# The Future of Medicine: Preventative mHealth Interventions and Health Digitization

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

submitted by
**Annette Barbara Isabell Mönninghoff**
from
Germany

Approved on the application of

**Prof. Dr. Andreas Herrmann**
and
**Prof. Dr. Dr. h.c. Torsten Tomczak**

Dissertation no. 5189

Difo-Druck GmbH, Untersiemau 2022

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

St.Gallen, October 29, 2021

The President:

Prof. Dr. Bernhard Ehrenzeller

“The doctor of the future will give no medicine, but will instruct his patients in care of the human frame, in diet and in the cause and prevention of disease.”

Thomas Edison

# Vorwort

Die Promotion an der Universität St.Gallen war für mich eine inspirierende und lehrreiche Zeit. Gerne möchte ich an dieser Stelle die Gelegenheit nutzen, um mich bei meinen Unterstützern und Weggefährten zu bedanken.

Zuallererst gebührt mein grosser Dank meinem Doktorvater Prof. Dr. Andreas Herrmann für seine Unterstützung und sein entgegengebrachtes Vertrauen. Mit unserem gemeinsamen Buchprojekt hast du mir die Tür zur Welt der digitalen Medizin geöffnet. Du hast mir Freiheiten in meiner Forschung gelassen und gleichzeitig meinen Ehrgeiz immer wieder angestachelt. Ebenfalls möchte ich meinem Korreferenten Prof. Dr. Dr. h.c. Torsten Tomczak für seinen Input und den spannenden, praxisnahen Austausch danken. Zudem danke ich meinen Co-Autoren, zuallererst Prof. Dr. Elgar Fleisch und Prof. Dr. Christoph Franz, für die vielen spannenden und bereichernden gemeinsamen Diskussionen zur Digitalisierung der Medizin und unsere erfolgreiche gemeinsame Buchpublikation. Ferner danke ich Prof. Dr. Simon Mayer, Dr. Tobias Kowatsch, Dr. Jan-Niklas Kramer, Dr. Klaus Fuchs, Kamila Ismailova, Jing Wu, Gisbert Teepe, Alexander Hess, Prof. Dr. Falk Müller-Riemenschneider, Dr. Lorraine Tudor Car und Jan Albert für die Zusammenarbeit an zwei spannenden Forschungsprojekten.

Desgleichen möchte ich meinen Kollegen am Institut für Customer Insight für die wundervolle Zeit danken, in der viele zu echten Freunden geworden sind: Dr. Johanna Hasenmaile und Hannah Leimert, vielen Dank für die vielen gemeinsam gerannten Kilometer bei diversen Läufen und die noch zahlreicheren gemeinsam verbrachten Stunden! Dr. Pia Burghartz, Dr. Lydia Ottlewski, Dr. Beatrice Hügler und Désirée Rehnert, danke für eine immer gut gelaunte Arbeitsatmosphäre in Büro 104. Dr. Matthias Fuchs und Prof. Dr. Emanuel de Bellis, danke für den anregenden inhaltlichen Austausch und die schöne gemeinsam verbrachte Zeit.

Mein ausserordentlicher Dank gilt zudem meiner Familie, die mich in meinem Dissertationsvorhaben immer unterstützt hat. Zuletzt danke ich meinem Mann Dr. Sebastian Mönninghoff, der massgeblich meine Dissertation von der ersten Idee an unterstützt hat. Ich bin unendlich glücklich, Dich an meiner Seite zu haben und freue mich auf die spannende Zeit, die wir nun als Familie vor uns haben.

St. Gallen, Juli 2021 Annette Mönninghoff

# Table of Contents

# Abstract

Rising healthcare costs are pushing healthcare systems to their limits. Non-communicable diseases (NCDs) such as diabetes or cancer cause more than half of healthcare spending. Research shows that NCDs depend on a set of modifiable risk factors and that unhealthy human lifestyles are responsible for 45%–75% of disease risk. Preventative health interventions supporting healthier lifestyles are thus needed to combat the rise in NCDs, to control health care spending, and to future-proof healthcare systems. However, currently only 3% of health spending are allocated to prevention. Moreover, the potential of mobile health (mHealth) technologies for health behavior change remains underresearched.

This dissertation consists of four articles that explore the potential of preventative mHealth interventions in supporting healthier lifestyles. It also investigates how digital technologies can help future-proof healthcare systems.

**Article I** provides a comprehensive summary of the existing research on the short- and long-term effectiveness of mHealth physical activity interventions. It quantitatively summarizes 117 studies in a meta-analysis and finds that mHealth physical activity interventions lead to small to moderate activity increases that are maintained up to 14 months post intervention.

Building on the first article, **Article II** investigates how gamification can increase the long-term effectiveness of preventative mHealth interventions. Based on two case studies, it develops a novel *motivational framework for gamification* and provides guidelines for gamifying activity tracking apps to prevent motivational crowding out.

**Article III** explores the use of future-self avatars to overcome present bias in preventative mHealth interventions. Using a randomized-controlled trial research design, the article finds that the future-self avatar intervention directionally increases physical activity and improves the nutritional quality of food purchases.

**Article IV** takes a broader perspective and explores the impact of digital technologies on future-proofing healthcare systems.

# Zusammenfassung

Steigende Kosten bringen unsere Gesundheitssysteme an ihre Grenzen. Nicht-übertragbare Krankheiten (NCDs) wie Diabetes oder Krebs verursachen mehr als die Hälfte dieser Gesundheitsausgaben. Studien zeigen, dass NCDs von einer Reihe beeinflussbarer Risikofaktoren begünstigt werden, welche für 45 bis 75 Prozent des Krankheitsrisikos verantwortlich sind. Präventive Gesundheitsmaßnahmen, die einen gesünderen Lebensstil unterstützen, sind daher notwendig, um den Anstieg der NCDs zu bekämpfen, die Gesundheitsausgaben zu kontrollieren und unser Gesundheitssystem zukunftssicher zu machen. Derzeit werden jedoch nur drei Prozent aller Ausgaben in Gesundheitsprävention investiert und es fehlen Erkenntnisse, wie digitale Technologien gesundes Verhalten fördern können.

Diese Dissertation beleuchtet das Potenzial von präventiven Mobile-Health-Interventionen (mHealth) in vier Artikeln und untersucht, wie digitale Technologien unser Gesundheitssystem zukunftsfähiger machen können.

**Artikel I** bietet eine umfassende Zusammenfassung der bestehenden Literatur über die kurz- und langfristige Wirksamkeit von bewegungsfördernden mHealth-Interventionen. Der Artikel fasst 117 Studien in einer Meta-Analyse quantitativ zusammen und kommt zu dem Ergebnis, dass mHealth-Bewegungsinterventionen zu kleinen bis moderaten Aktivitätssteigerungen führen, die bis zu 14 Monate nach der Intervention andauern.

Aufbauend auf dem ersten Artikel untersucht **Artikel II** anhand von zwei Fallstudien, wie Gamifizierung die langfristige Wirksamkeit von präventiven mHealth-Interventionen erhöhen kann. Der Artikel entwickelt ein neues Motivations-Modell und gibt Designempfehlungen, wie Aktivitätstracking-Apps gamifiziert werden sollten, damit Motivations-Crowding-Out verhindert werden kann.

**Artikel III** erforscht, ob Future-Self-Avatare als Instrument in präventiven mHealth-Interventionen dazu beitragen können, present bias zu vermindern. Unter Anwendung eines randomisierten, kontrollierten Studiendesigns kommt der Artikel zu dem Ergebnis, dass Future-Self-Avatare zu mehr Bewegung und einer verbesserten Ernährung führen könnten.

**Artikel IV** nimmt eine übergeordnete Perspektive ein und untersucht die Potenziale, die digitale Technologien bei der Zukunftssicherung unserer Gesundheitssysteme haben können.

# Introduction

## Preventative mHealth Interventions and Health Digitization

Globally, rising healthcare costs are pushing healthcare systems to their limits. In Switzerland, healthcare costs have doubled in the past 20 years (Bundesamt für Statistik, 2021). In the United States, medical costs have become the primary reason for private bankruptcy (Himmelstein et al., 2019), while in China government healthcare spending has quadrupled in the past 10 years (Yip et al., 2019). Increased life expectancy and the resulting rise in non-communicable diseases (NCDs) have been identified as key cost drivers, next to advances in medicine that offer ever better but often more expensive treatment options (Muka et al., 2015; Sorenson et al., 2013). Today, NCDs such as cardiovascular illnesses, diabetes, cancer, or dementia account for 59%–85% of healthcare spending in developed countries such as New Zealand or Switzerland (Blakely et al., 2019; Bundesamt für Gesundheit, 2016). Successfully tackling the healthcare cost challenge thus goes hand in hand with finding efficient and effective ways of lowering the financial and social burden of NCDs.

Research shows that NCDs largely depend on a set of modifiable risk factors. In its *Global Action Plan for the Prevention and Control of Non-Communicable Diseases*, the World Health Organization (WHO) identifies four core risk factors for NCDs: tobacco use; insufficient physical activity; unhealthy, sodium-rich diets; harmful use of alcohol (World Health Organization, 2013). Experts estimate that addressing these risk factors and complying with the resulting recommendations for healthier lifestyles can lower or delay the risk of contracting NCDs significantly by up to 45%–75% (Peters et al., 2019).

Preventative health interventions supporting healthier lifestyles are thus needed to combat the rise in NCDs, to curb rising health care spending, and to future-proof healthcare systems. However, today's healthcare systems are oriented toward treating diseases. The Organization for Economic Co-operation and Development (OECD) estimates that currently, 97% of healthcare spending are allocated to the treatment of diseases, while only 3% are invested in prevention (Gmeinder et al., 2017). Consequently, rethinking is urgently needed.

Research on preventative health interventions is multifaceted and has explored strategies for increasing physical activity (Dishman et al., 2007), reducing weight (Horvath et al., 2008), reducing smoking (Raw et al., 1998; Rigotti et al., 2008), promoting healthier diets (Bhattarai et al., 2013; De Ridder et al., 2017), and reducing alcohol consumption (Bertholet et al., 2005). Experts, however, conclude that interventions are too often implemented in small, controlled settings rather than being rolled out to larger populations (Hallal et al., 2012). In order to significantly reduce risky lifestyles, scalable preventative health interventions are needed (Reis et al., 2016). Scalability is defined as "the ability to scale up an intervention without requiring human resources" (Mönninghoff et al., 2021, p. 4). However, there are various barriers, including costs, resource restrictions, and poorly scalable intervention designs (Kowatsch et al., 2019; Reis et al., 2016). Owing to the increasing dissemination and ubiquity of digital technologies, mobile health (mHealth) interventions have been discussed as a solution for overcoming scalability challenges (Fairburn & Patel, 2017; Richard et al., 2019). mHealth interventions are defined as "programs that fully or partly deliver interventions using mobile technology such as pedometers or accelerometers with displays, activity trackers, smartphones, or tablets (Mönninghoff et al., 2021, p. 3)".

## Four Articles on mHealth Interventions and Health Digitization

This dissertation thus explores the potential of preventative mHealth interventions in supporting healthier lifestyles, and investigates how digital technologies can help future-proof healthcare systems.

Articles I and II focus on mHealth interventions targeting physical activity increases. Previous reviews have synthesized the effectiveness of mHealth physical activity interventions and found mixed effects (Bratava et al., 2007; Brickwood et al., 2019; De Vries et al., 2016; Direito et al., 2017; Franssen et al., 2020; Gal et al., 2018; Kirk et al., 2019; Qiu et al., 2014; Romeo et al., 2019; Schoeppe et al, 2016; Smith et al., 2020; Vaes et al., 2013). So far, however, only two reviews have quantitively investigated the longer-term effects of such interventions (Chaudhry et al., 2020; Yerrakalva et al., 2019), albeit with inconclusive results.

**Article I** addresses this research gap and is the most comprehensive study to date on the effectiveness of mHealth physical activity interventions. It systematically reviews the existing literature on mHealth physical activity interventions regarding their short- and long-term effectiveness. Using a random-effects model, this article quantitively summarizes 117 studies in a meta-analysis. It finds that such interventions are of small to moderate effectiveness (SMD 0.28-0.46). They increase activity levels at the end of the intervention by 1566 steps per day and by 36 incremental minutes of moderate to vigorous physical activity (MVPA) per day. Increases are maintained beyond the end of interventions, however with decreasing effect size. Long-term follow-up measurements taken on average 14 months after the end of an intervention still found 851 incremental steps per day and 24 minutes of incremental MVPA. The article also investigates multiple effect moderators and finds that population moderates the effectiveness of mHealth physical activity interventions, with interventions being more effective in sick or at-risk populations and less effective in healthy populations and preventative settings.

The following two articles build on the findings of Article I. **Article II** investigates how the long-term effectiveness of preventative mHealth interventions can be increased. Prior research has found that mHealth activity trackers act as an extrinsic incentive and increase short-term performance, but crowd out longer-term intrinsic motivation (Etkin et al., 2016; Kruglanski et al., 1975). While gamifying mHealth interventions potentially helps overcome motivational crowding out due to its intrinsic motivational qualities (Hamari et al., 2014; Teixeira et al., 2012), this area remains underresearched. Article II contributes to closing this research gap. Building on two case studies of gamified physical activity apps (Zombies, Run! and Jump, Jump Froggy 2), it presents a *motivational framework for gamification.* The article argues that owing to differences in the motivational impact of game elements, gamification should not be evaluated as a single, unified phenomenon but broken down into individual game elements. To promote long-term motivation, the novel framework suggests focusing on gamifying physical activity instead of post-tracking feedback, as gamifying ulterior motives to be active only results in short-term performance increases and may promote motivational crowding out.

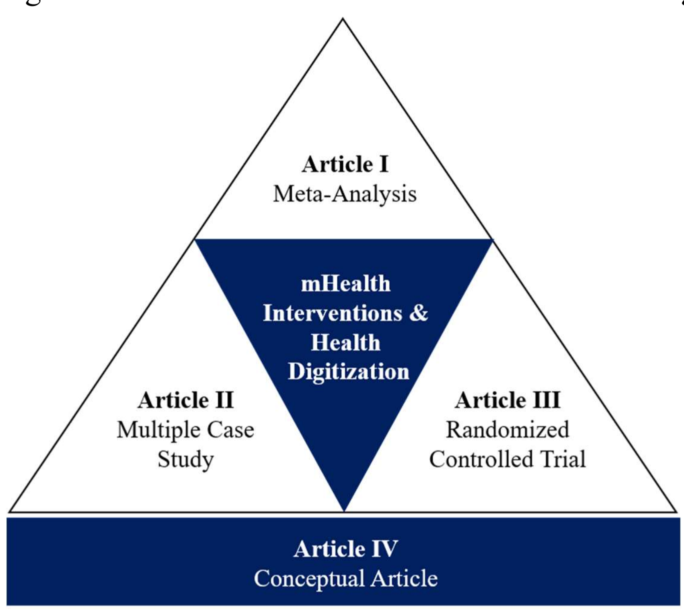
**Article III** also complements the findings of article I. Its systematic review (article I) shows that mHealth interventions are less effective when employed in preventative settings with healthy populations (Mönninghoff et al., 2021). Present bias theory (i.e., the tendency of consumers to take suboptimal decisions due to over-discounting future gains; cf., Ersner-Hershfield et al., 2009; Hershfield et al., 2011; Hunter et al., 2018; Wang & Sloan, 2018) contributes a potential theoretical explanation for the lower effectiveness of preventative mHealth interventions. Future-self avatar interventions (i.e., interventions that use age-progressed renderings of the user within the interface) have reduced present bias in related fields (Hershfield et al., 2011). Nevertheless, there is no evidence whether such interventions change health behavior.

Article III thus examines whether and, if so, how future-self avatars can help make mHealth interventions more effective. It investigates how a preventative future-self avatar mHealth intervention impacts physical activity and food purchasing behavior in a randomized-controlled trial research design. It finds that the future-self avatar mHealth intervention leads to small directional increases in physical activity and to directional improvements in the nutritional quality of food purchases. Compared to the control mHealth intervention, which uses a conventional number and text- based interface, the future-self avatar intervention significantly increases intrinsic motivation ($P$=.03) and directionally improves outcome expectancy. Overall, the article finds that future-self avatar mHealth interventions promise to encourage improvements in physical activity and food purchasing behavior in healthy population groups.

Finally, **Article IV** takes a broader perspective and explores the impact of digital technologies on future-proofing healthcare systems. Taking the current Covid-19 epidemic as an example, the article illustrates how digitized health care systems are better equipped to cope with unforeseen health crises. It identifies five reasons why digitization in healthcare must be accelerated beyond the Covid-19 epidemic.

**Figure 1** provides an overview of the structure and the respective research designs of the four articles gathered in this cumulative dissertation. For a more detailed overview (i.e., title, authors, abstract, keywords, status, publication reference), please refer to **Tables 1–4** (see **Appendix** below).

**Figure 1:** Exploring Preventative mHealth Interventions & Health Digitization



## Conclusion and Implications

The four articles of this dissertation advance our understanding of the effectiveness and design principles of preventative mHealth interventions, and of the overall opportunities of digital technologies for future healthcare systems. The findings are aimed at and will benefit a broad audience, including researchers, medical professionals, health policy makers, insurance companies, and commercial app developers.

### Practical Implications

**Article I** confirms the potential of mHealth interventions to increase physical activity short-term at the end of interventions. However, it also shows that higher physical activity levels are maintained at medically significant levels up to 14 months post-intervention. The results suggest that medical professionals and health policy makers should develop large-scale mHealth physical activity programs that leverage scalable mHealth intervention designs as these were found to be equally effective to intervention designs relying on human to human interaction, which are costly and complicate large-scale roll-out. Furthermore, medical professionals and health policy makers should prioritize sick and at-risk population groups when rolling-out mHealth physical activity intervention programs as these are more effective in such population groups.

**Article II** establishes a *motivational framework for gamification* that provides concrete design guidelines to commercial app developers and health decision-makers responsible for designing preventative mHealth interventions, such as activity trackers. The findings suggest that in order to stimulate intrinsic motivation and encourage long-term usage, decision-makers should gamify activities with game elements such as *sound*, *narrative context*, or *avatars,* among others. The findings also indicate that moderately active consumers in particular require intrinsically motivating game elements to promote behavior change. Commercial app developers might increase their consumer base by marketing intrinsically motivating gamified tracking solutions to moderately active consumers (cf., *Peloton* bike concept). These could supplement their current product solutions, which mainly use post-tracking game elements and target consumers with inherent intrinsic activity motivation.

The findings of **Article III** demonstrate that future-self avatars promise to improve the effectiveness of preventative mHealth interventions targeting physical activity increases and improvements in the nutritional quality of food purchases. Commercial app developers and health decision-makers should consider using such alternative interfaces when targeting young, healthy customer groups who are most likely to be present-biased. Additionally, the results show that leveraging digital receipts from loyalty cards is a feasible and acceptable means of tracking food purchases in a fully-automated way. Health insurance companies providing preventative health solutions to their customers should consider including such technologies, so as to offer their customers a convenient way of tracking their nutrition.

Lastly, **Article IV** provides five practical insights into why digitization should be accelerated to future-proof healthcare systems. The findings provide health decision-makers and consumers with concrete examples of the benefits of accelerated digitization, ranging from increased efficiency and cost saving potentials due to digital triage apps, to improved access to healthcare services in remote areas thanks to telemedicine solutions or online pharmacies.

### Theoretical Implications, Limitations and Future Research

The four articles of this dissertation primarily contribute to interdisciplinary research on preventative mHealth interventions and digital health technologies.

By quantitively synthesizing existing research on the effectiveness of mHealth physical activity interventions, **Article I** provides the most comprehensive assessment to date of the short- and long-term effects of such interventions on physical activity.

**Article II** establishes a conceptual framework for gamification in the context of activity tracking. It reviews the existing literature on gamification in the preventative health context and is the first study to conceptualize the motivational impact of different game elements. It conceptually combines the research stream of gamifying mHealth interventions (Hamari et al., 2014; Johnson et al., 2016; Lister et al., 2014; Thorsteinsen et al., 2014) with self-determination theory (Ryan & Deci, 2017), motivational crowding out (Etkin, 2016; Kruglanski et al., 1975), and goal systems theory (Kruglanski et al., 2018).

**Article III** contributes to the research on present bias theory in the context of health interventions. Building on Schwarzer's (2008) Health Action Process Approach, the article shows that a future-self avatar potentially promotes higher outcome expectancy and thus minimizes present bias.

The findings of this dissertation are obviously not without limitations. While **Article I** provides the most comprehensive review of mHealth physical activity interventions to date, the included evidence rates low to very low in the GRADE framework (Guyatt et al., 2011). Low-quality evidence is common in many systematic reviews for three reasons: (1) Qualitative outcome assessment via surveys is still a frequently used methodology in mHealth physical activity research; (2) the included studies have high attrition rates; (3) research personnel and participants are not blinded in terms of group allocation (Brickwood et al., 2019; Direito et al., 2017; Kirk et al., 2019; Qiu et al. 2014). Hence, more primary research using objective, quantitative outcome measurements and thorough research designs is needed to confirm that mHealth physical activity interventions can lead to long-term physical activity increases at medically significant levels.

Additionally, the *motivational framework for gamification* in **Article II** was developed based on two case studies of gamified physical activity tracking apps and on a thorough review of the existing literature. However, the framework has yet to be empirically validated. Further research is needed to empirically validate the novel framework using experimental quantitative research designs.

The strengths of **Article III** are its rigorous randomized-controlled trial research design and the objective measurement of all outcome measures. However, due to a higher than expected attrition rate, the sample size in the trial was small at the end of intervention, leading to directional yet insignificant results for the two primary outcomes. High attrition rates are a common problem in mHealth research (Eysenbach, 2005). Future research should confirm the directional research results, in order to fully evaluate the potential of future-self avatar interfaces for preventative mHealth interventions.

Lastly, **Article IV** is partly based on a book titled *Die Digitale Pille* (Fleisch et al., 2021). The article was written to address an audience of practitioners and is thus limited in length and detail. The book publication provides further context and detail, and thus helps to fully understand the potential of digital technologies for future healthcare systems.

Please refer to the respective articles for a detailed discussion of their practical and theoretical contributions, limitations, and future research directions.

## References

Bertholet, N., Daeppen, J. B., Wietlisbach, V., Fleming, M., & Burnand, B. (2005). Reduction of alcohol consumption by brief alcohol intervention in primary care: systematic review and meta-analysis. *Archives of internal medicine*, *165*(9), 986-995. https://doi.org/10.1001/archinte.165.9.986

Bhattarai, N., Prevost, A. T., Wright, A. J., Charlton, J., Rudisill, C., & Gulliford, M. C. (2013). Effectiveness of interventions to promote healthy diet in primary care: systematic review and meta-analysis of randomised controlled trials. *BMC public health, 13*(1), 1-14. https://doi.org/10.1186/1471-2458-13-1203

Blakely, T., Kvizhinadze, G., Atkinson, J., Dieleman, J., & Clarke, P. (2019). Health system costs for individual and comorbid noncommunicable diseases: An analysis of publicly funded health events from New Zealand. *PLoS medicine, 16*(1), e1002716. https://doi.org/10.1371/journal.pmed.1002716

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. https://doi.org/10.1001/jama.298.19.2296

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. https://doi.org/10.2196/11819

Bundesamt für Gesundheit. (2016). *Herausforderung nichtübertragbare Krankheiten. Nationale Strategie zur Prävention nichtübertragbarer Krankheiten 2017-2024*. https://www.bag.admin.ch/bag/de/home/strategie-und-politik/nationale-gesundheitsstrategien/strategie-nicht-uebertragbare-krankheiten.html

Bundesamt für Statistik. (2021). *Kosten und Finanzierung des Gesundheitswesens*. https://www.bfs.admin.ch/bfs/de/home/statistiken/querschnittsthemen/wohlfahrtsmessung/alle-indikatoren/gesellschaft/gesundheitsausgaben.assetdetail.16944797.html

Chaudhry, U. A., Wahlich, C., Fortescue, R., Cook, D. G., Knightly, R., & Harris, T. (2020). The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. *International Journal of Behavioral Nutrition and Physical Activity, 17*(1), 1-16. https://doi.org/10.1186/s12966-020-01020-8

De Ridder, D., Kroese, F., Evers, C., Adriaanse, M., & Gillebaart, M. (2017). Healthy diet: Health impact, prevalence, correlates, and interventions. *Psychology & health, 32*(8), 907-941. https://doi.org/10.1080/08870446.2017.1316849

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. https://doi.org/10.1002/oby.21619

Direito, A., Carraça, E., Rawstorn, J., Whittaker, R., & Maddison, R. (2017). 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. https://doi.org/10.1007/s12160-016-9846-0

Dishman, R. K., & Buckworth, J. (2007). Increasing physical activity: A quantitative synthesis. In D. Smith & M. Bar-Eli (Eds.), *Essential readings in sport and exercise psychology* (pp. 348–355). Human Kinetics.

Ersner-Hershfield, H., Garton, M. T., Ballard, K., Samanez-Larkin, G. R., & Knutson, B. (2009). Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. *Judgment and Decision making, 4*(4), 280.

Etkin, J. (2016). The hidden cost of personal quantification. *Journal of Consumer Research, 42*(6), 967–984. https://doi.org/10.1093/jcr/ucv095

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research, 7*(1):1–9 https://doi.org/10.2196/jmir.7.1.e11

Fairburn, C. G., & Patel, V. (2017). The impact of digital technology on psychological treatments and their dissemination. *Behaviour research and therapy, 88*, 19-25. https://doi.org/10.1016/j.brat.2016.08.012

Fleisch, E., Franz, C., Herrmann, A., & Mönninghoff, A. (2021) *Die Digitale Pille: Eine Reise in die Zukunft unseres Gesundheitssystems*. Campus Verlag.

Franssen, W. M., Franssen, G. H., Spaas, J., Solmi, F., & Eijnde, B. O. (2020). Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. *International Journal of Behavioral Nutrition and Physical Activity, 17*, 1-20. https://doi.org/10.1186/s12966-020-00955-2

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), 1-15. https://doi.org/10.1186/s40798-018-0157-9

Gmeinder, M., D. Morgan and M. Mueller (2017), How much do OECD countries spend on prevention?, *OECD Health Working Papers*, No. 101, OECD Publishing, Paris, https://doi.org/10.1787/f19e803c-en

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., ... & Schünemann, H. J. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of clinical epidemiology*, *64*(4), 383-394. https://doi.org/10.1016/j.jclinepi.2010.04.026

Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., Haskell, W., Ekelund, U., & Lancet Physical Activity Series Working Group. (2012). Global physical activity levels: surveillance progress, pitfalls, and prospects. *The Lancet, 380*(9838), 247-257. https://doi.org/10.1016/S0140-6736(12)60646-1

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. *Proceedings of the Annual Hawaii International Conference on System Sciences,* 3025–3034. https://doi.org/10.1109/HICSS.2014.377

Hershfield, H. E., Goldstein, D. G., Sharpe, W. F., Fox, J., Yeykelis, L., Carstensen, L. L., & Bailenson, J. N. (2011). Increasing saving behavior through age-progressed renderings of the future self. *Journal of Marketing Research, 48*(SPL), S23-S37. https://doi.org/10.1509/jmkr.48.SPL.S23

Himmelstein, D. U., Lawless, R. M., Thorne, D., Foohey, P., & Woolhandler, S. (2019). Medical bankruptcy: still common despite the Affordable Care Act. *Am J Public Health*, *109*(3), 431-433. doi: 10.2105/AJPH.2018.304901.

Hodkinson, A., Kontopantelis, E., Adeniji, C., Van Marwijk, H., McMillan, B., Bower, P., & Panagioti, M. (2019). Accelerometer-and Pedometer-Based physical activity interventions among adults with cardiometabolic conditions: a systematic review and meta-analysis. *JAMA network open, 2(*10). https://doi.org/10.1001/jamanetworkopen.2019.12895

Horvath, K., Jeitler, K., Siering, U., Stich, A. K., Skipka, G., Gratzer, T. W., & Siebenhofer, A. (2008). Long-term effects of weight-reducing interventions in hypertensive patients: systematic review and meta-analysis. *Archives of internal medicine, 168*(6), 571-580. https://doi.org/10.1001/archinte.168.6.571

Hunter, R. F., Tang, J., Hutchinson, G., Chilton, S., Holmes, D., & Kee, F. (2018). Association between time preference, present-bias and physical activity:

implications for designing behavior change interventions. *BMC public health, 18*(1), 1-12. https://doi.org/10.1186/s12889-018-6305-9

Johnson, D., Deterding, S., Kuhn, K. A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions, 6*, 89–106. https://doi.org/10.1016/j.invent.2016.10.002

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. https://doi.org/10.1177/0890117118816278

Kowatsch, T., Otto, L., Harperink, S., Cotti, A., & Schlieter, H. (2019). A design and evaluation framework for digital health interventions. *IT-Information Technology, 61*(5-6), 253-263. https://doi.org/10.1515/itit-2019-0019

Kruglanski, A. W., Riter, A., Amitai, A., Margolin, Bath-Shevah Shabtai, L., & Zaksh, D. (1975). Can money enhance intrinsic motivation? A test of the content-consequence hypothesis. *Journal of Personality and Social Psychology, 31*(4), 744–750. https://doi.org/10.1037/0022-3514.31.4.744

Kruglanski, A. W., Fishbach, A., Woolley, K., Bélanger, J. J., Chernikova, M., Molinario, E., & Pierro, A. (2018). A structural model of intrinsic motivation: On the psychology of means-ends fusion. *Psychological Review, 125*(2), 165–182. https://doi.org/0.1037/rev0000095

Lister, C., West, J. H., Cannon, B., Sax, T., & Brodegard, D. (2014). Just a fad? Gamification in health and fitness apps. *Journal of Medical Internet Research, 16*(8), e9. https://doi.org/10.2196/games.3413

Mönninghoff, A., Kramer, J. N., Hess, A. J., Ismailova, K., Teepe, G. W., Car, L. T., ... & Kowatsch, T. (2021). Long-term Effectiveness of mHealth Physical Activity Interventions: Systematic Review and Meta-analysis of Randomized Controlled Trials. *Journal of Medical Internet Research, 23*(4), e26699. https://doi.org/10.2196/26699

Muka, T., Imo, D., Jaspers, L., Colpani, V., Chaker, L., van der Lee, S. J., ... & Franco, O. H. (2015). The global impact of non-communicable diseases on healthcare spending and national income: a systematic review. *European journal of epidemiology, 30*(4), 251-277. https://doi.org/10.1007/s10654-014-9984-2

Peters, R., Ee, N., Peters, J., Beckett, N., Booth, A., Rockwood, K., & Anstey, K. J. (2019). Common risk factors for major noncommunicable disease, a systematic overview of reviews and commentary: the implied potential for targeted risk reduction. *Therapeutic advances in chronic disease*, *10*(2040622319880392), https://doi.org/10.1177/2040622319880392

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), 1-9. https://doi.org/10.1186/1741-7015-12-36

Raw, M., McNeill, A. & West, R. (1998). Smoking Cessation Guidelines for Health Professionals—A guide to effective smoking cessation interventions for the health care system. *Thorax, 53*(suppl 5), S1-S18. https://doi.org/10.1136/thx.53.2008.s1

Reis, R. S., Salvo, D., Ogilvie, D., Lambert, E. V., Goenka, S., Brownson, R. C., & Lancet Physical Activity Series 2 Executive Committee. (2016). Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. *The Lancet, 388*(10051), 1337-1348. https://doi.org/10.1016/S0140-6736(16)30728-0

Richard, E., van Charante, E. P. M., Hoevenaar-Blom, M. P., Coley, N., Barbera, M., van der Groep, A., ... & Van Gool, W. A. (2019). Healthy ageing through internet counselling in the elderly (HATICE): a multinational, randomised controlled trial. *The Lancet Digital Health, 1*(8), e424-e434. https://doi.org/10.1016/s2589-7500(19)30153-0

Rigotti, N. A., Munafo, M. R., & Stead, L. F. (2008). Smoking cessation interventions for hospitalized smokers: a systematic review. *Archives of internal medicine, 168*(18), 1950-1960. https://doi.org/10.1001/archinte.168.18.1950

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. https://doi.org/10.2196/12053

Schoeppe, S., Alley, S., Van Lippevelde, W., Bray, N. A., Williams, S. L., Duncan, M. J., & Vandelanotte, C. (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 13*(1), 1-26. https://doi.org/10.1186/s12966-016-0454-y

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied psychology, 57*(1), 1-29. https://doi.org/10.1111/j.1464-0597.2007.00325.x

Smith, D. M., Duque, L., Huffman, J. C., Healy, B. C., & Celano, C. M. (2020). Text message interventions for physical activity: a systematic review and meta-analysis. *American journal of preventive medicine, 58*(1), 142-151. https://doi.org/10.1016/j.amepre.2019.08.014

Sorenson, C., Drummond, M., & Khan, B. B. (2013). Medical technology as a key driver of rising health expenditure: disentangling the relationship. *Clinicoecon Outcomes Res.* 5, 223-234. https://doi.org/10.2147/CEOR.S39634

Teixeira, P. J., Silva, M. N., Mata, J., Palmeira, A. L., & Markland, D. (2012). Motivation, self-determination, and long-term weight control. *International Journal of Behavioral Nutrition and Physical Activity, 9*(21). https://doi.org/10.1186/1479-5868-9-22

Thorsteinsen, K., Vittersø, J., & Svendsen, G. B. (2014). Increasing physical activity efficiently: An experimental pilot study of a website and mobile phone intervention. *International Journal of Telemedicine and Applications*, 2014. https://doi.org/10.1155/2014/746232

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. https://doi.org/10.3109/07853890.2013.810891

Wang, Y., & Sloan, F. A. (2018). Present bias and health. Journal of risk and uncertainty, 57(2), 177-198. https://doi.org/10.1007/s11166-018-9289-z

World Health Organization. (2013). *Global Action Plan for the Prevention and Control of NCDs 2013-2020*. Geneva: World Health Organization.

Yerrakalva, D., Yerrakalva, D., Hajna, S., & Griffin, S. (2019). Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. *Journal of Medical Internet Research*, *21*(11), e14343. https://doi.org/10.2196/14343

Yip, W., Fu, H., Chen, A. T., Zhai, T., Jian, W., Xu, R., ... & Chen, W. (2019). 10 years of health-care reform in China: progress and gaps in universal health coverage. *The Lancet*, *394*(10204), 1192-1204. https://doi.org/10.1016/S0140-6736(19)32136-1

## Appendix

**Table 1:** Overview of Article I

| | |
|---|---|
| **Title** | Long-Term Effectiveness of mHealth Physical Activity Interventions: Systematic Review and Meta-Analysis of Randomized Controlled Trials |
| **Authors** | Annette Mönninghoff, Jan-Niklas Kramer, Alexander Jan Hess, Kamila Ismailova, Gisbert Teepe, Lorainne Tudor Car, Falk Müller-Riemenschneider, and Tobias Kowatsch |
| **Abstract** | ***Background:*** Mobile health (mHealth) interventions can increase physical activity (PA); however, their long-term impact is not well understood.<br>***Objective:*** The primary aim of this study is to understand the immediate and long-term effects of mHealth interventions on PA. The secondary aim is to explore potential effect moderators.<br>***Methods:*** We performed this study according to the Cochrane and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We searched PubMed, the Cochrane Library, SCOPUS, and PsychINFO in July 2020. Eligible studies included randomized controlled trials of mHealth interventions targeting PA as a primary outcome in adults. Eligible outcome measures were walking, moderate-to-vigorous physical activity (MVPA), total physical activity (TPA), and energy expenditure. Where reported, we extracted data for 3 time points (ie, end of intervention, follow-up ≤6 months, and follow-up >6 months). To explore effect moderators, we performed subgroup analyses by population, intervention design, and control group type. Results were summarized using random effects meta-analysis. Risk of bias was assessed using the Cochrane Collaboration tool.<br>***Results****:* Of the 2828 identified studies, 117 were included. These studies reported on 21,118 participants with a mean age of 52.03 years (SD 14.14), of whom 58.99% (12,459/21,118) were female. mHealth interventions significantly increased PA across all the 4 outcome measures at the end of intervention (walking standardized mean difference 0.46, 95% CI 0.36-0.55, *P*<.001; MVPA SMD 0.28, 95% CI 0.21-0.35, *P*<.001; TPA SMD 0.34, 95% CI 0.20-0.47, *P*<.001; energy expenditure SMD 0.44, 95% CI 0.13-0.75, *P*=.01). Only 33 studies reported short-term follow-up measurements, and 8 studies reported long-term follow-up measurements in addition to end-of-intervention results. In the short term, effects were sustained for walking (SMD 0.26, 95% CI 0.09-0.42; *P*=.002), MVPA (SMD 0.20, 95% CI 0.05-0.35; *P*=.008), and TPA (SMD 0.53, 95% CI 0.13-0.93; *P*=.009). In the long term, effects were also sustained for walking (SMD 0.25, 95% CI 0.10-0.39; *P*=.001) and MVPA (SMD 0.19, 95% CI 0.11-0.27; *P*<.001). We found study population to be an effect moderator, with higher effect scores in sick and at-risk populations. PA was increased both in scalable and nonscalable mHealth intervention designs and regardless of the control group type. The risk of bias was rated high in 80.3% (94/117) of the studies. Heterogeneity was significant, resulting in low to very low quality of evidence.<br>***Conclusions****:* mHealth interventions can foster small to moderate increases in PA. The effects are maintained long term; however, the effect size decreases over time. The results encourage using mHealth interventions in at-risk and sick populations and support the use of scalable mHealth intervention designs to affordably reach large populations. However, given the low evidence quality, further methodologically rigorous studies are warranted to evaluate the long-term effects. |
| **Keywords** | mHealth, physical activity, systematic review, meta-analysis, mobile phone |
| **Status** | Published |
| **Publication Reference** | Mönninghoff, A., Kramer, J. N., Hess, A. J., Ismailova, K., Teepe, G. W., Car, L. T., Müller-Riemenschneider, F., & Kowatsch, T. (2021). Long-term Effectiveness of mHealth Physical Activity Interventions: Systematic Review and Meta-analysis of Randomized Controlled Trials. *Journal of Medical Internet Research*, *23*(4), e26699. |

**Table 2:** Overview of Article II

| | |
|---|---|
| **Title** | Gamification 2.0: How Activity Tracking Apps Provide Long-Term Motivation |
| **Author** | Annette Mönninghoff |
| **Abstract** | One in four consumers is insufficiently physically active, resulting in substantial health risks. Conventional activity tracking apps motivate consumers but can promote motivational crowding out. Gamification can help overcome these challenges. This paper presents two case studies and introduces a new motivational framework for gamification. |
| **Keywords** | Gamification, mHealth, physical activity, crowding out, consumer motivation |
| **Status** | Published |
| **Publication Reference** | Mönninghoff, A. (2021). Gamification 2.0: How Activity Tracking Apps Provide Long-Term Motivation, *Marketing Review St.Gallen, (4),* 48-57. |

**Table 3:** Overview of Article III

| | |
|---|---|
| **Title** | The Effect of a Future-Self Avatar mHealth Intervention on Physical Activity and Food Purchases: The FutureMe Randomized Controlled Trial |
| **Authors** | Annette Mönninghoff, Klaus Fuchs, Jing Wu, Jan Albert, Simon Mayer |
| **Abstract** | ***Background*:** Insufficient physical activity and unhealthy diets are contributing to the rise in non-communicable diseases. Preventative mobile health (mHealth) interventions may enable reversing this trend, but present bias might reduce their effectiveness. Future-self avatar interventions have resulted in behavior change in related fields, yet evidence whether such interventions can change health behavior is lacking.<br>***Objective*:** Our primary objectives are to investigate the impact of a future-self avatar mHealth intervention on physical activity and food purchasing behavior, and to examine the feasibility of a novel automated nutrition tracking system. We also aim to understand how this intervention impacts related attitudinal and motivational constructs.<br>***Methods*:** We conducted a 12-week parallel randomized-controlled trial (RCT), followed by semi-structured interviews. German-speaking smartphone users aged ≥18 years living in Switzerland, and using at least one of the two leading Swiss grocery loyalty cards, were recruited for the trial. Data were collected from November 2020 to April 2021. The intervention group received the FutureMe intervention—a physical activity and food purchase tracking mobile phone application that uses a future-self avatar as the primary interface and provides participants with personalized food basket analysis and shopping tips. The control group received a conventional, text- and graphic-based primary interface intervention. We pioneered a novel system to track nutrition leveraging digital receipts from loyalty card data analyzing food purchases in a fully automated way. Data were consolidated in 4-week intervals and non-parametric tests were conducted to test for within- and between-group differences.<br>***Results*:** We recruited 167 participants; 95 eligible participants were randomized into either the intervention (n=42) or control group (n=53). The median age was 44.00 years (IQR 19.00), and the gender ratio was balanced (female 52/95, 55%). Attrition was unexpectedly high with only 30 participants completing the intervention, negatively impacting the statistical power of our study. The FutureMe intervention led to directional, small increases in physical activity (median +242 steps/day) and to directional improvements in the nutritional quality of food purchases (median –1.28 British Food Standards Agency Nutrient Profiling System Dietary Index points) at the end of the intervention. Intrinsic motivation significantly increased ($P$=.03) in the FutureMe group, but decreased in the control group. Outcome expectancy directionally increased for the FutureMe group, but decreased for the control group. Leveraging loyalty card data to track the nutritional quality of food purchases was found to be a feasible and an accepted fully automated nutrition tracking system.<br>***Conclusions*:** Preventative future-self avatar mHealth interventions promise to encourage improvements in physical activity and food purchasing behavior in healthy population groups. A full-powered RCT is needed to confirm this preliminary evidence and to investigate how future-self avatars might be modified to reduce attrition, overcome present bias, and promote sustainable behavior change. |
| **Keywords** | mHealth, preventative medicine, avatar, present bias, nutrition tracking, physical activity, randomized controlled trial |
| **Status** | Accepted for presentation at the Swiss Academy of Marketing Science (SAMS) Conference 2020.<br><br>Submitted to the Journal of Medical Internet Research; revision & re-review. |
| **Publication Reference** | Published as working paper with non-disclosure notice on Alexandria:<br>Mönninghoff, Annette; Fuchs, Klaus; Wu, Jing; Albert, Jan & Mayer, Simon. (2021). *The Effect of a Future-Self Avatar mHealth Intervention on Physical Activity and Food Purchases: The FutureMe Randomized Controlled Trial.* https://www.alexandria.unisg.ch/publications/263453 |

**Table 4:** Overview of Article IV

| | |
|---|---|
| **Title** | The Future of Medicine. Accelerating Digitization in Healthcare: Why We Need It and Where We Struggle |
| **Authors** | Christoph Franz, Annette Mönninghoff |
| **Abstract** | The Covid-19 pandemic has put the spotlight on the lack of digital infrastructure in today's healthcare systems and demonstrated why digitized healthcare is a competitive advantage. |
| **Keywords** | Digitization, digital health, Covid-19, health policy. |
| **Status** | Published |
| **Publication Reference** | Franz, C., Mönninghoff, A. (2021). The Future of Medicine. Accelerating Digitization in Healthcare. Whey We Need It and Where We Struggle, *The SternStewart Periodical*, 23, 16-21.<br><br>The article is partly based on extracts from the following book publication:<br>Fleisch, E., Franz, C., Herrmann, A., Mönninghoff, A. (2021*). Die Digitale Pille: Eine Reise in die Zukunft unseres Gesundheitssystems.* Campus Verlag. |

# Article I

## Long-term Effectiveness of mHealth Physical Activity Interventions: Systematic Review and Meta-analysis of Randomized Controlled Trials

Annette Mönninghoff[1,2], Jan Niklas Kramer[3,4], Alexander Jan Hess[3,5], Kamila Ismailova[3], Gisbert W Teepe[6], Lorainne Tudor Car[7,8], Falk Müller-Riemenschneider[9], Tobias Kowatsch[3,6,9,10]

[1]Institute for Customer Insight, University of St.Gallen, St.Gallen, Switzerland
[2]Institute for Mobility, University of St.Gallen, St.Gallen, Switzerland
[3]Centre for Digital Health Interventions, Institute of Technology Management, University of St.Gallen, St. Gallen, Switzerland
[4]CSS Insurance, Lucerne, Switzerland
[5]Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland
[6]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland
[7]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[8]School of Public Health, Department of Primary Care and Public Health, Imperial College London, London, United Kingdom
[9]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore
[10]Future Health Technologies Programme, Campus for Research Excellence and Technological Enterprise (CREATE), Singapore-ETH Centre, Singapore, Singapore

## Abstract

**Background:** Mobile health (mHealth) interventions can increase physical activity (PA); however, their long-term impact is not well understood.

