bad for your health. Try to interrupt sedentary periods and collect some steps.
7	If you travel from A to B, try to use public transport. Walking to and from stations will significantly increase your step count. If you have to take the car, park further away from your destination to collect some more steps
8	Tell a friend about your step goal today. If you dare, make a bet! Whoever loses has to pay for the next coffee.
9	Coffee breaks are a good opportunity to collect some steps. Just order your next coffee ‘to-go’.
10	Feeling stressed? Physical activity is a great way to relax. Take a walk today to clear your head and refill your energy.
11	If you are travelling with public transport, try to get off a stop or two earlier and walk the rest. That’s a great way to boost your step count.
12	Are you up for a challenge? Check the app’s dashboard and try to beat your step counts from last week! If you know a friend who is also counting steps: challenge them and see who collects more steps!
13	Did you know? When you are active, your body releases transmitters like dopamine and beta-endorphin who can considerably boost your mood! So, take a walk to brighten up your day!
14	Lunch breaks are a great opportunity to collect some steps. Plan to make 500 additional steps during your next lunch break. This will only take 5 minutes!
15	When you are at work, try to avoid sending emails to people who sit close to you. Just walk over and discuss matters personally. You not only collect some steps but also get a faster reply.
16	Make a list of all the reasons that motivate you to be more active. Place it on a spot you visit regularly, e.g. your desk at work, to stay motivated.
17	Change meetings at work to ‘walk-meetings’ where you discuss things while collecting some steps.
18	Most people have some spare time in the evening. Take a brief walk tonight. This helps you to reach your goal and is a good opportunity to unwind and reflect on your day.

Appendix K - Study II: Statistical Analyses Details

Multiple Imputation

Missing data was assumed to be missing at random and multiple imputations by chained equations (van Buuren & Groothuis-Oudshoorn, 2011) was used to impute missing data in order to perform an intention-to-treat analysis. Methodologists currently regard multiple imputation as a state-of-the-art technique because it improves accuracy and statistical power relative to other missing data techniques (Van Buuren, 2012). The percentage of missing data across variables in the data frame varied between 0% and 27.7% and most missing data occurred with regard to step count measurements. In total, 84 out of 274 participants (30.7%) had missing data. Missing step data occurred more often on weekends versus weekdays and for Android users compared to iOS users.

To make the missing at random assumption plausible, an imputation model for each variable with missing data was specified that included predictors, which were selected based on their relationship to the target variable and their ability to predict missingness. Specifically, the imputation models for step goal achievement and steps per day included all variables that predicted either step goal achievement or missingness in a univariate generalized mixed model with $p < .20$. Consequently, imputation models for step goal achievement and steps per day included baseline characteristics (gender, income, self-reported physical activity, steps per day and subjective health status), psychological measures (motivation, self-efficacy, action control) and indicators of app usage in addition to variables used in the statistical models (see below). Predictors in the imputation models of the remaining covariates were selected using the quickpred() function of the mice package in R. This function includes all variables as predictors in imputation models that have a minimum correlation of $r \geq .10$ with the target variable. Two-level (generalized) linear mixed models were used to impute missing data for step goal achievement and steps per day respectively and predictive mean matching was used to impute missing data on the remaining level 2 covariates.

Using 20 iteration cycles for imputation, 10 multiply imputed datasets were created. Step counts < 1000 were set to missing before starting the imputation process because previous research suggests that days with fewer than 1000 steps are unlikely to represent valid step count measurements (Rowe, Mahar, Raedeke, & Lore, 2004). Probability densities of observed and imputed data were compared to judge the quality of imputations and imputations were considered plausible. Visual inspection of trace lines (Figure 1) indicated convergence of the imputation algorithm (trace lines intermingle and show no apparent trend (van Buuren & Groothuis-Oudshoorn, 2011)). The intra-class correlations (ICCs) in the imputed datasets resembled the original ICC, suggesting that the imputations successfully recovered the hierarchical structure of the data.

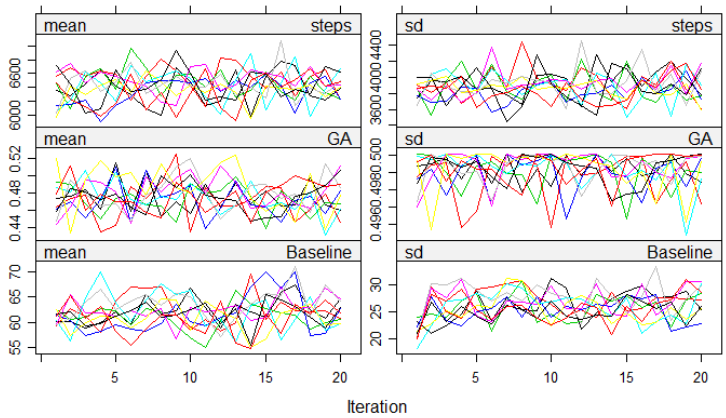


Figure K-1. Mean and standard deviations of imputations for steps, step goal achievement and baseline step count across all datasets and iteration cycles.