**Objective:** The primary aim of this study is to understand the immediate and long-term effects of mHealth interventions on PA. The secondary aim is to explore potential effect moderators.

**Methods:** We performed this study according to the Cochrane and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We searched PubMed, the Cochrane Library, SCOPUS, and PsychINFO in July 2020. Eligible studies included randomized controlled trials of mHealth interventions targeting PA as a primary outcome in adults. Eligible outcome measures were walking, moderate-to-vigorous physical activity (MVPA), total physical activity (TPA), and energy expenditure. Where reported, we extracted data for 3 time points (ie, end of intervention, follow-up ≤6 months, and follow-up >6 months). To explore effect moderators, we performed subgroup analyses by population, intervention design, and control group type. Results were summarized using random effects meta-analysis. Risk of bias was assessed using the Cochrane Collaboration tool.

**Results:** Of the 2828 identified studies, 117 were included. These studies reported on 21,118 participants with a mean age of 52.03 (SD 14.14) years, of whom 58.99% (n=12,459) were female. mHealth interventions significantly increased PA across all the 4 outcome measures at the end of intervention (walking standardized mean difference [SMD] 0.46, 95% CI 0.36-0.55; *P*<.001; MVPA SMD 0.28, 95% CI 0.21-0.35; *P*<.001; TPA SMD 0.34, 95% CI 0.20-0.47; *P*<.001; energy expenditure SMD 0.44, 95% CI 0.13-0.75; *P*=.01). Only 33 studies reported short-term follow-up measurements, and 8 studies reported long-term follow-up measurements in addition to end-of-intervention results. In the short term, effects were sustained for walking (SMD 0.26, 95% CI 0.09-0.42; *P*=.002), MVPA (SMD 0.20, 95% CI 0.05-0.35; *P*=.008), and TPA (SMD 0.53, 95% CI 0.13-0.93; *P*=.009). In the long term, effects were also sustained for walking (SMD 0.25, 95% CI 0.10-0.39; *P*=.001) and MVPA (SMD 0.19, 95% CI 0.11-0.27; *P*<.001). We found the study population to be an effect moderator, with higher effect scores in sick and at-risk populations. PA was increased both in scalable and nonscalable mHealth intervention designs and regardless of the control group type. The risk of bias was rated high in 80.3% (94/117) of the studies. Heterogeneity was significant, resulting in low to very low quality of evidence.

**Conclusions:** mHealth interventions can foster small to moderate increases in PA. The effects are maintained long term; however, the effect size decreases over time. The results encourage using mHealth interventions in at-risk and sick populations and support the use of scalable mHealth intervention designs to affordably reach large populations. However, given the low evidence quality, further methodologically rigorous studies are warranted to evaluate the long-term effects.

**Keywords:** mHealth; physical activity; systematic review; meta-analysis; mobile phone

# Introduction

## Background

In recent decades, populations have become increasingly sedentary. The World Health Organization (WHO) recommends 150 minutes of moderate-intensity physical activity (PA) or 75 minutes of vigorous-intensity PA per week for adults and 60 minutes of moderate-to-vigorous physical activity (MVPA) for adolescents per day [1]. An estimated 28% of adults worldwide do not meet these guidelines [2]. The prevalence of inactivity is high in Latin America and many high-income countries, with approximately every second adult inactive in Brazil or Saudi Arabia, and 40% of adults insufficiently active in the United States [2].

According to the WHO, physical inactivity is 1 of the 4 core modifiable risk factors for noncommunicable diseases (NCDs). As such, it is as important to be addressed as tobacco use or obesity and proven to increase the risk of cancer, cardiovascular diseases, diabetes, dementia, and depression [3-6].

In response to the high prevalence and substantial risk posed by physical inactivity, the WHO has formulated a target to reduce physical inactivity by 10% by 2025 as part of its strategy against NCDs [7]. Scaling up PA interventions is key to achieving the WHO target. However, there are various barriers, including cost, resource restrictions, and poorly scalable intervention designs [8,9]. Owing to the increasing dissemination and ubiquity of mobile technology, mobile technology–based interventions, that is, mobile health (mHealth), have been discussed as a solution for overcoming scalability challenges [10,11]. There are only a few examples of nationwide mHealth programs such as the *NHS Diabetes Prevention Program* [12] in the United Kingdom, the 10,000 steps program in Australia [13], and the *National Steps Challenge* and *Live Healthy SG* in Singapore [14,15]. Most governments and health organizations are still hesitant about rolling out mHealth PA programs, as clear evidence for the effectiveness of mHealth interventions for sustainable behavior change is lacking [16,17].

Previous evidence for the effectiveness of mHealth interventions on PA is mixed (Appendix 1 [18-31]). Most existing meta-analyses found significant positive effects on PA in sick and at-risk populations, with effect sizes ranging from small to large [18-28]. However, some studies did not find significant effects or reported conflicting results [29-31]. There is limited evidence for the sustainability of increased PA levels beyond the end of intervention. Only 2 studies quantitatively analyzed long-term effects: one review found that PA increases are maintained up to 3-4 years after the intervention [20] and the other did not find significant long-term results [31]. Kirk et al [25] and Romeo et al [30] found that shorter mHealth PA interventions (<16 weeks and <12 weeks, respectively) are more effective than longer ones, indicating that effects might not be maintained in the long term.

We also lack clarity on how population types, intervention design, and control group type moderate the impact of mHealth interventions on PA. Only 3 studies performed subgroup analyses according to population type with mixed results. A total of 2 studies found interventions to be equally effective in sick and healthy populations [23,30], and 1 review found mHealth interventions to be more effective in sick populations; however, the results were not statistically significant [27]. Most studies focused exclusively on sick or at-risk populations [21,22,24-26,28,31], making it difficult to draw clear conclusions.

The design of mHealth interventions influences the degree to which they are scalable. The promise of mHealth is that the technology itself (ie, without costly and limited human resources) promotes active lifestyles. However, these highly scalable interventions miss the element of human-to-human interaction, which is a potentially important active ingredient in behavior change interventions. Current evidence draws an inconclusive picture: existing studies have found no effects on PA when interventions are scalable [24,30] (ie, mHealth interventions without human-to-human interactions), stronger effects when interventions are nonscalable (ie, mHealth interventions with human-to-human interactions) [27,32], stronger effects in scalable interventions [20], or no moderating effects [22]. Thus, we need a comprehensive evaluation of scalable versus nonscalable designs to judge the potential of mHealth technologies in reaching large populations at low costs.

Furthermore, our current understanding of the effects of mHealth PA interventions is limited by the inclusion of different control groups in previous studies. Most previous studies included both minimal or no intervention control groups and control groups receiving an alternative intervention [18-20,22-29,31]. This makes it impossible to distinguish between the absolute effect of mHealth PA interventions on behavior and the degree to which mHealth interventions are superior to alternative nonmobile designs or the standard of care.

### Objectives

Accordingly, we sought to comprehensively collate and analyze trials evaluating mHealth interventions that promote PA in adult populations. Our primary aim is to understand the long-term impact of these interventions on PA. Our secondary aim is to explore potential effect moderators (population type, intervention design, and control group type), to understand which populations can benefit from mHealth interventions, to understand if scalable mHealth intervention designs are effective, and to understand if mHealth interventions produce superior results to nonmobile interventions.

## Methods

### Overview

This study was performed according to the Cochrane methodology, and the results were reported following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We searched PubMed, the Cochrane Library, SCOPUS, and PsychINFO for randomized controlled trials (RCTs) on mHealth interventions targeting PA increases (all search strategies are given in Appendices 2 and 3) published from database inception to July 3, 2020. We also searched the reference lists of the relevant existing systematic reviews for eligible studies. This study was registered with PROSPERO (CRD 42019124716).

### Eligibility Criteria

Studies were eligible if they assessed the impact of mHealth interventions on PA as a primary study outcome in individuals aged 18 years or more and were published in English. Eligible study designs were RCTs or cluster RCTs. Eligible comparators comprised no or minimal interventions and alternative interventions that did not include mobile technologies.

### Types of Interventions

mHealth interventions were defined as programs that fully or partly deliver interventions using mobile technology such as pedometers or accelerometers with displays, activity trackers, smartphones, or tablets. We excluded interventions where the use of a mobile device was unclear (eg, telephone or website interventions) or where increasing PA was not the primary outcome. This was to ensure that interventions genuinely aimed to increase PA and to avoid including studies measuring PA only as a supplemental outcome.

**Types of Outcomes**

Eligible outcome measures were walking, MVPA, total physical activity (TPA), and energy expenditure (EE), as these outcomes are most commonly reported. Multiple outcome units were eligible per outcome measure (eg, walking in minutes and walking in steps). Studies reporting objectively measured or self-reported outcome data were eligible.

**Data Collection Process**

Abstracts of all identified papers were exported and uploaded into Covidence Systematic Review software (Veritas Health Innovation Ltd, version accessed July 2020) for screening. Two reviewers independently screened the abstracts for eligibility (AM, JNK, or KI). If reviewers doubted whether an article was potentially relevant, it was included for full inspection. Next, full texts of potentially eligible papers were uploaded into Covidence and screened by 2 independent reviewers (AM, JNK, or KI). Conflicts were resolved by discussion or where required by a third reviewer. We contacted authors of potentially relevant articles for further information when needed. All reviewers were trained during a full-day workshop on eligibility criteria and software before screening.

**Data Extraction and Management**

Data for each study were extracted independently by 2 reviewers (AM, JNK, KI, AJH, or GWT) using standardized extraction forms. Conflicts were resolved by discussion between 2 primary reviewers or with a third, independent reviewer. All reviewers were trained to use the extraction form and Cochrane risk of bias criteria during a full-day workshop. Where reported, data were extracted for all 4 eligible outcome measures (walking, MVPA, TPA, and EE) and time points (end of intervention, short-term follow-up [≤6 months after the end of intervention], and long-term follow-up [>6 months after the end of intervention]). If studies reported both objectively measured and self-reported outcome data, the former were used for meta-analysis. If studies only reported self-reported outcome data, these were extracted, and the quality of evidence was rated as high risk of detection bias. Data were extracted as means and

SDs per outcome measure and time point. If SDs were not reported, they were calculated using the RevMan calculator and following the Cochrane Handbook [33]. Respective authors were contacted for any missing data.

**Assessment of Risk of Bias in Included Studies**

Two reviewers (AM, JNK, KI, AJH, or GWT) independently assessed the risk of bias for each study using the Cochrane Collaboration tool [34]. Additional criteria for cluster RCTs were assessed [35] and documented within the *other bias* domains of the Cochrane Collaboration tool. Discrepancies were resolved by consensus between reviewers or where needed by a third reviewer. We classified studies as overall high risk of bias if they scored high in any bias domain other than *performance bias*, as blinding of participants and personnel is almost impossible in mHealth intervention studies [23]. Blinding of outcome assessors was rated as high risk if outcomes were self-reported.

**Statistical Methods**

We summarized the intervention and sample characteristics of all the included studies. We quantitatively analyzed the data using RevMan software (Cochrane, version 5.4) and a DerSimonian and Laird random effects model for our meta-analysis [36]. We reported all 4 outcome measures using standardized mean differences (SMDs) and 95% CI. Where appropriate (eg, if one mHealth intervention was compared with a minimal and alternative nonmobile intervention), we combined means and SDs of control or intervention groups following the Cochrane Handbook [33].

We classified populations into 3 groups based on the reported recruitment criteria: sick, at-risk, and healthy. The sick group included populations experiencing illnesses such as diabetes, cancer, chronic obstructive pulmonary disease, and coronary heart disease. The at-risk group included inactive or sedentary, older, overweight, and obese populations.

We classified mHealth interventions into 2 designs: scalable and nonscalable. Scalability is defined as the ability to scale up an intervention without requiring human resources. Consequently, scalable mHealth interventions were defined as interventions that only leveraged automated components without any human-to-human interactions. Nonscalable mHealth interventions included human-to-human interactions, such as coaching, in-person feedback, or group activity sessions. We classified control groups into no or minimal interventions (no intervention or information material only) and alternative (nonmobile) interventions.

Following the recommendations of Richardson et al [37] and the Cochrane Handbook [33], a subgroup analysis was performed based on end-of-intervention values for all outcomes where at least 10 studies were available. We a priori defined 3 subgroup analyses following the population, intervention, comparison, and outcome framework [33] to identify possible effect moderators. Our aim is to understand the impact of population type (sick, at-risk, and healthy), intervention design (scalable and nonscalable), and control group type (no or minimal and alternative).

We present the primary results using forest plots for each outcome and time point. Subgroup analyses were displayed in forest plots using end-of-intervention data. We quantified inconsistencies between studies using the $I^2$ statistics (ie, the varying effect estimates towing to heterogeneity rather than chance) [33]. We classified $I^2$>50% as having substantial heterogeneity [38]. We examined the significance of heterogeneity using chi-square tests ($P \leq .05$). We assessed subgroup differences following the guidelines given by Richardson et al [37], which recommend testing for significant subgroup differences ($P \leq .10$) and covariate distribution and comparing heterogeneity and effect sizes between subgroups. Funnel plot analysis was used to detect sampling bias. We used end-of-intervention effect values in our funnel plot analyses, as all studies reported this time point.

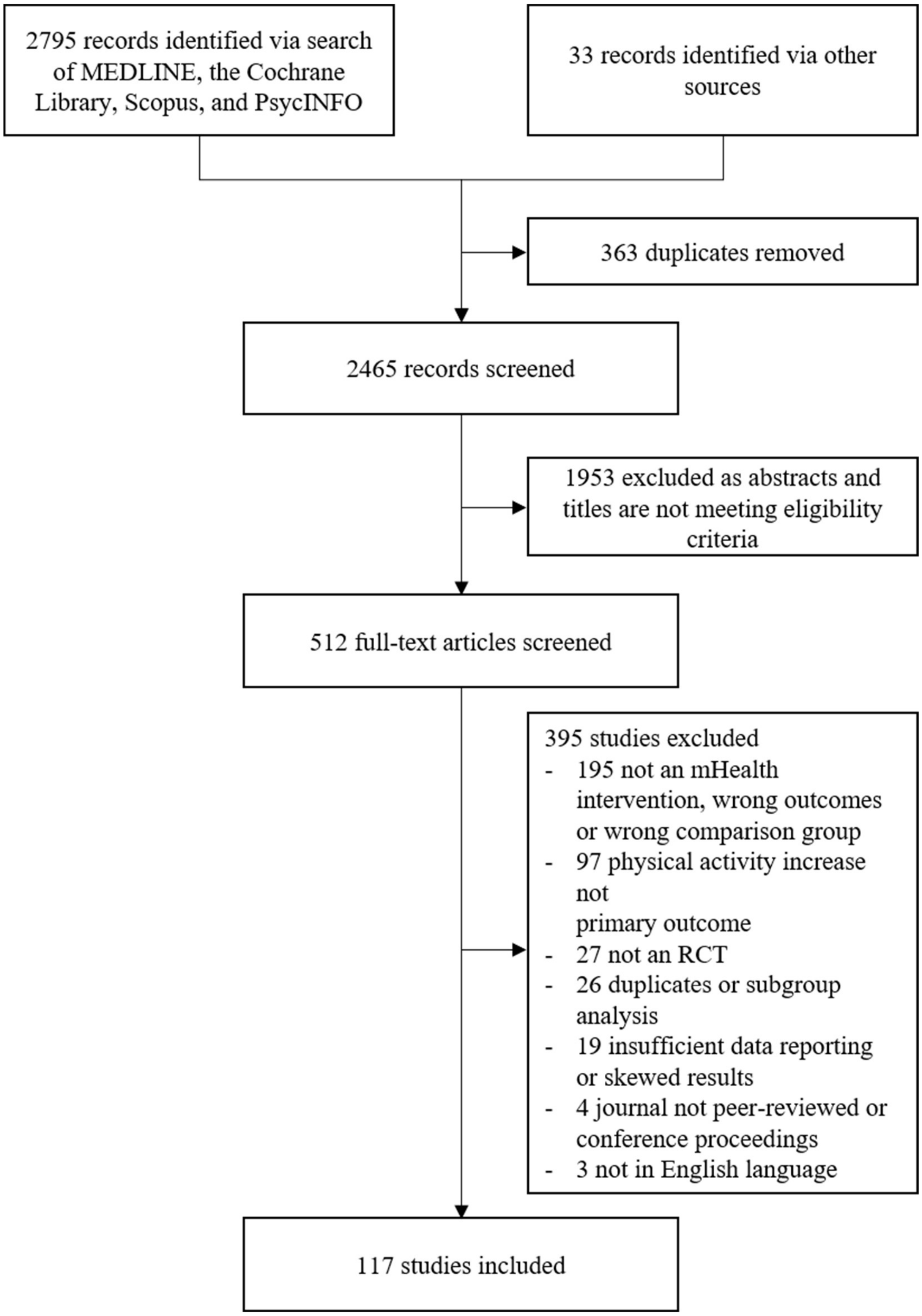
We conducted 3 sensitivity analyses to evaluate the robustness of our primary results: first, we excluded outlier studies; second, we excluded studies with high risk of bias; and third, we excluded studies not reporting long-term follow-up measurements to keep the study sample consistent across all time points. We used the grading of recommendations, assessment, development, and evaluation (GRADE) framework to assess the quality of evidence at the outcome measure level for the end-of-intervention time point and to report the standardized quality of evidence profiles, following the study by Guyatt et al [39].

## Results

### Overview

Of the 2828 identified studies, 512 full-text articles were screened, and 117 studies were included in the meta-analysis (Figure 1).

**Figure 1:** PRISMA Chart




## Study Characteristics

Appendix 4 [40-156] contains the included studies and their characteristics. The 117 trials represented 21,188 participants with a mean age of 52.03 years (SD 14.14), of whom 58.99% (12,459/21,118) were female. Most studies were conducted in high-income, developed regions such as North America (43/117, 36.8%), Europe (39/117, 33.3%), and Australia and New Zealand (24/117, 20.5%). Very few studies were conducted in Asia (7/117, 6.0%), Latin America (3/117, 2.6%), or Africa (1/117, 0.9%), limiting the representativeness of the evidence for low-income countries. Sample sizes ranged widely (from 15 to 1442), and the intervention duration ranged from 1 week to 2 years. All but one study [40] reported end-of-intervention results, 33 studies reported short-term follow-up results [40-72], and only 8 studies reported long-term follow-up results [61,72-78]. The mean time point for the short-term follow-up was 4.14 months (SD 2.08) after the end of intervention. The long-term follow-up measurement was taken on average after 13.96 months (SD 11.91).

Walking was the most reported outcome measure (77/117, 65.8%) [41,43,44,48,50-54,58-60,63-65,67-69,72-130], followed by MVPA (62/117, 53.0%) [42,44,46,47,49,51-55,57,59,61-64,68,70-76,78-80,82,84-87,89-91,94,98,104,107,109,112,114,117,125,126,131-147], TPA (33/117, 28.2%) [44,45,50,52,54,56,64,66,72,74,76-78,84,85,89,96,110-112,114,118,131,135,146,148-154,157], and EE (5/117, 4.3%) [61,103,137,155,156]. Most RCTs were conducted in at-risk (48/117, 41.0%) or sick populations (46/117, 39.3%).

Only 19.7% (23/117) of the studies tested mHealth interventions in healthy populations within a preventative setting. In most interventions, mHealth technologies were leveraged in nonscalable intervention designs (71/117, 60.7%). Human-to-human interactions included individual coaching, group coaching, PA classes, and physical education classes.

Of the 117 interventions, 45 (38.5%) used mHealth technologies without any human-to-human interactions and were thus classified as scalable. In 1 study [75], 2 mHealth interventions (scalable and nonscalable) were combined. Most mHealth interventions only leveraged basic technologies such as pedometers or accelerometers (86/117, 73.5%), text messages (20/117, 17.1%), or websites (20/117, 17.1%). Although some recent studies pioneered innovative mHealth technologies [49,96], overall, only a few studies used advanced mHealth technologies such as automated individualized feedback (19/117, 16.2%), mobile phone apps (15/117, 12.8%), social comparison (10/117, 8.5%), and automated coaching or virtual advisors (5/117, 4.3%).

Most studies had no or minimal intervention control groups (83/117, 70.9%). Only a few trials had alternative intervention control groups (22/117, 18.8%). Different control group types were combined into one control group in 12 cases.

**Meta-analysis of mHealth Interventions on PA**

Overall, mHealth interventions significantly increased PA across all 4 outcome measures at the end of intervention: walking SMD 0.46 (95% CI 0.36-0.55; *P*<.001; $I^2$=83%, *P*<.001); MVPA SMD 0.28 (95% CI 0.21-0.35; *P*<.001; $I^2$=62%, *P*<.001); TPA SMD 0.34 (95% CI 0.20-0.47; *P*<.001; $I^2$=77%, *P*<.001); and EE SMD 0.44 (95% CI 0.13-0.75; *P*=.005; $I^2$=60%, *P*=.04; Figures 2-9). Short-term effects were sustained (≤6 months after the end of intervention) for 3 of 4 outcome measures: walking SMD 0.26 (95% CI 0.09-0.42; *P*=.002; $I^2$=73%, *P*<.001); MVPA SMD 0.20 (95% CI 0.05-0.35; *P*=.008; $I^2$=72%, *P*<.001); and TPA SMD 0.53 (95% CI 0.13-0.93; *P*=.009; $I^2$=87%, *P*<.001). Only one study [61] reported short-term follow-up measurements for EE, and the results were not statistically significant.

In addition, long-term (>6 months after the end of intervention) effects were sustained for 2 of 4 outcome measures: walking SMD 0.25 (95% CI 0.10-0.39; *P*=.001; $I^2$=68%, *P*=.004) and MVPA SMD 0.19 (95% CI 0.11-0.27; *P*<.001; $I^2$=0%, *P*=.44). TPA results were sustained, but the effects were just below the significance threshold (SMD 0.19, 95% CI 0.00-0.38; *P*=.05; $I^2$=72%, *P*=.003). Again, only one study [61] reported long-term follow-up effects for EE, which were not statistically significant. Effect sizes decreased over time, from almost moderate at the end of intervention to small during the long-term follow-up measurement. We found substantial and significant heterogeneity across all outcomes and most time points, with $I^2$ ranging from 60% to 83% for end-of-intervention measurements (Figures 2-9).

**Figure 2:** Primary Outcome Analysis for The Outcome Walking at Timepoint End of Intervention

| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Standardized mean difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **End of intervention** | | | | | | | | |
| Aittasalo, 2012 [41] | 457 | 306 | 123 | 431 | 403 | 118 | 1.3% | 0.07 [-0.18, 0.33] |
| Alsaleh, 2016 [79] | 36.7 | 109.8 | 66 | 16.5 | 58.73 | 79 | 1.2% | 0.23 [-0.09, 0.56] |
| Ashton, 2017 [80] | 1588.2 | 2740.6 | 26 | 575.4 | 2735.1 | 24 | 0.8% | 0.36 [-0.20, 0.92] |
| Baker, 2008 [81] | 9977 | 4669 | 39 | 7078 | 2911 | 40 | 1.0% | 0.74 [0.28, 1.20] |
| Barwais, 2013 [82] | 1625 | 553.8 | 18 | 483 | 175.9 | 15 | 0.5% | 2.61 [1.65, 3.57] |
| Butler, 2004 [83] | 56,396 | 3315 | 17 | 53,666 | 6543 | 16 | 0.7% | 0.52 [-0.18, 1.21] |
| Cadmus-Bertram, 2019 [84] | 6697 | 2878 | 24 | 4853 | 1703 | 23 | 0.8% | 0.76 [0.17, 1.36] |
| Coelho, 2018 [43] | 8853 | 3320 | 20 | 6248 | 2030 | 17 | 0.7% | 0.91 [0.23, 1.59] |
| Compernolle, 2015 [85] | 9484 | 4875 | 86 | 8589 | 4380 | 93 | 1.2% | 0.19 [-0.10, 0.49] |
| Creel, 2016 [86] | 6084 | 873 | 52 | 5253 | 421 | 35 | 1.0% | 1.13 [0.67, 1.60] |
| Croteau, 2004 [129] | 2419 | 6074.6 | 7 | 2320 | 4732 | 8 | 0.4% | 0.02 [-1.00, 1.03] |
| Croteau, 2007 [130] | 6180 | 3530 | 79 | 6378 | 2994 | 68 | 1.2% | -0.06 [-0.38, 0.26] |
| Cruz, 2016 [44] | 10,440 | 4012 | 13 | 6430 | 2613 | 13 | 0.5% | 1.15 [0.31, 1.99] |
| Dadaczynski, 2017 [87] | 526.48 | 240.56 | 80 | 442.87 | 200.36 | 64 | 1.2% | 0.37 [0.04, 0.70] |
| De Blok, 2006 [88] | 3927 | 2617 | 8 | 3554 | 2019.6 | 8 | 0.4% | 0.15 [-0.83, 1.13] |
| DeGreef, 2010 [73] | 9601 | 5002 | 20 | 5538 | 3877 | 21 | 0.7% | 0.89 [0.25, 1.54] |
| DeGreef, 2011[89] | 7703 | 2729 | 60 | 3883 | 2537 | 32 | 0.9% | 1.42 [0.94, 1.90] |
| DeGreef, 2011a [74] | 6771 | 3889 | 43 | 5173 | 3094 | 24 | 0.9% | 0.44 [-0.07, 0.94] |
| Demeyer, 2017 [90] | 7 | 32.5 | 129 | -10 | 23.5 | 132 | 1.3% | 0.60 [0.35, 0.85] |
| Dishman, 2009 [91] | 14.8 | 17.3 | 564 | 11.2 | 17.3 | 265 | 1.4% | 0.21 [0.06, 0.35] |
| Duru, 2010 [92] | 9883 | 14,014 | 34 | 2426 | 8429 | 28 | 0.9% | 0.62 [0.11, 1.14] |
| Engel, 2006 [93] | 150 | 117 | 24 | 216 | 87 | 30 | 0.8% | -0.64 [-1.19, -0.09] |
| Finkelstein, 2016 [48] | -480 | 2568 | 201 | 43.7 | 2640.7 | 599 | 1.4% | -0.20 [-0.36, -0.04] |
| Fjeldsoe, 2010 [94] | 16.67 | 89.4 | 45 | 0.34 | 89.4 | 43 | 1.0% | 0.18 [-0.24, 0.60] |
| Furber, 2010 [50] | 249.9 | 196 | 97 | 202.6 | 189.5 | 107 | 1.2% | 0.24 [-0.03, 0.52] |
| Gell, 2015 [95] | 6867.7 | 2227 | 41 | 6189 | 2297 | 46 | 1.0% | 0.30 [-0.13, 0.72] |
| Gill, 2019 [96] | 1646 | 3302 | 59 | -1485 | 3171.5 | 59 | 1.1% | 0.96 [0.58, 1.34] |
| Glynn, 2014 [97] | 5855 | 4264 | 31 | 4859 | 3474 | 35 | 0.9% | 0.25 [-0.23, 0.74] |
| Hardeman, 2020 [98] | 8419 | 3224 | 417 | 8191 | 3003 | 442 | 1.4% | 0.07 [-0.06, 0.21] |
| Harris, 2018 [75] | 8306 | 3140 | 778 | 7198 | 2809 | 456 | 1.4% | 0.37 [0.25, 0.48] |
| Hornikx, 2015 [99] | 984 | 1208 | 12 | 1013 | 1275 | 15 | 0.6% | -0.02 [-0.78, 0.74] |
| Hospes, 2009 [100] | 7872 | 3962 | 18 | 6172 | 3194 | 17 | 0.7% | 0.46 [-0.21, 1.13] |
| Houle, 2011 [101] | 9850 | 3282 | 23 | 7970 | 3433 | 22 | 0.8% | 0.55 [-0.05, 1.15] |
| Hultquist, 2005 [102] | 10,159 | 292 | 27 | 8270 | 354 | 31 | 0.3% | 5.70 [4.51, 6.90] |
| Izawa, 2012 [103] | 8609.6 | 3064.5 | 52 | 5512.9 | 2571.8 | 51 | 1.0% | 1.09 [0.67, 1.50] |
| James, 2015 [51] | 800.8 | 2820.6 | 57 | -1294 | 3304.4 | 52 | 1.1% | 0.68 [0.29, 1.07] |
| Katzmarzyk, 2011[104] | 7248 | 2737 | 20 | 6637 | 2418 | 23 | 0.8% | 0.23 [-0.37, 0.83] |
| Kawagoshi, 2015 [105] | 51.3 | 63.7 | 12 | 12.3 | 25.5 | 15 | 0.6% | 0.82 [0.02, 1.61] |
| Kernot, 2019 [52] | 276.3 | 221.7 | 74 | 235 | 218.9 | 33 | 1.0% | 0.19 [-0.23, 0.60] |
| King, 2013 [106] | 253.5 | 248.7 | 20 | 26.8 | 67 | 19 | 0.7% | 1.21 [0.52, 1.89] |
| Kolt, 2012 [76] | 107.4 | 164.9 | 130 | 92.2 | 120.5 | 123 | 1.3% | 0.10 [-0.14, 0.35] |
| Li, 2020 [107] | 6673 | 3462 | 55 | 5819 | 2860 | 57 | 1.1% | 0.27 [-0.10, 0.64] |
| Long, 2013 [108] | 65,983 | 18,069 | 38 | 47,596 | 13,900 | 33 | 0.9% | 1.12 [0.61, 1.62] |
| Lynch, 2019 [109] | 8193 | 3301 | 37 | 7539 | 3404 | 40 | 1.0% | 0.19 [-0.26, 0.64] |
| Lyons, 2017 [110] | 6193 | 3183 | 20 | 4586 | 2476 | 20 | 0.7% | 0.55 [-0.08, 1.19] |
| Maher, 2015 [53] | 332 | 289 | 51 | 160 | 185 | 59 | 1.1% | 0.72 [0.33, 1.10] |
| Mansi, 2015 [54] | 9792 | 2053 | 29 | 6551 | 1154 | 29 | 0.7% | 1.92 [1.29, 2.55] |
| Martin, 2015 [111] | 1067 | 2098 | 32 | -1042 | 2202 | 16 | 0.7% | 0.97 [0.34, 1.61] |
| Melville, 2015 [112] | 4823 | 2059 | 42 | 4784 | 2613 | 40 | 1.0% | 0.02 [-0.42, 0.45] |
| Mendoza, 2015 [113] | 3080 | 3254 | 52 | 138 | 1950 | 50 | 1.0% | 1.08 [0.67, 1.50] |
| Merom, 2007 [114] | 52 | 107.17 | 105 | 48.9 | 126.83 | 209 | 1.3% | 0.03 [-0.21, 0.26] |
| Mutrie, 2012 [58] | 9351 | 2017 | 20 | 7138 | 2169 | 19 | 0.7% | 1.04 [0.36, 1.71] |
| Nolan, 2017 [59] | 272 | 2275.9 | 63 | 155 | 2557.1 | 59 | 1.1% | 0.05 [-0.31, 0.40] |
| Oliveira, 2019 [60] | 7507 | 3077 | 54 | 7401 | 2841 | 55 | 1.1% | 0.04 [-0.34, 0.41] |
| Paul, 2016 [115] | 5791 | 2952 | 15 | 2947 | 2399 | 8 | 0.5% | 0.99 [0.07, 1.90] |
| Poirier, 2016 [116] | 5411 | 2277 | 107 | 4751 | 1834 | 110 | 1.2% | 0.32 [0.05, 0.59] |
| Pope, 2018 [117] | 5175 | 2308 | 12 | 4746 | 2045 | 8 | 0.5% | 0.19 [-0.71, 1.08] |
| Prestwich, 2010 [118] | 1.98 | 1.9 | 88 | 1.17 | 1.58 | 46 | 1.1% | 0.45 [0.09, 0.81] |
| Reijonsaari, 2012 [119] | 2047 | 1650 | 264 | 2338 | 1762 | 257 | 1.4% | -0.17 [-0.34, 0.00] |
| Ribeiro, 2014 [63] | 969.7 | 1464.1 | 101 | -181.5 | 2284 | 94 | 1.2% | 0.60 [0.32, 0.89] |
| Roos, 2014 [120] | 3810 | 10,993.5 | 29 | 817.3 | 2349.2 | 22 | 0.8% | 0.35 [-0.21, 0.91] |
| Rowley, 2019 [121] | 9015 | 2842 | 97 | 4654 | 1447 | 32 | 1.0% | 1.69 [1.23, 2.14] |
| Simons, 2018 [64] | 7741 | 4553 | 55 | 8061 | 5112 | 63 | 1.1% | -0.07 [-0.43, 0.30] |
| Spence, 2009 [122] | 690 | 450.2 | 32 | 592.5 | 368.9 | 31 | 0.9% | 0.23 [-0.26, 0.73] |
| Stacey, 2016 [65] | 10,849 | 5127 | 75 | 8014 | 4568 | 58 | 1.1% | 0.58 [0.23, 0.93] |
| Tabak, 2014 [123] | 5603 | 3475.8 | 13 | 4617 | 3460 | 16 | 0.6% | 0.28 [-0.46, 1.01] |
| Talbot, 2003 [67] | 4337 | 2,903 | 17 | 3972 | 2563 | 17 | 0.7% | 0.13 [-0.54, 0.80] |
| Ter Hoeve, 2018 [68] | 8 | 2.6 | 121 | 7.7 | 3 | 252 | 1.3% | 0.10 [-0.11, 0.32] |
| Thorndike, 2014 [124] | 7886 | 3622 | 50 | 7600 | 3492 | 49 | 1.1% | 0.08 [-0.31, 0.47] |
| Tudor-Locke, 2004 [69] | 9123 | 4539 | 24 | 5622 | 2405 | 23 | 0.8% | 0.94 [0.34, 1.55] |
| Vallance, 2016 [125] | 5923 | 3109 | 41 | 6885 | 3576 | 37 | 1.0% | -0.29 [-0.73, 0.16] |
| Van Blarigan, 2019 [126] | 10,047 | 4461 | 20 | 12,541 | 5535 | 19 | 0.7% | -0.49 [-1.13, 0.15] |
| Van Hoye, 2018 [72] | 13.2 | 29.6 | 157 | 6.7 | 26 | 50 | 1.2% | 0.23 [-0.09, 0.54] |
| Warren, 2014 [127] | 8371 | 3069 | 37 | 7576 | 2993 | 69 | 1.1% | 0.26 [-0.14, 0.66] |
| Wyke, 2019 [77] | 9801 | 3730 | 464 | 8518 | 3254 | 471 | 1.4% | 0.37 [0.24, 0.50] |
| Yamada, 2012 [128] | 3726 | 1607 | 40 | 2267 | 1837 | 42 | 1.0% | 0.84 [0.38, 1.29] |
| Yates, 2017 [78] | -486 | 2252 | 287 | -690 | 2041 | 272 | 1.4% | 0.09 [-0.07, 0.26] |
| **Subtotal (95% CI)** | | | **6390** | | | **6073** | **73.6%** | **0.46 [0.36, 0.55]** |

Heterogeneity: $Tau^2=0.14$; $Chi^2=458.16$, $df=76$ ($P<.001$); $I^2=83\%$ Test for overall effect: $Z=8.98$ ($P<.001$)