Statistical Analysis

For each main effect and interaction, effects were estimated in an unadjusted complete case analysis as well as in two sensitivity analyses that adjusted for missing data (intention-to-treat analysis) and for covariates of physical activity (adjusted intention-to-treat analysis). The complete case analyses included data from $N = 217$ participants for incentives and planning interventions and data from $N = 226$ participants for the analysis of physical activity prompts. Numbers of participants differ slightly, because participants with less than 3 days of activity data were excluded when aggregating the outcome as described in the main text (Tudor-Locke et al., 2005).

Incentives

The total proportion of step goals achieved and the number of steps per day averaged over the complete intervention period were the primary and secondary outcome respectively in the analysis of incentive effects. Using a linear regression framework, the following adjusted model was fit to the data to estimate the effects of cash and charity incentives reported in the paper:

$$Y_i = \beta_0 + \beta_1(Age_i) + \beta_2(Gender_i) + \beta_3(Baseline\ Steps/100_i) + \beta_4(Device_i) + \beta_5(Employment_i) + \beta_6(Cash_i) + \beta_7(Charity_i) \quad \textbf{(1)}$$

for $i = 1, \ldots, 274$

where

- Y_i is the proportion of step goals achieved/the number of steps per day averaged over the complete intervention period for participant i,
- Age_i indicates the mean-centered age of participant i,
- $Gender_i$ indicates the gender of participant i (0: female, 1: male),
- $Baseline\ steps/100_i$ indicates the mean-centered average daily step count at baseline of participant i (divided by 100 for better interpretability of coefficients),
- $Employment_i$ indicates whether participant i is employed (1) or not (0) at the time of the study,
- $Cash_i$ indicates whether participant i was randomized to the cash incentive group (1) or not (0) and
- $Charity_i$ indicates whether participant i was randomized to the charity incentive group (1) or not (0).

Note that $Cash_i = Charity_i = 0$ when the participant was randomized to the incentive control group at the beginning of the study.

Planning

For the analysis of micro-randomized intervention components, the approach by Boruvka et al. (Boruvka et al., 2017) for analyzing data from micro-randomized trials was used. Boruvka et al. recommend an analysis with weighted and centered treatment indicators that guarantees robust and unbiased causal effect estimates when treatments and potential moderators are time-varying. In our case, this strategy simplifies to a regular regression analysis using generalized estimating equations (GEE) without weighted treatment indicators (Zeger, Liang, & Albert, 1988) because all randomization probabilities are constant and time-varying covariates are not impacted by prior treatment. Similar to mixed models, GEE models are able to account for within-person correlations of longitudinal outcome data by using robust standard errors.

The weekly proportion of step goals achieved and the weekly average number of steps per day, calculated separately for weeks one to six of the intervention period, were the primary and secondary outcome respectively in the analysis of planning effects. Using generalized estimating equations, the following adjusted model was fit to the data to estimate the effects of action planning and coping planning reported in the paper:

$$\begin{aligned} Y_{it} = {} & \beta_0 + \beta_1(\text{Week in study}_{it}) + \beta_2(\text{Age}_i) + \beta_3(\text{Gender}_i) + \\ & \beta_4(\text{Baseline steps}/100_i) + \beta_5(\text{Device}_i) + \beta_6(\text{Employment}_i) + \\ & \beta_7(\text{Action planning}_{it}) + \beta_8(\text{Coping planning}_{it}) \end{aligned} \qquad (\mathbf{2})$$

for $i = 1, \ldots, 274$ and $t = 0, \ldots, 5$

where

- Y_{it} indicates the proportion of step goals achieved / the average steps per day of participant i in week t,
- $Week\ in\ study_{it}$ is the index for week in study for participant i with t ranging from 0 to 5,
- $Action\ planning_{it}$ indicates whether participant i was randomized to receiving (1) or not receiving (0) an action planning intervention in week t,
- $Coping\ planning_{it}$ indicates whether participant i was randomized to receiving (1) or not receiving (0) a coping planning intervention in week t and

all other model components are the same as in (1). Note that $Action\ planning_{it} = Coping\ planning_{it} = 0$ when a participant was randomized to receiving no planning intervention in week *t.* To investigate whether the effect of planning interventions is enhanced by incentives, the interaction between planning interventions and incentives was added to the model. To investigate the time-varying effect of planning interventions, the interactions between the planning terms and week in study was added to the model.

Physical activity prompts

The binary indicator of daily step goal achievement and number of steps per day were the primary and secondary outcome respectively in the analysis of physical activity prompts. As outlined in the main paper, physical activity prompts were not delivered on Sundays and these days were excluded from the analysis when estimating the treatment effect of physical activity prompts. Using generalized estimating equations, the following adjusted model was fit to the data to estimate the effect of physical activity prompts reported in the paper:

$$Y_{it} = \beta_0+\beta_1(\mathrm{Day_{it}})+\beta_2(\mathrm{Weekend_{it}})+\beta_3(\mathrm{Age_i})+\beta_4(\mathrm{Gender_i})+ \beta_5(\mathrm{Baseline\ steps/100_i})+\beta_6(\mathrm{Device_i})+ \beta_7(\mathrm{Employment_i})+\beta_8(\mathrm{Prompt_{it}}) \qquad \mathbf{(3)}$$

for $i = 1, \ldots, 274$ and $t = 0, \ldots, 35$

where

- Y_{it} is the logit of the probability of step goal achievement/the number of steps on day t for participant i,
- Day_{it} is the index for day in study for participant i with t ranging from 0 to 41,
- $Weekend_{it}$ indicates whether day t for participant i is a weekend (1) or a weekday (0),
- $Prompt_{it}$ indicates whether participant i was randomized to receiving (1) or not receiving (0) a physical activity prompt at day t and

all other model components are the same as in (1). Like for planning interventions, the interaction between incentives with physical activity prompts was added to the model to investigate whether incentives enhance the effect of physical activity prompts. To investigate the time-varying effect of physical activity prompts, the interaction between physical activity prompts and day in study was added to the model.

Exploratory Analysis

For the exploratory analyses of notification dependent intervention components, i.e. physical activity prompts and planning interventions, the treatment variables were recoded into a new variable that differentiated whether a participant reacted to the treatment notification or not. A reaction to a treatment notification was defined as a response by the participant to the conversation initiated by the chatbot that triggered the respective notification. Consequently, in the statistical model of the exploratory analysis the effect of the physical activity prompts is now represented by two dummy-coded variables:

- $Prompt - reacted_{it}$ takes the value 1 if participant i was randomized to receiving a physical activity prompt on day t and reacted to the respective notification or takes the value 0 otherwise and

- $Prompt - not\ reacted_{it}$ takes the value 1 if participant i was randomized to receiving a physical activity prompt on day *t* and did not react to the respective notification or takes the value 0 otherwise.

Note that $Prompt - reacted_{it} = Prompt - not\ reacted_{it} = 0$ when the participant was randomized to not receiving a physical activity prompt on day *t*. Likewise, the effect of the action planning (AP) and coping planning (CP) intervention is now represented by two dummy-coded variables each:

- $AP - reacted_{it}$ takes the value 1 if participant *i* was randomized to receiving an action planning intervention in the week of day t and reacted to the respective notification or takes the value 0 otherwise,
- $AP - not\ reacted_{it}$ takes the value 1 if participant i was randomized to receiving a action planning intervention in the week of day *t* and did not react to the respective notification or takes the value 0 otherwise,
- $CP - reacted_{it}$ takes the value 1 if participant *i* was randomized to receiving a coping planning intervention in the week of day t and reacted to the respective notification or takes the value 0 otherwise,
- $CP - not\ reacted_{it}$ takes the value 1 if participant i was randomized to receiving a coping planning intervention in the week of day *t* and did not react to the respective notification or takes the value 0 otherwise.

Note that $AP - reacted_{it} = AP\ not\ reacted_{it} = CP\ reacted_{it} = CP\ not\ reacted_{it} = 0$ when the participant was randomized to not receiving an action or coping planning intervention in the week of day *t*. The relationship between engaging with a treatment conversation and subsequent physical activity is likely confounded by variables affecting both, the selection into treatment and the likelihood of step goal achievement. For example, participants who are more motivated to increase their activity levels may have a higher likelihood of reaching daily step goals and may simultaneously be more likely to react to intervention notifications. To adjust for possible confounding, the effect estimates were adjusted for covariates of physical activity.

Appendix L - Study II: Result Tables

Table L-1. Linear models of incentive main effects on step goal achievement.

	Dependent variable: total proportion of step goals achieved		
	Model 1 (*n* = 217)	Model 2 (*n* = 274)	Model 3 (*n* = 274)
(Intercept)	**0.533 [0.489, 0.577]**	**0.508 [0.461, 0.555]**	**0.672 [0.583, 0.761]**
Age			**0.003 [0.001, 0.005]**
Sex: male			-0.036 [-0.089, 0.018]
Baseline PA / 100			0.0004 [-0.001, 0.001]
Operating system: iOS			**-0.068 [-0.12, -0.016]**
Employment			**-0.117 [-0.192, -0.041]**
Cash incentives	**0.081 [0.021, 0.141]**	**0.092 [0.029, 0.155]**	**0.089 [0.030, 0.148]**
Charity incentives	**0.069 [0.010, 0.128]**	0.054 [-0.010, 0.117]	0.050 [-0.009, 0.109]

Note. Boldface indicates statistical significance ($p < .05$). Table depicts point estimates and associated 95% confidence intervals. Model 1: complete case analysis, Model 2: intention-to-treat analysis, Model 3: adjusted intention-to-treat analysis. PA: physical activity.

Table L-2. Linear models of incentive main effects on step count.