◆=Subtotal or total standardized mean difference

Standardized mean difference IV, Random, 95% CI

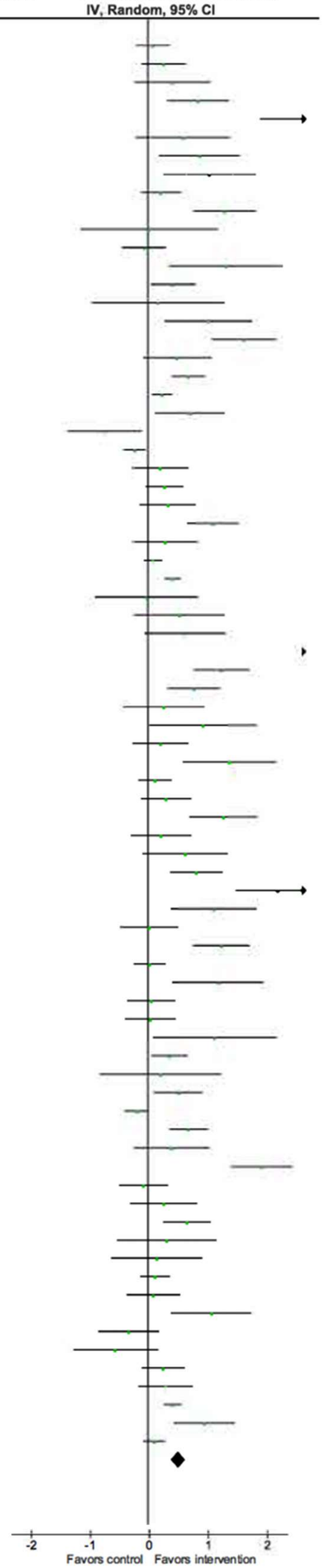


**Figure 3:** Primary Outcome Analysis for The Outcome Walking at Timepoint Short-term Follow-up

| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Standardized mean difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **Short-term follow-up** | | | | | | | | |
| Aittasalo, 2012 [41] | 521 | 468 | 123 | 395 | 319 | 118 | 1.3% | 0.31 [0.06, 0.57] |
| Coelho, 2018 [43] | 8996 | 3551 | 12 | 6963 | 3857 | 13 | 0.6% | 0.53 [-0.27, 1.33] |
| Cruz, 2016 [44] | 9747 | 3511 | 13 | 6481 | 3454 | 13 | 0.6% | 0.91 [0.09, 1.72] |
| Finkelstein, 2016 [48] | -480 | 2893 | 201 | -23.6 | 2878.9 | 599 | 1.4% | -0.16 [-0.32, 0.00] |
| Furber, 2010 [50] | 231.5 | 189.1 | 95 | 168.3 | 171.8 | 106 | 1.2% | 0.35 [0.07, 0.63] |
| James, 2015 [51] | 478.8 | 3851 | 46 | -1282 | -3828 | 48 | | Not estimable |
| Kernot, 2019 [52] | 191 | 107.3 | 66 | 192 | 165.7 | 30 | 1.0% | -0.01 [-0.44, 0.42] |
| Maher, 2015 [53] | 165 | 186 | 51 | 133 | 137 | 59 | 1.1% | 0.20 [-0.18, 0.57] |
| Mansi, 2015 [54] | 9645 | 1906 | 29 | 6266 | 1648 | 29 | 0.8% | 1.87 [1.25, 2.49] |
| Mutrie, 2012 [58] | 9161 | 2631 | 19 | 9100 | 3175 | 17 | 0.7% | 0.02 [-0.63, 0.67] |
| Nolan, 2017 [59] | -263 | 1861.3 | 56 | -461 | 2130.1 | 57 | 1.1% | 0.10 [-0.27, 0.47] |
| Oliveira, 2019 [60] | 7010 | 3163 | 46 | 6584 | 2612 | 52 | 1.1% | 0.15 [-0.25, 0.54] |
| Ribeiro, 2014 [63] | 535.2 | 1805.6 | 101 | 128.5 | 1723 | 94 | 1.2% | 0.23 [-0.05, 0.51] |
| Simons, 2018 [64] | 7767 | 4470 | 53 | 8543 | 4862 | 57 | 1.1% | -0.16 [-0.54, 0.21] |
| Stacey, 2016 [65] | 10,307 | 4446 | 75 | 8026 | 4698 | 58 | 1.1% | 0.50 [0.15, 0.85] |
| Talbot, 2003 [67] | 3729 | 2347 | 17 | 4175 | 2655 | 17 | 0.7% | -0.17 [-0.85, 0.50] |
| Ter Hoeve, 2018 [68] | 8 | 2.8 | 112 | 7.5 | 2.8 | 247 | 1.3% | 0.18 [-0.05, 0.40] |
| Tudor-Locke, 2004 [69] | 7924 | 3308 | 16 | 6557 | 2742 | 22 | 0.7% | 0.45 [-0.21, 1.10] |
| Van Hoye, 2018 [72] | 11.3 | 24.4 | 157 | 2.8 | 23 | 50 | 1.2% | 0.35 [0.03, 0.67] |
| **Subtotal (95% CI)** | | | 1288 | | | 1686 | 18.1% | 0.26 [0.09, 0.42] |

Heterogeneity: $Tau^2$=0.08; $Chi^2$=62.03, df=17 (P<.001); $I^2$=73% Test for overall effect: Z=3.08 (P=.002)

◆=Subtotal or total standardized mean difference


**Figure 4:** Primary Outcome Analysis for The Outcome Walking at Timepoint Long-term Follow-up

| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Standardized mean difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **Long-term follow-up** | | | | | | | | |
| DeGreef, 2010 [73] | 8024 | 5331 | 20 | 4350 | 3214 | 21 | 0.7% | 0.82 [0.18, 1.46] |
| DeGreef, 2011[89] | 6831 | 3164 | 60 | 3864 | 3440 | 32 | 1.0% | 0.90 [0.45, 1.35] |
| Harris, 2018 [75] | 7971 | 3099 | 770 | 7135 | 2709 | 459 | 1.4% | 0.28 [0.17, 0.40] |
| Kolt, 2012 [76] | 94 | 153 | 140 | 89.4 | 172.8 | 130 | 1.3% | 0.03 [-0.21, 0.27] |
| Van Hoye, 2018 [72] | 7.2 | 25.1 | 157 | 3.9 | 22.6 | 50 | 1.2% | 0.13 [-0.18, 0.45] |
| Wyke, 2019 [77] | 9234 | 3530 | 451 | 8494 | 3168 | 470 | 1.4% | 0.22 [0.09, 0.35] |
| Yates, 2017 [78] | -599 | 2144 | 277 | -769 | 2192 | 274 | 1.4% | 0.08 [-0.09, 0.25] |
| **Subtotal (95% CI)** | | | 1875 | | | 1436 | 8.4% | 0.25 [0.10, 0.39] |

Heterogeneity: $Tau^2$=0.02; $Chi^2$=18.89, df=6 (P=.004); $I^2$ = 68%
Test for overall effect: Z=3.28 (P=.001)

◆=Subtotal or total standardized mean difference

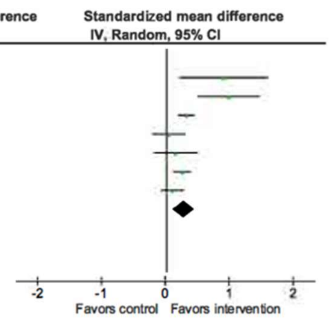


**Figure 5:** Primary Outcome Analysis for The Outcome Moderate-to-vigorous Physical Activity at Timepoint End of Intervention

| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Standardized mean difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **End of intervention** | | | | | | | | |
| Aittasalo, 2006 [131] | 95 | 83.7 | 203 | 94 | 70.9 | 62 | 1.4% | 0.01 [-0.27, 0.30] |
| Alsaleh, 2016 [79] | 237.9 | 121.1 | 66 | 108.9 | 198 | 79 | 1.2% | 0.77 [0.43, 1.11] |
| Ashton, 2017 [80] | 154.1 | 216.1 | 26 | 26.1 | 210.2 | 24 | 0.7% | 0.59 [0.02, 1.16] |
| Barnes, 2015 [42] | 0.2 | 1.6 | 23 | -0.19 | 1.5 | 19 | 0.6% | 0.25 [-0.36, 0.86] |
| Barwais, 2013 [82] | 649 | 494.9 | 18 | 192 | 221.3 | 15 | 0.5% | 1.13 [0.38, 1.87] |
| Cadmus-Bertram, 2019 [84] | 230 | 120 | 24 | 143 | 107 | 23 | 0.7% | 0.75 [0.16, 1.35] |
| Compernolle, 2015 [85] | 32.4 | 37.7 | 62 | 38.4 | 60 | 79 | 1.3% | -0.12 [-0.45, 0.22] |
| Creel, 2016 [86] | 37 | 41.6 | 52 | 20.4 | 17.1 | 35 | 1.0% | 0.48 [0.05, 0.92] |
| Cruz, 2016 [44] | 57.8 | 32.8 | 13 | 26.7 | 19.6 | 13 | 0.4% | 1.11 [0.28, 1.95] |
| Dadaczynski, 2017 [87] | 310.9 | 129.4 | 80 | 291.8 | 114 | 64 | 1.3% | 0.15 [-0.17, 0.48] |
| DeGreef, 2010 [73] | 44 | 38 | 20 | 24 | 29 | 21 | 0.6% | 0.58 [-0.04, 1.21] |
| DeGreef, 2011[89] | 23 | 20 | 60 | 20 | 25 | 32 | 1.0% | 0.14 [-0.29, 0.57] |
| DeGreef, 2011a [74] | 82 | 84 | 43 | 27 | 39 | 24 | 0.8% | 0.76 [0.24, 1.28] |
| Demeyer, 2017 [90] | 8 | 21.1 | 140 | -3 | 17.51 | 140 | 1.6% | 0.57 [0.33, 0.80] |
| Dishman, 2009 [91] | 10.6 | 15.5 | 564 | 7.9 | 15.1 | 265 | 1.9% | 0.18 [0.03, 0.32] |
| Eakin, 2014 [46] | 163.6 | 155.6 | 119 | 114.6 | 108.6 | 127 | 1.5% | 0.37 [0.11, 0.62] |
| Edney, 2020 [47] | 108.1 | 74.1 | 272 | 108.9 | 52.6 | 130 | 1.7% | -0.01 [-0.22, 0.20] |
| Fischer, 2019 [132] | 228.4 | 157.8 | 63 | 194 | 136 | 128 | 1.4% | 0.24 [-0.06, 0.54] |
| Fjeldsoe, 2010 [94] | 18.26 | 167.3 | 45 | 16.4 | 167.4 | 43 | 1.0% | 0.01 [-0.41, 0.43] |
| Fjeldsoe, 2015 [49] | -8 | 262.4 | 104 | -29.2 | 196 | 107 | 1.5% | 0.09 [-0.18, 0.36] |
| Golsteijn, 2018 [133] | 331 | 234 | 208 | 301 | 219 | 211 | 1.7% | 0.13 [-0.06, 0.32] |
| Hardeman, 2020 [98] | 77.3 | 36.5 | 417 | 76.7 | 35.4 | 442 | 1.9% | 0.02 [-0.12, 0.15] |
| Harris, 2018 [75] | 374 | 179 | 778 | 310 | 168 | 456 | 2.0% | 0.37 [0.25, 0.48] |
| James, 2015 [51] | 33.9 | 145.3 | 57 | 9.6 | 128.6 | 52 | 1.1% | 0.18 [-0.20, 0.55] |
| Katzmarzyk, 2011[104] | 16.3 | 17.3 | 20 | 16.2 | 17.1 | 23 | 0.7% | 0.01 [-0.59, 0.60] |
| Kendzor, 2017 [134] | 76 | 67 | 17 | 46 | 43 | 15 | 0.5% | 0.51 [-0.19, 1.22] |
| Kernot, 2019 [52] | 204.1 | 131 | 74 | 150 | 96.9 | 33 | 1.0% | 0.44 [0.03, 0.86] |
| Keyserling, 2008 [135] | 14 | 3.71 | 86 | 13 | 2.8 | 89 | 1.4% | 0.30 [0.01, 0.60] |
| Kim, 2018 [136] | 51.2 | 4.4 | 41 | 50.3 | 4.4 | 42 | 1.0% | 0.20 [-0.23, 0.63] |
| King, 2008 [137] | 310.6 | 298.3 | 19 | 135 | 208.2 | 18 | 0.6% | 0.66 [0.00, 1.33] |
| Koizumi, 2009 [138] | 27.2 | 14.7 | 34 | 19 | 9.8 | 34 | 0.9% | 0.65 [0.16, 1.14] |
| Kolt, 2012 [76] | 121.2 | 184.4 | 130 | 111.4 | 110.6 | 123 | 1.6% | 0.06 [-0.18, 0.31] |
| Li, 2017 [139] | 64.2 | 70.5 | 17 | 56 | 60.1 | 17 | 0.6% | 0.12 [-0.55, 0.80] |
| Li, 2020 [107] | 44.7 | 41.2 | 55 | 31.6 | 32.4 | 57 | 1.1% | 0.35 [-0.02, 0.73] |
| Lynch, 2019 [109] | 249.6 | 159.7 | 40 | 170.1 | 127.9 | 40 | 1.0% | 0.54 [0.10, 0.99] |
| Maher, 2015 [53] | 528 | 391 | 51 | 391 | 371 | 59 | 1.1% | 0.36 [-0.02, 0.74] |
| Mansi, 2015 [54] | 1469 | 524 | 29 | 538 | 254 | 29 | 0.6% | 2.23 [1.57, 2.90] |
| Maselli, 2019 [55] | 134.4 | 83.4 | 11 | 185.5 | 126.3 | 21 | 0.5% | -0.44 [-1.18, 0.30] |
| Maxwell Smith, 2019 [140] | 312 | 229.3 | 34 | 240 | 177.7 | 33 | 0.9% | 0.35 [-0.14, 0.83] |
| Melville, 2015 [112] | 3 | 2.6 | 42 | 3.1 | 2.1 | 40 | 1.0% | -0.04 [-0.47, 0.39] |
| Merom, 2007 [114] | 79 | 543.7 | 105 | 32.8 | 171.3 | 209 | 1.6% | 0.13 [-0.10, 0.37] |
| Murawski, 2019 [57] | 428.4 | 523.4 | 59 | 319.7 | 378.2 | 66 | 1.2% | 0.24 [-0.11, 0.59] |
| Nolan, 2017 [59] | 11 | 68.8 | 63 | 11 | 58.8 | 59 | 1.2% | 0.00 [-0.36, 0.36] |
| Pekmezi, 2017 [141] | 31.5 | 58.9 | 39 | 20.8 | 38.2 | 37 | 0.9% | 0.21 [-0.24, 0.66] |
| Pinto, 2013 [61] | 214 | 147.3 | 19 | 97 | 148 | 24 | 0.6% | 0.78 [0.15, 1.40] |
| Pinto, 2015 [62] | 70.3 | 65.9 | 36 | 16.5 | 31.9 | 32 | 0.8% | 1.01 [0.50, 1.52] |
| Pope, 2018 [117] | 34.2 | 18.7 | 12 | 37.8 | 20.4 | 8 | 0.4% | -0.18 [-1.07, 0.72] |
| Prestwich, 2009 [142] | 1.3 | 1.16 | 60 | 1 | 1 | 94 | 1.3% | 0.28 [-0.05, 0.61] |
| Ribeiro, 2014 [63] | 473.5 | 913.2 | 101 | 58.5 | 484.3 | 94 | 1.4% | 0.56 [0.27, 0.85] |
| Samuels, 2011 [143] | 34.8 | 13.6 | 12 | 26.5 | 16.1 | 25 | 0.5% | 0.53 [-0.17, 1.23] |
| Sharp, 2016 [144] | 100 | 109 | 72 | 90 | 119 | 65 | 1.3% | 0.09 [-0.25, 0.42] |
| Simons, 2018 [64] | 27.3 | 26.7 | 55 | 30.1 | 31 | 63 | 1.2% | -0.10 [-0.46, 0.27] |
| Ter Hoeve, 2018 [68] | 6.7 | 3 | 121 | 6.5 | 3.1 | 252 | 1.7% | 0.07 [-0.15, 0.28] |
| Unick, 2012 [145] | 133 | 217 | 11 | 44.8 | 124.6 | 12 | 0.4% | 0.49 [-0.35, 1.32] |
| Vallance, 2008 [70] | 213 | 173 | 172 | 180 | 141 | 166 | 1.7% | 0.21 [-0.01, 0.42] |
| Vallance, 2016 [125] | 143 | 132 | 41 | 146 | 155 | 37 | 1.0% | -0.02 [-0.47, 0.42] |
| Van Blarigan, 2019 [126] | 46.6 | 48.4 | 20 | 54.5 | 24.9 | 19 | 0.6% | -0.20 [-0.83, 0.43] |
| Vandelanotte, 2018 [146] | 148.8 | 181.1 | 78 | 79.8 | 77.1 | 46 | 1.2% | 0.45 [0.08, 0.82] |
| Van der Weegen, 2015 [71] | 48.2 | 23.8 | 65 | 42.9 | 25.8 | 134 | 1.4% | 0.21 [-0.09, 0.51] |
| Van Hoye, 2018 [72] | 17.8 | 49.6 | 157 | 9.6 | 34 | 50 | 1.3% | 0.18 [-0.14, 0.49] |
| Wijsman, 2013 [147] | 11.1 | 21.8 | 108 | -0.1 | 15.4 | 105 | 1.5% | 0.59 [0.32, 0.86] |
| Yates, 2017 [78] | -4.1 | 19 | 287 | -5.9 | 17.7 | 272 | 1.8% | 0.10 [-0.07, 0.26] |
| **Subtotal (95% CI)** | | | **5850** | | | **5133** | **68.0%** | **0.28 [0.21, 0.35]** |

Heterogeneity: Tau²=0.04; Chi²=161.93, df=61 (P<.001); I²=62%

Test for overall effect: Z=7.81 (P<.001)

◆=Subtotal or total standardized mean difference

**Standardized mean Difference**
**IV, Random, 95% CI**

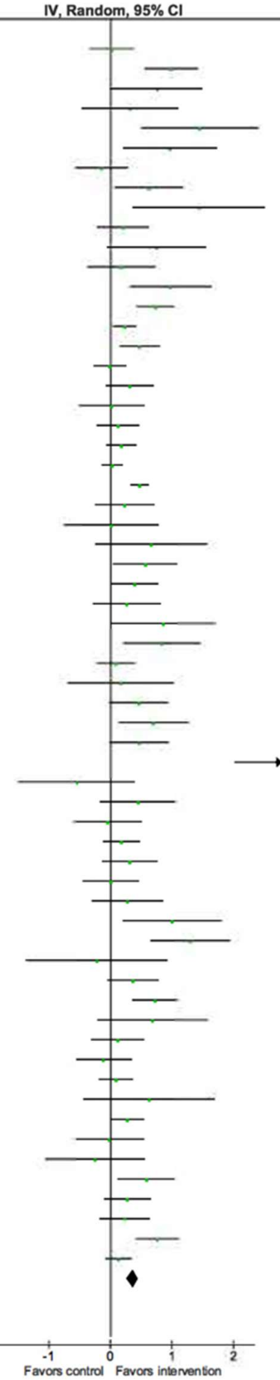
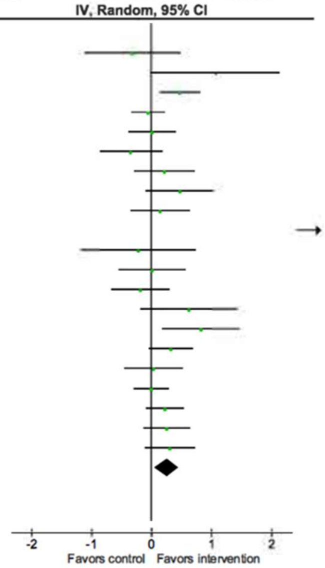


**Figure 6:** Primary Outcome Analysis for The Outcome Moderate-to-vigorous Physical Activity at Timepoint Short-term Follow-up



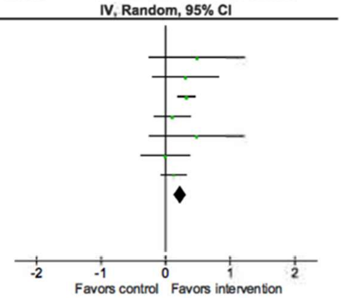
| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Standardized mean difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **Short-term follow-up** | | | | | | | | |
| Barnes, 2015 [42] | -0.5 | 1.7 | 23 | -0.1 | 1.6 | 19 | 0.6% | -0.24 [-0.85, 0.37] |
| Cruz, 2016 [44] | 51.6 | 29.4 | 13 | 28 | 26 | 13 | 0.4% | 0.82 [0.02, 1.63] |
| Eakin, 2014 [46] | 137.8 | 139.4 | 115 | 95.6 | 89.9 | 124 | 1.5% | 0.36 [0.11, 0.62] |
| Edney, 2020 [47] | 109.5 | 53.4 | 261 | 111.7 | 52.6 | 122 | 1.7% | -0.04 [-0.26, 0.17] |
| Fjeldsoe, 2015 [49] | -54.8 | 285.1 | 83 | -57.2 | 231.5 | 87 | 1.4% | 0.01 [-0.29, 0.31] |
| James, 2015 [51] | -16.2 | 79.6 | 46 | 8.7 | 104 | 48 | 1.1% | -0.27 [-0.67, 0.14] |
| Kangasniemi, 2015 [40] | 29.5 | 17.6 | 54 | 26.6 | 16.8 | 49 | 1.1% | 0.17 [-0.22, 0.55] |
| Kernot, 2019 [52] | 199.3 | 118.6 | 66 | 160 | 75 | 30 | 1.0% | 0.36 [-0.07, 0.80] |
| Maher, 2015 [53] | 376 | 377 | 51 | 335 | 342 | 59 | 1.1% | 0.11 [-0.26, 0.49] |
| Mansi, 2015 [54] | 1,383 | 402 | 29 | 520 | 246 | 29 | 0.5% | 2.55 [1.85, 3.26] |
| Maselli, 2019 [55] | 185.6 | 158 | 11 | 206.2 | 98.9 | 21 | 0.5% | -0.16 [-0.90, 0.57] |
| Murawski, 2019 [57] | 405.3 | 491.5 | 35 | 400.2 | 497.8 | 54 | 1.0% | 0.01 [-0.42, 0.44] |
| Nolan, 2017 [59] | 2 | 70.6 | 56 | 12 | 73.2 | 57 | 1.2% | -0.14 [-0.51, 0.23] |
| Pinto, 2013 [61] | 160 | 145.2 | 19 | 89 | 145.7 | 23 | 0.6% | 0.48 [-0.14, 1.10] |
| Pinto, 2015 [62] | 54.6 | 81.6 | 36 | 13.4 | 35.2 | 31 | 0.9% | 0.63 [0.14, 1.12] |
| Ribeiro, 2014 [63] | 254.2 | 1135.5 | 101 | -23.5 | 1084.9 | 94 | 1.4% | 0.25 [-0.03, 0.53] |
| Simons, 2018 [64] | 32.6 | 26.2 | 53 | 31.8 | 29.4 | 57 | 1.1% | 0.03 [-0.35, 0.40] |
| Ter Hoeve, 2018 [68] | 6.6 | 3.1 | 112 | 6.6 | 3.2 | 247 | 1.6% | 0.00 [-0.22, 0.22] |
| Vallance, 2008 [70] | 184 | 201 | 136 | 153 | 148 | 130 | 1.6% | 0.17 [-0.07, 0.42] |
| Van der Weegen, 2015 [71] | 48.8 | 23.8 | 65 | 43.85 | 25.7 | 134 | 1.4% | 0.20 [-0.10, 0.49] |
| Van Hoye, 2018 [72] | 22.3 | 57.2 | 157 | 9.3 | 45.5 | 50 | 1.3% | 0.24 [-0.08, 0.56] |
| **Subtotal (95% CI)** | | | 1522 | | | 1478 | 23.1% | 0.20 [0.05, 0.35] |

Heterogeneity: Tau²=0.08; Chi²=71.70, df=20 (P<.001); I² = 72%

Test for overall effect: Z=2.63 (P=.008)

◆=Subtotal or total standardized mean difference

**Figure 7:** Primary Outcome Analysis for The Outcome Moderate-to-vigorous Physical Activity at Timepoint Long-term Follow-up



| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Standardized mean difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **Long-term follow-up** | | | | | | | | |
| DeGreef, 2010 [73] | 32 | 34 | 20 | 20 | 24 | 21 | 0.6% | 0.40 [-0.22, 1.02] |
| DeGreef, 2011[89] | 19 | 19 | 60 | 14 | 20 | 32 | 1.0% | 0.26 [-0.17, 0.69] |
| Harris, 2018 [75] | 354 | 179 | 770 | 308 | 160 | 459 | 2.0% | 0.27 [0.15, 0.38] |
| Kolt, 2012 [76] | 114.2 | 186.8 | 140 | 99.4 | 157.4 | 130 | 1.6% | 0.09 [-0.15, 0.32] |
| Pinto, 2013 [61] | 146 | 142.1 | 19 | 88 | 143.4 | 23 | 0.6% | 0.40 [-0.22, 1.01] |
| Van Hoye, 2018 [72] | 10.3 | 50.5 | 157 | 10.5 | 42 | 50 | 1.3% | -0.00 [-0.32, 0.31] |
| Yates, 2017 [78] | -4.7 | 17.4 | 278 | -6.6 | 19.4 | 274 | 1.8% | 0.10 [-0.06, 0.27] |
| **Subtotal (95% CI)** | | | 1444 | | | 989 | 9.0% | 0.19 [0.11, 0.27] |

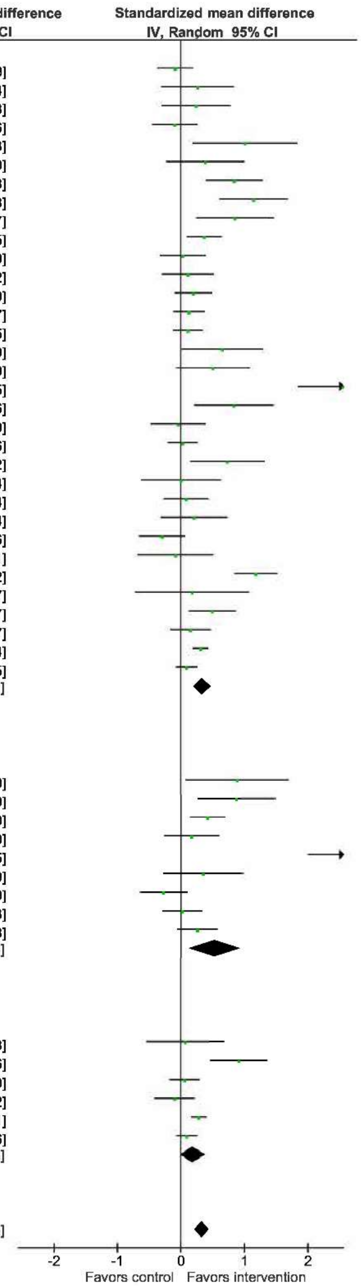
Heterogeneity: Tau²=0.00; Chi²=5.88, df=6 (P=0.44); I²=0%

Test for overall effect: Z=4.58 (P<.001)

◆=Subtotal or total standardized mean difference

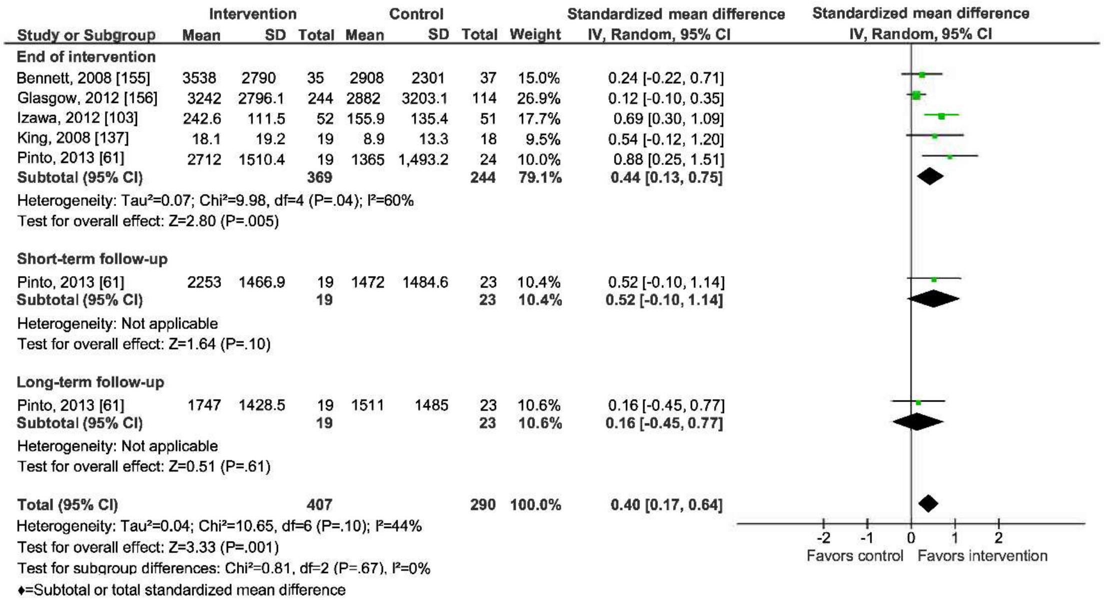
**Figure 8:** Primary Outcome Analysis by Measurement Timepoint for The Outcome Total Physical Activity

| | Intervention | | | Control | | | | Standardized mean difference |
|---|---|---|---|---|---|---|---|---|
| **Study or Subgroup** | **Mean** | **SD** | **Total** | **Mean** | **SD** | **Total** | **Weight** | **IV, Random, 95% CI** |
| **End of intervention** | | | | | | | | |
| Aittasalo, 2006 [131] | 509.5 | 479.4 | 203 | 555 | 519.7 | 62 | 2.6% | -0.09 [-0.38, 0.19] |
| Cadmus-Bertram, 2019 [84] | 1463 | 489 | 24 | 1343 | 395 | 23 | 1.7% | 0.26 [-0.31, 0.84] |
| Carr, 2013 [148] | 171.4 | 234.5 | 25 | 121.8 | 174.6 | 28 | 1.8% | 0.24 [-0.30, 0.78] |
| Compernolle, 2015 [85] | 81.6 | 78 | 52 | 90.6 | 100.2 | 71 | 2.3% | -0.10 [-0.46, 0.26] |
| Cruz, 2016 [44] | 279.5 | 74 | 13 | 212 | 53.9 | 13 | 1.1% | 1.01 [0.19, 1.83] |
| DeGreef, 2010 [73] | 301 | 106 | 20 | 260 | 104 | 21 | 1.6% | 0.38 [-0.24, 1.00] |
| DeGreef, 2011[89] | 93 | 66 | 60 | 40 | 56 | 32 | 2.1% | 0.84 [0.39, 1.28] |
| DeGreef, 2011a [74] | 176 | 108 | 43 | 65 | 68 | 24 | 1.8% | 1.15 [0.61, 1.68] |
| Dlugonski, 2012 [45] | 28.2 | 15.6 | 22 | 15.4 | 13.9 | 23 | 1.6% | 0.85 [0.24, 1.47] |
| Furber, 2010 [50] | 366.5 | 270.8 | 97 | 270.9 | 244.4 | 107 | 2.6% | 0.37 [0.09, 0.65] |
| Gill, 2019 [96] | 2.13 | 25.8 | 59 | 1.37 | 25.7 | 59 | 2.3% | 0.03 [-0.33, 0.39] |
| Kernot, 2019 [52] | 298.9 | 97.7 | 74 | 288.8 | 68.6 | 33 | 2.2% | 0.11 [-0.30, 0.52] |
| Keyserling, 2008 [135] | 29.8 | 6.5 | 86 | 28.6 | 5.7 | 90 | 2.5% | 0.20 [-0.10, 0.49] |
| Kolt, 2012 [76] | 168.9 | 190.5 | 130 | 146.3 | 163.5 | 123 | 2.7% | 0.13 [-0.12, 0.37] |
| Lane, 2015 [149] | 84.9 | 50.3 | 125 | 80.3 | 32.8 | 159 | 2.7% | 0.11 [-0.12, 0.35] |
| Lyons, 2017 [110] | 117 | 121 | 20 | 58 | 33 | 20 | 1.5% | 0.65 [0.01, 1.29] |
| Mailey, 2010 [150] | 299,792 | 102,800 | 24 | 251,625 | 83,081 | 23 | 1.7% | 0.51 [-0.08, 1.09] |
| Mansi, 2015 [54] | 1,035 | 444 | 29 | 188 | 135 | 29 | 1.4% | 2.55 [1.84, 3.25] |
| Martin, 2015 [111] | 10.5 | 21.1 | 32 | -8 | 23 | 16 | 1.5% | 0.84 [0.21, 1.46] |
| Melville, 2015 [112] | 33.5 | 10 | 42 | 34 | 12 | 40 | 2.1% | -0.04 [-0.48, 0.39] |
| Merom, 2007 [114] | 91 | 243.1 | 105 | 84.2 | 231.8 | 209 | 2.7% | 0.03 [-0.21, 0.26] |
| Motl, 2011 [151] | 24.7 | 18.8 | 23 | 12.4 | 14.2 | 25 | 1.6% | 0.73 [0.14, 1.32] |
| Müller, 2016 [56] | 383.4 | 843.4 | 18 | 377.4 | 842.6 | 21 | 1.5% | 0.01 [-0.62, 0.64] |
| Prestwich, 2010 [118] | 3 | 1.8 | 88 | 2.3 | 14.3 | 46 | 2.3% | 0.08 [-0.27, 0.44] |
| Schwerdtfeger, 2012 [154] | 738.6 | 245.7 | 21 | 690.5 | 217.1 | 41 | 1.8% | 0.21 [-0.32, 0.74] |
| Simons, 2018 [64] | 317 | 135 | 55 | 359.5 | 146 | 63 | 2.3% | -0.30 [-0.66, 0.06] |
| Sugden, 2008 [152] | 108.7 | 54.7 | 26 | 113.8 | 62.3 | 18 | 1.6% | -0.09 [-0.69, 0.51] |
| Suggs, 2013 [66] | 49.5 | 39.6 | 79 | 5.8 | 33.9 | 79 | 2.4% | 1.18 [0.84, 1.52] |
| Thorsteinsen, 2014 [153] | 574 | 435 | 12 | 502 | 323 | 8 | 1.0% | 0.17 [-0.72, 1.07] |
| Vandelanotte, 2018 [146] | 387.7 | 377.7 | 78 | 230 | 164.1 | 46 | 2.3% | 0.50 [0.13, 0.87] |
| Van Hoye, 2018 [72] | 1.8 | 8.3 | 157 | 0.6 | 7 | 50 | 2.5% | 0.15 [-0.17, 0.47] |
| Wyke, 2019 [77] | 3717 | 3307 | 499 | 2741 | 2951 | 505 | 3.0% | 0.31 [0.19, 0.44] |
| Yates, 2017 [78] | -32.3 | 94.6 | 287 | -40.3 | 85 | 272 | 2.9% | 0.09 [-0.08, 0.25] |
| **Subtotal (95% CI)** | | | **2628** | | | **2379** | **67.8%** | **0.34 [0.20, 0.47]** |
| Heterogeneity: Tau² =0.10; Chi² =140.09, df =32 (P <.001); I² =77% | | | | | | | | |
| Test for overall effect: Z =4.96 (P <.001) | | | | | | | | |
| **Short-term follow-up** | | | | | | | | |
| Cruz, 2016 [44] | 269.3 | 61.5 | 13 | 202.9 | 82.5 | 13 | 1.1% | 0.88 [0.07, 1.70] |
| Dlugonski, 2012 [45] | 26.9 | 16.2 | 22 | 14.5 | 11.4 | 23 | 1.6% | 0.87 [0.26, 1.49] |
| Furber, 2010 [50] | 358.7 | 260 | 95 | 250.9 | 250.6 | 106 | 2.6% | 0.42 [0.14, 0.70] |
| Kernot, 2019 [52] | 289.7 | 102.3 | 66 | 273.2 | 83.3 | 30 | 2.1% | 0.17 [-0.26, 0.60] |
| Mansi, 2015 [54] | 972 | 383 | 29 | 180 | 133 | 29 | 1.3% | 2.73 [2.00, 3.45] |
| Müller, 2016 [56] | 434.7 | 728.7 | 18 | 172 | 728.2 | 21 | 1.5% | 0.35 [-0.28, 0.99] |
| Simons, 2018 [64] | 326.9 | 132.5 | 53 | 364.8 | 138.9 | 57 | 2.3% | -0.28 [-0.65, 0.10] |
| Suggs, 2013 [66] | 48.8 | 33.1 | 79 | 48.1 | 28.7 | 79 | 2.5% | 0.02 [-0.29, 0.33] |
| Van Hoye, 2018 [72] | 2.8 | 8.8 | 157 | 0.6 | 7.4 | 50 | 2.5% | 0.26 [-0.06, 0.58] |
| **Subtotal (95% CI)** | | | **532** | | | **408** | **17.5%** | **0.53 [0.13, 0.93]** |
| Heterogeneity: Tau² =0.31; Chi² =61.29, df =8 (P <.001); I² = 87% Test for overall effect: Z =2.61 (P =.009) | | | | | | | | |
| **Long-term follow-up** | | | | | | | | |
| DeGreef, 2010 [73] | 253 | 99 | 20 | 246 | 109 | 21 | 1.6% | 0.07 [-0.55, 0.68] |
| DeGreef, 2011[89] | 77 | 51 | 60 | 35 | 33 | 32 | 2.0% | 0.91 [0.46, 1.36] |
| Kolt, 2012 [76] | 160.4 | 243 | 140 | 147.1 | 213.2 | 130 | 2.7% | 0.06 [-0.18, 0.30] |
| Van Hoye, 2018 [72] | 0.7 | 7.86 | 157 | 1.5 | 8.4 | 50 | 2.5% | -0.10 [-0.42, 0.22] |
| Wyke, 2019 [77] | 3523 | 3158 | 489 | 2670 | 2899 | 504 | 3.0% | 0.28 [0.16, 0.41] |
| Yates, 2017 [78] | -9.3 | 446.6 | 278 | -39.5 | 93.7 | 274 | 2.9% | 0.09 [-0.07, 0.26] |
| **Subtotal (95% CI)** | | | **1144** | | | **1011** | **14.7%** | **0.19 [-0.00, 0.38]** |
| Heterogeneity: Tau² =0.04; Chi² =17.79, df =5 (P =.003); I² = 72% | | | | | | | | |
| Test for overall effect: Z =1.92 (P =.05) | | | | | | | | |
| **Total (95% CI)** | | | **4304** | | | **3798** | **100.0%** | **0.34 [0.23, 0.45]** |




Heterogeneity: Tau² =0.10; Chi² =221.44, df =47 (P<.001); I² = 79%
Test for overall effect: Z =6.10 (P<.001)
Test for subgroup differences: Chi² =2.81, df =2 (P =.24), I² =28.9%
◆ =Subtotal or total standardized mean difference

**Figure 9:** Primary Outcome Analysis by Measurement Timepoint for The Outcome Energy Expenditure




## Publication Bias Assessment

Publication bias was assessed using funnel plot analysis for end-of-intervention measurements, as all but one study [40] reported this time point (Appendix 5 [54,82,102]). No systematic publication bias was observed. However, the funnel plot analysis revealed 3 outlier studies [54,82,102]. We identified unusually high adherence rates [54], possibly because the research team and the study participants were based on the same campus, and a short intervention duration (only 4 weeks) [82,102] as potential reasons for the high effect scores in the outlier studies. We conducted a sensitivity analysis and excluded these studies across all outcome measures. All effects were found to be stable, and heterogeneity was substantially reduced (Appendix 6).

## Sensitivity Analyses

Sensitivity analysis by risk of bias was conducted for all outcome measures but for EE, as only one study [155] measuring EE classified as low risk of bias. Results at the end of intervention were found to be robust across outcome measures (Appendix 6), with effect sizes substantially increasing for walking, MVPA, and TPA to moderate effect sizes. Short- and long-term follow-up effects were not statistically significant when only studies with low risk of bias were included. Heterogeneity increased and remained substantial. Sensitivity analysis of studies reporting long-term follow-up measurements was conducted for all outcome measures but for EE, as only one study [61] measuring EE reported a long-term follow-up measurement. The results across all time points were robust for all outcome measures.

**Subgroup Analysis by Population Type**

We used subgroup analysis to evaluate the effect moderators. Table 1 summarizes all results, and Appendices 7 [41-154], 8 [41-154], and 9 [41-154] provide detailed forest plots for each analysis. We found that population type moderates the effect of mHealth interventions on PA. The intervention design and control group type were not found to be significant effect moderators. Subgroup analysis by population type revealed statistically significant ($P \leq .10$) quantitative subgroup effects for all outcome measures.

The treatment effect at the end of intervention was greater in sick populations (walking SMD 0.44, 95% CI 0.29-0.60, $P$<.001, $I^2$=71%, $P$<.001; MVPA SMD 0.33, 95% CI 0.21-0.45, $P$<.001, $I^2$=55%, $P$<.001; TPA SMD 0.59, 95% CI 0.36-0.81, $P$<.001, $I^2$=49%, $P$=.03; Appendix 7) than in healthy populations (walking SMD 0.20, 95% CI 0.04-0.35, $P$=.01, $I^2$=78%, $P$<.001; MVPA SMD 0.14, 95% CI 0.06-0.23, $P$=.001, $I^2$=15%, $P$=.29; TPA SMD 0.29, 95% CI −0.10 to 0.67, $P$=.14, $I^2$=85%, $P$<.001; Appendix 7). Within the healthy subgroup, summary effects were only statistically significant for the outcome measures walking and MVPA. The results for at-risk populations were mixed. The outcome measures walking and MVPA exhibited high effect scores, similar to the high effect scores of sick populations (walking SMD 0.59, 95% CI 0.42-0.76, $P$<.001, $I^2$=87%, $P$<.001; MVPA SMD 0.30, 95% CI 0.18-0.43, $P$<.001, $I^2$=72%, $P$<.001; Multimedia Appendix 7), whereas effect scores for at-risk populations were lower for TPA (SMD 0.21, 95% CI 0.04-0.38; $P$=.02; $I^2$=78%, $P$<.001).

Although heterogeneity was somewhat reduced within most subgroups compared with the overall outcome heterogeneity, it remained high and significant. The covariate distribution between sick, at-risk, and healthy population subgroups was uneven, as fewer studies investigated preventative mHealth PA interventions in healthy populations.

**Table 1:** Summary of Subgroup Analyses Results

| Analysis by | Test for subgroup differences *P*-value | Outcome Measure | Timepoint | No. of studies | SMD 95% CI | *P*-value | Heterogeneity $I^2$ | Heterogeneity *P*-value |
|---|---|---|---|---|---|---|---|---|
| **Population Type** | .003 | Walking | Healthy | 14 | 0.20 [0.04, 0.35] | .01 | 78% | .001 |
| | | | At-risk | 30 | 0.59 [0.42, 0.76] | <.001 | 87% | <.001 |
| | | | Sick | 33 | 0.44 [0.29, 0.60] | <.001 | 71% | <.001 |
| | .02 | MVPA | Healthy | 12 | 0.14 [0.06, 0.23] | .001 | 15% | .29 |
| | | | At-risk | 25 | 0.30 [0.18, 0.43] | <.001 | 72% | <.001 |
| | | | Sick | 25 | 0.33 [0.21, 0.45] | <.001 | 55% | <.001 |
| | .03 | TPA | Healthy | 6 | 0.29 [-0.10, 0.67] | .14 | 85% | <.001 |
| | | | At-risk | 16 | 0.21 [0.04, 0.38] | .02 | 78% | <.001 |
| | | | Sick | 11 | 0.59 [0.36, 0.81] | <.001 | 49% | .03 |
| **Intervention Design** | .35 | Walking | Scalable | 31 | 0.54 [0.34, 0.74] | <.001 | 89% | <.001 |
| | | | Non-scalable | 45 | 0.42 [0.31, 0.54] | <.001 | 77% | <.001 |
| | | | Combined | 1 | 0.37 [0.25, 0.48] | <.001 | n/a[b] | n/a[b] |
| | .12 | MVPA | Scalable | 23 | 0.20 [0.08, 0.32] | .001 | 67% | <.001 |
| | | | Non-scalable | 38 | 0.33 [0.24, 0.43] | <.001 | 57% | <.001 |
| | | | Combined | 1 | 0.37 [0.25, 0.48] | <.001 | n/a[b] | n/a[b] |
| | .60 | TPA | Scalable | 12 | 0.39 [0.06, 0.73] | .02 | 89% | <.001 |
| | | | Non-scalable | 21 | 0.30 [0.18, 0.42] | <.001 | 53% | .002 |
| | | | Combined[a] | - | - | - | - | - |
| **Control Group Type** | .15 | Walking | No/minimal intervention | 62 | 0.47 [0.36, 0.59] | <.001 | 85% | <.001 |
| | | | Alternative intervention | 10 | 0.48 [0.12, 0.83] | .009 | 80% | <.001 |
| | | | Combined | 5 | 0.23 [0.01, 0.45] | .04 | 62% | .03 |
| | .26 | MVPA | No/minimal intervention | 43 | 0.29 [0.21, 0.38] | <.001 | 67% | <.001 |
| | | | Alternative intervention | 9 | 0.39 [0.14, 0.65] | .002 | 62% | .007 |
| | | | Combined | 10 | 0.20 [0.08, 0.32] | <.001 | 34% | .13 |
| | .006 | TPA | No/minimal intervention | 24 | 0.34 [0.19, 0.50] | <.001 | 76% | <.001 |
| | | | Alternative intervention | 6 | 0.48 [0.06, 0.91] | .03 | 83% | <.001 |
| | | | Combined | 3 | 0.00 [-0.17, 0.17] | .97 | 0% | .59 |

[a]No studies with combined subgroups, thus no numbers reported. b In subgroups where n=1, heterogeneity cannot be calculated.

## Subgroup Analysis by Intervention Design

Subgroup analysis by intervention design revealed no significant subgroup differences across 3 outcome measures (walking, *P*=.35; MVPA, *P*=.12; TPA, *P*=.60; Table 1; Appendix 8) and did not identify mHealth intervention design as a significant effect moderator. Heterogeneity within subgroups was substantial and significant across all outcome measures (Table 1). Both scalable and nonscalable mHealth intervention designs significantly increased PA at similar levels (scalable walking SMD 0.54, 95% CI 0.34-0.74, *P*<.001, $I^2$=89%, *P*<.001; scalable MVPA SMD 0.20, 95% CI 0.08-0.32, *P*=.001, $I^2$=67%, *P*<.001; scalable TPA SMD 0.39, 95% CI 0.06-0.73, *P*=.02, $I^2$=89%, *P*<.001; nonscalable walking SMD 0.42, 95% CI 0.31-0.54, *P*<.001, $I^2$=77%, *P*<.001; nonscalable MVPA SMD 0.33, 95% CI 0.24-0.43, *P*<.001, $I^2$=57%, *P*<.001; nonscalable TPA SMD 0.30, 95% CI 0.18-0.42, *P*<.001, $I^2$=53%, *P*=.002; Appendix 8).

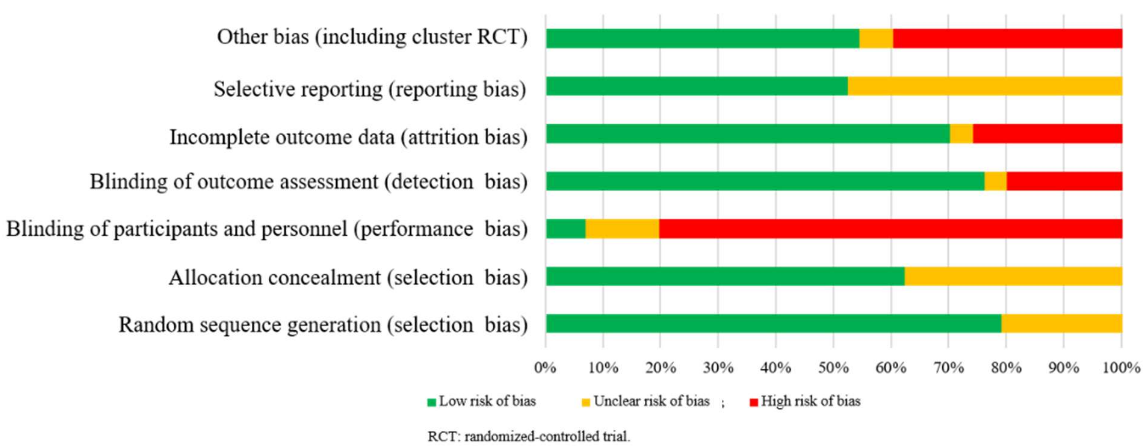
### Subgroup Analysis by Control Group Type

Subgroup analysis by control group type found no statistically significant subgroup effect for the outcome measures walking (*P*=.15) and MVPA (*P*=.26). Subgroup differences were only significant for TPA (*P*=.006), where mHealth interventions led to larger effects in studies compared with alternative control groups (SMD 0.48, 95% CI 0.06-0.91; *P*=.03; $I^2$=83%, *P*<.001; Appendix 9) than in studies with no or minimal control groups (SMD 0.34, 95% CI 0.19-0.50; *P*<.001; $I^2$=76%, *P*<.001; Appendix 9). Subgroup analysis did not significantly reduce heterogeneity, and the covariate distribution between the no or minimal intervention subgroup and the alternative intervention subgroup was extremely uneven).

### Risk of Bias in Included Studies

Figure 10 shows the overall risk of bias assessment across all included studies. Overall, 94 studies were classified as high risk because of selection bias (14/117, 11.9%), detection bias (37/117, 31.6%), attrition bias (42/117, 35.9%), reporting bias (13/117, 11.1%), and other biases (56/117, 47.9%). These mostly included baseline group indifferences or biases resulting from the respective study design (including potential cluster RCT biases) [35]. Appendix 10 [40-151,153-157] displays the individual risk of bias assessment by study. GRADE analysis of all 4 outcomes (Appendix 11) revealed no evidence of publication bias but evidence of inconsistency for the outcome measure EE. Thus, the overall quality of evidence rating ranged from low (walking, MPVA, and TPA) to very low (EE).

**Figure 10:** Summary of The Overall Risk of Bias Assessment for Included Studies



## Discussion

### Principal Findings

This systematic review is the most comprehensive study to date of mHealth PA interventions in adult populations. The aims of this study are to understand the long-term impacts of mHealth interventions on PA and to identify important effect moderators.

Overall, our analysis confirms the potential of mHealth interventions to increase PA at the end of intervention. We found small to moderate positive effects (SMD 0.28-0.46), which concur with previous research that reported small to large effect sizes [19,21-24,27,28]. Transforming our results into mean differences based on a representative low risk of bias study [109], we found mHealth interventions to result in 1566 incremental steps per day and an additional 36 minutes of MVPA per week. Previous research found that 1000 incremental steps per day can result in a 10% lower risk of having metabolic syndrome (MetS) and a 6% risk reduction of all-cause mortality, substantiating that mHealth interventions can result in significant health benefits [158,159].

This study is among the first to find that activity increases are sustained beyond the end of intervention. Increased PA levels remained significant in short-term follow-ups taken on average 4.14 months after the end of intervention for the outcome measures walking, MVPA, and TPA. Long-term follow-up measurements, taken on average after 13.96 months, confirmed these results. However, effect sizes decreased over time and ranged from 0.19 to 0.25 at the long-term follow-up time point, which is equivalent to an incremental 851 steps per day and 24 minutes per week of MVPA. Our results concur with the recent review by Chaudhry et al [20], who also found maintained but decreasing effects of step-count monitoring interventions on PA; however, as Chaudhry et al [20] defined time frames from the start of intervention and this study looks at follow-up measurements after the end of intervention, absolute effect scores cannot be compared.

Given the inverse relationship of PA with the prevalence of MetS [158], it can be assumed that mHealth interventions still yield health benefits in the long term. These observations are encouraging and provide initial evidence that mHealth interventions can support sustainable behavioral changes. However, our follow-up effects were not robust when only low risk of bias studies were analyzed because of the limited number of high-quality studies with longitudinal designs. Thus, as the current evidence base for studies with long-term follow-up measurements is very limited, further primary research is needed to confirm the sustained effects of mHealth on PA beyond the end of intervention.

Our analysis of effect moderators found that population type moderates the effect of mHealth on PA, whereas intervention design and control group type were not found to be effect moderators. Our evidence suggests that mHealth interventions might be most effective when targeting sick or at-risk populations, thereby supporting the indicative results by Smith et al [27]—effect sizes in sick and at-risk populations were about twice as high as in healthy populations. However, we still found mHealth interventions to be effective in all population types. These results challenge previous findings by Gal et al [23] and Romeo et al [30], who found no differences in effectiveness by population type, likely owing to the small number of studies reviewed. Previous studies found that baseline activity levels are negatively correlated with activity increases in mHealth interventions [144,160,161].

An underlying driver for the higher effectiveness of mHealth interventions in sick and at-risk populations could thus be lower baseline activity levels usually seen within these populations. However, there could also be further underlying factors, such as higher expectations that increases in PA lead to improved health outcomes (outcome expectancy). Further research is thus needed to understand the variety of underlying factors driving higher effectiveness in sick and at-risk populations. Our results provide helpful guidance to policy makers developing scaled-up mHealth intervention programs. Our results suggest that technology-enabled preventative, population-wide programs (eg, *The National Steps Challenge* [14]) might maximize their public health impact if they specifically target at-risk populations (eg, older or overweight groups). Focusing on at-risk groups should also increase the cost-effectiveness of large-scale mHealth programs.

mHealth technologies are cost-effective and scalable. However, this holds true only if technologies are effective without additional nonscalable intervention components (eg, face-to-face coaching). Previous research has found no effects in scalable mHealth intervention designs [24,30], stronger effects in nonscalable designs that combined technology with human-to-human interactions [27,32], and stronger effects when technology was used stand-alone [20]. We found preliminary evidence that mHealth interventions could be effective in scalable intervention designs. Our analysis found no significant subgroup differences between scalable and nonscalable intervention designs, suggesting that both designs can be equally effective in increasing PA. These results are promising and encourage the development of scalable mHealth intervention designs to efficiently increase PA in large population groups. Within our sample, most scalable mHealth interventions leveraged basic technologies (eg, texting, pedometers, or accelerometers), without taking advantage of more advanced mobile technologies (eg, automated individualized coaching, social comparison, and mobile apps), which could have further increased intervention effectiveness [162,163].

Our analysis is among the first to explore whether mHealth PA interventions produce results superior to alternative nonmobile interventions. We found that across the outcome measures walking, MPVA, and TPA, mHealth interventions led to increased levels of PA compared with alternative nonmobile interventions and no or minimal control groups, which accords with previous findings [21]. These results encourage the addition of mHealth technology to nonmobile PA interventions to increase their effectiveness.

**Strengths and Limitations**

The strengths of this study are the large number of mHealth interventions analyzed and its rigorous methodology. However, this study has several limitations. First, in line with other studies [164], we encountered large and significant heterogeneity in our results, despite performing several subgroup analyses. Our wide inclusion criteria led us to expect high heterogeneity because of the diverse multicomponent interventions, settings, and intervention durations. In addition, the uneven covariate distribution between subgroups limits the validity of our findings on effect moderators.

Second, most of the included studies were classified as having a high risk of bias, and the overall quality of evidence was graded low to very low. The quality of evidence could be improved if future research agreed on standardized reporting of PA outcomes (eg, MVPA in minutes per day) and objective outcome measurement [21]. When replicating our primary results with low risk of bias studies, we could not confirm the effectiveness of mHealth interventions to increase PA beyond the end of interventions, as the available high-quality evidence was limited.

Third, we did not attempt to identify unpublished reports or gray literature. Previous research has shown that excluding gray literature might exaggerate the results of a meta-analysis [165]. We tried to mitigate this limitation by conducting a funnel plot analysis to detect potential publication bias. Furthermore, we performed sensitivity analyses to assess the robustness of our results. We detected no systematic publication bias and sensitivity analyses that excluded outlier studies, confirming that our results were robust.

Fourth, some studies included in this study allowed intervention participants to keep mHealth devices after the end of intervention. This might have positively skewed the follow-up effects of our review. Finally, although this study provides initial evidence on the long-term effects of mHealth interventions, it only presents results for follow-up measurements taken on average 13.96 months after intervention, and our analysis included only 8 studies. We found that summary effects decrease over time, thus raising the question about the sustainability of positive effects. Further research is required to evaluate whether behavior change—toward a more active lifestyle—is truly sustainable in long term.

**Conclusions**

We conclude that mHealth interventions can moderately increase PA in adults at the end of intervention, both compared with alternative nonmobile control groups and no or minimal control groups. PA increases are maintained in follow-up measurements taken after intervention but decrease over time. Population type seems to moderate the effect of mHealth intervention on PA, with higher effectiveness in sick and at-risk populations compared with healthy population samples. mHealth interventions with scalable and nonscalable intervention designs seem to be equivalent in terms of effectiveness. Further high-quality studies investigating scalable mHealth interventions with long-term follow-up measurements are needed to confirm our results. This study concludes that mHealth technologies might not only support sustainable behavior change toward more active lifestyles but also contribute to preventing and controlling chronic disease risk.

**Acknowledgments**

This study was supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise Programme and by the CSS Insurance (Switzerland). We thank Iva Milhaylova for her support during screening and extraction.

**Authors' Contributions**

JNK wrote the initial study protocol with inputs from AM and KI. AM wrote the adjusted study protocol. JNK designed and implemented the search strategy. AM, KI, GWT, and AJH screened and coded the primary studies and extracted data. AM analyzed the data and drafted the initial manuscript, supervised by TK. FM and LTC provided methodological guidance and feedback on the manuscript. All authors reviewed and approved the final manuscript.

**Conflicts of Interest**

JNK, AJH, KI, GWT and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which was funded in part by CSS Insurance, Switzerland. TK is also a cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, Pathmate Technologies was not involved in this research. Since January 2021, JNK is associated with CSS Insurance, Switzerland.

**Abbreviations**

EE: energy expenditure
GRADE: grading of recommendations, assessment, development, and evaluation
MetS: metabolic syndrome
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
NCD: noncommunicable disease
PA: physical activity
RCT: randomized controlled trial
SMD: standardized mean difference
TPA: total physical activity
WHO: World Health Organization

## References


1. Mendis S, Armstrong T, Bettcher D, Branca F, Lauer J, Mace C, et al. Global Status Report on Noncommunicable Diseases 2014. Geneva: World Health Organization; 2014.
2. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):1077-1086 [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
3. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
4. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, Lancet Physical Activity Series 2 Executive Committee. Progress in physical activity over the Olympic quadrennium. Lancet 2016 Sep 24;388(10051):1325-1336. [doi:10.1016/S0140-6736(16)30581-5] [Medline: 27475270]
5. Bianchini F, Kaaks R, Vainio H. Weight control and physical activity in cancer prevention. Obes Rev 2002 Feb;3(1):5-8. [doi: 10.1046/j.1467-789x.2002.00046.x] [Medline: 12119660]
6. Choi KW, Chen C, Stein MB, Klimentidis YC, Wang M, Koenen KC, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample Mendelian randomization study. JAMA Psychiatry 2019 Apr 01;76(4):399-408 [doi: 10.1001/jamapsychiatry.2018.4175] [Medline: 30673066]
7. World Health Organization. Global Action Plan for the Prevention and Control of NCDs 2013-2020. Geneva: World Health Organization; 2013:1-103.
8. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. Inf Technol 2019;61(5-6):253-263 [doi: 10.1515/itit-2019-0019]
9. Reis RS, Salvo D, Ogilvie D, Lambert EV, Goenka S, Brownson RC, Lancet Physical Activity Series 2 Executive Committee. Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving.

Lancet 2016 Sep 24;388(10051):1337-1348 [doi: 10.1016/S0140-6736(16)30728-0] [Medline: 27475273]
10. Richard E, van Charante EP, Hoevenaar-Blom MP, Coley N, Barbera M, van der Groep A, et al. Healthy ageing through internet counselling in the elderly (HATICE): a multinational, randomised controlled trial. Lancet Digit Health 2019 Dec;1(8):424-434. [doi: 10.1016/s2589-7500(19)30153-0]
11. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
12. NHS Diabetes Prevention Programme – digital stream. NHS England. 2016. URL: https://tinyurl.com/u58h67ry [accessed 2020-06-03]
13. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [doi: 10.2196/jmir.4339] [Medline: 26180040]
14. Yao J, Tan CS, Chen C, Tan J, Lim N, Müller-Riemenschneider F. Bright spots, physical activity investments that work: National Steps Challenge, Singapore: a nationwide mHealth physical activity programme. Br J Sports Med 2020 Sep;54(17):1047-1048. [doi: 10.1136/bjsports-2019-101662] [Medline: 31857340]
15. Fitbit collaborates with Singapore's Health Promotion Board on population-based public health initiative in Singapore. Fitbit Press Release. 2019. URL: https://investor.fitbit.com/press/press-releases/press-release-details/2019/Fitbit-Collaborates-with-Singapores-Health-Promotion-Board-on-Population-Based-Public-Health-Initiative-in-Singapore/default.aspx [accessed 2020-07-01]
16. The Lancet Digital Health. Wearable technology and lifestyle management: the fight against obesity and diabetes. Lancet Digit Health 2019 Oct;1(6):e243 [doi: 10.1016/S2589-7500(19)30135-9] [Medline: 33323244]
17. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
18. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]

19. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):11819 [doi: 10.2196/11819] [Medline: 30977740]
20. Chaudhry UA, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020 Oct 09;17(1):129 [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]
21. De Vries HJ, Kooiman TJ, Van Ittersum MW, Van Brussel M, De Groot M. Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. Obesity (Silver Spring) 2016 Oct;24(10):2078-2091 [doi: 10.1002/oby.21619] [Medline: 27670401]
22. Franssen WM, Franssen GH, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2020 May 11;17(1):57 [doi: 10.1186/s12966-020-00955-2] [Medline: 32393357]
23. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
24. Hodkinson A, Kontopantelis E, Adeniji C, van Marwijk H, McMillan B, Bower P, et al. Accelerometer- and pedometer-based physical activity interventions among adults with cardiometabolic conditions: a systematic review and meta-analysis. JAMA Netw Open 2019 Oct 02;2(10):e1912895 [doi: 10.1001/jamanetworkopen.2019.12895] [Medline: 31596494]
25. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable technology and physical activity behavior change in adults with chronic cardiometabolic disease: a systematic review and meta-analysis. Am J Health Promot 2019 Jun;33(5):778-791. [doi: 10.1177/0890117118816278] [Medline: 30586996]
26. Qiu S, Cai X, Chen X, Yang B, Sun Z. Step counter use in type 2 diabetes: a meta-analysis of randomized controlled trials. BMC Med 2014 Feb 27;12:36 [doi: 10.1186/1741-7015-12-36] [Medline: 24571580]

27. Smith DM, Duque L, Huffman JC, Healy BC, Celano CM. Text message interventions for physical activity: a systematic review and meta-analysis. Am J Prev Med 2020 Jan;58(1):142-151 [doi: 10.1016/j.amepre.2019.08.014] [Medline: 31759805]

28. Vaes AW, Cheung A, Atakhorrami M, Groenen MT, Amft O, Franssen FM, et al. Effect of 'activity monitor-based' counseling on physical activity and health-related outcomes in patients with chronic diseases: a systematic review and meta-analysis. Ann Med 2013 Sep;45(5-6):397-412. [doi: 10.3109/07853890.2013.810891] [Medline: 23952917]

29. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]

30. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):12053 [doi: 10.2196/12053] [Medline: 30888321]

31. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):14343 [doi: 10.2196/14343] [Medline: 31778121]

32. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]

33. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Hoboken, New Jersey, United States: John Wiley & Sons Inc; 2008:1-649.

34. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [doi: 10.1136/bmj.d5928] [Medline: 22008217]

35. Eldridge S, Campbell M, Campbell M, Dahota A, Giraudeau B, Higgins J, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), additional

considerations for cluster-randomized trials. 2016. URL: https://sites.google.com/ site/riskofbiastool/welcome/rob-2-0-tool/archive-rob-2-0-cluster-randomized-trials-2016 [accessed 2020-07-01]

36. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
37. Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: a tutorial. Clin Epidemiol Glob Health 2019 Jun;7(2):192-198. [doi: 10.1016/j.cegh.2018.05.005]
38. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 06;327(7414):557-560 [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
39. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011 Apr;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
40. Kangasniemi AM, Lappalainen R, Kankaanpää A, Tolvanen A, Tammelin T. Towards a physically more active lifestyle based on one's own values: the results of a randomized controlled trial among physically inactive adults. BMC Public Health 2015 Mar 18;15:260 [doi: 10.1186/s12889-015-1604-x] [Medline: 25848812]
41. Aittasalo M, Rinne M, Pasanen M, Kukkonen-Harjula K, Vasankari T. Promoting walking among office employees - evaluation of a randomized controlled intervention with pedometers and e-mail messages. BMC Public Health 2012 Jun 06;12:403 [doi: 10.1186/1471-2458-12-403] [Medline: 22672576]
42. Barnes AT, Plotnikoff RC, Collins CE, Morgan PJ. Feasibility and preliminary efficacy of the MADE4Life Program: a pilot randomized controlled trial. J Phys Act Health 2015 Oct;12(10):1378-1393. [doi: 10.1123/jpah.2014-0331] [Medline: 25599119]
43. Coelho CM, Reboredo MM, Valle FM, Malaguti C, Campos LA, Nascimento LM, et al. Effects of an unsupervised pedometer-based physical activity program on daily steps of adults with moderate to severe asthma: a randomized controlled trial. J Sports Sci 2018 May;36(10):1186-1193. [doi: 10.1080/02640414.2017.1364402] [Medline: 28799458]
44. Cruz J, Brooks D, Marques A. Walk2Bactive: a randomised controlled trial of a physical activity-focused behavioural intervention beyond pulmonary

rehabilitation in chronic obstructive pulmonary disease. Chron Respir Dis 2016 Feb;13(1):57-66 [doi: 10.1177/1479972315619574] [Medline: 26703921]

45. Dlugonski D, Motl RW, Mohr DC, Sandroff BM. Internet-delivered behavioral intervention to increase physical activity in persons with multiple sclerosis: sustainability and secondary outcomes. Psychol Health Med 2012;17(6):636-651. [doi: 10.1080/13548506.2011.652640] [Medline: 22313192]
46. Eakin EG, Winkler EA, Dunstan DW, Healy GN, Owen N, Marshall AM, et al. Living well with diabetes: 24-month outcomes from a randomized trial of telephone-delivered weight loss and physical activity intervention to improve glycemic control. Diabetes Care 2014 Aug;37(8):2177-2185. [doi: 10.2337/dc13-2427] [Medline: 24658390]
47. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):51-62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
48. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
49. Fjeldsoe BS, Miller YD, Graves N, Barnett AG, Marshall AL. Randomized controlled trial of an improved version of mobilemums, an intervention for increasing physical activity in women with young children. Ann Behav Med 2015 Aug;49(4):487-499. [doi: 10.1007/s12160-014-9675-y] [Medline: 25582987]
50. Furber S, Butler L, Phongsavan P, Mark A, Bauman A. Randomised controlled trial of a pedometer-based telephone intervention to increase physical activity among cardiac patients not attending cardiac rehabilitation. Patient Educ Couns 2010 Aug;80(2):212-218. [doi: 10.1016/j.pec.2009.11.012] [Medline: 20022201]
51. James EL, Stacey FG, Chapman K, Boyes AW, Burrows T, Girgis A, et al. Impact of a nutrition and physical activity intervention (ENRICH: Exercise and Nutrition Routine Improving Cancer Health) on health behaviors of cancer survivors and carers: a pragmatic randomized controlled trial. BMC Cancer 2015 Oct 15;15:710 [doi: 10.1186/s12885-015-1775-y] [Medline: 26471791]
52. Kernot J, Lewis L, Olds T, Maher C. Effectiveness of a Facebook-delivered physical activity intervention for postpartum women: a randomized controlled trial. J Phys

Act Health 2019 Feb 01;16(2):125-133. [doi: 10.1123/jpah.2017-0573] [Medline: 30541378]

53. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De Bourdeaudhuij I, Thomas S, et al. A web-based, social networking physical activity intervention for insufficiently active adults delivered via Facebook app: randomized controlled trial. JMed Internet Res 2015 Jul 13;17(7):e174 [doi: 10.2196/jmir.4086] [Medline: 26169067]
54. Mansi S, Milosavljevic S, Tumilty S, Hendrick P, Higgs C, Baxter DG. Investigating the effect of a 3-month workplace-based pedometer-driven walking programme on health-related quality of life in meat processing workers: a feasibility study within a randomized controlled trial. BMC Public Health 2015 Apr 22;15:410 [doi: 10.1186/s12889-015-1736-z] [Medline: 25895747]
55. Maselli M, Gobbi E, Carraro A. Effectiveness of individual counseling and activity monitors to promote physical activity among university students. J Sports Med Phys Fitness 2019 Jan;59(1):132-140. [doi: 10.23736/S0022-4707.17.07981-6] [Medline: 29199784]
56. Müller AM, Khoo S, Morris T. Text messaging for exercise promotion in older adults from an upper-middle-income country: randomized controlled trial. J Med Internet Res 2016 Jan 07;18(1):e5 [doi: 10.2196/jmir.5235] [Medline: 26742999]
57. Murawski B, Plotnikoff RC, Rayward AT, Oldmeadow C, Vandelanotte C, Brown WJ, et al. Efficacy of an m-Health physical activity and sleep health intervention for adults: a randomized waitlist-controlled trial. Am J Prev Med 2019 Oct;57(4):503-514. [doi: 10.1016/j.amepre.2019.05.009] [Medline: 31542128]
58. Mutrie N, Doolin O, Fitzsimons CF, Grant PM, Granat M, Grealy M, et al. Increasing older adults' walking through primary care: results of a pilot randomized controlled trial. Fam Pract 2012 Dec;29(6):633-642 [doi: 10.1093/fampra/cms038] [Medline: 22843637]
59. Nolan CM, Maddocks M, Canavan JL, Jones SE, Delogu V, Kaliaraju D, et al. Pedometer step count targets during pulmonary rehabilitation in chronic obstructive pulmonary disease. A randomized controlled trial. Am J Respir Crit Care Med 2017 May 15;195(10):1344-1352. [doi: 10.1164/rccm.201607-1372oc]
60. Oliveira JS, Sherrington C, Paul SS, Ramsay E, Chamberlain K, Kirkham C, et al. A combined physical activity and fall prevention intervention improved mobility-related goal attainment but not physical activity in older adults: a randomized

trial. J Physiother 2019 Jan;65(1):16-22 [doi: 10.1016/j.jphys.2018.11.005] [Medline: 30581138]

61. Pinto BM, Papandonatos GD, Goldstein MG, Marcus BH, Farrell N. Home-based physical activity intervention for colorectal cancer survivors. Psychooncology 2013 Jan;22(1):54-64. [doi: 10.1002/pon.2047] [Medline: 21905158]
62. Pinto BM, Stein K, Dunsiger S. Peers promoting physical activity among breast cancer survivors: a randomized controlled trial. Health Psychol 2015 May;34(5):463-472 [doi: 10.1037/hea0000120] [Medline: 25110844]
63. Ribeiro MA, Martins MA, Carvalho CRF. Interventions to increase physical activity in middle-age women at the workplace: a randomized controlled trial. Med Sci Sports Exerc 2014;46(5):1008-1015. [doi: 10.1249/MSS.0000000000000190] [Medline: 24126967]
64. Simons D, De Bourdeaudhuij I, Clarys P, De Cocker K, Vandelanotte C, Deforche B. Effect and process evaluation of a smartphone app to promote an active lifestyle in lower educated working young adults: cluster randomized controlled trial. JMIR Mhealth Uhealth 2018 Aug 24;6(8):e10003 [doi: 10.2196/10003] [Medline: 30143477]
65. Stacey FG, James EL, Chapman K, Lubans DR. Social cognitive theory mediators of physical activity in a lifestyle program for cancer survivors and carers: findings from the ENRICH randomized controlled trial. Int J Behav Nutr Phys Act 2016 Apr 14;13:49 [doi: 10.1186/s12966-016-0372-z] [Medline: 27075417]
66. Suggs S, Blake H, Bardus M, Lloyd S. Effects of text messaging in addition to emails on physical activity among university and college employees in the UK. J Health Serv Res Policy 2013 Apr;18(1 Suppl):56-64. [doi: 0.1177/1355819613478001] [Medline: 27552780]
67. Talbot LA, Gaines JM, Huynh TN, Metter EJ. A home-based pedometer-driven walking program to increase physical activity in older adults with osteoarthritis of the knee: a preliminary study. J Am Geriatr Soc 2003 Mar;51(3):387-392. [doi: 10.1046/j.1532-5415.2003.51113.x] [Medline: 12588583]
68. Hoeve NT, Sunamura M, Stam HJ, Boersma E, Geleijnse ML, van Domburg RT, et al. Effects of two behavioral cardiac rehabilitation interventions on physical activity: a randomized controlled trial. Int J Cardiol 2018 Mar 15;255:221-228. [doi: 10.1016/j.ijcard.2017.12.015] [Medline: 29425564]
69. Tudor-Locke C, Bell RC, Myers AM, Harris SB, Ecclestone NA, Lauzon N, et al. Controlled outcome evaluation of the First Step Program: a daily physical activity

intervention for individuals with type II diabetes. Int J Obes Relat Metab Disord 2004 Jan;28(1):113-119. [doi: 10.1038/sj.ijo.0802485] [Medline: 14569279]

70. Vallance JK, Courneya KS, Plotnikoff RC, Dinu I, Mackey JR. Maintenance of physical activity in breast cancer survivors after a randomized trial. Med Sci Sports Exerc 2008 Jan;40(1):173-180. [doi: 10.1249/mss.0b013e3181586b41] [Medline: 18091007]
71. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. It's LiFe! Mobile and web-based monitoring and feedback tool embedded in primary care increases physical activity: a cluster randomized controlled trial. J Med Internet Res 2015 Jul 24;17(7):e184 [doi: 10.2196/jmir.4579] [Medline: 26209025]
72. Van Hoye K, Wijtzes AI, Lefevre J, De Baere S, Boen F. Year-round effects of a four-week randomized controlled trial using different types of feedback on employees' physical activity. BMC Public Health 2018 Apr 12;18(1):492 [doi: 10.1186/s12889-018-5402-0] [Medline: 29649997]
73. De Greef K, Deforche B, Tudor-Locke C, De Bourdeaudhuij I. A cognitive-behavioural pedometer-based group intervention on physical activity and sedentary behaviour in individuals with type 2 diabetes. Health Educ Res 2010 Oct;25(5):724-736 [doi: 10.1093/her/cyq017] [Medline: 20338978]
74. De Greef KP, Deforche BI, Ruige JB, Bouckaert JJ, Tudor-Locke CE, Kaufman J, et al. The effects of a pedometer-based behavioral modification program with telephone support on physical activity and sedentary behavior in type 2 diabetes patients. Patient Educ Couns 2011 Aug;84(2):275-279. [doi: 0.1016/j.pec.2010.07.010] [Medline: 20732776]
75. Harris T, Kerry SM, Limb ES, Furness C, Wahlich C, Victor CR, et al. Physical activity levels in adults and older adults 3-4 years after pedometer-based walking interventions: Long-term follow-up of participants from two randomised controlled trials in UK primary care. PLoS Med 2018 Mar;15(3):e1002526 [doi: 10.1371/journal.pmed.1002526] [Medline: 29522529]
76. Kolt GS, Schofield GM, Kerse N, Garrett N, Ashton T, Patel A. Healthy Steps trial: pedometer-based advice and physical activity for low-active older adults. Ann Fam Med 2012;10(3):206-212 [doi: 10.1370/afm.1345] [Medline: 22585884]
77. Wyke S, Bunn C, Andersen E, Silva MN, Van Nassau F, McSkimming P, et al. Correction: the effect of a programme to improve men's sedentary time and physical activity: the European Fans in Training (EuroFIT) randomised

controlled trial. PLoS Med 2019 Mar;16(3):e1002772 [doi: 10.1371/journal.pmed.1002772] [Medline: 30870433]

78. Yates T, Edwardson CL, Henson J, Gray LJ, Ashra NB, Troughton J, et al. Walking away from type 2 diabetes: a cluster randomized controlled trial. Diabet Med 2017 May;34(5):698-707. [doi: 10.1111/dme.13254] [Medline: 27589017]
79. Alsaleh E, Windle R, Blake H. Behavioural intervention to increase physical activity in adults with coronary heart disease in Jordan. BMC Public Health 2016 Jul 26;16:643 [doi: 10.1186/s12889-016-3313-5] [Medline: 27460675]
80. Ashton LM, Morgan PJ, Hutchesson MJ, Rollo ME, Collins CE. Feasibility and preliminary efficacy of the 'HEYMAN' healthy lifestyle program for young men: a pilot randomised controlled trial. Nutr J 2017 Jan 13;16(1):2 [doi: 10.1186/s12937-017-0227-8] [Medline: 28086890]
81. Baker G, Gray SR, Wright A, Fitzsimons C, Nimmo M, Lowry R, Scottish Physical Activity Research Collaboration (SPARColl). The effect of a pedometer-based community walking intervention "Walking for Wellbeing in the West" on physical activity levels and health outcomes: a 12-week randomized controlled trial. Int J Behav Nutr Phys Act 2008 Sep 05;5:44 [doi: 10.1186/1479-5868-5-44] [Medline: 18775062]
82. Barwais FA, Cuddihy TF, Tomson LM. Physical activity, sedentary behavior and total wellness changes among sedentary adults: a 4-week randomized controlled trial. Health Qual Life Outcomes 2013 Oct 29;11:183 [doi: 10.1186/1477-7525-11-183] [Medline: 24168638]
83. Butler L, Dwyer D. Pedometers may not provide a positive effect on walking activity. Health Promot J Aust 2004 Aug 01;15(2):134-136. [doi: 10.1071/he04134]
84. Cadmus-Bertram L, Tevaarwerk AJ, Sesto ME, Gangnon R, Van Remortel B, Date P. Building a physical activity intervention into clinical care for breast and colorectal cancer survivors in Wisconsin: a randomized controlled pilot trial. J Cancer Surviv 2019 Aug;13(4):593-602 [doi: 10.1007/s11764-019-00778-6] [Medline: 31264183]
85. Compernolle S, Vandelanotte C, Cardon G, De Bourdeaudhuij I, De Cocker K. Effectiveness of a web-based, computer-tailored, pedometer-based physical activity intervention for adults: a cluster randomized controlled trial. J Med Internet Res 2015 Feb 09;17(2):e38 [doi: 10.2196/jmir.3402] [Medline: 25665498]

86. Creel DB, Schuh LM, Reed CA, Gomez AR, Hurst LA, Stote J, et al. A randomized trial comparing two interventions to increase physical activity among patients undergoing bariatric surgery. Obesity (Silver Spring) 2016 Aug;24(8):1660-1668 [doi: 10.1002/oby.21548] [Medline: 27367821]
87. Dadaczynski K, Schiemann S, Backhaus O. Promoting physical activity in worksite settings: results of a German pilot study of the online intervention Healingo fit. BMC Public Health 2017 Sep 08;17(1):696 [doi: 10.1186/s12889-017-4697-6] [Medline: 28886734]
88. de Blok BM, de Greef MH, Hacken NH, Sprenger SR, Postema K, Wempe JB. The effects of a lifestyle physical activity counseling program with feedback of a pedometer during pulmonary rehabilitation in patients with COPD: a pilot study. Patient Educ Couns 2006 Apr;61(1):48-55. [doi: 10.1016/j.pec.2005.02.005] [Medline: 16455222]
89. De Greef K, Deforche B, Tudor-Locke C, De Bourdeaudhuij I. Increasing physical activity in Belgian type 2 diabetes patients: a three-arm randomized controlled trial. Int J Behav Med 2011 Sep;18(3):188-198. [doi: 10.1007/s12529-010-9124-7] [Medline: 21052886]
90. Demeyer H, Louvaris Z, Frei A, Rabinovich RA, de Jong C, Gimeno-Santos E, Mr Papp PROactive study group the PROactive consortium. Physical activity is increased by a 12-week semiautomated telecoaching programme in patients with COPD: a multicentre randomised controlled trial. Thorax 2017 May;72(5):415-423 [doi: 10.1136/thoraxjnl-2016-209026] [Medline: 28137918]
91. Dishman RK, DeJoy DM, Wilson MG, Vandenberg RJ. Move to improve: a randomized workplace trial to increase physical activity. Am J Prev Med 2009 Feb;36(2):133-141. [doi: 10.1016/j.amepre.2008.09.038] [Medline: 19135905]
92. Duru OK, Sarkisian CA, Leng M, Mangione CM. Sisters in motion: a randomized controlled trial of a faith-based physical activity intervention. J Am Geriatr Soc 2010 Oct;58(10):1863-1869 [doi: 10.1111/j.1532-5415.2010.03082.x] [Medline: 20929464]
93. Engel L, Lindner H. Impact of using a pedometer on time spent walking in older adults with type 2 diabetes. Diabetes Educ 2006;32(1):98-107. [doi: 10.1177/0145721705284373] [Medline: 16439498]
94. Fjeldsoe BS, Miller YD, Marshall AL. MobileMums: a randomized controlled trial of an SMS-based physical activity intervention. Ann Behav Med 2010 May;39(2):101-111. [doi: 10.1007/s12160-010-9170-z] [Medline: 20174902]