	Dependent variable: total mean steps per day		
	Model 1 (n = 217)	Model 2 (n = 274)	Model 3 (n = 274)
(Intercept)	**6599 [5932, 7267]**	**6476 [5789, 7162]**	**6580 [5334, 7826]**
Age			17 [-9, 44]
Sex: male			-292 [-1020, 437]
Baseline PA / 100			**64 [52, 76]**
Operating system: iOS			162 [-541, 865]
Employment			-134 [-1092, 825]
Cash incentives	783 [-135, 1701]	877 [-91, 1844]	887 [-50, 1823]
Charity incentives	602 [-305, 1509]	553 [-330, 1435]	635 [-154, 1424]

Note. Boldface indicates statistical significance ($p < .05$). Table depicts point estimates and associated 95% confidence intervals. Model 1: complete case analysis, Model 2: intention-to-treat analysis, Model 3: adjusted intention-to-treat analysis. PA: physical activity

Table L-3. GEE models of physical activity prompt main effects (models 1-3) and effects by incentive group (models 4-6) on step goal achievement.

	Dependent variable: daily step goal achievement					
	Model 1 (n = 217)	Model 2 (n = 274)	Model 3 (n = 274)	Model 4 (n = 217)	Model 5 (n = 274)	Model 6 (n = 274)
(Intercept)	**0.38 [0.28, 0.49]**	**0.22 [0.11, 0.33]**	**1.12 [0.71, 1.54]**	**0.20 [0.03, 0.36]**	0.05 [-0.11, 0.21]	**0.94 [0.51, 1.37]**
Day in study			**-0.006 [-0.01, -0.002]**			**-0.006 [-0.01, -0.002]**
Weekend			**-0.17 [-0.32, -0.03]**			**-0.17 [-0.32, -0.03]**
Age			**0.01 [0.004, 0.02]**			**0.01 [0.004, 0.02]**
Sex: male			-0.21 [-0.43, 0.002]			-0.18 [-0.39, 0.03]
Baseline PA / 100			**0.006 [0.001, 0.01]**			**0.006 [0.001, 0.01]**
Operating system: iOS			-0.23 [-0.46, 0.01]			-0.23 [-0.47, 0.002]
Employment			**-0.58 [-0.93, -0.22]**			**-0.56 [-0.92, -0.21]**
Cash incentives				**0.28 [0.03, 0.54]**	**0.31 [0.07, 0.54]**	**0.30 [0.07, 0.52]**
Charity incentives				0.27 [0.01, 0.53]	0.21 [-0.05, 0.46]	0.20 [-0.04, 0.434]
Prompt	0.04 [-0.05, 0.13]	0.07 [-0.02, 0.15]	0.07 [-0.02, 0.16]	-0.0005 [-0.18, 0.17]	0.04 [-0.11, 0.20]	0.04 [-0.13, 0.20]
Prompt ∗ cash inc.				0.10 [-0.13, 0.32]	0.08 [-0.17, 0.32]	0.08 [-0.17, 0.33]
Prompt ∗ charity inc.				0.03 [-0.20, 0.26]	-0.001 [-0.22, 0.22]	0.01 [-0.21, 0.24]

Note: Boldface indicates statistical significance ($p < .05$). Table depicts point estimates and associated 95% confidence intervals on the logit scale. Models 1&4: complete case analysis, models 2&5: intention-to-treat analysis, models 3&6: adjusted intention-to-treat analysis. PA: physical activity.

Table L-4. GEE models of physical activity prompt main effects (models 1-3) and effects by incentive group (models 4-6) on steps per day.

	Dependent variable: steps per day					
	Model 1 (n = 217)	Model 2 (n = 274)	Model 3 (n = 274)	Model 4 (n = 217)	Model 5 (n = 274)	Model 6 (n = 274)
(Intercept)	**7464 [7081, 7847]**	**7108 [6732, 7484]**	**7283 [6174, 8391]**	**6850 [6241, 7460]**	**6678 [6014, 7342]**	**6768 [5439, 8097]**
Day in study			3 [-5, 11]			3 [-5, 11]
Weekend			91 [-152, 334]			91 [-152, 335]
Age			18 [-10, 46]			18 [-9, 46]
Sex: male			-372 [-1115, 372]			-317 [-1084, 449]
Baseline PA / 100			**62 [50, 75]**			**62 [50, 74]**
Operating system: iOS			116 [-613, 844]			108 [-611, 827]
Employment			-199 [-1203, 805]			-157 [-1149, 835]
Cash incentives				834 [-78, 1746]	722 [-265, 1708]	730 [-273, 1732]
Charity incentives				**968 [69, 1867]**	552 [-287, 1392]	635 [-130, 1399]
Prompt	43 [-114, 200]	31 [-107, 168]	32 [-106, 169]	108 [-191, 406]	48 [-218, 314]	52 [-214, 318]
Prompt ∗ cash inc.				-67 [-458, 325]	-17 [-382, 348]	-18 [-383, 346]
Prompt ∗ charity inc.				-135 [-538, 267]	-34 [-413, 345]	-41 [-420, 338]

Note: Boldface indicates statistical significance (p < .05). Table depicts point estimates and associated 95% confidence intervals. Models 1&4: complete case analysis, models 2&5: intention-to-treat analysis, models 3&6: adjusted intention-to-treat analysis. PA: physical activity.