95. Gell NM, Wadsworth DD. The use of text messaging to promote physical activity in working women: a randomized controlled trial. J Phys Act Health 2015 Jun;12(6):756-763 [doi: 10.1123/jpah.2013-0144] [Medline: 25110303]
96. Gill DP, Blunt W, Silva NC, Stiller-Moldovan C, Zou GY, Petrella RJ. The HealtheSteps™ lifestyle prescription program to improve physical activity and modifiable risk factors for chronic disease: a pragmatic randomized controlled trial. BMC Public Health 2019 Jun 28;19(1):841 [doi: 10.1186/s12889-019-7141-2] [Medline: 31253112]
97. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):384-391 [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
98. Hardeman W, Mitchell J, Pears S, Van Emmenis M, Theil F, Gc VS, VBI Research Team. Evaluation of a very brief pedometer-based physical activity intervention delivered in NHS Health Checks in England: the VBI randomised controlled trial. PLoS Med 2020 Mar;17(3):e1003046 [doi: 10.1371/journal.pmed.1003046] [Medline: 32142507]
99. Hornikx M, Demeyer H, Camillo CA, Janssens W, Troosters T. The effects of a physical activity counseling program after an exacerbation in patients with Chronic Obstructive Pulmonary Disease: a randomized controlled pilot study. BMC Pulm Med 2015 Nov 04;15:136 [doi: 10.1186/s12890-015-0126-8] [Medline: 26530543]
100. Hospes G, Bossenbroek L, Hacken NH, Van Hengel P, De Greef MH. Enhancement of daily physical activity increases physical fitness of outclinic COPD patients: results of an exercise counseling program. Patient Educ Couns 2009 May;75(2):274-278. [doi: 10.1016/j.pec.2008.10.005] [Medline: 19036552]
101. Houle J, Doyon O, Vadeboncoeur N, Turbide G, Diaz A, Poirier P. Innovative program to increase physical activity following an acute coronary syndrome: randomized controlled trial. Patient Educ Couns 2011 Dec;85(3):237-244. [doi: 10.1016/j.pec.2011.03.018] [Medline: 21546203]
102. Hultquist CN, Albright C, Thompson DL. Comparison of walking recommendations in previously inactive women. Med Sci Sports Exerc 2005 Apr;37(4):676-683. [doi: 10.1249/01.mss.0000158993.39760.1b] [Medline: 15809569]

103. Izawa KP, Watanabe S, Hiraki K, Morio Y, Kasahara Y, Takeichi N, et al. Determination of the effectiveness of accelerometer use in the promotion of physical activity in cardiac patients: a randomized controlled trial. Arch Phys Med Rehabil 2012 Nov;93(11):1896-1902. [doi: 10.1016/j.apmr.2012.06.015] [Medline: 22750166]
104. Katzmarzyk PT, Champagne CM, Tudor-Locke C, Broyles ST, Harsha D, Kennedy BM, et al. A short-term physical activity randomized trial in the Lower Mississippi Delta. PLoS One 2011;6(10):e26667 [doi: 10.1371/journal.pone.0026667] [Medline: 22046325]
105. Kawagoshi A, Kiyokawa N, Sugawara K, Takahashi H, Sakata S, Satake M, et al. Effects of low-intensity exercise and home-based pulmonary rehabilitation with pedometer feedback on physical activity in elderly patients with chronic obstructive pulmonary disease. Respir Med 2015 Mar;109(3):364-371 [doi: 10.1016/j.rmed.2015.01.008] [Medline: 25682543]
106. King AC, Bickmore TW, Campero MI, Pruitt LA, Yin JL. Employing virtual advisors in preventive care for underserved communities: results from the COMPASS study. J Health Commun 2013;18(12):1449-1464 [doi: 10.1080/10810730.2013.798374] [Medline: 23941610]
107. Li LC, Feehan LM, Xie H, Lu N, Shaw C, Gromala D, et al. Efficacy of a physical activity counseling program with use of a wearable tracker in people with inflammatory arthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2020 Dec;72(12):1755-1765. [doi: 10.1002/acr.24199] [Medline: 32248626]
108. Long JE, Ring C, Bosch JA, Eves F, Drayson MT, Calver R, et al. A life-style physical activity intervention and the antibody response to pneumococcal vaccination in women. Psychosom Med 2013 Oct;75(8):774-782. [doi: 10.1097/PSY.0b013e3182a0b664] [Medline: 23922400]
109. Lynch BM, Nguyen NH, Moore MM, Reeves MM, Rosenberg DE, Boyle T, et al. A randomized controlled trial of a wearable technology-based intervention for increasing moderate to vigorous physical activity and reducing sedentary behavior in breast cancer survivors: the ACTIVATE trial. Cancer 2019 Aug 15;125(16):2846-2855 [doi: 10.1002/cncr.32143] [Medline: 31012970]
110. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and acceptability of a wearable technology physical activity intervention with telephone counseling for mid-aged and older adults: a randomized controlled

pilot trial. JMIR Mhealth Uhealth 2017 Mar 06;5(3):e28 [doi: 10.2196/mhealth.6967] [Medline: 28264796]

111. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: a randomized clinical trial of an automated mhealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 09;4(11):1-9 [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
112. Melville CA, Mitchell F, Stalker K, Matthews L, McConnachie A, Murray HM, et al. Effectiveness of a walking programme to support adults with intellectual disabilities to increase physical activity: walk well cluster-randomised controlled trial. Int J Behav Nutr Phys Act 2015 Sep 29;12:125 [doi: 10.1186/s12966-015-0290-5] [Medline: 26416606]
113. Mendoza L, Horta P, Espinoza J, Aguilera M, Balmaceda N, Castro A, et al. Pedometers to enhance physical activity in COPD: a randomised controlled trial. Eur Respir J 2015 Feb;45(2):347-354 [doi: 10.1183/09031936.00084514] [Medline: 25261324]
114. Merom D, Rissel C, Phongsavan P, Smith BJ, Van Kemenade C, Brown WJ, et al. Promoting walking with pedometers in the community: the step-by-step trial. Am J Prev Med 2007 Apr;32(4):290-297. [doi: 10.1016/j.amepre.2006.12.007] [Medline: 17303369]
115. Paul L, Wyke S, Brewster S, Sattar N, Gill JMR, Alexander G, et al. Increasing physical activity in stroke survivors using STARFISH, an interactive mobile phone application: a pilot study. Top Stroke Rehabil 2016 Jun;23(3):170-177. [doi: 10.1080/10749357.2015.1122266] [Medline: 27077973]
116. Poirier J, Bennett WL, Jerome GJ, Shah NG, Lazo M, Yeh H, et al. Effectiveness of an activity tracker- and internet-based adaptive walking program for adults: a randomized controlled trial. J Med Internet Res 2016 Feb 09;18(2):e34 [doi: 10.2196/jmir.5295] [Medline: 26860434]
117. Pope ZC, Zeng N, Zhang R, Lee HY, Gao Z. Effectiveness of combined smartwatch and social media intervention on breast cancer survivor health outcomes: a 10-week pilot randomized trial. J Clin Med 2018 Jun 07;7(6):1-17 [doi: 10.3390/jcm7060140] [Medline: 29880779]
118. Prestwich A, Perugini M, Hurling R. Can implementation intentions and text messages promote brisk walking? A randomized trial. Health Psychol 2010 Jan;29(1):40-49. [doi: 10.1037/a0016993] [Medline: 20063934]

119. Reijonsaari K, Vehtari A, Kahilakoski O, Van Mechelen W, Aro T, Taimela S. The effectiveness of physical activity monitoring and distance counseling in an occupational setting - results from a randomized controlled trial (CoAct). BMC Public Health 2012 May 11;12:344 [doi: 10.1186/1471-2458-12-344] [Medline: 22578104]

120. Roos R, Myezwa H, Van Aswegen H, Musenge E. Effects of an education and home-based pedometer walking program on ischemic heart disease risk factors in people infected with HIV: a randomized trial. J Acquir Immune Defic Syndr 2014 Nov 01;67(3):268-276. [doi: 10.1097/QAI.0000000000000299] [Medline: 25140908]

121. Rowley TW, Lenz EK, Swartz AM, Miller NE, Maeda H, Strath SJ. Efficacy of an individually tailored, internet-mediated physical activity intervention in older adults: a randomized controlled trial. J Appl Gerontol 2019 Jul;38(7):1011-1022. [doi: 10.1177/0733464817735396] [Medline: 29165018]

122. Spence JC, Burgess J, Rodgers W, Murray T. Effect of pretesting on intentions and behaviour: a pedometer and walking intervention. Psychol Health 2009 Sep;24(7):777-789. [doi: 10.1080/08870440801989938] [Medline: 20205026]

123. Tabak M, Vollenbroek-Hutten MM, Van der Valk PD, Van der Palen J, Hermens HJ. A telerehabilitation intervention for patients with Chronic Obstructive Pulmonary Disease: a randomized controlled pilot trial. Clin Rehabil 2014 Jun;28(6):582-591. [doi: 10.1177/0269215513512495] [Medline: 24293120]

124. Thorndike AN, Mills S, Sonnenberg L, Palakshappa D, Gao T, Pau CT, et al. Activity monitor intervention to promote physical activity of physicians-in-training: randomized controlled trial. PLoS One 2014;9(6):e100251 [doi: 10.1371/journal.pone.0100251] [Medline: 24950218]

125. Vallance JK, Friedenreich CM, Lavallee CM, Culos-Reed N, Mackey JR, Walley B, et al. Exploring the feasibility of a broad-reach physical activity behavior change intervention for women receiving chemotherapy for breast cancer: a randomized trial. Cancer Epidemiol Biomarkers Prev 2015 Dec 16;25(2):391-398. [doi: 10.1158/1055-9965.epi-15-0812]

126. Van Blarigan EL, Chan H, Van Loon K, Kenfield SA, Chan JM, Mitchell E, et al. Self-monitoring and reminder text messages to increase physical activity in colorectal cancer survivors (Smart Pace): a pilot randomized controlled trial. BMC Cancer 2019 Mar 11;19(1):218 [doi: 10.1186/s12885-019-5427-5] [Medline: 30866859]

127. Warren FC, Stych K, Thorogood M, Sharp DJ, Murphy M, Turner KM, et al. Evaluation of different recruitment and randomisation methods in a trial of general practitioner-led interventions to increase physical activity: a randomized controlled feasibility study with factorial design. Trials 2014 Apr 21;15:134 [doi: 10.1186/1745-6215-15-134] [Medline: 24746263]
128. Yamada M, Mori S, Nishiguchi S, Kajiwara Y, Yoshimura K, Sonoda T, et al. Pedometer-based behavioral change program can improve dependency in sedentary older adults: a randomized controlled trial. J Frailty Aging 2012;1(1):39-44. [doi: 10.14283/jfa.2012.7] [Medline: 27092936]
129. Croteau KA, Richeson NE, Vines SW, Jones DB. Effects of a pedometer-based physical activity program on older adults' mobility-related self-efficacy and physical performance. Act Adapt Aging 2004 Apr 21;28(2):19-33. [doi: 10.1300/j016v28n02_02]
130. Croteau KA, Richeson NE, Farmer BC, Jones DB. Effect of a pedometer-based intervention on daily step counts of community-dwelling older adults. Res Q Exerc Sport 2007 Dec;78(5):401-406. [doi: 10.1080/02701367.2007.10599439] [Medline: 18274211]
131. Aittasalo M, Miilunpalo S, Kukkonen-Harjula K, Pasanen M. A randomized intervention of physical activity promotion and patient self-monitoring in primary health care. Prev Med 2006 Jan;42(1):40-46. [doi: 10.1016/j.ypmed.2005.10.003] [Medline: 16297442]
132. Fischer X, Kreppke J, Zahner L, Gerber M, Faude O, Donath L. Telephone-based coaching and prompting for physical activity: short- and long-term findings of a randomized controlled trial (Movingcall). Int J Environ Res Public Health 2019 Jul 23;16(14):1-16 [doi: 10.3390/ijerph16142626] [Medline: 31340528]
133. Golsteijn RH, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Short-term efficacy of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2018 Oct 30;15(1):106 [doi: 10.1186/s12966-018-0734-9] [Medline: 30376857]
134. Kendzor DE, Allicock M, Businelle MS, Sandon LF, Gabriel KP, Frank SG. Evaluation of a shelter-based diet and physical activity intervention for homeless adults. J Phys Act Health 2017 Feb;14(2):88-97. [doi: 10.1123/jpah.2016-0343] [Medline: 27775471]

135. Keyserling TC, Samuel Hodge CD, Jilcott SB, Johnston LF, Garcia BA, Gizlice Z, et al. Randomized trial of a clinic-based, community-supported, lifestyle intervention to improve physical activity and diet: the North Carolina enhanced WISEWOMAN project. Prev Med 2008 Jun;46(6):499-510. [doi: 10.1016/j.ypmed.2008.02.011] [Medline: 18394692]

136. Kim Y, Lumpkin A, Lochbaum M, Stegemeier S, Kitten K. Promoting physical activity using a wearable activity tracker in college students: a cluster randomized controlled trial. J Sports Sci 2018 Aug;36(16):1889-1896. [doi: 10.1080/02640414.2018.1423886] [Medline: 29318916]

137. King AC, Ahn DK, Oliveira BM, Atienza AA, Castro CM, Gardner CD. Promoting physical activity through hand-held computer technology. Am J Prev Med 2008 Feb;34(2):138-142 [doi: 10.1016/j.amepre.2007.09.025] [Medline: 18201644]

138. Koizumi D, Rogers NL, Rogers ME, Islam MM, Kusunoki M, Takeshima N. Efficacy of an accelerometer-guided physical activity intervention in community-dwelling older women. J Phys Act Health 2009 Jul;6(4):467-474. [doi: 10.1123/jpah.6.4.467] [Medline: 19842461]

139. Li LC, Sayre EC, Xie H, Clayton C, Feehan LM. A community-based physical activity counselling program for people with knee osteoarthritis: feasibility and preliminary efficacy of the Track-OA study. JMIR Mhealth Uhealth 2017 Jun 26;5(6):e86 [doi: 10.2196/mhealth.7863] [Medline: 28652228]

140. Maxwell-Smith C, Hince D, Cohen PA, Bulsara MK, Boyle T, Platell C, et al. A randomized controlled trial of WATAAP to promote physical activity in colorectal and endometrial cancer survivors. Psychooncology 2019 Jul;28(7):1420-1429. [doi: 10.1002/pon.5090] [Medline: 30980691]

141. Pekmezi D, Ainsworth C, Joseph RP, Williams V, Desmond R, Meneses K, et al. Pilot trial of a home-based physical activity program for African American women. Med Sci Sports Exerc 2017 Dec;49(12):2528-2536 [doi: 10.1249/MSS.0000000000001370] [Medline: 28704343]

142. Prestwich A, Perugini M, Hurling R. Can the effects of implementation intentions on exercise be enhanced using text messages? Psychol Health 2009 Jul;24(6):677-687. [doi: 10.1080/08870440802040715] [Medline: 20205020]

143. Samuels TY, Raedeke TD, Mahar MT, Karvinen KH, DuBose KD. A randomized controlled trial of continuous activity, short bouts, and a 10,000 step guideline in inactive adults. Prev Med 2011 Feb;52(2):120-125. [doi: 10.1016/j.ypmed.2010.12.001] [Medline: 21147156]

144. Sharp P, Caperchione C. The effects of a pedometer-based intervention on first-year university students: a randomized control trial. J Am Coll Health 2016;64(8):630-638. [doi: 10.1080/07448481.2016.1217538] [Medline: 27471879]
145. Unick JL, O'Leary KC, Bond DS, Wing RR. Physical activity enhancement to a behavioral weight loss program for severely obese individuals: a preliminary investigation. ISRN Obes 2012 Sep 05;2012:1-4 [doi: 10.5402/2012/465158] [Medline: 24379985]
146. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The effectiveness of a web-based computer-tailored physical activity intervention using Fitbit activity trackers: randomized trial. J Med Internet Res 2018 Dec 18;20(12):e11321 [doi: 10.2196/11321] [Medline: 30563808]
147. Wijsman CA, Westendorp RG, Verhagen EA, Catt M, Slagboom PE, De Craen AJ, et al. Effects of a web-based intervention on physical activity and metabolism in older adults: randomized controlled trial. J Med Internet Res 2013 Nov 06;15(11):e233 [doi: 10.2196/jmir.2843] [Medline: 24195965]
148. Carr LJ, Dunsiger SI, Lewis B, Ciccolo JT, Hartman S, Bock B, et al. Randomized controlled trial testing an internet physical activity intervention for sedentary adults. Health Psychol 2013 Mar;32(3):328-336 [doi: 10.1037/a0028962] [Medline: 22823069]
149. Lane A, Murphy N, Bauman A. An effort to 'leverage' the effect of participation in a mass event on physical activity. Health Promot Int 2015 Sep 13;30(3):542-551. [doi: 10.1093/heapro/dat077] [Medline: 24226297]
150. Mailey EL, Wójcicki TR, Motl RW, Hu L, Strauser DR, Collins KD, et al. Internet-delivered physical activity intervention for college students with mental health disorders: a randomized pilot trial. Psychol Health Med 2010 Dec;15(6):646-659. [doi: 10.1080/13548506.2010.498894] [Medline: 21154018]
151. Motl RW, Dlugonski D, Wójcicki TR, McAuley E, Mohr DC. Internet intervention for increasing physical activity in persons with multiple sclerosis. Mult Scler 2011 Jan;17(1):116-128. [doi: 10.1177/1352458510383148] [Medline: 20921239]
152. Sugden JA, Sniehotta FF, Donnan PT, Boyle P, Johnston DW, McMurdo ME. The feasibility of using pedometers and brief advice to increase activity in sedentary older women-a pilot study. BMC Health Serv Res 2008 Aug 08;8:169 [doi: 10.1186/1472-6963-8-169] [Medline: 18691392]

153. Thorsteinsen K, Vittersø J, Svendsen GB. Increasing physical activity efficiently: an experimental pilot study of a website and mobile phone intervention. Int J Telemed Appl 2014;2014 [FREE Full text] [doi: 10.1155/2014/746232] [Medline: 24963290]
154. Schwerdtfeger AR, Schmitz C, Warken M. Using text messages to bridge the intention-behavior gap? A pilot study on the use of text message reminders to increase objectively assessed physical activity in daily life. Front Psychol 2012;3:270 [doi: 10.3389/fpsyg.2012.00270] [Medline: 22876237]
155. Bennett JA, Young HM, Nail LM, Winters-Stone K, Hanson G. A telephone-only motivational intervention to increase physical activity in rural adults: a randomized controlled trial. Nurs Res 2008;57(1):24-32. [doi: 10.1097/01.NNR.0000280661.34502.c1] [Medline: 18091289]
156. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Twelve-month outcomes of an internet-based diabetes self-management support program. Patient Educ Couns 2012 Apr;87(1):81-92 [doi: 10.1016/j.pec.2011.07.024] [Medline: 21924576]
157. McNab F, Hillebrand A, Swithenby SJ, Rippon G. Combining temporal and spectral information with spatial mapping to identify differences between phonological and semantic networks: a magnetoencephalographic approach. Front Psychol 2012;3:273 [doi: 10.3389/fpsyg.2012.00273] [Medline: 22908001]
158. Sisson SB, Camhi SM, Church TS, Tudor-Locke C, Johnson WD, Katzmarzyk PT. Accelerometer-determined steps/day and metabolic syndrome. Am J Prev Med 2010 Jun;38(6):575-582. [doi: 10.1016/j.amepre.2010.02.015] [Medline: 20494233]
159. Oftedal S, Holliday EG, Attia J, Brown WJ, Collins CE, Ewald B, et al. Daily steps and diet, but not sleep, are related to mortality in older Australians. J Sci Med Sport 2020 Mar;23(3):276-282. [doi: 10.1016/j.jsams.2019.09.018] [Medline: 31615727]
160. Chan CB, Ryan DA, Tudor-Locke C. Health benefits of a pedometer-based physical activity intervention in sedentary workers. Prev Med 2004 Dec;39(6):1215-1222. [doi: 10.1016/j.ypmed.2004.04.053] [Medline: 15539058]

161. Althoff T, Sosic R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Jul 20;547(7663):336-339 [doi: 10.1038/nature23018] [Medline: 28693034]
162. Shcherbina A, Hershman SG, Lazzeroni L, King AC, O'Sullivan JW, Hekler E, et al. The effect of digital physical activity interventions on daily step count: a randomised controlled crossover substudy of the MyHeart Counts Cardiovascular Health Study. Lancet Digit Health 2019 Nov;1(7):344-352. [doi: 10.1016/s2589-7500(19)30129-3] [Medline: 33323209]
163. Kramer J, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, et al. Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
164. Cook R, Davidson P, Martin R, NIHR Dissemination Centre. Pedometers can help people get more active as part of an exercise programme. Br Med J 2020 Apr 06;369. [doi: 10.1136/bmj.m877] [Medline: 32253188]
165. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet 2000 Oct 07;356(9237):1228-1231. [doi: 10.1016/S0140-6736(00)02786-0] [Medline: 11072941]

**Appendix 1:** Overview of Existing Meta-analyses on The Effect of Mobile Health Intervention on Physical Activity.

| Review | NOS | N | Inter-vention duration, wks | CGT | Population | Results | Analysis of long-term effects | Subgroup analysis by population type | Subgroup analysis by CGT | Subgroup analysis by mHealth intervention design | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bravata et al. (2007) [18] | 26 | 2767 | 18 (24) | Active | Outpatient, mostly at-risk and sick | Pedometers significantly increase PA by 2491 steps. Significant decrease in BMI and systolic blood pressure. | No | No | No | Yes 10 000 step goal significant predictor of step increase, activity diary predictor, PA counselling no predictor but due to heterogeneity. | Numbers refer to RCTs only. Short duration and small sample size. High heterogeneity in effects, interventions and samples. |
| Brickwood et al. (2019) [19] | 26 | 3636 | 20.0 (21.0) | Mixed | Mixed mostly at-risk or sick | Significant increase in daily step count (SMD 0.24, 627 steps), energy expenditure (SMD 0.28, 300kcal), and MVPA (SMD 0.27, 75min per day). No significant effect on sedentary behavior. | No | No | No | Yes Comparison of wearable only (scalable) and multi-faceted (nonscalable) interventions. For walking, both significant positive effects. For MVPA, only the multi-faceted interventions produce significant effects. | Large heterogeneity of samples, interventions. |
| Chaudhry et al. (2020) [20] | 70 | 12491 | NR | Mixed | Mixed, healthy and at-risk; but all community dwelling | Significant short- and long-term PA increases (≤4 months +1126 steps/day; 1 year +464 steps/day). Newer devices (accelerometers, smartphone apps) are less advantageous than simpler pedometers. Adding coaching or financial incentives to step tracking interventions did not lead to further step increases. | Yes At ≤4 months follow-up increases of 1126 steps/day; +1050 at 6mths; +464 steps/day after 1 year; +121 steps/day at 2 years (but non-significant); +434 stpes/day at 3-4 years follow-ups. | No | No | Yes Comparison by type of monitor and intervention intensity. Studies with body-worn trackers or apps reported lower step increases than pedometer interventions (-927 steps/day). Interventions that combined tracking devices with counseling or financial incentives led to lower step increases than simple step tracking interventions (-931 steps/day). | Large degree of heterogeneity, limited sample of studies with long-term follow-ups, sub-group results/meta-regression is a between trial analysis and results are thus less rigorous than within trial comparisons. |
| De Vries et al. (2016) [21] | 14 | 1157 | 16.6 (11.6) | Mixed | At-risk, all overweight or obese | Significant increase vs. waitlist control group or usual care for walking (SMD 0.9) and MVPA (SMD 0.5). Significant increase vs. alternative intervention found for walking (MD 282 MET-minutes per week) and energy expenditure (SMD 0.45), no significant effect on MVPA. | No | n/a | Yes, differentiated between waitlist/usual care and alternative intervention. | No | Small sample size, high level of heterogenicity between studies, potential of publication bias. |

| Review | NOS | N | Inter-vention duration, wks | CGT | Population | Results | Analysis of long-term effects | Subgroup analysis by population type | Subgroup analysis by CGT | Subgroup analysis by mHealth intervention design | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Direito et al. (2017) [29] | 21 | 1701 | 15.5 (13.7) | Mixed | Mixed, mostly at-risk | No significant differences found for PA, MVPA, walking. Moderate, significant effect on reduction of sedentary behavior (SMD -0.26) | No | No | No | No | Small sample size, limited duration, high levels of heterogeneity, all studies performed in high income countries. |
| Franssen et al. (2020) [22] | 2858 | 35 | 22 (17) | Mixed | Sick and at-risk, adults with chronic diseases, obese or elderly adults | Significant increase in PA by 2123 steps/day. Larger effects found in younger populations. Duration and type of intervention (tracker only or combined) had no impact on the effect size. | No | No<br>Analysis by chronic diseases type and risk-group. CVD patients had lower effect compared to other populations. | No | Yes<br>Intervention type (tracker only or tracker combined) had no impact on effect size. | High heterogeneity, inclusion of studies with high risk of bias and small sample sizes, publication bias detected for some populations. |
| Gal et al. (2018) [23] | 18 | 2734 | 13.7 (11.6) | Mixed | Mixed, healthy, at-risk and sick | Small to moderate increase in objectively measured PA (SMD 0.43) and moderate increase in objectively measured walking (SMD 0.51). No significant effect found on subjectively measured MVPA. | No | Yes<br>No differences found between healthy and sick study populations. | No | No | Small sample size, potential of publication bias, short duration of studies, large heterogeneity between studies. |
| Hodkinson et al. (2019) [24] | 5208 | 36 | 30 (-) | Mixed | Sick & at-risk, adults w/ cardio-metabolic conditions | Significant small to medium effect (walking SMD 0.52, 1703 steps/day; MVPA SMD 0.22) Improvements in PA levels of patients with cardiometabolic conditions. | No<br>Narrative synthesis of long-term follow-up studies. | n/a<br>Adults with cardiometabolic conditions only | No<br>Sensitivity analysis for usual care group. | Yes<br>Face-to-face sessions with facilitators increased PA. No results found for simple self-monitored (ie, scalable) accelerometer interventions. | Heterogeneity amongst studies remained substantial, limiting the validity of the meta-regression results, potential miss of studies not explicitly stating cardiometabolic conditions. |
| Kirk et al. (2019) [25] | 35 | 4528 | 20.5 (15.9) | Mixed | Sick, chronic cardio-metabolic diseases | Large effect on walking and small effect on MVPA. Statistically significant increases in walking measured in steps/day (MD 2 592) and MVPA measured in min/week (MD 36.31). | No | n/a | No | Yes<br>Health coaching had no further impact on MVPA or steps/day. | Large heterogeneity between studies. |

| Review | NOS | N | Inter-vention duration, wks | CGT | Population | Results | Analysis of long-term effects | Subgroup analysis by population type | Subgroup analysis by CGT | Subgroup analysis by mHealth intervention design | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Qiu et al. (2014) [26][a] | 7 | 861 | 23.9 (15.3) | Mixed | Sick, Type 2 diabetes patients | Significant increase in physical activity (walking) by 1822 steps/day. When step counter was combined with goal setting, effect increased to 3 200 steps/day. | No | n/a | No | Yes<br>mHealth plus goal setting increases effect size. | Potential publication bias, small study sample, large heterogenicity between studies, interventions had multiple components making it difficult to evaluate the contribution of the step counter. |
| Romeo et al. (2019) [30] | 9 | 1740 | 13.2 (14.9) | Mini-mal or no-interven tion | Mixed | No significant impact of smartphone apps on PA (MVPA and walking) found. When only interventions of less than 3 months were included, significant effect on walking in steps/day was found (2 075 steps). When apps targeting only PA were used, significant effects on walking in steps/day were found (717 steps). | No | Yes<br>Subgroup analysis performed for healthy vs. sick populations but no significant differences were found. For both groups, results were insignificant. | n/a<br>Only waitlist and minimal intervention. | n/a<br>Only scalable interventions included (smartphone app only). | Small sample, high level of heterogeneity. |
| Smith et al. (2020) [27] | 8742 | 59 | - | Mixed | Mixed | Text message interventions led to significantly greater objectively measured steps/day (SMD 0.38), effects for MVPA were not statistically significant (SMD 0.31). Interventions with more components and interventions in medical populations led to larger effect sizes, however differences were not statistically significant. | No | Yes<br>Text interventions were more effective in sick populations, yet results were not statistically significant. | No | Yes<br>Interventions with more components or tailored text messages were more effective, yet results were not statistically significant. | Small sample, significant heterogeneity in included studies, quality of evidence moderate. |
| Vaes et al. (2013) [28][a] | 8 | 633 | 21.1 (15.3) | Mixed | Sick, Patients with diabetes | Large significant effect (SMD 0.81) on walking. Analysis showed that activity monitor-based interventions in combination with counseling resulted in a significantly greater number of steps/day (2 042 steps). | No | n/a<br>Diabetes patients only. | No | No | Small sample, large heterogeneity of study samples in duration and intervention design, among others. |

| Review | NOS | N | Inter-vention duration, wks | CGT | Population | Results | Analysis of long-term effects | Subgroup analysis by population type | Subgroup analysis by CGT | Subgroup analysis by mHealth intervention design | Limitations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Yerra-kalva et al. (2019) [31] | 486 | 6 | 14.4 (5.4) | Mixed | At-risk, elderly adults | Mobile app interventions may be effective in increasing PA in trials 3 months or shorter and in trials 6 months or longer, but results cannot be concluded with certainty. No statistically significant increases in PA short- and long-term were found, but results indicate short-term and long-term increase (506 and 753 steps/day respectively). | Yes Quantitative analysis of follow-up measure-ments, however very small sample. | No | No | No | Small sample, limited generalizability, review includes non-randomized trials. |

*Abbreviations:* NOS: number of studies included in the review; N: number of participants included in the review, CGT: control group type; MD: mean difference; SMD: standardized mean difference; PA: physical activity; MVPA: moderate to vigorous PA; TPA: total physical activity; EE: energy expenditure. *Comments:* at-risk populations include elderly, sedentary overweight and obese populations. Duration: mean duration and standard deviation of duration of studies included in the respective review. [a]We only display results relevant to PA, as these reviews also cover other health related outcomes.

## **Appendix 2:** Overview of The Search Strategy and Keywords

| Outcomes | Interventions | Exclusion |
|---|---|---|
| physical activity, walking, steps, moderate to vigorous physical activity, moderate-to-vigorous physical activity, exercise, leisure time activity, leisure-time activity, fitness, running, sitting, sedentary, sedentary behavior, sedentary behaviour, inactive, inactivity, active lifestyle, sedentary lifestyle, training, sport*, cycling, bicycle, aerobic*, aerobic exercise, aerobic physical activity | smartphone application, smartphone app, mobile app, mobile application, app, mobile phone, smartphone*, mobile device, PDA, tablet, cell phone, text message*, sms, short message service, mobile health, mHealth, m-health, internet, telehealth, telemedicine, eHealth, e-health iPod, Fitbit, Garmin, Jawbone, Nike, Withings, ambulatory monitoring, ambulatory assessment, wireless technology, accelerometer, pedometer, wearable*, wearable activity tracker, intervention, program*, support, education, therapy | child*, adolescent*, protocol |

## **Appendix 3:** Search Algorithms

**PubMed search strategy**

((physical activity [Title/Abstract]) OR (steps[Title/Abstract]) OR (exercise[Title/Abstract]) OR (fitness[Title/Abstract]) OR (running[Title/Abstract]) OR (sitting[Title/Abstract]) OR (sedentary*[Title/Abstract]) OR (inactive[Title/Abstract]) OR (inactivity[Title/Abstract]) OR (active lifestyle[Title/Abstract]) OR (training[Title/Abstract])) **AND** ((mobile application[Title/Abstract]) OR (app[Title/Abstract]) OR (mobile phone[Title/Abstract]) OR (smartphone[Title/Abstract]) OR (mobile device[Title/Abstract]) OR PDA[Title/Abstract]) OR (tablet[Title/Abstract]) OR (cell phone[Title/Abstract]) OR (text messag*[Title/Abstract]) OR sms[Title/Abstract]) OR (short message service[Title/Abstract]) OR (mobile health[Title/Abstract]) OR (mHealth[Title/Abstract]) OR (internet[Title/Abstract]) OR (telehealth[Title/Abstract]) OR (telemedicine[Title/Abstract]) OR (eHealth[Title/Abstract]) OR (iPod[Title/Abstract]) OR (Fitbit[Title/Abstract]) OR (accelerometer[Title/Abstract]) OR (pedometer[Title/Abstract]) OR (wearable*[Title/Abstract])) **AND** ((intervention[Title/Abstract]) OR (program*[Title/Abstract]) OR (support[Title/Abstract]) OR (education[Title/Abstract]) OR (therapy[Title/Abstract])) **NOT** ((child*[Title]) OR (adolescent*[Title]) OR (protocol[Title]))

Limitations applied: Clinical Trial, Randomized Controlled Trial, Humans

**Cochrane search strategy**

(physical activity or steps or exercise or fitness or running or sitting or sedentary* or inactive or inactivity or active lifestyle or training):ti AND (physical activity or steps or exercise or fitness or running or sitting or sedentary* or inactive or inactivity or active lifestyle or training):ab AND (mobile application or app or mobile phone or smartphone or mobile device or PDA or tablet or cell phone or text messag* or sms or short message service or mobile health or mHealth or internet or telehealth or telemedicine or eHealth or iPod or Fitbit or accelerometer or pedometer or wearable*):ti AND (mobile application or app or mobile phone or smartphone or mobile device or PDA or tablet or cell phone or text messag* or sms or short message service or mobile health or mHealth or internet or telehealth or telemedicine or eHealth or iPod or Fitbit or accelerometer or pedometer or wearable*):ab AND (Intervention or program* or support or education or therapy):ti AND (Intervention or program* or support or education or therapy):ab NOT (child* or adolescent* or protocol):ti

Limitations applied: EMBASE database

**PsychInfo search strategy**

TI ( physical activity or steps or exercise or fitness or running or sitting or sedentary* or inactive or inactivity or active lifestyle or training ) AND AB ( physical activity or steps or exercise or fitness or running or sitting or sedentary* or inactive or inactivity or active lifestyle or training ) **AND** TI ( mobile application or app or mobile phone or smartphone or mobile device or PDA or tablet or cell phone or text messag* or sms or short message service or mobile health or mHealth or internet or telehealth or telemedicine or eHealth or iPod or Fitbit or accelerometer or pedometer or wearable* ) AND AB ( mobile application or app or mobile phone or smartphone or mobile device or PDA or tablet or cell phone or text messag* or sms or short message service or mobile health or mHealth or internet or telehealth or telemedicine or eHealth or iPod or Fitbit or accelerometer or pedometer or wearable* ) **AND** TI ( Intervention or program* or support or education or therapy ) AND AB ( Intervention or program* or support or education or therapy ) **NOT** TI ( child* or adolescent* or protocol )

Limitations applied: no limits

**Scopus search strategy**

( ( TITLE ( "physical activity" OR steps OR exercise OR fitness OR running OR sitting OR sedentary* OR inactivity OR inactive OR "active lifestyle" OR training ) ) ) AND ( ( TITLE ( "mobile application" OR app OR "mobile phone" OR smartphone OR "mobile device" OR pda OR tablet OR "cell phone" OR "text messag*" OR sms OR "short message service" OR "mobile health" OR mhealth OR internet OR telehealth OR telemedicine OR ehealth OR ipod OR fitbit OR accelerometer OR pedometer OR wearable* ) ) ) AND ( ( TITLE ( intervention OR program* OR support OR education OR therapy ) ) ) AND NOT ( ( TITLE ( child* OR adolescent* OR protocol ) ) ) AND ( LIMIT-TO ( EXACTKEYWORD,"Human" ) OR LIMIT-TO ( EXACTKEYWORD,"Humans" ) OR LIMIT-TO ( EXACTKEYWORD,"Controlled Study" ) )