Table L-5. GEE models of the time-varying effect of physical activity prompts on step goal achievement.

	Dependent variable: daily step goal achievement		
	Model 1 (n = 217)	Model 2 (n = 274)	Model 3 (n = 274)
(Intercept)	**0.616 [0.455, 0.778]**	**0.423 [0.258, 0.587]**	**1.178 [0.736, 1.62]**
Day in study	-0.011 [-0.02, -0.01]	**-0.01 [-0.02, -0.004]**	**-0.009 [-0.01, -0.003]**
Weekend			**-0.173 [-0.32, -0.026]**
Age			**0.011 [0.004, 0.018]**
Sex: male			-0.212 [-0.426, 0.003]
Baseline PA / 100			**0.006 [0.001, 0.01]**
Operating system: iOS			-0.224 [-0.463, 0.014]
Employment			**-0.576 [-0.93, -0.222]**
Prompt	-0.153 [-0.34, 0.03]	-0.071 [-0.26, 0.117]	-0.044 [-0.234, 0.146]
Prompt ∗ day in study	**0.01 [0.002, 0.017]**	0.007 [-0.001, 0.014]	0.005 [-0.002, 0.013]

Note. Boldface indicates statistical significance ($p < .05$). Table depicts point estimates and associated 95% confidence intervals on the logit scale. Model 1: complete case analysis, Model 2: intention-to-treat analysis, Model 3: adjusted intention-to-treat analysis. PA: physical activity.

Table L-6. GEE models of the time-varying effect of physical activity prompts on steps per day.

	Dependent variable: steps per day		
	Model 1 (*n* = 217)	Model 2 (*n* = 274)	Model 3 (*n* = 274)
(Intercept)	**7282 [6840, 7724]**	**7082 [6632, 7533]**	**7321 [6200, 8443]**
Day in study	9 [-4, 22]	1 [-10, 12]	1 [-9, 12]
Weekend			91 [-152, 334]
Age			18 [-10, 46]
Sex: male			-371 [-1115, 372]
Baseline PA / 100			**62 [50, 75]**
Operating system: iOS			117 [-611, 845]
Employment			-198 [-1202, 807]
Prompt	-43 [-407, 321]	-60 [-378, 259]	-49 [-337, 240]
Prompt ∗ day in study	4 [-12, 20]	4 [-10, 19]	4 [-9, 17]

Note. Boldface indicates statistical significance ($p < .05$). Table depicts point estimates and associated 95% confidence intervals. Model 1: complete case analysis, Model 2: intention-to-treat analysis, Model 3: adjusted intention-to-treat analysis. PA: physical activity.

Table L-7. GEE models of planning exercise main effects (models 1-3) and effects by incentive group (models 4-6) on step goal achievement.

	Dependent variable: weekly proportion of step goals achieved					
	Model 1 (*n* = 217)	Model 2 (*n* = 274)	Model 3 (*n* = 274)	Model 4 (*n* = 217)	Model 5 (*n* = 274)	Model 6 (*n* = 274)
(Intercept)	**0.582 [0.553, 0.611]**	**0.55 [0.521, 0.58]**	**0.731 [0.651, 0.812]**	**0.519 [0.478, 0.56]**	**0.482 [0.436, 0.527]**	**0.659 [0.57, 0.747]**
Week in study			**-0.01 [-0.02, -0.002]**			**-0.009 [-0.02, -0.002]**
Age			**0.003 [0.001, 0.005]**			**0.003 [0.001, 0.005]**
Sex: male			-0.048 [-0.098, 0.002]			-0.041 [-0.09, 0.008]
Baseline PA / 100			**0.001 [0, 0.002]**			**0.001 [0, 0.002]**
Operating system: iOS			-0.05 [-0.104, 0.004]			-0.051 [-0.104, 0.002]
Employment			**-0.124 [-0.20, -0.052]**			**-0.12 [-0.19, -0.049]**
Cash incentives				**0.103 [0.037, 0.169]**	**0.12 [0.052, 0.188]**	**0.117 [0.054, 0.181]**
Charity incentives				**0.082 [0.016, 0.147]**	**0.084 [0.014, 0.153]**	**0.082 [0.018, 0.146]**
AP	0.011 [-0.02, 0.041]	0.018 [-0.011, 0.047]	0.018 [-0.011, 0.047]	0.033 [-0.018, 0.083]	**0.056 [0.01, 0.102]**	**0.058 [0.012, 0.104]**
CP	-0.003 [-0.031, 0.025]	-0.006 [-0.036, 0.023]	-0.006 [-0.036, 0.023]	0.011 [-0.034, 0.057]	0.03 [-0.013, 0.074]	0.032 [-0.012, 0.076]
AP ∗ cash incentives				-0.052 [-0.126, 0.023]	-0.069 [-0.14, 0.002]	**-0.071 [-0.14, -0.001]**
AP ∗ charity incentives				-0.01 [-0.083, 0.063]	-0.045 [-0.114, 0.024]	-0.049 [-0.118, 0.02]
CP ∗ cash incentives				-0.014 [-0.08, 0.053]	-0.046 [-0.112, 0.019]	-0.049 [-0.115, 0.018]
CP ∗ charity incentives				-0.028 [-0.095, 0.039]	-0.062 [-0.126, 0.002]	-0.063 [-0.127, 0.002]