**Limitations applied:** Human, Humans, Controlled Trial

## **Appendix 4:** Overview of The Study Characteristics

| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measurement | Population type | Ease of Scalability |
|---|---|---|---|---|---|---|
| Aittasalo 2006 [131] | RCT; Finland; n=265; Age: 47 (SD 11); 76% female; Duration: 26wks | MVPA & TPA (in min/week; questionnaire) | I: Pedometer & PA-log.<br><br>C: Combined; usual care group and counseling group. | Short-term after 6mths | At-risk | High |
| Aittasalo 2012 [41] | RCT; Finland; n=241; Age: 44.6 (SD 9.2); 68.4% female; Duration: 24wks | Walking (in min/week; IPAQ) | I: Pedometer use, step-log and monthly e-mail messages from occupational health care provider.<br><br>C: No intervention; waitlist. | - | At-risk | High |
| Alsaleh 2016 [79] | RCT; Jordan; n=156; Age: 57.8 (SD 9.5); 46.2% female; Duration: 26wks | Walking & MVPA (in min/week; IPAQ) | I: Usual care plus 6-month behavioral intervention by a cardiac nurse, PA diary, 6 telephone consultations, mobile text alerts.<br><br>C: Minimal intervention; information from physician about general benefits of PA. | - | Sick | Low |
| Ashton 2017 [80] | RCT; Australia; n=50; Age: 22.1 (SD 2); 0% female; Duration: 12wks | Walking (in steps/day vs. baseline; Yamax Digi-Walker SW200); MVPA (in min/week; GLTEQ) | I: Website, wearable device, Facebook support group, face-to-face sessions (group and individual), personalized food and nutrient report, home-based resistance training equipment and portion control tool.<br><br>C: No intervention; waitlist. | - | Healthy | Low |
| Baker 2008 [81] | RCT; United Kingdom; n=79; Age: 49.2 (SD 8); 79.7% female; Duration: 12wks | Walking (in steps/day; Omron HJ-109E Step-O-Meter) | I: PA consultation, followed by 12-week pedometer-based walking program.<br><br>C: Minimal intervention; instruction to maintain normal walking levels. | - | At-risk | High |
| Barnes 2015 [155] | RCT; Australia; n=40; Age: 39.1 (SD 4.8); 100% female; Duration: 8wks | MVPA (in % MVPA/time tracked vs. baseline; ActiGraph GT3X and GT3X+) | I: Pedometer, 5-minute mothers and daughter education sessions, 60-minutes PA session.<br><br>C: No intervention; waitlist. | Short-term after 3mths | At-risk | Low |
| Barwais 2013 [82] | RCT; Australia; n=30; Age: 27.7 (SD 4.1); 33.3% female; Duration: 4wks | Walking & MVPA (in min/week; IPAQ-SF) | I: Personal activity monitor (Gruve Solution), goal setting, motivational emails.<br><br>C: No intervention; waitlist. | - | Healthy | High |
| Bennett 2008 [156] | RCT; New Zealand; n=72; Age: 57.9 (SD 11.2); 90% female; Duration: 24wks | EE (in kcal/week; CHAMPS) | I: Pedometer, telephone coaching with motivational interviewing.<br><br>C: Alternative intervention; equal number of telephone calls but no motivational interviewing. | - | At-risk | Low |
| Butler 2004 [83] | RCT; Australia; n=33; Age: 52 (SD 1.2); 84.8% female; Duration: 4wks | Walking (in steps/week; pedometer) | I: Pedometer, information package, goal setting.<br><br>C: Minimal intervention; blinded pedometer, information package, goal setting. | - | At-risk | High |
| Cadmus-Bertram 2019 [84] | RCT; USA; n=50; Age: 54.4 (SD 11.2); 96% female; Duration: 12wks | Walking (steps/day; ActiGraph); MVPA & TPA (min/week; ActiGraph) | I: Fitbit tracker, educational handbook, in-person coaching, social support, email coaching, monitoring by clinic.<br><br>C: Minimal intervention; educational handbook, emails. | - | Sick | Low |
| Carr 2013 [148] | RCT; USA; n=66; Age: 37.6 (SD 11.9); 75.4% female; Duration: 24wks | TPA (min/week; PAR) | I: Pedometer, website (step into motion), immediate, individually tailored PA messages generated by a computerized expert system.<br><br>C: Alternative intervention; list of 6 reputable PA websites, email prompts on same schedule as IG. | - | At-risk | High |

| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measure-ment | Popu-lation type | Ease of Scala-bility |
|---|---|---|---|---|---|---|
| Coelho 2018 [43] | RCT; Brazil; n=37; Age: 45.9 (SD 16.7); 86.4% female; Duration: 12wks | Walking (in steps/day; Yamax Digi-Walker SW200) | I: Pedometer, individualized daily step target, weekly phone consult.<br><br>C: No intervention; only usual asthma follow-ups and telephone check-ins. | Short-term after 2.8mths | Sick | High |
| Compernolle 2015 [85] | Cluster-RCT; Belgium; n=274; Age: 42 (SD 11); 62.5% female; Duration: 12wks | Walking (in steps/day; Omron HJ-203-ED); MVPA & TPA (in min/week; IPAQ) | I: Pedometer, information booklet, opportunity to request online tailored advice.<br><br>C: No intervention. | - | Healthy | High |
| Creel 2016 [86] | RCT; USA; n=150; Age: 43.2 (SD 11.2); 84% female; Duration: 26wks | Walking (in steps/day; ActiGraph GT3X); MVPA (in min/week; ActiGraph GT3X) | I: Pedometer, information sheet, printed manual, PA goals, exercise counseling.<br><br>C: Minimal intervention; usual care, educational materials. | - | At-risk | Low |
| Croteau 2004 [129] | RCT; USA; n=15; Age: 80.7 (SD 7.3); 93.3% female; Duration: 4wks | Walking (in steps/week; Yamax Digi-Walker SW200) | I: Counseling session, pedometer, weekly follow-ups.<br><br>C: No intervention; instruction to continue with daily activities. | - | At-risk | Low |
| Croteau 2007 [130] | RCT; USA; n=179; Age: 72.8 (SD 8.8); 78.2% female; Duration: 12wks | Walking (in stpes/day; Yamax Digi-Walker SW 200) | I: Counseling, pedometer, self-monitoring.<br><br>C: No intervention; waitlist. | - | At-risk | Low |
| Cruz 2016 [44] | RCT; Portugal; n=32; Age: 66.4 (SD 8.4); 15.6% female; Duration: 12wks | Walking (in steps/day; ActiGraph GT3X+); MVPA & TPA (in min/day; ActiGraph GT3X+) | I: Pulmonary rehabilitation (exercise training, weekly 90-min psychosocial support and education sessions); PA intervention (pedometer, log diary, goal setting, individual feedback by therapists).<br><br>C: Alternative intervention same pulmonary rehabilitation as IG. | Short-term after 3mths | Sick | Low |
| Dadaczynski 2017 [87] | RCT; Germany; n=232; -; 35% female; Duration: 6wks | Walking & MVPA (in min /week; IPAQ-SF) | I: Online-based program (goal setting, quizzes, PA challenges, social comparison), Fitbit.<br><br>C: No intervention; waitlist. | - | Healthy | High |
| De Blok 2006 [88] | RCT; Netherlands; n=21; Age: 64 (SD 11.3); 57.1% female; Duration: 9wks | Walking (in steps/day; Yamax Digi-Walker SW200) | I: Pedometer, 4 individual PA counselling sessions, pulmonary rehabilitation program (exercise training, dietary intervention, psycho-educational modules).<br><br>C: Minimal intervention; usual care, same pulmonary rehabilitation program as IG. | - | Sick | Low |
| De Greef 2010 [73] | RCT; Belgium; n=41; Age: 61.3 (SD -); 31.7% female; Duration: 12wks | Walking (in steps/day; Yamax Digi-Walker SW200); MVPA & TPA (in min/day; Yamax Digi-Walker SW200) | I: Pedometer, 5 cognitive-behavioral group coaching sessions, pedometer diary.<br><br>C: Minimal intervention; informational materials, usual endocrinologist visits. | Long-term after 9.2mths | Sick | Low |
| De Greef 2011 [74] | RCT; Belgium; n=92; Age: 62 (SD 9); 31.5% female; Duration: 24wks | Walking (in steps/day; accelerometer); MVPA & TPA (in min/day; IPAQ) | I: Face-to-face session, pedometer, diary, 7 telephone follow-ups.<br><br>C: Minimal intervention; usual diabetes care. | Long-term after 6.5mths | Sick | Low |
| De Greef 2011 [89] | RCT; Belgium; n=67; Age: 67.4 (SD 9.3); 29.9% female; Duration: 12wks | Walking (in steps/day; Yamax Digi-Walker SW200); MVPA & TPA (in min/day, IPAQ) | I: Pedometer, 3 counselling session (delivered by general practitioner or in groups), diary, goal setting.<br><br>C: No intervention. | - | Sick | Low |

| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measure-ment | Popu-lation type | Ease of Scala-bility |
|---|---|---|---|---|---|---|
| Demeyer 2017 [90] | RCT; Belgium, Greece, UK, Switzerland, Netherlands; n=343; Age: 66.5 (SD 8); 36.2% female; Duration: 12wks | Walking & MVPA (in min/day vs. baseline; Dynaport Movemonitor & ActiGraph GT3x) | I: Usual care, tele-coaching intervention (pedometer, one-to-one interview with investigator, automated coaching app, individualized automated goal setting, booklet containing home exercises, weekly text message with activity proposals, telephone contacts when non-compliant or failure to progress.<br><br>C: Minimal intervention; educational materials explaining importance of PA in COPD patients, 10-min fact-to-face discussion with investigator. | - | Sick | Low |
| Dishman 2009 [91] | RCT; USA; n=1442; Age: 36.2 (SD 9.8); 69% female; Duration: 12wks | Walking & MVPA (in MET-hrs/week; IPAQ) | I: Pedometer, individual goal setting, handbook, management endorsement, employee–management steering committees, group and organizational goal setting, environmental prompts.<br><br>C: Minimal intervention; CDC health risk appraisal, monthly newsletter. | - | Healthy | High |
| Dlugonski 2012 [45] | RCT; USA; n=45; Age: 46.6 (SD 9.7); 86.7% female; Duration: 12wks | TPA (in MET-min/week; GLTEQ) | I: Website with educational videos, pedometer, self-monitoring & goal setting, logbook, CD with instructional videos, 7 individual video coaching sessions.<br><br>C: No intervention; waitlist. | Short-term after 3mths | Sick | Low |
| Duru 2010 [92] | RCT; USA; n=71; Age: 72.8 (SD 8.1); 100% female; Duration: 8wks | Walking (in steps/week vs. baseline; Yamax Digi-Walker SW200) | I: Goal setting, pedometer self-monitoring, exercise classes.<br><br>C: Alternative intervention; lectures on non-PA topics (e.g. memory loss), exercise classes. | - | At-risk | Low |
| Eakin 2014 [46] | RCT; Australia; n=302; Age: 58 (SD 8.6); 43.7% female; Duration: 78wks | MVPA (in min/week; ActiGraph) | I: Pedometer, telephone counselling (up to 27 calls), motivational interviewing, digital scale, self-monitoring and goal setting.<br><br>C: Minimal intervention; usual primary care visits, educational brochure. | Short-term after 6mths | Sick | Low |
| Edney 2020 [47] | RCT; Australia; n=444; Age: 41.3 (SD 11.6); 74.1% female; Duration: 39wks | MVPA (in min/day; GENEActiv) | I: Combined intervention group; Active Team app and a wrist-worn pedometer; or a weekly e-mail and basic app.<br><br>C: Minimal intervention; educational materials. | Short-term after 6mths | At-risk | High |
| Engel 2006 [93] | RCT; Australia; n=57; Age: 62.4 (SD 7.2); 48.1% female; Duration: 26wks | Walking (in min/week; weekly logbook) | I: Pedometer, coaching (goal setting, education and motivational tactics).<br><br>C: Alternative intervention; same coaching as IG. | - | Sick | Low |
| Finkelstein 2016 [48] | Cluster-RCT; Singapore; n=800; Age: 35.7 (SD 8.5); 53.7% female; Duration: 26wks | Walking (in steps/day vs. baseline; ActiGraph GT-3x+); MVPA (in min/week vs. baseline; ActiGraph GT-3x+) | I: Combined intervention group; all participants received Fitbit and educational booklets; 1/3rd received charity incentives; 1/3rd received cash incentives.<br><br>C: Minimal intervention; educational materials. | Short-term after 6mths | Healthy | High |
| Fischer 2019 [132] | RCT; Switzerland; n=288; Age: 42.2 (SD 11.4); 68.4% female; Duration: 26wks | MVPA (in min/day; ActiGraph wGT3X-BT) | I: Telephone coaching, text messages.<br><br>C: Combined; telephone coaching or educational materials. | - | At-risk | Low |
| Fjeldsoe 2010 [94] | RCT; Australia; n=88; Age: 29.5 (SD 6.2); 100% female; Duration: 12wks | Walking & MVPA (in min/week vs. baseline; AWAS) | I: Face-to-face PA goal setting consultation, a goal setting magnet, 3-5 personally tailored SMS/week, a nominated support person who received two SMS weekly.<br><br>C: Minimal intervention; one face-to-face PA goal setting consultation and educational materials. | - | At-risk | Low |

| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measure-ment | Popu-lation type | Ease of Scala-bility |
|---|---|---|---|---|---|---|
| Fjeldsoe 2015 [49] | RCT; Australia; n=263; Age: 31.9 (SD 9.5); 100% female; Duration: 13wks | MVPA (in min/week vs. baseline; ActiGraph GT1M) | I: Face-to-face coaching, individually tailored text messages, telephone coaching, educational materials, website, Facebook group.<br><br>C: Alternative intervention, educational materials, website, Facebook group. | Short-term after 6mths | Healthy | Low |
| Furber 2010 [50] | RCT; Australia; n=222; Age: 66 (SD 11.1); 29.8% female; Duration: 6wks | Walking & TPA (in min/week; AAS) | I: Pedometer, step calendar for self-monitoring, telephone support including goal setting, informational brochures.<br><br>C: Minimal intervention; informational brochures. | Short-term after 4.5mths | Sick | Low |
| Gell 2015 [95] | RCT; USA; n=87; Age: 47 (SD 11); 100% female; Duration: 24wks | Walking (in steps/day; Omron HJ-720ITC) | I: Informational website, map with 5 suggested walking routes, 3 text messages per week.<br><br>C: Alternative intervention, informational website, map with 5 suggested walking routes. | - | Healthy | High |
| Gill 2019 [96] | RCT; Canada; n=118; Age: 57.7 (SD 13.5); 78.8% female; Duration: 26wks | Walking (in steps/day vs. baseline; Yamax Digi-Walker SW200); TPA (in MET-min/week vs. baseline; IPAQ) | I: In-person coaching, pedometer, personalized feedback, goal setting, social online community, telephone coaching, smartphone app, virtual coach, website.<br><br>C: No intervention; provided publicly available resources related to healthy lifestyles. | - | At-risk | Low |
| Glasgow 2012 [156] | RCT; USA; n=463; Age: 58.4 (SD 9.2); 49.8% female; Duration: 52wks | EE (in cals/week; CHAMPS) | I: Combined intervention group; educational website, pedometer, goal setting, self-monitoring, motivational calls, follow-up calls, group coaching.<br><br>C: Alternative intervention; educational website, computer-based health risk appraisal, feedback and recommended preventive care behaviors. | - | Sick | Low |
| Glynn 2014 [97] | RCT; Ireland; n=139; Age: 44.1 (SD 11.5); 64% female; Duration: 8wks | Walking (in steps/day; smartphone) | I: PA app, goal setting, PA brochure.<br><br>C: No intervention; waitlist. | - | Healthy | High |
| Golsteijn 2018 [133] | RCT; Netherlands; n=510; Age: 66.4 (SD 7.6); 13% female; Duration: 12wks | MVPA (in min/week; ActiGraph GT3X-BT) | I: Computer-tailored PA advice, pedometer, access to informational website.<br><br>C: No intervention; waitlist. | - | Sick | High |
| Hardeman 2020 [98] | RCT; United Kingdom; n=1007; Age: 56.1 (SD 9.5); 61.7% female; Duration: 13wks | Walking (in steps/day; ActiGraph GT3X); MVPA (in min/day; ActiGraph GT3X) | I: NHS health check, 5-min.coaching, educational materials, pedometer, step chart.<br><br>C: Minimal intervention; NHS health check. | - | Healthy | Low |
| Harris 2018 [75] | RCT; United Kingdom; n=1321; Age: 60.8 (SD 3.3); 61.8% female; Duration: 12wks | Walking (in steps/day; ActiGraph GT3X+); MVPA (in min/week; ActiGraph GT3X+) | I: Combined intervention; pedometer, activity diary, informational leaflet, goal setting; or 3-4 in-person consultations with nurse, pedometer, activity diary, informational leaflet, goal setting.<br><br>C: Minimal intervention; usual primary care. | Long-term after 45.2mths | At-risk | Com-bined |
| Hornikx 2015 [99] | RCT; Belgium; n=30; Age: 67 (SD 6.5); 43.3% female; Duration: 4wks | Walking (in steps/day vs. baseline; Dynaport MoveMonitor) | I: Fitbit ultra, telephone PA counselling (3x per week), goal setting.<br><br>C: Minimal intervention; PA consultation. | - | Sick | Low |

| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measure-ment | Popu-lation type | Ease of Scala-bility |
|---|---|---|---|---|---|---|
| Hospes 2009 [100] | RCT; Netherlands; n=39; Age: 62.2 (SD 8.6); 40% female; Duration: 12wks | Walking (in steps/day; Yamax Digi-Walker SW200) | I: Pedometer, 5 counselling sessions with motivational interviewing.<br><br>C: Minimal intervention; usual care. | - | Sick | Low |
| Houle 2011 [101] | RCT; Canada; n=65; Age: 58.5 (SD 8.5); 21.5% female; Duration: 52wks | Walking (in steps/day; Yamax Digi-Walker NL-2000) | I: Pedometer, activity diary, 5 coaching sessions.<br><br>C: Minimal intervention; standard PA and medication advice at discharge, access to center-based cardiac rehabilitation program or health professional. | - | Sick | Low |
| Hultquist 2005 [102] | RCT; USA; n=62; Age: 45 (SD 6); 100% female; Intervention duration: 4wks | Walking (in steps/day; New Lifestyles NL-2000) | I: Pedometer, daily step log, goal setting.<br><br>C: Minimal intervention; sealed pedometer, daily step log, goal setting. | - | At-risk | High |
| Izawa 2012 [103] | RCT; Japan; n=126; Age: 59.2 (SD 10.7); 20% female; Duration: 3wks | Walking (in steps/day; Kenz Lifecorder EXa1); EE (in cal/day; Kenz Lifecorder EXa1) | I: Accelerometer, group exercise sessions (5x weekly), dietary and medication advice, self-monitoring.<br><br>C: Alternative intervention; group exercise sessions (5x weekly), dietary and medication advice. | - | Sick | Low |
| James 2015 [51] | RCT; Australia; n=176; Age: 57 (SD 12); 77.4% female; Duration: 8wks | Walking (in steps/day vs. baseline; Yamax Digi-Walker SW200); MVPA (in min/week vs. baseline; AAS) | I: Pedometer and activity log, 6 educational sessions on nutrition and activity, gymstick for resistance training, goal setting, information about community-based programs and support groups.<br><br>C: No intervention; waitlist. | Short-term after 2.8mths | Sick | Low |
| Kangasniemi 2015 [40] | RCT; Finland; n=138; Age: 43.5 (SD 5); 82.3% female; Duration: 9wks | Walking (in steps/day; ActiGraph GT1M, GT3X); MVPA (in min/day; questionnaire) | I: Pedometer, PA diary, written feedback, 6 group sessions including goal setting and mindfulness exercises.<br><br>C: Alternative intervention; written feedback on PA, PA diary during measurement periods. | Short-term after 6mths | At-risk | Low |
| Katzmarzyk 2011 [104] | RCT; USA; n=43; Age: 51.4 (SD 8.2); 83.7% female; Duration: 1wk | Walking (in steps/day; ActiGraph GT3X); MVPA (in min/day; ActiGraph GT3X) | I: Informational brochure, pedometer, 10-min introductory walk and coaching on walking strategies.<br><br>C: Minimal intervention; educational brochure. | - | At-risk | Low |
| Kawagoshi 2015 [105] | RCT; Japan, n=39; Age: 74.6 (SD 8.4); 11.1% female; Duration 52wks | Walking (in min/day vs. baseline; A-MES accelerometer) | I: Pedometer and pulmonary rehab (education program including lectures about equipment use, nutrition, stress management, relaxation techniques, home exercises and the benefits of PA).<br><br>C: Alternative intervention; pulmonary rehab. | - | Sick | Low |
| Kendzor 2017 [134] | RCT; USA; n=32; Age: 48.4 (SD 8.1); 25% female; Duration: 4wks | MVPA (in min/day; ActiGraph GT3X) | I: Pedometer, automated personalized informational newsletters, inspirational testimonial, fruit/vegetable snack.<br><br>C: No intervention; waitlist. | - | Healthy | Low |
| Kernot 2019 [52] | RCT; Australia; n=120; Age: 31.8 (SD 4.5); 100% female; Duration: 6wks. | Walking (in min/week; AAS); MVPA (in min/week; ActiGraph GT3X+); TPA (in total activity counts; ActiGraph GT3X+) | I: Combined group: moms step it up Facebook app, daily PA tip, emails, walking challenge; or pedometer and step-log.<br><br>C: Minimal intervention; PA advice emails. | Short-term after 4.6mths | Healthy | High |

| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measure-ment | Popu-lation type | Ease of Scala-bility |
|---|---|---|---|---|---|---|
| Keyserling 2008 [135] | RCT; USA; n=236; Age: 53 (SD 1.2); 100% female; Duration: 52wks | MVPA & TPA (in PAA score; New Leaf PAA) | I: Pedometer, individual and group counselling, counselling calls, reinforcement mailings, information about community support services.<br><br>C: Minimal intervention; informational brochures. | - | Healthy | Low |
| Kim 2018 [136] | Cluster-RCT; USA; n=187; Age: 20.2 (SD 1.7); 62% female; Duration: 15wks | MVPA (in min/week; ActiGraph Actitrainer) | I: Activity tracker, course on PA and nutrition (1 semester).<br><br>C: Alternative intervention; course on PA and nutrition (1 semester). | - | Healthy | Low |
| King 2008 [137] | RCT; USA; n=37; Age: 60.2 (SD 7.1); 43.2% female; Intervention duration: 8wks | MVPA (in min/week; CHAMPS); EE (in kcal/kg/week; CHAMPS) | I: Educational materials, PDA, pedometer.<br><br>C: Minimal intervention, educational materials. | - | At-risk | High |
| King 2013 [106] | RCT; USA; n=40; Age: 68.3 (SD 8.2); 72% female; Duration: 17wks | Walking (in min/week vs. baseline; CHAMPS) | I: Virtual advisor, pedometer, accompanied walks, goal setting, monetary incentives to walk (raffle).<br><br>C: Minimal intervention; general health program. | - | At-risk | Low |
| Koizumi 2009 [138] | RCT; Japan; n=68; Age: 66.5 (SD 4); 100% female; Duration: 12wks | MVPA (in min/day; Kenz Lifecorder) | I: Accelerometer, in person meetings at community center to set step goals and get PA feedback.<br><br>C: No intervention; blinded accelerometer. | - | At-risk | Low |
| Kolt 2012 [76] | RCT; New Zealand; n=330; Age: 74.1 (SD 6.1); 53.9% female; Duration: 12wks | Walking, MVPA & TPA (in min/week; AHSPAQ) | I: Pedometer, initial face-to-face advice on PA from physician, 3 telephone counselling sessions by trained PA counsellors, goal setting.<br><br>C: Alternative intervention; initial face-to-face advice on PA from physician, 3 telephone counselling sessions by trained PA counsellors, goal setting. | Long-term after 9.2mths | At-risk | Low |
| Lane 2015 [149] | Cluster-RCT; Ireland; n=402; Age: 52.6 (SD -); 100% female; Intervention duration: 9wks | TPA (in min/week; IPAQ) | I: Pedometer, informational booklets, PA classes, individualized training plans.<br><br>C: Minimal intervention informational booklet on healthy eating. | - | Healthy | Low |
| Li 2017 [139] | RCT; Canada; n=34; Age: 55.5 (SD 8.6); 82% female; Duration: 8wks | MVPA (in MET-in/day; SenseWear Mini) | I: Group education session about PA, Fitbit Flex, individual weekly activity counselling with a physical therapist by telephone.<br><br>C: No intervention; delayed group. | - | Sick | Low |
| Li 2020 [107] | RCT; Canada; n=118; Age: 53.3 (SD 13.6); 89% female; Duration: 8wks | Walking (in steps/day; SenseWear) MVPA (in min/day; SenseWear) | I: Education and counselling, Fitbit, web-application, feedback, 4 follow-up calls.<br><br>C: Minimal intervention; monthly emails. | - | Sick | Low |
| Long 2013 [108] | RCT; United Kingdom; n=89; Age: 47.3 (SD 7); 100% female; Duration: 16wks | Walking (in steps/week; New Lifestyles NL 1000) | I: PA consultation by practitioner (including booklet), pedometer, weekly prompts (telephone, e-mail or texts).<br><br>C: Minimal intervention; advisory leaflet. | - | At-risk | Low |
| Lynch 2019 [109] | RCT; Australia; n=83; Age: 61.6 (SD 6.4); 100% female; Duration: 8wks | Walking (in steps/day; ActivPAL); MVPA (in min/week; ActiGraph GT3X+) | I: Garmin Vivofit 2 activity monitor, feedback, goal-setting session, telephone coaching sessions, app, individualized automated feedback.<br><br>C: No intervention; waitlist. | - | Sick | Low |

| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measure-ment | Popu-lation type | Ease of Scala-bility |
|---|---|---|---|---|---|---|
| Lyons 2017 [110] | RCT; USA; n=40; Age: 61.5 (SD 5.6); 85% female; Duration: 12wks | Walking (in steps/day; ActivPAL); TPA (in stepping-min/day; ActivPAL) | I: Mini tablet mobile device, wearable activity monitor, app, orientation visit, weekly telephone counselling.<br><br>C: No intervention; waitlist. | - | At-risk | Low |
| Maher 2015 [53] | RCT; Australia; n=110; 74.5% female; Duration: 8wks | Walking & MVPA (in min/week; AAS) | I: Pedometer, Facebook app (active team) used in group of 3-8 Facebook friends, calendar to log daily (team) step counts, dashboard, team tally board, social comparison, gamification features.<br><br>C: No intervention; waitlist | Short-term after 2.8mths | At-risk | High |
| Mailey 2010 [150] | RCT; USA; n=51; Age: 25 (SD -); 68.1% female; Duration: 10wks | TPA (in activity counts/day; ActiGraph) | I: Pedometer, website, activity log, reminder email to submit an activity log, 2 monthly meetings with PA counsellor.<br><br>C: Minimal intervention; mental health counseling, waitlist. | - | Sick | Low |
| Mansi 2015 [54] | RCT; New Zealand; n=58; Age: 41.5 (SD 13.6); 58.6% female; Duration: 12wks | Walking (in steps/day; Yamax Digi-Walker SW200); MVPA & TPA (in MET-min/week; IPAQ-SF) | I: Pedometer, goal setting, feedback, educational material, self-monitoring.<br><br>C: Minimal intervention; educational materials. | Short-term after 3mths | At-risk | High |
| Martin 2015 [111] | RCT; USA; n=48; Age: 58 (SD 8); 46% female; Duration: 5wks | Walking (in steps/day vs. baseline; Fitbug Orb); TPA (in min/day vs. baseline; Fitbug Orb) | I; Fitbug orb+, smartphone and web interface, smart texts (half of group).<br><br>C: Minimal intervention; blinded Fitbug Orb. | - | Sick | High |
| Maselli 2019 [55] | RCT; Italy; n=33; Age: 22 (SD 2); 60.6% female; Duration: 12wks | MVPA (in min/week; ActiGraph) | I: MyWellness Key accelerometer, goal setting.<br><br>C: Combined; counseling and goal setting or waitlist. | Short-term after 3mths | At-risk | High |
| Maxwell-Smith 2019 [140] | RCT; Australia; n=68; Age: 64.1 (SD 7.9); 50% female; Duration: 12wks | MVPA (in min/week; ActiGraph GT9X) | I: Wearable tracker, group coaching, supportive phone calls.<br><br>C: Minimal intervention; educational materials. | - | Sick | Low |
| Melville 2015 [112] | Cluster-RCT; United Kingdom; n=102; Age: 46.2 (SD 13.0); 44% female; Duration: 12wks | Walking (in steps/day; ActiGraph GT3X); MVPA & TPA (in percentage time/day; ActiGraph GT3X) | I: Pedometer, individual PA consultations, structured walking program, booklet, goal setting, choice-based involvement of advisors.<br><br>C: No intervention; waitlist. | - | Sick | Low |
| Mendoza 2015 [113] | RCT; Chile; n=102; Age: 68.6 (SD 8.5); 39.2% female; Duration: 12wks | Waling (in steps/day vs. baseline; Tanita PD724) | I: Pedometer, daily diary, regular monthly COPD visit.<br><br>C: Alternative intervention; received counselling at each COPD visit and were advised to walk for at least 30-min per day, diary. | - | Sick | High |
| Merom 2007 [114] | RCT; Australia; n=369; Age: 49.1 (SD 9.3); 85% female; Duration: 12wks | Walking, MVPA & TPA (in min/week vs. baseline; AAS) | I: Pedometer, brochure, diary, booklet.<br><br>C: Combined; self-help walking program and weekly diaries; or no intervention. | - | At-risk | High |
| Motl 2011 [151] | RCT; USA; n=54; Age: 45.8 (SD 9.7); 89.6% female; Duration: 12wks | TPA (in MET-min/week; GLTEQ) | I: Informational video, text messages, pedometer, automated email announcements about new information, chats with PA coaches.<br><br>C: No intervention; waitlist. | - | Sick | Low |
| Müller 2016 [56] | RCT; Malaysia; n=43; Age: 63.3 (SD 4.5); 74% female; Duration: 12wks | TPA (in MET-min/week; IPAQ) | I: Exercise booklet, personal exercise instruction, 60 text messages, home visits.<br><br>C: Minimal intervention; exercise booklet, personal exercise instructions. | Short-term after 2.8mths | At-risk | Low |

| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measurement | Population type | Ease of Scalability |
|---|---|---|---|---|---|---|
| Murawski 2019 [57] | RCT; Australia; n=160; Age: 41.7 (SD 9.9); 80.0% female; Duration: 12wks | MVPA (in min/week; AAS) | I: Balanced smartphone app, personalized feedback, educational materials, text messages, emails, pedometer.<br><br>C: No intervention; waitlist. | Short-term after 3.2mths | Healthy | High |
| Mutrie 2012 [58] | RCT; United Kingdom; n=41; Age: 70.8 (SD 5.2); 68.3% female; Duration: 12wks | Walking (in steps/day; ActivPAL) | I: PA consultations, walking program (booklet and pedometer), option to participate in walking groups.<br><br>C: No intervention; waitlist. | Short-term after 2.8mths | At-risk | Low |
| Nolan 2017 [59] | RCT; United Kingdom; n=152; Age: 68 (SD 9); 28% female; Duration: 8wks | Walking (in steps/day vs. baseline; SenseWear); MVPA (in MET-min/day vs. baseline) | I: Pulmonary rehabilitation, pedometer, step-count diary.<br><br>C: Minimal intervention; pulmonary rehabilitation. | Short-term after 6mths | Sick | High |
| Oliveira 2019 [60] | RCT; Australia; n=131; Age: 71.5 (SD 6.5); 71% female; Duration: 26wks | Walking (in steps/day; ActiGraph GT3X-BT) | I: Fact-to-face health coaching, continuous telephone coaching, goal setting, pedometer, individualized advice, educational materials.<br><br>C: Minimal intervention; educational materials. | Short-term after 6mths | At-risk | Low |
| Paul 2016 [115] | RCT; United Kingdom; n=24; Age: 56 (SD 10.0); 52% female; Intervention duration: 6wks | Walking (in steps/day; ActivPAL) | I: Smartphone, Starfish app, pedometer.<br><br>C: Minimal intervention; usual care. | - | Sick | High |
| Pekmezi 2017 [141] | RCT; USA; n=84; Age: 57 (SD 4.7); 100% female; Duration: 26wks | MVPA (in min/week; ActiGraph GT3X) | I: Pedometer, regular mailings, computer generated individualized feedback-reports addressing psychosocial and environmental factors affecting PA.<br><br>C: Minimal intervention; cancer prevention, information on topics other than PA. | - | At-risk | High |
| Pinto 2013 [61] | RCT; USA; n=46; Age: 57.3 (SD 9.7); 54% female; Duration: 12wks | MVPA (in in/week; 7-day PAR); EE (in cals/week; CHAMPS) | I: Pedometer, walking log, weekly phone calls.<br><br>C: Alternative intervention; weekly phone calls. | Short-term after 3mths, Long-term after 9mths | Sick | Low |
| Pinto 2015 [62] | RCT; USA; n=76; Age: 55.6 (SD 9.6); 100% female; Duration: 12wks | MVPA (in min/week; ActiGraph GT3X) | I: Telephone-based PA counseling, pedometer, heart rate monitor, information booklets, frequent feedback reports.<br><br>C: Alternative intervention, 12 calls, informational booklets, PA tip sheets. | Short-term after 2.8mths | Sick | Low |
| Poirier 2016 [116] | RCT; USA; n=265; Age: 39.9 (SD 11.7); 66% female; Duration: 6wks | Walking (in steps/day; Pebble+) | I: Activity tracker, internet-based program (Walkadoo).<br><br>C: Minimal intervention; asked to continue usual routine. | - | Healthy | High |
| Pope 2018 [117] | RCT; USA; n=32; Age: 52.6 (SD 9.3); 100% female; Duration: 10wks | Walking (in steps/day; ActiGraph GT3X+); MVPA (in min/day; ActiGraph GT3X+) | I: Polar M400 smartwatch, Facebook page, strength and aerobic training program via Facebook.<br><br>C: Alternative intervention; Facebook group, strength and aerobic training program via Facebook, weekly tips. | - | Sick | Low |
| Prestwich 2009 [142] | RCT; United Kingdom; n=155; Age: 23.8 (SD 4.6); 58% female; Duration: 4wks | MVPA (in exercise frequency, questionnaire) | I: Motivational text messages, implementation intention plan.<br><br>C: Combined; motivational messages, implementation intention plan; or no intervention. | - | At-risk | High |

| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measure-ment | Popu-lation type | Ease of Scala-bility |
|---|---|---|---|---|---|---|
| Prestwich 2010 [118] | RCT; United Kingdom; n=149; Age: 23.4 (SD 5.6); 66% female; Duration: 4wks | Walking (in days/week w/>30min; SWET); TPA (in days/week w/>30min exercise; SWET) | I: PA guidelines, implementation intention plan, website, specific PA plan, SMS reminders (on goals or plan).<br><br>C: Minimal intervention; PA guidelines. | - | At-risk | High |
| Reijonsaari 2012 [119] | RCT; Finland; n=544; Age: 43.5 (SD 10); 64% female; Duration: 52wks | Walking (in MET-min/week; IPAQ-SF) | I: Accelerometer, goal setting, online service helping to track their activity levels, telephone or web counselling.<br><br>C: Minimal intervention; fitness test informational leaflet on PA. | - | Healthy | Low |
| Ribeiro 2014 [63] | RCT; Brazil; n=195; Age: 45 (SD 3); 100% female; Duration: 13wks | Walking (in steps/week vs. baseline; Digi-Walker PW610); MVPA (in moderate intensity steps/week vs. baseline; Digital-Walker PW 610) | I: Counseling sessions (individual or group), booklet, pedometer, diary.<br><br>C: Combined; counseling sessions, booklet; or aerobic exercise sessions (twice per week). | Short-term after 3mths | At-risk | Low |
| Roos 2014 [120] | RCT; South Africa; n=84; Age: 39.1 (SD 9.2); 79% female; Duration: 52wks | Walking (in steps/day vs. baseline; Yamax Digi-Walker SW200) | I: Pedometer, activity diary, contact sessions, SMS.<br><br>C: Minimal intervention; standard HIV care, monthly call from principal investigator. | - | Sick | Low |
| Rowley 2019 [121] | RCT; USA; n=170; Age: 67.3 (SD 6.3); 79.4% female; Duration: 12wks | Walking (in steps/day; Omron HJ-720ITC) | I: Combined; individually tailored feedback through interactive website, goal setting, pedometer; or pedometer and goal setting.<br><br>C: No intervention; instruction to maintain current behavior. | - | At-risk | High |
| Samuels 2011 [143] | RCT; USA; n=50; Age: 48.7 (SD 9.1); 81% female; Duration: 4wks | MVPA (in min/day; ActiGraph 7164) | I: Pedometers, instructed to achieve 10 000 steps daily, weekly meetings, activity logs.<br><br>C: Combined; sealed pedometers, weekly meetings with principal investigator, activity logs, instructed to engage in 30-minutes of MVPA daily; or instructed to accumulate 30-minutes of MVPA in bouts of at least 10-minutes or longer on a daily basis. | - | At-risk | Low |
| Schwerdt-feger 2012 [154] | RCT; Austria; n=63; Age: 23.7 (SD 4); 68% female; Duration: 1wk | TPA (in counts/min; ActiGraph GT1M) | I: Individual PA information and planning sessions, goal setting, daily SMS reminders (only half of participants).<br><br>C: Minimal intervention: government PA guidelines. | - | At-risk | Low |
| Sharp 2016 [144] | RCT; Canada; n=184; Age: 18 (SD 0.7); 53% female; Duration: 12wks | MVPA (in min/week, GLTEQ) | I: Pedometer, monthly tracking logs, follow-up e-mails.<br><br>C: No intervention; instruction to continue with usual activity. | - | Healthy | High |
| Simons 2018 [64] | Cluster-RCT; Belgium; n=130; Age: 25 (SD 3.0); 51.5% female; Duration: 9wks | Walking (in steps/day; ActiGraph GT3X+); MVPA & TPA (in min/day; ActiGraph GT3X+) | I: Active Coach app, Fitbit, goal setting, educational materials.<br><br>C: Minimal intervention; educational materials. | Short-term after 3.2mths | At-risk | High |
| Spence 2009 [122] | RCT; Canada; n=63; -; 100% female; Duration: 1wk | Walking (in min/week; IPAQ-SF) | I: Pedometer, log sheet, walking intentions and self-efficacy questionnaire encouraging 12 500 steps/day (only half of participants).<br><br>C: Combined; walking intentions and self-efficacy questionnaire encouraging 12 500 steps/day; or no intervention. | - | Healthy | High |
| Stacey 2016 [65] | RCT; Australia; n=174; Age: 57 (SD 12); 77% female; Duration: 8wks | Walking (in steps/day; Yamax Digi-Walker SW200) | I: Group education sessions, workbook, pedometer, gymstick.<br><br>C: No intervention; waitlist. | Short-term after 2.8mths | Sick | Low |

| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measure-ment | Popu-lation type | Ease of Scala-bility |
|---|---|---|---|---|---|---|
| Sugden 2008 [152] | RCT; United Kingdom; n=54; Age: 76 (SD -); 100% female; Intervention duration: 12wks | TPA (in activity count/day; Stay Healthy RT3) | I: Pedometer, individualized activity action plans, daily activity diary, counseling session and phone calls.<br><br>C: Alternative intervention; individualized activity action plans, daily activity diary, counseling session and phone calls. | - | At-risk | Low |
| Suggs 2013 [66] | RCT; United Kingdom; n=331; Age: 39.5 (SD 14.8); 82.5% female; Duration: 12wks | TPA (in MET hrs/week; IPAQ-L) | I: 12 Emails, 24 text messages.<br><br>C: Alternative intervention; 12 emails. | Short-term after 0.9mths | Healthy | High |
| Tabak 2014 [123] | RCT; Netherlands; n=34; Age: 66.6 (SD 7.4); 37% female; Duration: 3wks | Walking (in steps/ay; Yamax Digi-Walker SW200) | I: Activity coach (smartphone and accelerometer), web portal, feedback text messages, usual COPD rehabilitation care (medication and physiotherapy).<br><br>C: Minimal intervention; usual COPD rehabilitation care (medication and physiotherapy). | - | Sick | High |
| Talbot 2003 [67] | RCT; USA; n=34; Age: 70.2 (SD 5.8); 76.5% female; Duration: 12wks | Walking (in steps/day; Yamax Digi-Walker SW200) | I: Arthritis self-management education, pedometer, activity logs, brief individual counseling, exercise booklet.<br><br>C: Alternative intervention; 12-hour arthritis self-management program. | Short-term after 2.8mths | Sick | Low |
| Ter Hoeve 2018 [68] | RCT; Netherlands; n=731; Age: 58.7 (SD 8.7); 18% female; Duration: 52wks | Walking (in steps/wear time; ActiGraph); MVPA (in MVPA/wear time; Actigraph) | I: Pedometer, group exercise and counseling sessions, booklet.<br><br>C: Combined; group exercise sessions; or exercise sessions and 9 month telephonic after-care program. | Short-term after 6mths | Sick | Low |
| Thorndike 2014 [124] | RCT; USA; n=103; Age: 29 (SD -); 54% female; Duration: 12wks | Walking (in steps/day; Fitbit) | I: Fitbit, Fitbit website.<br><br>C: No intervention; blinded Fitbit. | - | Healthy | High |
| Thorsteinsen 2014 [153] | RCT; Norway; n=21; Age: 55.3 (SD 11.2); 47.6% female; Duration: 12wks | TPA (in min/week; questionnaire) | I: Information meeting, website, goal setting, personalized text messages, game elements (competition and social comparison, group goal setting).<br><br>C: Minimal intervention; information meeting, daily reporting of PA levels through survey. | - | Healthy | High |
| Tudor-Locke 2004 [69] | RCT; USA; n=60; Age: 52.7 (SD 5.2); 45% female; Duration: 16wks | Walking (in steps/day; Yamax Digi-Walker SW200) | I: Group meetings, pedometers, program manual, calendars, postcards.<br><br>C: No intervention; waitlist. | Short-term after 1.8mths | Sick | Low |
| Unick 2012 [145] | RCT; USA; n=29; Age: 42.3 (SD 9.8); 82% female; Duration: 26wks | MVPA (in min/week; SenseWear) | I: Body media FIT system (mobile tracker, website), group meetings, exercise and nutrition goals, paper diaries, feedback.<br><br>C: Alternative intervention; group meetings, exercise and nutrition goals, paper diaries, feedback. | - | Sick | Low |
| Vallance 2008 [70] | RCT; Canada; n=377; Age: 58 (SD -); 100% female; Duration: 12wks | MVPA (in min/week; GLTEQ) | I: Pedometer, step calendar, PA guidebook (only half of participants), instruction to perform 30-min MVPA 5x per week.<br><br>C: Combined; instruction to perform 30min MVPA 5x per week; or instructions plus PA guidebook. | Short-term after 6mths | Sick | High |
| Vallance 2016 [125] | RCT; Canada; n=95; Age: 52.8 (SD 9.8); 100% female; Duration: 13wks | Walking (in steps/day; StepsCount SC-01); MVPA (in min/week; GLTEQ) | I: Informational materials, pedometer, activity diary, goal setting.<br><br>C: Minimal intervention; informational materials. | - | Sick | High |

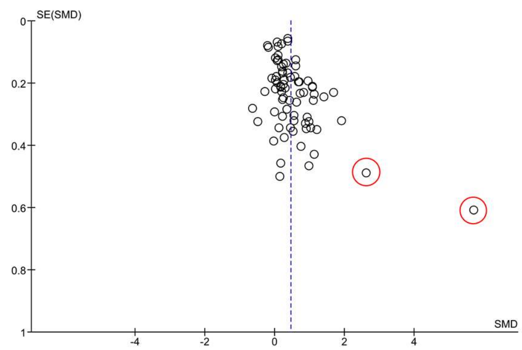
| | Study design, key sample demographics | Outcome | Intervention and control description | Follow-up Measure-ment | Popu-lation type | Ease of Scala-bility |
|---|---|---|---|---|---|---|
| Van Blarigan 2019 [126] | RCT; USA; n=42; Age: 54 (SD 11); 58.5% female; Duration: 21wks | Walking (in steps/day ActiGraph GTX3+); MVPA in min/day; ActiGraph GTX3+) | I: Print materials on A after cancer, Fitbit Flex, daily text messages.<br><br>C: Minimal intervention; print materials on PA after cancer. | - | Sick | High |
| Vandelanotte 2018 [146] | RCT; Australia; n=243; Age: 51.5 (SD 11.1); 74.9% female; Duration: 13wks | MVPA & TPA (in min/week; AAS) | I: Fitbit; Taylor Active web-based intervention, educational materials, goal setting, individualized feedback.<br><br>C: Alternative intervention; TaylorActive web-based intervention, educational materials, goal setting, individualized feedback. | - | At-risk | High |
| Van der Weegen 2015 [71] | Cluster-RCT; Netherlands; n=199; Age: 57.8 (SD 7.7); 51.2% female; Duration: 26wks | MVPA (in min/day; Personal Activity Monitor AM300) | I: It's LiFe! smartphone and web-app, monitoring, self-management support program including feedback, booklet, and consultation sessions.<br><br>C: Combined; self-management support program including feedback, booklet, and consultation sessions; or usual primary care. | Short-term after 3mths | Sick | Low |
| Van Hoye 2018 [72] | RCT; Belgium; n=227; Age: 42.4 (SD 10.4); 17% female; Duration: 4wks | Walking (in steps/day vs. baseline; SenseWear); MVPA (in min/day vs. baseline; SenseWear); TPA (in METs/day vs. baseline; SenseWear) | I: Combined intervention group; 1/3rd pedometer & feedback, 1/3rd feedback on daily steps, daily minutes of MVPA and total daily EE in real-time from a SWA display; 1/3rd feedback on daily steps, daily minutes of MVPA and total daily EE in real-time from a SenseWear display plus weekly face-to-face coaching.<br><br>C: Minimal intervention; information on the energy expenditure of familiar activities (e.g. housework, walking, cycling). | Short-term after 6mths, Long-term after 12mths | At-risk | Low |
| Warren 2014 [127] | RCT; United Kingdom; n=131; Age: 59.9 (SD 9.3); 33% female; Duration: 12wks | Walking (in steps/day; New Lifestyles VL-800) | I: Pedometer, informational booklet, PA advice by general practitioner to walk at least 1 mile a day, self-monitoring.<br><br>C: Combined; informational booklet and sealed pedometer; or PA advice by general practitioner to walk at least 1 mile a day and sealed pedometer. | - | At-risk | High |
| Wijsman 2013 [147] | RCT; Netherlands; n=235; Age: 64.8 (SD 2.9); 41% female; Duration: 12wks | MVPA (in min/day; GENEActiv) | I: Web-based physical activity program (accelerometer and monitor, personal website and eCoach).<br><br>C: No intervention; waitlist | - | At-risk | Low |
| Wyke 2019 [77] | RCT; England, Netherlands, Norway, Protugal; n=1113; Age: 45.8 (SD 8.8); 0% female; Duration: 12wks | Walking (in steps/day; ActivPAL); TPA (in MET-min/week; IPAQ-SF) | I: Educational materials, group coaching at soccer club, pocket-worn tracking device (SitFIT), game-based app (MatchFIT).<br><br>C: Minimal intervention; educational materials. | Long-term after 9.2mths | At-risk | Low |
| Yamada 2012 [128] | RCT; Japan; n=87; Age: 75.6 (SD 6.7); 46% female; Duration: 26wks | Walking (in steps/day; Yamax Power-Walker EX-510) | I: Pedometer, step log grid, feedback.<br><br>C: No intervention. | - | At-risk | High |
| Yates 2017 [78] | Cluster-RCT; United Kingdom; n=808; Age: 63.1 (SD 8.2); 36.4% female; Duration: 104wks | Walking (in steps/day vs. baseline; ActiGraph GT3X); MVPA (in min/day vs. baseline; ActiGraph GT3X); TPA (in 1000 counts/day vs. baseline; ActiGraph GT3X) | I: Pedometer, action plan and step diary, yearly group sessions, telephone contact.<br><br>C: Minimal intervention; informational booklet. | Long-term after 12mths | At-risk | Low |

*Abbreviations:* RCT, randomized controlled trial; SD, standard deviation; MVPA, moderate to vigorous physical activity; TPA, total physical activity; EE, energy expenditure; PA, physical activity; GLTEQ, Godin Leisure-Time Exercise Questionnaire; IPAQ-L or IPAQ, International Physical Activity Questionnaire long-form; IPAQ-SF, International Physical Activity Questionnaire short-form; AAS, Active Australia Survey, PAR, 7-day Physical Activity Recall; CHAMPS, Community Health Activities Model Program for Seniors; SWET, Self-report Walking and Exercise Tables; AWAS, Australian Women's Activity Study; AHSPAQ, Auckland Heart Physical Activity Questionnaire; CDC, Center for Disease Control and Prevention; NHS, National Health Service; COPD, Chronic Obstructive Pulmonary Disease; MET, metabolic equivalent of task .

*Comments:* At-risk group includes elderly, obese, and inactive populations; sick group includes populations that currently experienced or previously experienced illnesses such as diabetes, cancer, chronic obstructive pulmonary disease, coronary heart disease or others; High ease of scalability defined as the ability to scale-up an intervention without human resource requirement; short-term follow up includes follow ups performed ≤6mths after end of intervention; long-term follow-up defined as measurements taken >6mths after end of intervention.

**Appendix 5:** Funnel Plot Analysis to Detect Publication Bias

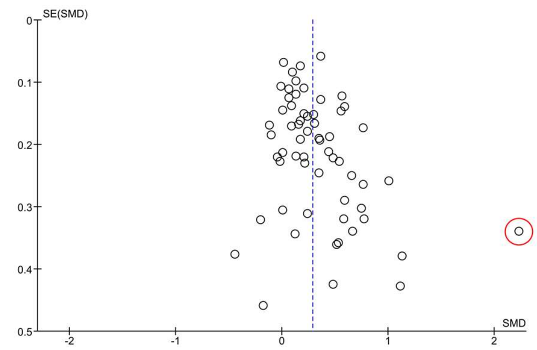
**Studies reporting Walking as outcome (N=77)**




*No publication bias detected apart from outlier.*

*Outlier was visually identified and marked in red: Barwais et al., 2013*[82]*; Hultquist et al., 2013.*[102]

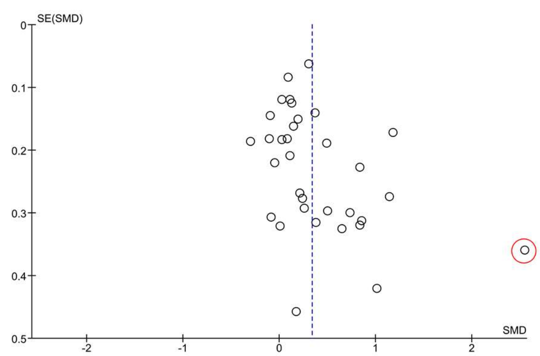
**Studies reporting MVPA as outcome (N=62)**




*No publication bias detected apart from outliers.*

*Outliers were visually identified and marked in red: Mansi et al., 2015.*[54]

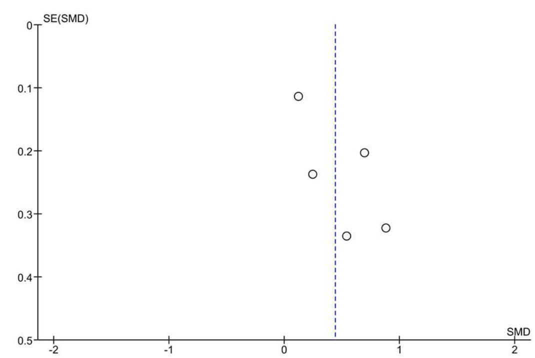
**Studies reporting Total Physical Activity as outcome (N=33)**




*No publication bias detected apart from outlier.*

*Outlier was visually identified and marked in red: Mansi et al., 2015.* [54]

**Studies reporting Energy Expenditure as outcome (N=5)**




*No publication bias detected, small sample.*

## **Appendix 6:** Sensitivity Analysis

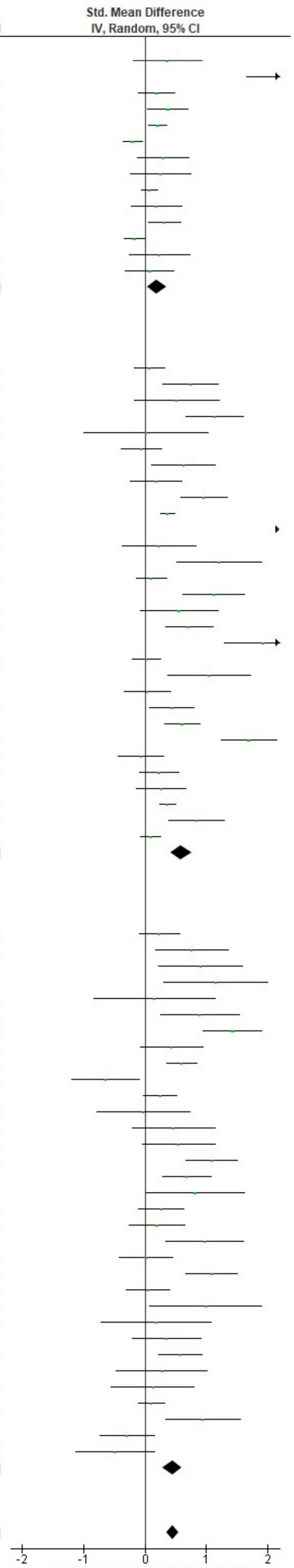
| Scenarios | Outcome Measure | Timepoint | No. of studies | SMD 95% CI | P-value | $I^2$ | P-value |
|---|---|---|---|---|---|---|---|
| **Baseline scenario** | Walking | End of intervention | 77 | 0.46 [0.36, 0.55] | <.001 | 83% | <.001 |
| | | Short-term follow-up (≤6 months) | 19 | 0.26 [0.09, 0.42] | .002 | 73% | <.001 |
| | | Long-term follow-up (>6 months) | 7 | 0.25 [0.10, 0.39] | .001 | 68% | .004 |
| | MVPA | End of intervention | 62 | 0.28 [0.21, 0.35] | <.001 | 62% | <.001 |
| | | Short-term follow-up (≤6 months) | 21 | 0.20 [0.05, 0.35] | .008 | 72% | <.001 |
| | | Long-term follow-up (>6 months) | 7 | 0.19 [0.11, 0.27] | <.001 | 0% | .44 |
| | TPA | End of intervention | 33 | 0.34 [0.20, 0.47] | <.001 | 77% | <.001 |
| | | Short-term follow-up (≤6 months) | 9 | 0.53 [0.13, 0.93] | .009 | 87% | <.001 |
| | | Long-term follow-up (>6 months) | 6 | 0.19 [-0.00, 0.38] | .05 | 72% | .003 |
| | EE | End of intervention | 5 | 0.44 [0.13, 0.75] | .05 | 60% | .04 |
| | | Short-term follow-up (≤6 months) | 1 | 0.52 [-0.10, 1.14] | .10 | n/a[a] | n/a[a] |
| | | Long-term follow-up (>6 months) | 1 | 0.16 [-0.45, 0.77] | .61 | n/a[a] | n/a[a] |
| **Without outlier studies** | Walking | End of intervention | 74 | 0.39 [0.30, 0.48] | <.001 | 78% | <.001 |
| | | Short-term follow-up (≤6 months) | 18 | 0.18 [0.06, 0.31] | .004 | 51% | .008 |
| | | Long-term follow-up (>6 months) | 7 | 0.25 [0.10, 0.39] | .001 | 68% | .004 |
| | MVPA | End of intervention | 60 | 0.25 [0.19, 0.32] | <.001 | 52% | <.001 |
| | | Short-term follow-up (≤6 months) | 20 | 0.12 [0.03, 0.22] | .008 | 28% | .12 |
| | | Long-term follow-up (>6 months) | 7 | 0.19 [0.11, 0.27] | <.001 | 0 | .44 |
| | TPA | End of intervention | 32 | 0.28 [0.17, 0.40] | <.001 | 69% | <.001 |
| | | Short-term follow-up (≤6 months) | 8 | 0.26 [0.03, 0.50] | .03 | 60% | .01 |
| | | Long-term follow-up (>6 months) | 6 | 0.19 [-0.00, 0.38] | .05 | 72% | .003 |
| | EE | End of intervention | 5 | 0.44 [0.13, 0.75] | .005 | 60% | .04 |
| | | Short-term follow-up (≤6 months) | 1 | 0.52 [-0.10, 1.14] | .10 | n/a[a] | n/a[a] |
| | | Long-term follow-up (>6 months) | 1 | 0.16 [-0.45, 0.77] | .61 | n/a[a] | n/a[a] |
| **Only low risk of bias studies** | Walking | End of intervention | 17 | 0.59 [0.37, 0.82] | <.001 | 84% | <.001 |
| | | Short-term follow-up (≤6 months) | 4 | 0.58 [-0.04, 1.21] | .07 | 87% | <.001 |
| | | Long-term follow-up (>6 months) | 2 | 0.38 [-0.33, 1.10] | .29 | 79% | .03 |
| | MVPA | End of intervention | 15 | 0.51 [0.29, 0.72] | <.001 | 83% | <.001 |
| | | Short-term follow-up (≤6 months) | 5 | 0.56 [-0.13, 1.25] | .11 | 92% | <.001 |
| | | Long-term follow-up (> 6 months) | 2 | 0.12 [-0.04, 0.28] | .13 | 0% | .36 |
| | TPA | End of intervention | 5 | 0.66 [0.10, 1.23] | .02 | 91% | <.001 |
| | | Short-term follow-up (≤6 months) | 2 | 1.55 [-0.71, 3.81] | .18 | 97% | <.001 |
| | | Long-term follow-up (>6 months) | 2 | 0.09 [-0.07, 0.25] | .27 | 0% | .93 |
| | EE | End of intervention | 5 | 0.44 [0.13, 0.75] | .005 | 60% | .04 |
| | | Short-term follow-up (≤6 months) | 1 | 0.52 [-0.10, 1.14] | .10 | n/a[a] | n/a[a] |
| | | Long-term follow-up (>6 months) | 1 | 0.16 [-0.45, 0.77] | .61 | n/a[a] | n/a[a] |
| **Only studies reporting long-term follow-ups** | Walking | End of intervention | 7 | 0.39 [0.19, 0.59] | <.001 | 83% | <.001 |
| | | Short-term follow-up (≤6 months) | 1 | 0.35 [0.03, 0.67] | .03 | n/a[a] | n/a[a] |
| | | Long-term follow-up (>6 months) | 7 | 0.25 [0.10, 0.39] | .001 | 68% | .004 |
| | MVPA | End of intervention | 7 | 0.24 [0.08, 0.39] | .003 | 55% | .04 |
| | | Short-term follow-up (≤6 months) | 2 | 0.29 [0.00, 0.57] | .05 | 0% | .50 |
| | | Long-term follow-up (>6 months) | 7 | 0.19 [0.11, 0.27] | <.001 | 0% | .44 |
| | TPA | End of intervention | 6 | 0.25 [0.09, 0.42] | .003 | 61% | .03 |
| | | Short-term follow-up (≤6 months) | 1 | 0.26 [-0.06, 0.58] | .11 | n/a[a] | n/a[a] |
| | | Long-term follow-up (>6 months) | 6 | 0.19 [-0.00, 0.38] | .05 | 72% | .003 |
| | EE | End of intervention | 1 | 0.88 [0.25, 1.51] | .006 | n/a[a] | n/a[a] |
| | | Short-term follow-up (≤6 months) | 1 | 0.52 [-0.10, 1.14] | .10 | n/a[a] | n/a[a] |
| | | Long-term follow-up (>6 months) | 1 | 0.16 [-0.45, 0.77] | .61 | n/a[a] | n/a[a] |

[a] In subgroups where n=1, heterogeneity cannot be calculated.

## **Appendix 7:** Subgroup Analysis by Population Type

Subgroup Analysis by Population Type for The Outcome Walking

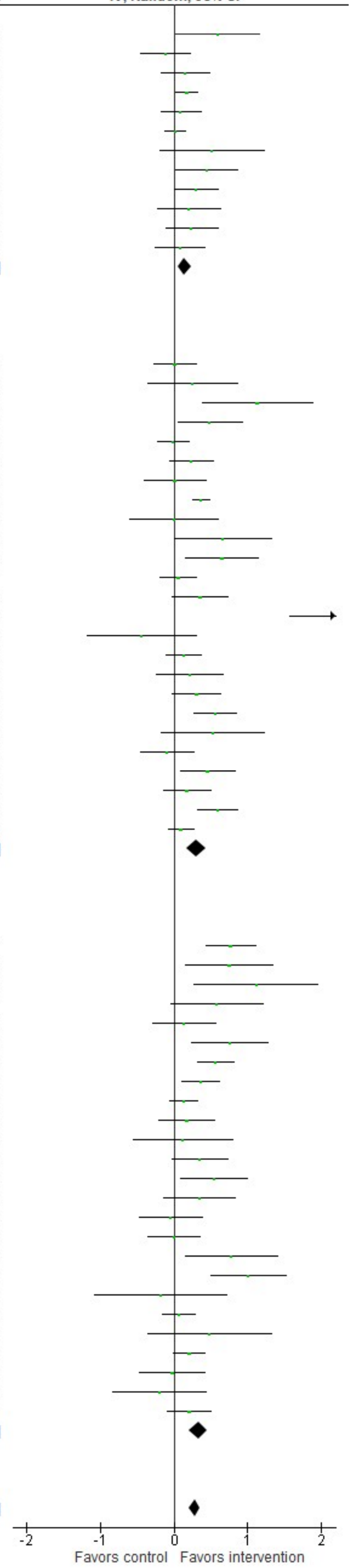
| | Intervention | | | Control | | | | Std. Mean Difference |
|---|---|---|---|---|---|---|---|---|
| **Study or Subgroup** | **Mean** | **SD** | **Total** | **Mean** | **SD** | **Total** | **Weight** | **IV, Random, 95% CI** |
| **2.1.1 Healthy** | | | | | | | | |
| Ashton, 2017 [80] | 1,588.2 | 2,740.6 | 26 | 575.4 | 2,735.1 | 24 | 1.2% | 0.36 [-0.20, 0.92] |
| Barwais, 2013 [82] | 1,625 | 553.8 | 18 | 483 | 175.9 | 15 | 0.7% | 2.61 [1.65, 3.57] |
| Compernolle, 2015 [85] | 9,484 | 4,875 | 86 | 8,589 | 4,380 | 93 | 1.6% | 0.19 [-0.10, 0.49] |
| Dadaczynski, 2017 [87] | 526.5 | 240.6 | 80 | 442.9 | 200.4 | 64 | 1.5% | 0.37 [0.04, 0.70] |
| Dishman, 2009 [91] | 14.8 | 17.3 | 564 | 11.2 | 17.3 | 265 | 1.8% | 0.21 [0.06, 0.35] |
| Finkelstein, 2016 [48] | -480 | 2,568 | 201 | 43.7 | 2,640.7 | 599 | 1.8% | -0.20 [-0.36, -0.04] |
| Gell, 2015 [95] | 6,867.7 | 2,227 | 41 | 6,189 | 2,297 | 46 | 1.4% | 0.30 [-0.13, 0.72] |
| Glynn, 2014 [97] | 5,855 | 4,264 | 31 | 4,859 | 3,474 | 35 | 1.3% | 0.25 [-0.23, 0.74] |
| Hardeman, 2020 [98] | 8,419 | 3,224 | 417 | 8,191 | 3,003 | 442 | 1.8% | 0.07 [-0.06, 0.21] |
| Kernot, 2019 [52] | 276.3 | 221.7 | 74 | 235 | 218.9 | 33 | 1.4% | 0.19 [-0.23, 0.60] |
| Poirier, 2016 [116] | 5,411 | 2,277 | 107 | 4,751 | 1,834 | 110 | 1.6% | 0.32 [0.05, 0.59] |
| Reijonsaari, 2012 [119] | 2,047 | 1,650 | 264 | 2,338 | 1,762 | 257 | 1.8% | -0.17 [-0.34, 0.00] |
| Spence, 2009 [122] | 690 | 450.2 | 32 | 592.5 | 368.9 | 31 | 1.3% | 0.23 [-0.26, 0.73] |
| Thorndike, 2014 [124] | 7,886 | 3,622 | 50 | 7,600 | 3,492 | 49 | 1.4% | 0.08 [-0.31, 0.47] |
| **Subtotal (95% CI)** | | | **1991** | | | **2063** | **20.5%** | **0.20 [0.04, 0.35]** |
| Heterogeneity: Tau² = 0.06; Chi² = 59.23, df = 13 (P < 0.00001); I² = 78% | | | | | | | | |
| Test for overall effect: Z = 2.44 (P = 0.01) | | | | | | | | |
| **2.1.2 At-risk** | | | | | | | | |
| Aittasalo, 2012 [41] | 457 | 306 | 123 | 431 | 403 | 118 | 1.7% | 0.07 [-0.18, 0.33] |
| Baker, 2008 [81] | 9,977 | 4,669 | 39 | 7,078 | 2,911 | 40 | 1.3% | 0.74 [0.28, 1.20] |
| Butler, 2004 [83] | 56,396 | 3,315 | 17 | 53,666 | 6,543 | 16 | 1.0% | 0.52 [-0.18, 1.21] |
| Creel, 2016 [86] | 6,084 | 873 | 52 | 5,253 | 421 | 35 | 1.3% | 1.13 [0.67, 1.60] |
| Croteau, 2004 [129] | 2,419 | 6,074.6 | 7 | 2,320 | 4,732 | 8 | 0.6% | 0.02 [-1.00, 1.03] |
| Croteau, 2007 [130] | 6,180 | 3,530 | 79 | 6,378 | 2,994 | 68 | 1.5% | -0.06 [-0.38, 0.26] |
| Duru, 2010 [92] | 9,883 | 14,014 | 34 | 2,426 | 8,429 | 28 | 1.2% | 0.62 [0.11, 1.14] |
| Fjeldsoe, 2010 [94] | 16.67 | 89.4 | 45 | 0.34 | 89.4 | 43 | 1.4% | 0.18 [-0.24, 0.60] |
| Gill, 2019 [96] | 1,646 | 3,302 | 59 | -1,485 | 3,171.5 | 59 | 1.5% | 0.96 [0.58, 1.34] |
| Harris, 2018 [75] | 8,306 | 3,140 | 778 | 7,198 | 2,809 | 456 | 1.8% | 0.37 [0.25, 0.48] |
| Hultquist, 2005 [102] | 10,159 | 292 | 27 | 8,270 | 354 | 31 | 0.5% | 5.70 [4.51, 6.90] |
| Katzmarzyk, 2011[104] | 7,248 | 2,737 | 20 | 6,637 | 2,418 | 23 | 1.1% | 0.23 [-0.37, 0.83] |
| King, 2013 [106] | 253.5 | 248.7 | 20 | 26.8 | 67 | 19 | 1.0% | 1.21 [0.52, 1.89] |
| Kolt, 2012 [76] | 107.4 | 164.9 | 130 | 92.2 | 120.5 | 123 | 1.7% | 0.10 [-0.14, 0.35] |
| Long, 2013 [108] | 65,983 | 18,069 | 38 | 47,596 | 13,900 | 33 | 1.3% | 1.12 [0.61, 1.62] |
| Lyons, 2017 [110] | 6,193 | 3,183 | 20 | 4,586 | 2,476 | 20 | 1.1% | 0.55 [-0.08, 1.19] |
| Maher, 2015 [53] | 332 | 289 | 51 | 160 | 185 | 59 | 1.4% | 0.72 [0.33, 1.10] |
| Mansi, 2015 [54] | 9,792 | 2,053 | 29 | 6,551 | 1,154 | 29 | 1.1% | 1.92 [1.29, 2.55] |
| Merom, 2007 [114] | 52 | 107.2 | 105 | 48.9 | 126.8 | 209 | 1.7% | 0.03 [-0.21, 0.26] |
| Mutrie, 2012 [58] | 9,351 | 2,017 | 20 | 7,138 | 2,169 | 19 | 1.0% | 1.04 [0.36, 1.71] |
| Oliveira, 2019 [60] | 7,507 | 3,077 | 54 | 7,401 | 2,841 | 55 | 1.5% | 0.04 [-0.34, 0.41] |
| Prestwich, 2010 [118] | 2 | 1.9 | 88 | 1.2 | 1.58 | 46 | 1.5% | 0.44 [0.08, 0.80] |
| Ribeiro, 2014 [63] | 969.7 | 1,464.1 | 101 | -181.5 | 2,284 | 94 | 1.6% | 0.60 [0.32, 0.89] |
| Rowley, 2019 [121] | 9,015 | 2,842 | 97 | 4,654 | 1,447 | 32 | 1.3% | 1.69 [1.23, 2.14] |
| Simons, 2018 [64] | 7,741 | 4,553 | 55 | 8,061 | 5,112 | 63 | 1.5% | -0.07 [-0.43, 0.30] |
| Van Hoye, 2018 [72] | 13.2 | 29.6 | 157 | 6.7 | 26 | 50 | 1.6% | 0.23 [-0.09, 0.54] |
| Warren, 2014 [127] | 8,371 | 3,069 | 37 | 7,576 | 2,993 | 69 | 1.4% | 0.26 [-0.14, 0.66] |
| Wyke, 2019 [77] | 9,801 | 3,730 | 464 | 8,518 | 3,254 | 471 | 1.8% | 0.37 [0.24, 0.50] |
| Yamada, 2012 [128] | 3,726 | 1,607 | 40 | 2,267 | 1,837 | 42 | 1.3% | 0.84 [0.38, 1.29] |
| Yates, 2017 [78] | -486 | 2,252 | 287 | -690 | 2,041 | 272 | 1.8% | 0.09 [-0.07, 0.26] |
| **Subtotal (95% CI)** | | | **3073** | | | **2630** | **40.3%** | **0.59 [0.42, 0.76]** |
| Heterogeneity: Tau² = 0.17; Chi² = 230.72, df = 29 (P < 0.00001); I² = 87% | | | | | | | | |
| Test for overall effect: Z = 6.81 (P < 0.00001) | | | | | | | | |
| **2.1.3 Sick** | | | | | | | | |
| Alsaleh, 2016 [79] | 36.67 | 109.76 | 66 | 16.46 | 58.73 | 79 | 1.5% | 0.23 [-0.09, 0.56] |
| Cadmus-Bertram, 2019 [84] | 6,697 | 2,878 | 24 | 4,853 | 1,703 | 23 | 1.1% | 0.76 [0.17, 1.36] |
| Coelho, 2018 [43] | 8,853 | 3,320 | 20 | 6,248 | 2,030 | 17 | 1.0% | 0.91 [0.23, 1.59] |
| Cruz, 2016 [44] | 10,440 | 4,012 | 13 | 6,430 | 2,613 | 13 | 0.8% | 1.15 [0.31, 1.99] |
| De Blok, 2006 [88] | 3,927 | 2,617 | 8 | 3,554 | 2,019.6 | 8 | 0.7% | 0.15 [-0.83, 1.13] |
| DeGreef, 2010 [73] | 9,601 | 5,002 | 20 | 5,538 | 3,877 | 21 | 1.0% | 0.89 [0.25, 1.54] |
| DeGreef, 2011[89] | 7,703 | 2,729 | 60 | 3,883 | 2,537 | 32 | 1.3% | 1.42 [0.94, 1.90] |
| DeGreef, 2011a [74] | 6,771 | 3,889 | 43 | 5,173 | 3,094 | 24 | 1.3% | 0.44 [-0.07, 0.94] |
| Demeyer, 2017 [90] | 7 | 32.5 | 129 | -10 | 23.5 | 132 | 1.7% | 0.60 [0.35, 0.85] |
| Engel, 2006 [93] | 150 | 117 | 24 | 216 | 87 | 30 | 1.2% | -0.64 [-1.19, -0.09] |
| Furber, 2010 [50] | 249.9 | 196 | 97 | 202.6 | 189.5 | 107 | 1.6% | 0.24 [-0.03, 0.52] |
| Hornikx, 2015 [99] | 984 | 1,208 | 12 | 1,013 | 1,275 | 15 | 0.9% | -0.02 [-0.78, 0.74] |
| Hospes, 2009 [100] | 7,872 | 3,962 | 18 | 6,172 | 3,194 | 17 | 1.0% | 0.46 [-0.21, 1.13] |
| Houle, 2011 [101] | 9,850 | 3,282 | 23 | 7,970 | 3,433 | 22 | 1.1% | 0.55 [-0.05, 1.15] |
| Izawa, 2012 [103] | 8,609.6 | 3,064.5 | 52 | 5,512.9 | 2,571.8 | 51 | 1.4% | 1.09 [0.67, 1.50] |
| James, 2015 [51] | 800.8 | 2,820.6 | 57 | -1,294 | 3,304.4 | 52 | 1.4% | 0.68 [0.29, 1.07] |
| Kawagoshi, 2015 [105] | 51.3 | 63.7 | 12 | 12.3 | 25.5 | 15 | 0.8% | 0.82 [0.02, 1.61] |
| Li, 2020 [107] | 6,673 | 3,462 | 55 | 5,819 | 2,860 | 57 | 1.5% | 0.27 [-0.10, 0.64] |
| Lynch, 2019 [109] | 8,193 | 3,301 | 37 | 7,539 | 3,404 | 40 | 1.3% | 0.19 [-0.26, 0.64] |
| Martin, 2015 [111] | 1,067 | 2,098 | 32 | -1,042 | 2,202 | 16 | 1.1% | 0.97 [0.34, 1.61] |
| Melville, 2015 [112] | 4,823 | 2,059 | 42 | 4,784 | 2,613 | 40 | 1.4% | 0.02 [-0.42, 0.45] |
| Mendoza, 2015 [113] | 3,080 | 3,254 | 52 | 138 | 1,950 | 50 | 1.4% | 1.08 [0.67, 1.50] |
| Nolan, 2017 [59] | 272 | 2,275.9 | 63 | 155 | 2,557.1 | 59 | 1.5% | 0.05 [-0.31, 0.40] |
| Paul, 2016 [115] | 5,791 | 2,952 | 15 | 2,947 | 2,399 | 8 | 0.7% | 0.99 [0.07, 1.90] |
| Pope, 2018 [117] | 5,175 | 2,308 | 12 | 4,746 | 2,045 | 8 | 0.7% | 0.19 [-0.71, 1.08] |
| Roos, 2014 [120] | 3,810 | 10,993.5 | 29 | 817.3 | 2,349.2 | 22 | 1.2% | 0.35 [-0.21, 0.91] |
| Stacey, 2016 [65] | 10,849 | 5,127 | 75 | 8,014 | 4,568 | 58 | 1.5% | 0.58 [0.23, 0.93] |
| Tabak, 2014 [123] | 5,603 | 3,475.8 | 13 | 4,617 | 3,460 | 16 | 0.9% | 0.28 [-0.46, 1.01] |
| Talbot, 2003 [67] | 4,337 | 2,903 | 17 | 3,972 | 2,563 | 17 | 1.0% | 0.13 [-0.54, 0.80] |
| Ter Hoeve, 2018 [68] | 8 | 2.6 | 121 | 7.7 | 3 | 252 | 1.7% | 0.10 [-0.11, 0.32] |
| Tudor-Locke, 2004 [69] | 9,123 | 4,539 | 24 | 5,622 | 2,405 | 23 | 1.1% | 0.94 [0.34, 1.55] |
| Vallance, 2016 [125] | 5,923 | 3,109 | 41 | 6,885 | 3,576 | 37 | 1.3% | -0.29 [-0.73, 0.16] |
| Van Blarigan, 2019 [126] | 10,047 | 4,461 | 20 | 12,541 | 5,535 | 19 | 1.0% | -0.49 [-1.13, 0.15] |
| **Subtotal (95% CI)** | | | **1326** | | | **1380** | **39.2%** | **0.44 [0.29, 0.60]** |
| Heterogeneity: Tau² = 0.13; Chi² = 111.87, df = 32 (P < 0.00001); I² = 71% | | | | | | | | |
| Test for overall effect: Z = 5.55 (P < 0.00001) | | | | | | | | |
| **Total (95% CI)** | | | **6390** | | | **6073** | **100.0%** | **0.46 [0.36, 0.55]** |
| Heterogeneity: Tau² = 0.14; Chi² = 458.11, df = 76 (P < 0.00001); I² = 83% | | | | | | | | |
| Test for overall effect: Z = 8.98 (P < 0.00001) | | | | | | | | |
| Test for subgroup differences: Chi² = 11.50, df = 2 (P = 0.003), I² = 82.6% | | | | | | | | |



Subgroup Analysis by Population Type for The Outcome Moderate-to-vigorous Physical Activity

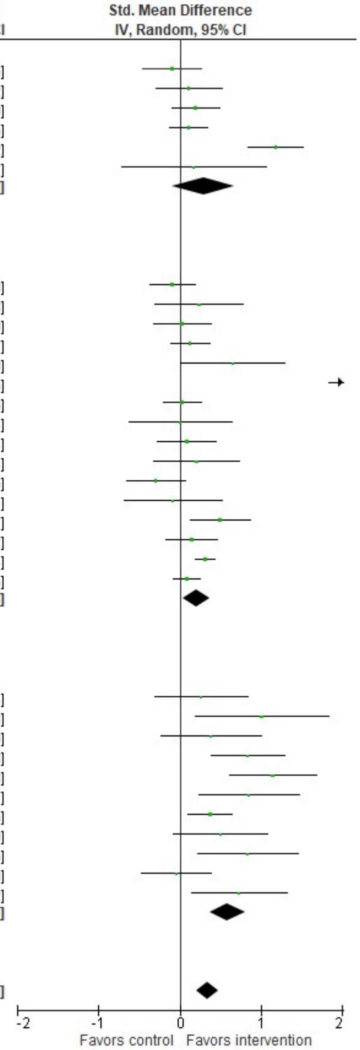
| Study or Subgroup | Intervention Mean | SD | Total | Control Mean | SD | Total | Weight | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **2.2.1 Healthy** | | | | | | | | |
| Ashton, 2017 [80] | 154.1 | 216.1 | 26 | 26.1 | 210.2 | 24 | 1.0% | 0.59 [0.02, 1.16] |
| Compernolle, 2015 [85] | 32.4 | 37.7 | 62 | 38.4 | 60 | 79 | 1.9% | -0.12 [-0.45, 0.22] |
| Dadaczynski, 2017 [87] | 310.9 | 129.4 | 80 | 291.81 | 114 | 64 | 1.9% | 0.15 [-0.17, 0.48] |
| Dishman, 2009 [91] | 10.6 | 15.5 | 564 | 7.9 | 15.1 | 265 | 2.8% | 0.18 [0.03, 0.32] |
| Fjeldsoe, 2015 [49] | -8 | 262.4 | 104 | -29.2 | 196 | 107 | 2.2% | 0.09 [-0.18, 0.36] |
| Hardeman, 2020 [98] | 77.3 | 36.5 | 417 | 76.7 | 35.4 | 442 | 2.8% | 0.02 [-0.12, 0.15] |
| Kendzor, 2017 [134] | 76 | 67 | 17 | 46 | 43 | 15 | 0.8% | 0.51 [-0.19, 1.22] |
| Kernot, 2019 [52] | 204.1 | 131 | 74 | 150 | 96.9 | 33 | 1.5% | 0.44 [0.03, 0.86] |
| Keyserling, 2008 [135] | 14 | 3.71 | 86 | 13 | 2.8 | 89 | 2.0% | 0.30 [0.01, 0.60] |
| Kim, 2018 [136] | 51.2 | 4.43 | 41 | 50.3 | 4.4 | 42 | 1.5% | 0.20 [-0.23, 0.63] |
| Murawski, 2019 [57] | 428.4 | 523.4 | 59 | 319.7 | 378.2 | 66 | 1.8% | 0.24 [-0.11, 0.59] |
| Sharp, 2016 [144] | 100 | 109 | 72 | 90 | 119 | 65 | 1.9% | 0.09 [-0.25, 0.42] |
| **Subtotal (95% CI)** | | | **1602** | | | **1291** | **22.0%** | **0.14 [0.06, 0.23]** |
| Heterogeneity: Tau² = 0.00; Chi² = 12.99, df = 11 (P = 0.29); I² = 15% | | | | | | | | |
| Test for overall effect: Z = 3.19 (P = 0.001) | | | | | | | | |
| **2.2.2 At-risk** | | | | | | | | |
| Aittasalo, 2006 [131] | 95 | 83.7 | 203 | 94 | 70.9 | 62 | 2.1% | 0.01 [-0.27, 0.30] |
| Barnes, 2015 [42] | 0.2 | 1.6 | 23 | -0.2 | 1.5 | 19 | 0.9% | 0.25 [-0.36, 0.86] |
| Barwais, 2013 [82] | 649 | 494.9 | 18 | 192 | 221.3 | 15 | 0.7% | 1.13 [0.38, 1.87] |
| Creel, 2016 [86] | 37 | 41.6 | 52 | 20.4 | 17.1 | 35 | 1.4% | 0.48 [0.05, 0.92] |
| Edney, 2020 [47] | 108.1 | 74.1 | 272 | 108.9 | 52.6 | 130 | 2.5% | -0.01 [-0.22, 0.20] |
| Fischer, 2019 [132] | 228.4 | 157.8 | 63 | 194 | 136 | 128 | 2.0% | 0.24 [-0.06, 0.54] |
| Fjeldsoe, 2010 [94] | 18.3 | 167.3 | 45 | 16.4 | 167.4 | 43 | 1.5% | 0.01 [-0.41, 0.43] |
| Harris, 2018 [75] | 374 | 179 | 778 | 310 | 168 | 456 | 2.9% | 0.37 [0.25, 0.48] |
| Katzmarzyk, 2011[104] | 16.3 | 17.3 | 20 | 16.2 | 17.1 | 23 | 1.0% | 0.01 [-0.59, 0.60] |
| King, 2008 [137] | 310.6 | 298.3 | 19 | 135 | 208.2 | 18 | 0.8% | 0.66 [0.00, 1.33] |
| Koizumi, 2009 [138] | 27.2 | 14.7 | 34 | 19 | 9.7 | 34 | 1.3% | 0.65 [0.16, 1.14] |
| Kolt, 2012 [76] | 121.2 | 184.4 | 130 | 111.4 | 110.6 | 123 | 2.3% | 0.06 [-0.18, 0.31] |
| Maher, 2015 [53] | 528 | 391 | 51 | 391 | 371 | 59 | 1.7% | 0.36 [-0.02, 0.74] |
| Mansi, 2015 [54] | 1,469 | 524 | 29 | 538 | 254 | 29 | 0.8% | 2.23 [1.57, 2.90] |
| Maselli, 2019 [55] | 134.4 | 83.4 | 11 | 185.5 | 126.3 | 21 | 0.7% | -0.44 [-1.18, 0.30] |
| Merom, 2007 [114] | 79 | 543.7 | 105 | 32.8 | 171.3 | 209 | 2.4% | 0.13 [-0.10, 0.37] |
| Pekmezi, 2017 [141] | 31.5 | 58.9 | 39 | 20.8 | 38.2 | 37 | 1.4% | 0.21 [-0.24, 0.66] |
| Prestwich, 2009 [142] | 1.3 | 1.16 | 60 | 0.97 | 1 | 94 | 1.9% | 0.31 [-0.02, 0.63] |
| Ribeiro, 2014 [63] | 473.5 | 913.2 | 101 | 58.5 | 484.3 | 94 | 2.1% | 0.56 [0.27, 0.85] |
| Samuels, 2011 [143] | 34.8 | 13.6 | 12 | 26.5 | 16.1 | 25 | 0.8% | 0.53 [-0.17, 1.23] |
| Simons, 2018 [64] | 27.3 | 26.7 | 55 | 30.1 | 31 | 63 | 1.7% | -0.10 [-0.46, 0.27] |
| Vandelanotte, 2018 [146] | 148.8 | 181.1 | 78 | 79.8 | 77.1 | 46 | 1.7% | 0.45 [0.08, 0.82] |
| Van Hoye, 2018 [72] | 17.8 | 49.6 | 157 | 9.6 | 34 | 50 | 1.9% | 0.18 [-0.14, 0.49] |
| Wijsman, 2013 [147] | 11.1 | 21.8 | 108 | -0.1 | 15.4 | 105 | 2.1% | 0.59 [0.32, 0.86] |
| Yates, 2017 [78] | -4.1 | 19 | 287 | -5.9 | 17.7 | 272 | 2.7% | 0.10 [-0.07, 0.26] |
| **Subtotal (95% CI)** | | | **2750** | | | **2190** | **41.4%** | **0.30 [0.18, 0.43]** |
| Heterogeneity: Tau² = 0.06; Chi² = 84.90, df = 24 (P < 0.00001); I² = 72% | | | | | | | | |
| Test for overall effect: Z = 4.80 (P < 0.00001) | | | | | | | | |
| **2.2.3 Sick** | | | | | | | | |
| Alsaleh, 2016 [79] | 237.9 | 121.1 | 66 | 108.9 | 198 | 79 | 1.8% | 0.77 [0.43, 1.11] |
| Cadmus-Bertram, 2019 [84] | 230 | 120 | 24 | 143 | 107 | 23 | 1.0% | 0.75 [0.16, 1.35] |
| Cruz, 2016 [44] | 57.8 | 32.8 | 13 | 26.7 | 19.6 | 13 | 0.6% | 1.11 [0.28, 1.95] |
| DeGreef, 2010 [73] | 44 | 38 | 20 | 24 | 29 | 21 | 0.9% | 0.58 [-0.04, 1.21] |
| DeGreef, 2011[89] | 23 | 20 | 60 | 20 | 25 | 32 | 1.5% | 0.14 [-0.29, 0.57] |
| DeGreef, 2011a [74] | 82 | 84 | 43 | 27 | 39 | 24 | 1.2% | 0.76 [0.24, 1.28] |
| Demeyer, 2017 [90] | 8 | 21.1 | 140 | -3 | 17.5 | 140 | 2.3% | 0.57 [0.33, 0.80] |
| Eakin, 2014 [46] | 163.6 | 155.6 | 119 | 114.6 | 108.6 | 127 | 2.3% | 0.37 [0.11, 0.62] |
| Golsteijn, 2018 [133] | 331 | 234 | 208 | 301 | 219 | 211 | 2.6% | 0.13 [-0.06, 0.32] |
| James, 2015 [51] | 33.9 | 145.3 | 57 | 9.6 | 128.6 | 52 | 1.7% | 0.18 [-0.20, 0.55] |
| Li, 2017 [139] | 64.2 | 70.5 | 17 | 56 | 60.1 | 17 | 0.8% | 0.12 [-0.55, 0.80] |
| Li, 2020 [107] | 44.7 | 41.2 | 55 | 31.6 | 32.4 | 57 | 1.7% | 0.35 [-0.02, 0.73] |
| Lynch, 2019 [109] | 249.6 | 159.7 | 40 | 170.1 | 127.9 | 40 | 1.4% | 0.54 [0.10, 0.99] |
| Maxwell Smith, 2019 [140] | 312 | 229.3 | 34 | 240 | 177.7 | 33 | 1.3% | 0.35 [-0.14, 0.83] |
| Melville, 2015 [112] | 3 | 2.6 | 42 | 3.1 | 2.1 | 40 | 1.4% | -0.04 [-0.47, 0.39] |
| Nolan, 2017 [59] | 11 | 68.8 | 63 | 11 | 58.8 | 59 | 1.8% | 0.00 [-0.36, 0.36] |
| Pinto, 2013 [61] | 214 | 147.3 | 19 | 97 | 148 | 24 | 0.9% | 0.78 [0.15, 1.40] |
| Pinto, 2015 [62] | 70.3 | 65.9 | 36 | 16.5 | 31.9 | 32 | 1.2% | 1.01 [0.50, 1.52] |
| Pope, 2018 [117] | 34.2 | 18.7 | 12 | 37.8 | 20.4 | 8 | 0.5% | -0.18 [-1.07, 0.72] |
| Ter Hoeve, 2018 [68] | 6.7 | 3 | 121 | 6.5 | 3.1 | 252 | 2.4% | 0.07 [-0.15, 0.28] |
| Unick, 2012 [145] | 133 | 217 | 11 | 44.8 | 124.6 | 12 | 0.6% | 0.49 [-0.35, 1.32] |
| Vallance, 2008 [70] | 213 | 173 | 172 | 180 | 141 | 166 | 2.5% | 0.21 [-0.01, 0.42] |
| Vallance, 2016 [125] | 143 | 132 | 41 | 146 | 155 | 37 | 1.4% | -0.02 [-0.47, 0.42] |
| Van Blarigan, 2019 [126] | 46.6 | 48.4 | 20 | 54.5 | 24.9 | 19 | 0.9% | -0.20 [-0.83, 0.43] |
| Van der Weegen, 2015 [71] | 48.2 | 23.8 | 65 | 42.9 | 25.8 | 134 | 2.0% | 0.21 [-0.09, 0.51] |
| **Subtotal (95% CI)** | | | **1498** | | | **1652** | **36.6%** | **0.33 [0.21, 0.45]** |
| Heterogeneity: Tau² = 0.04; Chi² = 53.53, df = 24 (P = 0.0005); I² = 55% | | | | | | | | |
| Test for overall effect: Z = 5.47 (P < 0.00001) | | | | | | | | |
| **Total (95% CI)** | | | **5850** | | | **5133** | **100.0%** | **0.28 [0.21, 0.35]** |
| Heterogeneity: Tau² = 0.04; Chi² = 162.08, df = 61 (P < 0.00001); I² = 62% | | | | | | | | |
| Test for overall effect: Z = 7.82 (P < 0.00001) | | | | | | | | |
| Test for subgroup differences: Chi² = 7.73, df = 2 (P = 0.02), I² = 74.1% | | | | | | | | |