Note: Boldface indicates statistical significance (p < .05). Table depicts point estimates and associated 95% confidence intervals. Models 1&4: complete case analysis, models 2&5: intention-to-treat analysis, models 3&6: adjusted intention-to-treat analysis; PA: physical activity , AP: action planning, CP: coping planning.

Table L-8. GEE models of planning exercise main effects (models 1-3) and effects by incentive group (models 4-6) on steps per day.

	Dependent variable: weekly average of steps per day					
	Model 1 (n = 217)	Model 2 (n = 274)	Model 3 (n = 274)	Model 4 (n = 217)	Model 5 (n = 274)	Model 6 (n = 274)
(Intercept)	**7285 [6889, 7680]**	**6962 [6571, 7354]**	**7098 [5977, 8219]**	**6755 [6108, 7401]**	**6479 [5779, 7180]**	**6530 [5202, 7857]**
Week in study			31 [-23, 85]			32 [-22, 86]
Age			17 [-10, 45]			18 [-10, 45]
Sex: male			-406 [-1125, 314]			-343 [-1083, 397]
Baseline PA / 100			**63 [50, 75]**			**63 [50, 75]**
Operating system: iOS			182 [-530, 895]			174 [-529, 877]
Employment			-184 [-1207, 840]			-139 [-1147, 868]
Cash incentives				908 [-39, 1856]	885 [-144, 1914]	882 [-148, 1913]
Charity incentives				633 [-311, 1578]	552 [-319, 1423]	626 [-178, 1431]
AP	101 [-163, 366]	148 [-67, 364]	148 [-67, 363]	-38 [-488, 412]	142 [-236, 520]	134 [-244, 513]
CP	-113 [-351, 125]	-44 [-242, 155]	-44 [-243, 154]	-139 [-585, 308]	86 [-274, 446]	82 [-278, 442]
AP ∗ cash incentives				-105 [-721, 510]	-113 [-621, 394]	-106 [-614, 402]
AP ∗ charity incentives				538 [-143, 1219]	122 [-386, 631]	136 [-373, 645]
CP ∗ cash incentives				-98 [-682, 487]	-206 [-682, 271]	-197 [-673, 279]
CP ∗ charity incentives				176 [-438, 789]	-178 [-667, 311]	-176 [-664, 313]

Note: Boldface indicates statistical significance (p < .05). Table depicts point estimates and associated 95% confidence intervals. Models 1&4: complete case analysis, models 2&5: intention-to-treat analysis, models 3&6: adjusted intention-to-treat analysis; PA: physical activity, AP: action planning, CP: coping planning.

Table L-9. GEE models of time-varying effects of planning exercises on step goal achievement.

	Dependent variable: weekly proportion of step goals achieved		
	Model 1 (n = 217)	Model 2 (n = 274)	Model 3 (n = 274)
(Intercept)	**0.608 [0.565, 0.651]**	**0.575 [0.532, 0.617]**	**0.732 [0.647, 0.818]**
Week in study	-0.011 [-0.025, 0.003]	-0.01 [-0.024, 0.004]	-0.009 [-0.022, 0.005]
Age			**0.003 [0.001, 0.004]**
Sex: male			-0.048 [-0.098, 0.001]
Baseline PA / 100			0.001 [>-0.0001, 0.002]
Operating system: iOS			-0.05 [-0.104, 0.004]
Employment			**-0.124 [-0.196, -0.052]**
AP	0.011 [-0.052, 0.073]	0.017 [-0.047, 0.082]	0.023 [-0.037, 0.083]
CP	-0.02 [-0.078, 0.038]	-0.015 [-0.074, 0.044]	-0.013 [-0.069, 0.042]
AP ∗ week in study	0.0002 [-0.022, 0.023]	0.0001 [-0.022, 0.023]	-0.002 [-0.023, 0.019]
CP ∗ week in study	0.007 [-0.014, 0.028]	0.003 [-0.018, 0.025]	0.003 [-0.018, 0.023]

Note: Boldface indicates statistical significance (p < .05). Table depicts point estimates and associated 95% confidence intervals. Model 1: complete case analysis, Model 2: intention-to-treat analysis, Model 3: adjusted intention-to-treat analysis. PA: physical activity, AP: action planning, CP: coping planning.

Table L-10. GEE models of time-varying effects of planning exercises on steps per day.

	Dependent variable: weekly average of steps per day		
	Model 1 (n = 217)	Model 2 (n = 274)	Model 3 (n = 274)
(Intercept)	**7337 [6787, 7887]**	**6971 [6391, 7550]**	**7150 [5953, 8346]**
Week in study	-22 [-191, 148]	-3 [-158, 152]	11 [-136, 157]
Age			17 [-11, 45]
Sex: male			-406 [-1128, 316]
Baseline PA / 100			**63 [50, 75]**
Operating system: iOS			181 [-531, 893]
Employment			-185 [-1207, 838]
AP	-326 [-1047, 395]	-61 [-748, 625]	73 [-532, 677]
CP	-357 [-1014, 300]	-90 [-772, 591]	-120 [-786, 545]
AP ∗ week in study	176 [-98, 450]	84 [-164, 332]	30 [-186, 247]
CP ∗ week in study	102 [-169, 373]	19 [-247, 284]	30 [-225, 286]

Note: Boldface indicates statistical significance ($p < .05$). Table depicts point estimates and associated 95% confidence intervals. Model 1: complete case analysis, Model 2: intention-to-treat analysis, Model 3: adjusted intention-to-treat analysis. PA: physical activity, AP: action planning, CP: coping planning.

Table L-11. GEE models of exploratory analysis with recoded treatment indicators for physical activity prompts.

	Dependent variable:	
	Daily step goal achievement (*n* = 188)	Steps per day (*n* = 188)
(Intercept)	**1.408 [0.873, 1.943]**	**7894 [7086, 8702]**
Day in study	**-0.005 [-0.009, -0.001]**	8 [-2, 18]
Weekend (yes/no)	-0.082 [-0.284, 0.12]	210 [-151, 571]
Age	**0.014 [0.004, 0.024]**	**32 [10, 54]**
Sex: male	**-0.244 [-0.46, -0.028]**	**-636 [-1201, -71]**
Baseline PA / 100	**0.007 [0.003, 0.011]**	**78 [69, 87]**
Operating system: iOS	**-0.279 [-0.532, -0.026]**	-227 [-863, 409]
Employment (yes/no)	**-0.673 [-1.161, -0.185]**	-194 [-936, 548]
Prompt – not reacted	**-0.394 [-0.545, -0.243]**	**-754 [-1077, -431]**
Prompt - reacted	**0.26 [0.123, 0.397]**	**405 [189, 621]**

Note: Boldface indicates statistical significance ($p < .05$). Table depicts point estimates and associated 95% confidence intervals. PA: physical activity.

Table L-12. GEE models of exploratory analysis with recoded treatment indicators for planning exercises.

	Dependent variable:	
	Weekly proportion of step goals achieved (n = 182)	Weekly average of steps per day (n = 182)
(Intercept)	**0.766 [0.676, 0.856]**	**7725 [6957, 8493]**
Week in study	-0.005 [-0.013, 0.003]	65 [-6, 136]
Age	**0.003 [0.001, 0.005]**	**32 [12, 52]**
Sex: male	**-0.055 [-0.102, -0.008]**	**-670 [-1199, -141]**
Baseline PA / 100	**0.001 [0.001, 0.001]**	**77 [67, 87]**
Operating system: iOS	**-0.06 [-0.113, -0.007]**	-145 [-755, 465]
Employment (yes/no)	**-0.13 [-0.216, -0.044]**	-226 [-930, 478]
Action planning - not reacted	-0.02 [-0.065, 0.025]	-250 [-701, 201]
Action planning - reacted	**0.039 [0.002, 0.076]**	**421 [127, 715]**
Coping planning - not reacted	-0.034 [-0.073, 0.005]	**-579 [-942, -216]**
Coping planning - reacted	0.035 [-0.0003, 0.07]	**475 [128, 822]**

Note: Boldface indicates statistical significance ($p < .05$). Table depicts point estimates and associated 95% confidence intervals. PA: physical activity.

Table L-13. Sensitivity analyses of incentive effects on non-usage attrition.

	Dependent variable: non-usage attrition (yes/no)	
	Model 1 (*n* = 1547)	**Model 2 (*n* = 1160)**
Age		-0.002 [-0.03, 0.03]
Sex: male		-0.49 [-1.39, 0.41]
Baseline PA		-0.00006 [-0.0002, 0.00009]
Operating system: iOS		-0.41 [-1.21, 0.39]
Intention to increase PA: yes		-0.37 [-1.30, 0.56]
Incentive condition (reference group: control)		
Cash incentives	0.46 [-0.18, 1.10]	0.38 [-0.52, 1.28]
Charity incentives	-0.45 [-1.25, 0.35]	-0.35 [-1.38, 0.68]

Note: Boldface indicates statistical significance (p<0.05). Table depicts point estimates and associated 95% confidence intervals on the logit scale. PA: physical activity.

Appendix M- Study II: App Engagement and Physical Activity

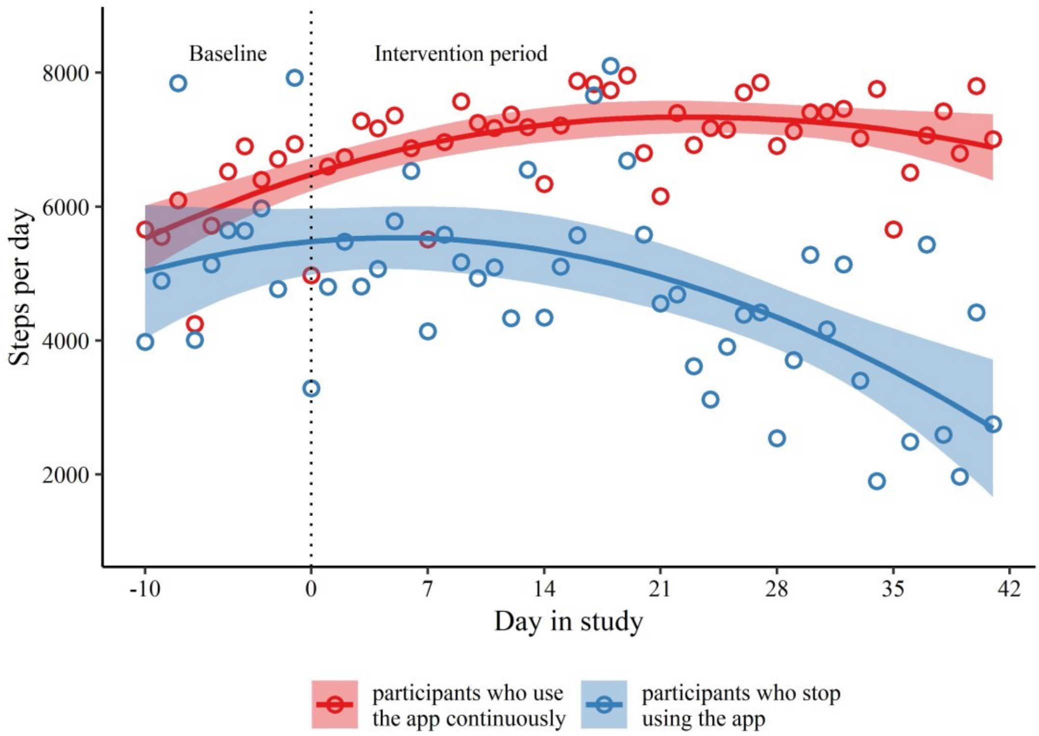


Figure M-1. Steps per day over the course of the study for continuously engaged and non-engaged participants.

Appendix N - Study II: Evaluation of the Ally App

Table N-1. Evaluation of the Ally app.

Construct	**Item**	***M***	***SD***
Overall evaluation (answers given on a 5-point Likert scale)			
Satisfaction	How satisfied are you with the Ally app overall?	3.45	1.15
Technology acceptance (answers given on a 7-point Likert scale)			
Performance expectancy	The Ally app helped me to increase my daily physical activity.	4.89	1.72
Effort expectancy	The Ally app is easy to use.	5.94	1.27
Hedonic motivation	Using the Ally app is fun.	4.78	1.70
Habit	Using the Ally app has become a habit for me.	5.18	1.68
Intention for continued use	If given the chance, I would continue using the Ally app	4.49	1.86
Attitudes towards the digital coach (answers given on a 7-point scale)			
Continuance	How much would you like continue working with Ally? [a]	4.39	1.74
Relationship	How would you describe your relationship with Ally? [b]	4.04	1.88
Preference	Would you rather have talked to a human coach than to Ally? [c]	3.94	2.03
Adherence	How likely is it that you follow Ally's advice? [d]	4.38	1.55

[a] anchor 1: not at all, anchor 7: very much
[b] anchor 1: complete stranger, anchor 7: close friend
[c] anchor 1: definitely to a human coach, anchor 7: definitely to Ally
[d] anchor 1: not at all likely, anchor 7: very likely

Curriculum Vitae

Name	Jan-Niklas Kramer
Date of Birth	19. November 1988 in Berlin, Germany
Contact	janniklas.kramer@googlemail.com

Education

2016 – 2020	*University of St.Gallen, Switzerland* PhD in Business Innovation
2013 – 2015	*Ruprecht-Karls-University Heidelberg, Germany* MSc Psychology
2009 – 2012	*University of Potsdam, Germany* BSc Psychology

Professional Experience

2016 – 2020	*University of St.Gallen, Switzerland* PhD candidate and doctoral researcher
2015	*Robert Bosch GmbH, Germany* Working student, HR Steering & Controlling
2014 – 2015	*Hays AG, Germany* Working student, Marketing & Market Research
2013	*University of Auckland, New Zealand* Research Assistant