Std. Mean Difference IV, Random, 95% CI



Subgroup Analysis by Population Type for The Outcome Total Physical Activity

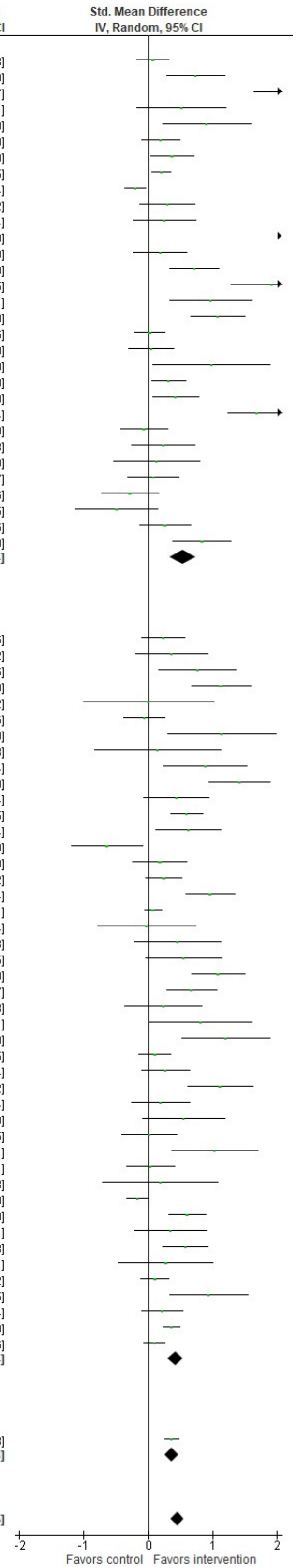
| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **2.3.1 Healthy** | | | | | | | | |
| Compernolle, 2015 [85] | 81.6 | 78 | 52 | 90.6 | 100.2 | 71 | 3.4% | -0.10 [-0.46, 0.26] |
| Kernot, 2019 [52] | 298.9 | 97.7 | 74 | 288.8 | 68.586 | 33 | 3.2% | 0.11 [-0.30, 0.52] |
| Keyserling, 2008 [135] | 29.8 | 6.5 | 86 | 28.6 | 5.69 | 90 | 3.7% | 0.20 [-0.10, 0.49] |
| Lane, 2015 [149] | 84.9 | 50.3 | 125 | 80.3 | 32.8 | 159 | 4.0% | 0.11 [-0.12, 0.35] |
| Suggs, 2013 [66] | 49.5 | 39.6 | 79 | 5.8 | 33.9 | 79 | 3.5% | 1.18 [0.84, 1.52] |
| Thorsteinsen, 2014 [153] | 574 | 435 | 12 | 502 | 323 | 8 | 1.5% | 0.17 [-0.72, 1.07] |
| **Subtotal (95% CI)** | | | **428** | | | **440** | **19.4%** | **0.29 [-0.10, 0.67]** |
| Heterogeneity: Tau² = 0.19; Chi² = 34.42, df = 5 (P < 0.00001); I² = 85% | | | | | | | | |
| Test for overall effect: Z = 1.47 (P = 0.14) | | | | | | | | |
| **2.3.2 At-risk** | | | | | | | | |
| Aittasalo, 2006 [131] | 509.5 | 479.4 | 203 | 555 | 519.7 | 62 | 3.8% | -0.09 [-0.38, 0.19] |
| Carr, 2013 [148] | 171.4 | 234.5 | 25 | 121.8 | 174.6 | 28 | 2.6% | 0.24 [-0.30, 0.78] |
| Gill, 2019 [96] | 2.13 | 25.8 | 59 | 1.37 | 25.7 | 59 | 3.4% | 0.03 [-0.33, 0.39] |
| Kolt, 2012 [76] | 168.9 | 190.5 | 130 | 146.3 | 163.5 | 123 | 4.0% | 0.13 [-0.12, 0.37] |
| Lyons, 2017 [110] | 117 | 121 | 20 | 58 | 33 | 20 | 2.2% | 0.65 [0.01, 1.29] |
| Mansi, 2015 [54] | 1,035 | 444 | 29 | 188 | 135 | 29 | 2.0% | 2.55 [1.84, 3.25] |
| Merom, 2007 [114] | 91 | 243.1 | 105 | 84.2 | 231.8 | 209 | 4.0% | 0.03 [-0.21, 0.26] |
| Müller, 2016 [56] | 383.4 | 843.4 | 18 | 377.4 | 842.6 | 21 | 2.3% | 0.01 [-0.62, 0.64] |
| Prestwich, 2010 [118] | 3 | 1.8 | 88 | 2.28 | 14.25 | 46 | 3.5% | 0.08 [-0.27, 0.44] |
| Schwerdtfeger, 2012 [154] | 738.6 | 245.7 | 21 | 690.5 | 217.1 | 41 | 2.7% | 0.21 [-0.32, 0.74] |
| Simons, 2018 [64] | 317 | 135 | 55 | 359.5 | 146 | 63 | 3.4% | -0.30 [-0.66, 0.06] |
| Sugden, 2008 [152] | 108.7 | 54.7 | 26 | 113.8 | 62.337 | 18 | 2.4% | -0.09 [-0.69, 0.51] |
| Vandelanotte, 2018 [146] | 387.7 | 377.7 | 78 | 230 | 164.1 | 46 | 3.4% | 0.50 [0.13, 0.87] |
| Van Hoye, 2018 [72] | 1.8 | 8.3 | 157 | 0.6 | 7 | 50 | 3.6% | 0.15 [-0.17, 0.47] |
| Wyke, 2019 [77] | 3,717 | 3,307 | 499 | 2,741 | 2,951 | 505 | 4.4% | 0.31 [0.19, 0.44] |
| Yates, 2017 [78] | -32.3 | 94.6 | 287 | -40.3 | 85 | 272 | 4.3% | 0.09 [-0.08, 0.25] |
| **Subtotal (95% CI)** | | | **1800** | | | **1592** | **52.0%** | **0.21 [0.04, 0.38]** |
| Heterogeneity: Tau² = 0.08; Chi² = 67.84, df = 15 (P < 0.00001); I² = 78% | | | | | | | | |
| Test for overall effect: Z = 2.40 (P = 0.02) | | | | | | | | |
| **2.3.3 Sick** | | | | | | | | |
| Cadmus-Bertram, 2019 [84] | 1,463 | 489 | 24 | 1,343 | 395 | 23 | 2.5% | 0.26 [-0.31, 0.84] |
| Cruz, 2016 [44] | 279.5 | 74 | 13 | 212 | 53.9 | 13 | 1.7% | 1.01 [0.19, 1.83] |
| DeGreef, 2010 [73] | 301 | 106 | 20 | 260 | 104 | 21 | 2.3% | 0.38 [-0.24, 1.00] |
| DeGreef, 2011[89] | 93 | 66 | 60 | 40 | 56 | 32 | 3.0% | 0.84 [0.39, 1.28] |
| DeGreef, 2011a [74] | 176 | 108 | 43 | 65 | 68 | 24 | 2.6% | 1.15 [0.61, 1.68] |
| Dlugonski, 2012 [45] | 28.2 | 15.6 | 22 | 15.4 | 13.9 | 23 | 2.3% | 0.85 [0.24, 1.47] |
| Furber, 2010 [50] | 366.5 | 270.8 | 97 | 270.9 | 244.4 | 107 | 3.8% | 0.37 [0.09, 0.65] |
| Mailey, 2010 [150] | 299,792 | 102,800 | 24 | 251,625 | 83,080 | 23 | 2.5% | 0.51 [-0.08, 1.09] |
| Martin, 2015 [111] | 10.5 | 21.1 | 32 | -8 | 23 | 16 | 2.3% | 0.84 [0.21, 1.46] |
| Melville, 2015 [112] | 33.5 | 10 | 42 | 34 | 12 | 40 | 3.1% | -0.04 [-0.48, 0.39] |
| Motl, 2011 [151] | 24.7 | 18.8 | 23 | 12.4 | 14.2 | 25 | 2.4% | 0.73 [0.14, 1.32] |
| **Subtotal (95% CI)** | | | **400** | | | **347** | **28.5%** | **0.59 [0.36, 0.81]** |
| Heterogeneity: Tau² = 0.07; Chi² = 19.68, df = 10 (P = 0.03); I² = 49% | | | | | | | | |
| Test for overall effect: Z = 5.18 (P < 0.00001) | | | | | | | | |
| **Total (95% CI)** | | | **2628** | | | **2379** | **100.0%** | **0.34 [0.20, 0.47]** |
| Heterogeneity: Tau² = 0.10; Chi² = 140.07, df = 32 (P < 0.00001); I² = 77% | | | | | | | | |
| Test for overall effect: Z = 4.97 (P < 0.00001) | | | | | | | | |
| Test for subgroup differences: Chi² = 7.14, df = 2 (P = 0.03), I² = 72.0% | | | | | | | | |



## **Appendix 8:** Subgroup Analysis by Intervention Design

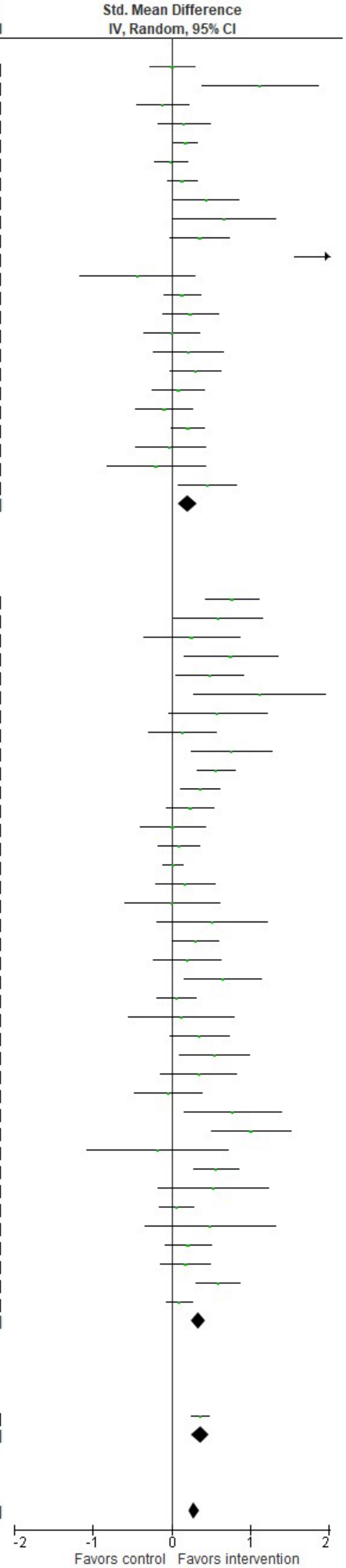
Subgroup Analysis by Intervention Design for The Outcome Walking

| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **4.1.1 Scalable (no human interaction)** | | | | | | | | |
| Aittasalo, 2012 [41] | 457 | 306 | 123 | 431 | 403 | 118 | 1.7% | 0.07 [-0.18, 0.33] |
| Baker, 2008 [81] | 9,977 | 4,669 | 39 | 7,078 | 2,911 | 40 | 1.3% | 0.74 [0.28, 1.20] |
| Barwais, 2013 [82] | 1,625 | 553.8 | 18 | 483 | 175.9 | 15 | 0.7% | 2.61 [1.65, 3.57] |
| Butler, 2004 [83] | 56,396 | 3,315 | 17 | 53,666 | 6,543 | 16 | 1.0% | 0.52 [-0.18, 1.21] |
| Coelho, 2018 [43] | 8,853 | 3,320 | 20 | 6,248 | 2,030 | 17 | 1.0% | 0.91 [0.23, 1.59] |
| Compernolle, 2015 [85] | 9,484 | 4,875 | 86 | 8,589 | 4,380 | 93 | 1.6% | 0.19 [-0.10, 0.49] |
| Dadaczynski, 2017 [87] | 526.5 | 240.6 | 80 | 442.9 | 200.4 | 64 | 1.5% | 0.37 [0.04, 0.70] |
| Dishman, 2009 [91] | 14.8 | 17.3 | 564 | 11.2 | 17.3 | 265 | 1.8% | 0.21 [0.06, 0.35] |
| Finkelstein, 2016 [48] | -480 | 2,568 | 201 | 43.7 | 2,640.7 | 599 | 1.8% | -0.20 [-0.36, -0.04] |
| Gell, 2015 [95] | 6,867.7 | 2,227 | 41 | 6,189 | 2,297 | 46 | 1.4% | 0.30 [-0.13, 0.72] |
| Glynn, 2014 [97] | 5,855 | 4,264 | 31 | 4,859 | 3,474 | 35 | 1.3% | 0.25 [-0.23, 0.74] |
| Hultquist, 2005 [102] | 10,159 | 292 | 27 | 8,270 | 354 | 31 | 0.5% | 5.70 [4.51, 6.90] |
| Kernot, 2019 [52] | 276.3 | 221.7 | 74 | 235 | 218.9 | 33 | 1.4% | 0.19 [-0.23, 0.60] |
| Maher, 2015 [53] | 332 | 289 | 51 | 160 | 185 | 59 | 1.4% | 0.72 [0.33, 1.10] |
| Mansi, 2015 [54] | 9,792 | 2,053 | 29 | 6,551 | 1,154 | 29 | 1.1% | 1.92 [1.29, 2.55] |
| Martin, 2015 [111] | 1,067 | 2,098 | 32 | -1,042 | 2,202 | 16 | 1.1% | 0.97 [0.34, 1.61] |
| Mendoza, 2015 [113] | 3,080 | 3,254 | 52 | 138 | 1,950 | 50 | 1.4% | 1.08 [0.67, 1.50] |
| Merom, 2007 [114] | 52 | 107.2 | 105 | 48.9 | 126.8 | 209 | 1.7% | 0.03 [-0.21, 0.26] |
| Nolan, 2017 [59] | 272 | 2,275.9 | 63 | 155 | 2,557.1 | 59 | 1.5% | 0.05 [-0.31, 0.40] |
| Paul, 2016 [115] | 5,791 | 2,952 | 15 | 2,947 | 2,399 | 8 | 0.7% | 0.99 [0.07, 1.90] |
| Poirier, 2016 [116] | 5,411 | 2,277 | 107 | 4,751 | 1,834 | 110 | 1.6% | 0.32 [0.05, 0.59] |
| Prestwich, 2010 [118] | 1.98 | 1.9 | 88 | 1.2 | 1.6 | 46 | 1.5% | 0.43 [0.07, 0.79] |
| Rowley, 2019 [121] | 9,015 | 2,842 | 97 | 4,654 | 1,447 | 32 | 1.3% | 1.69 [1.23, 2.14] |
| Simons, 2018 [64] | 7,741 | 4,553 | 55 | 8,061 | 5,112 | 63 | 1.5% | -0.07 [-0.43, 0.30] |
| Spence, 2009 [122] | 690 | 450.2 | 32 | 592.5 | 368.9 | 31 | 1.3% | 0.23 [-0.26, 0.73] |
| Talbot, 2003 [67] | 4,337 | 2,903 | 17 | 3,972 | 2,563 | 17 | 1.0% | 0.13 [-0.54, 0.80] |
| Thorndike, 2014 [124] | 7,886 | 3,622 | 50 | 7,600 | 3,492 | 49 | 1.4% | 0.08 [-0.31, 0.47] |
| Vallance, 2016 [125] | 5,923 | 3,109 | 41 | 6,885 | 3,576 | 37 | 1.3% | -0.29 [-0.73, 0.16] |
| Van Blarigan, 2019 [126] | 10,047 | 4,461 | 20 | 12,541 | 5,535 | 19 | 1.0% | -0.49 [-1.13, 0.15] |
| Warren, 2014 [127] | 8,371 | 3,069 | 37 | 7,576 | 2,993 | 69 | 1.4% | 0.26 [-0.14, 0.66] |
| Yamada, 2012 [128] | 3,726 | 1,607 | 40 | 2,267 | 1,837 | 42 | 1.3% | 0.84 [0.38, 1.29] |
| **Subtotal (95% CI)** | | | **2252** | | | **2317** | **40.5%** | **0.54 [0.34, 0.74]** |
| Heterogeneity: $Tau^2 = 0.26$; $Chi^2 = 266.67$, df = 30 ($P < 0.00001$); $I^2 = 89\%$ | | | | | | | | |
| Test for overall effect: $Z = 5.27$ ($P < 0.00001$) | | | | | | | | |
| **4.1.2 Non-scalable (mHealth combined with human interactions)** | | | | | | | | |
| Alsaleh, 2016 [79] | 36.7 | 109.8 | 66 | 16.5 | 58.73 | 79 | 1.5% | 0.23 [-0.09, 0.56] |
| Ashton, 2017 [80] | 1,588.2 | 2,740.6 | 26 | 575.4 | 2,735.1 | 24 | 1.2% | 0.36 [-0.20, 0.92] |
| Cadmus-Bertram, 2019 [84] | 6,697 | 2,878 | 24 | 4,853 | 1,703 | 23 | 1.1% | 0.76 [0.17, 1.36] |
| Creel, 2016 [86] | 6,084 | 873 | 52 | 5,253 | 421 | 35 | 1.3% | 1.13 [0.67, 1.60] |
| Croteau, 2004 [129] | 2,419 | 6,074.6 | 7 | 2,320 | 47,312 | 8 | 0.6% | 0.00 [-1.01, 1.02] |
| Croteau, 2007 [130] | 6,180 | 3,530 | 79 | 6,378 | 2,994 | 68 | 1.5% | -0.06 [-0.38, 0.26] |
| Cruz, 2016 [44] | 10,440 | 4,012 | 13 | 6,430 | 2,613 | 13 | 0.8% | 1.15 [0.31, 1.99] |
| De Blok, 2006 [88] | 3,927 | 2,617 | 8 | 3,554 | 2,019.6 | 8 | 0.7% | 0.15 [-0.83, 1.13] |
| DeGreef, 2010 [73] | 9,601 | 5,002 | 20 | 5,538 | 3,877 | 21 | 1.0% | 0.89 [0.25, 1.54] |
| DeGreef, 2011[89] | 7,703 | 2,729 | 60 | 3,883 | 2,537 | 32 | 1.3% | 1.42 [0.94, 1.90] |
| DeGreef, 2011a [74] | 6,771 | 3,889 | 43 | 5,173 | 3,094 | 24 | 1.3% | 0.44 [-0.07, 0.94] |
| Demeyer, 2017 [90] | 7 | 32.5 | 129 | -10 | 23.5 | 132 | 1.7% | 0.60 [0.35, 0.85] |
| Duru, 2010 [92] | 9,883 | 14,014 | 34 | 2,426 | 8,429 | 28 | 1.2% | 0.62 [0.11, 1.14] |
| Engel, 2006 [93] | 150 | 117 | 24 | 216 | 87 | 30 | 1.2% | -0.64 [-1.19, -0.09] |
| Fjeldsoe, 2010 [94] | 16.7 | 89.4 | 45 | 0.3 | 89.4 | 43 | 1.4% | 0.18 [-0.24, 0.60] |
| Furber, 2010 [50] | 249.9 | 196 | 97 | 202.6 | 189.5 | 107 | 1.6% | 0.24 [-0.03, 0.52] |
| Gill, 2019 [96] | 1,646 | 3,302 | 59 | -1,485 | 3,171.5 | 59 | 1.5% | 0.96 [0.58, 1.34] |
| Hardeman, 2020 [98] | 8,419 | 3,224 | 417 | 8,191 | 3,003 | 442 | 1.8% | 0.07 [-0.06, 0.21] |
| Hornikx, 2015 [99] | 984 | 1,208 | 12 | 1,013 | 1,275 | 15 | 0.9% | -0.02 [-0.78, 0.74] |
| Hospes, 2009 [100] | 7,872 | 3,962 | 18 | 6,172 | 3,194 | 17 | 1.0% | 0.46 [-0.21, 1.13] |
| Houle, 2011 [101] | 9,850 | 3,282 | 23 | 7,970 | 3,433 | 22 | 1.1% | 0.55 [-0.05, 1.15] |
| Izawa, 2012 [103] | 8,609.6 | 3,064.5 | 52 | 5,512.9 | 2,571.8 | 51 | 1.4% | 1.09 [0.67, 1.50] |
| James, 2015 [51] | 800.8 | 2,820.6 | 57 | -1,294 | 3,304.4 | 52 | 1.4% | 0.68 [0.29, 1.07] |
| Katzmarzyk, 2011[104] | 7,248 | 2,737 | 20 | 6,637 | 2,418 | 23 | 1.1% | 0.23 [-0.37, 0.83] |
| Kawagoshi, 2015 [105] | 51.3 | 63.7 | 12 | 12.3 | 25.5 | 15 | 0.8% | 0.82 [0.02, 1.61] |
| King, 2013 [106] | 253.5 | 248.7 | 20 | 26.8 | 67 | 19 | 1.0% | 1.21 [0.52, 1.89] |
| Kolt, 2012 [76] | 107.4 | 164.9 | 130 | 92.2 | 120.5 | 123 | 1.7% | 0.10 [-0.14, 0.35] |
| Li, 2020 [107] | 6,673 | 3,462 | 55 | 5,819 | 2,860 | 57 | 1.5% | 0.27 [-0.10, 0.64] |
| Long, 2013 [108] | 65,983 | 18,069 | 38 | 47,596 | 13,900 | 33 | 1.3% | 1.12 [0.61, 1.62] |
| Lynch, 2019 [109] | 8,193 | 3,301 | 37 | 7,539 | 3,404 | 40 | 1.3% | 0.19 [-0.26, 0.64] |
| Lyons, 2017 [110] | 6,193 | 3,183 | 20 | 4,586 | 2,476 | 20 | 1.1% | 0.55 [-0.08, 1.19] |
| Melville, 2015 [112] | 4,823 | 2,059 | 42 | 4,784 | 2,613 | 40 | 1.4% | 0.02 [-0.42, 0.45] |
| Mutrie, 2012 [58] | 9,351 | 2,017 | 20 | 7,138 | 2,169 | 19 | 1.0% | 1.04 [0.36, 1.71] |
| Oliveira, 2019 [60] | 7,507 | 3,077 | 54 | 7,401 | 2,841 | 55 | 1.5% | 0.04 [-0.34, 0.41] |
| Pope, 2018 [117] | 5,175 | 2,308 | 12 | 4,746 | 2,045 | 8 | 0.7% | 0.19 [-0.71, 1.08] |
| Reijonsaari, 2012 [119] | 2,047 | 1,650 | 264 | 2,338 | 1,762 | 257 | 1.8% | -0.17 [-0.34, 0.00] |
| Ribeiro, 2014 [63] | 969.7 | 1,464.1 | 101 | -181.5 | 2,284 | 94 | 1.6% | 0.60 [0.32, 0.89] |
| Roos, 2014 [120] | 3,810 | 10,993.5 | 29 | 817.3 | 2,349.2 | 22 | 1.2% | 0.35 [-0.21, 0.91] |
| Stacey, 2016 [65] | 10,849 | 5,127 | 75 | 8,014 | 4,568 | 58 | 1.5% | 0.58 [0.23, 0.93] |
| Tabak, 2014 [123] | 5,603 | 3,475.8 | 13 | 4,617 | 3,460 | 16 | 0.9% | 0.28 [-0.46, 1.01] |
| Ter Hoeve, 2018 [68] | 8 | 2.6 | 121 | 7.7 | 3 | 252 | 1.7% | 0.10 [-0.11, 0.32] |
| Tudor-Locke, 2004 [69] | 9,123 | 4,539 | 24 | 5,622 | 2,405 | 23 | 1.1% | 0.94 [0.34, 1.55] |
| Van Hoye, 2018 [72] | 13.2 | 29.6 | 157 | 6.7 | 26 | 50 | 1.6% | 0.23 [-0.09, 0.54] |
| Wyke, 2019 [77] | 9,801 | 3,730 | 464 | 8,518 | 3,254 | 471 | 1.8% | 0.37 [0.24, 0.50] |
| Yates, 2017 [78] | -486 | 2,252 | 287 | -690 | 2,041 | 272 | 1.8% | 0.09 [-0.07, 0.26] |
| **Subtotal (95% CI)** | | | **3360** | | | **3300** | **57.7%** | **0.42 [0.31, 0.54]** |
| Heterogeneity: $Tau^2 = 0.10$; $Chi^2 = 188.72$, df = 44 ($P < 0.00001$); $I^2 = 77\%$ | | | | | | | | |
| Test for overall effect: $Z = 7.19$ ($P < 0.00001$) | | | | | | | | |
| **4.1.3 Combined intervention** | | | | | | | | |
| Harris, 2018 [75] | 8,306 | 3,140 | 778 | 7,198 | 2,809 | 456 | 1.8% | 0.37 [0.25, 0.48] |
| **Subtotal (95% CI)** | | | **778** | | | **456** | **1.8%** | **0.37 [0.25, 0.48]** |
| Heterogeneity: Not applicable | | | | | | | | |
| Test for overall effect: $Z = 6.16$ ($P < 0.00001$) | | | | | | | | |
| **Total (95% CI)** | | | **6390** | | | **6073** | **100.0%** | **0.45 [0.36, 0.55]** |
| Heterogeneity: $Tau^2 = 0.14$; $Chi^2 = 458.02$, df = 76 ($P < 0.00001$); $I^2 = 83\%$ | | | | | | | | |
| Test for overall effect: $Z = 8.98$ ($P < 0.00001$) | | | | | | | | |
| Test for subgroup differences: $Chi^2 = 2.11$, df = 2 ($P = 0.35$), $I^2 = 5.2\%$ | | | | | | | | |



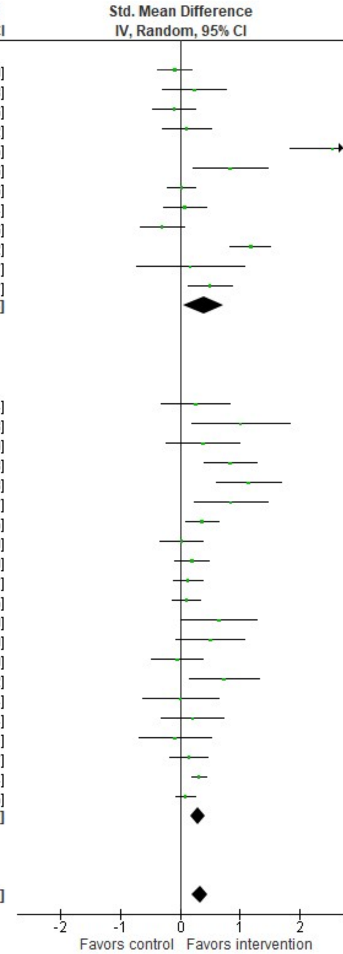
Subgroup Analysis by Intervention Design for The Outcome Moderate-to-vigorous Physical Activity

| | Intervention | | | Control | | | | Std. Mean Difference |
|---|---|---|---|---|---|---|---|---|
| **Study or Subgroup** | **Mean** | **SD** | **Total** | **Mean** | **SD** | **Total** | **Weight** | **IV, Random, 95% CI** |
| **4.2.1 Scalable (no human interaction)** | | | | | | | | |
| Aittasalo, 2006 [131] | 95 | 83.7 | 203 | 94 | 70.9 | 62 | 2.1% | 0.01 [-0.27, 0.30] |
| Barwais, 2013 [82] | 649 | 494.9 | 18 | 192 | 221.3 | 15 | 0.7% | 1.13 [0.38, 1.87] |
| Compernolle, 2015 [85] | 32.4 | 37.7 | 62 | 38.4 | 60 | 79 | 1.9% | -0.12 [-0.45, 0.22] |
| Dadaczynski, 2017 [87] | 310.9 | 129.4 | 80 | 291.8 | 114 | 64 | 1.9% | 0.15 [-0.17, 0.48] |
| Dishman, 2009 [91] | 10.6 | 15.5 | 564 | 7.9 | 15.1 | 265 | 2.8% | 0.18 [0.03, 0.32] |
| Edney, 2020 [47] | 108.1 | 74.1 | 272 | 108.9 | 52.6 | 130 | 2.5% | -0.01 [-0.22, 0.20] |
| Golsteijn, 2018 [133] | 331 | 234 | 208 | 301 | 219 | 211 | 2.6% | 0.13 [-0.06, 0.32] |
| Kernot, 2019 [52] | 204.1 | 131 | 74 | 150 | 96.9 | 33 | 1.5% | 0.44 [0.03, 0.86] |
| King, 2008 [137] | 310.6 | 298.3 | 19 | 135 | 208.2 | 18 | 0.8% | 0.66 [0.00, 1.33] |
| Maher, 2015 [53] | 528 | 391 | 51 | 391 | 371 | 59 | 1.7% | 0.36 [-0.02, 0.74] |
| Mansi, 2015 [54] | 1,469 | 524 | 29 | 538 | 254 | 29 | 0.8% | 2.23 [1.57, 2.90] |
| Maselli, 2019 [55] | 134.4 | 83.4 | 11 | 185.5 | 126.3 | 21 | 0.7% | -0.44 [-1.18, 0.30] |
| Merom, 2007 [114] | 79 | 543.7 | 105 | 32.8 | 171.3 | 209 | 2.4% | 0.13 [-0.10, 0.37] |
| Murawski, 2019 [57] | 428.4 | 523.4 | 59 | 319.7 | 378.2 | 66 | 1.8% | 0.24 [-0.11, 0.59] |
| Nolan, 2017 [59] | 11 | 68.8 | 63 | 11 | 58.8 | 59 | 1.8% | 0.00 [-0.36, 0.36] |
| Pekmezi, 2017 [141] | 31.5 | 58.9 | 39 | 20.8 | 38.2 | 37 | 1.4% | 0.21 [-0.24, 0.66] |
| Prestwich, 2009 [142] | 1.3 | 1.2 | 60 | 0.97 | 1 | 94 | 1.9% | 0.30 [-0.02, 0.63] |
| Sharp, 2016 [144] | 100 | 109 | 72 | 90 | 119 | 65 | 1.9% | 0.09 [-0.25, 0.42] |
| Simons, 2018 [64] | 27.3 | 26.7 | 55 | 30.1 | 31 | 63 | 1.7% | -0.10 [-0.46, 0.27] |
| Vallance, 2008 [70] | 213 | 173 | 172 | 180 | 141 | 166 | 2.5% | 0.21 [-0.01, 0.42] |
| Vallance, 2016 [125] | 143 | 132 | 41 | 146 | 155 | 37 | 1.4% | -0.02 [-0.47, 0.42] |
| Van Blarigan, 2019 [126] | 46.6 | 48.4 | 20 | 54.5 | 24.9 | 19 | 0.9% | -0.20 [-0.83, 0.43] |
| Vandelanotte, 2018 [146] | 148.8 | 181.1 | 78 | 79.8 | 77.1 | 46 | 1.7% | 0.45 [0.08, 0.82] |
| **Subtotal (95% CI)** | | | **2355** | | | **1847** | **39.2%** | **0.20 [0.08, 0.32]** |
| Heterogeneity: Tau² = 0.05; Chi² = 66.16, df = 22 (P < 0.00001); I² = 67% | | | | | | | | |
| Test for overall effect: Z = 3.30 (P = 0.0010) | | | | | | | | |
| **4.2.2 Non-scalable (mHealth combined with human interactions)** | | | | | | | | |
| Alsaleh, 2016 [79] | 237.9 | 121.1 | 66 | 108.9 | 198 | 79 | 1.8% | 0.77 [0.43, 1.11] |
| Ashton, 2017 [80] | 154.1 | 216.1 | 26 | 26.1 | 210.2 | 24 | 1.0% | 0.59 [0.02, 1.16] |
| Barnes, 2015 [42] | 0.2 | 1.6 | 23 | -0.2 | 1.5 | 19 | 0.9% | 0.25 [-0.36, 0.86] |
| Cadmus-Bertram, 2019 [84] | 230 | 120 | 24 | 143 | 107 | 23 | 1.0% | 0.75 [0.16, 1.35] |
| Creel, 2016 [86] | 37 | 41.6 | 52 | 20.4 | 17.1 | 35 | 1.4% | 0.48 [0.05, 0.92] |
| Cruz, 2016 [44] | 57.8 | 32.8 | 13 | 26.7 | 19.6 | 13 | 0.6% | 1.11 [0.28, 1.95] |
| DeGreef, 2010 [73] | 44 | 38 | 20 | 24 | 29 | 21 | 0.9% | 0.58 [-0.04, 1.21] |
| DeGreef, 2011[89] | 23 | 20 | 60 | 20 | 25 | 32 | 1.5% | 0.14 [-0.29, 0.57] |
| DeGreef, 2011a [74] | 82 | 84 | 43 | 27 | 39 | 24 | 1.2% | 0.76 [0.24, 1.28] |
| Demeyer, 2017 [90] | 8 | 21.1 | 140 | -3 | 17.5 | 140 | 2.3% | 0.57 [0.33, 0.80] |
| Eakin, 2014 [46] | 163.6 | 155.6 | 119 | 114.6 | 108.6 | 127 | 2.3% | 0.37 [0.11, 0.62] |
| Fischer, 2019 [132] | 228.4 | 157.8 | 63 | 194 | 136 | 128 | 2.0% | 0.24 [-0.06, 0.54] |
| Fjeldsoe, 2010 [94] | 18.3 | 167.3 | 45 | 16.4 | 167.4 | 43 | 1.5% | 0.01 [-0.41, 0.43] |
| Fjeldsoe, 2015 [49] | -8 | 262.4 | 104 | -29.2 | 196 | 107 | 2.2% | 0.09 [-0.18, 0.36] |
| Hardeman, 2020 [98] | 77.3 | 36.5 | 417 | 76.7 | 35.4 | 442 | 2.8% | 0.02 [-0.12, 0.15] |
| James, 2015 [51] | 33.9 | 145.3 | 57 | 9.6 | 128.6 | 52 | 1.7% | 0.18 [-0.20, 0.55] |
| Katzmarzyk, 2011[104] | 16.3 | 17.3 | 20 | 16.2 | 17.1 | 23 | 1.0% | 0.01 [-0.59, 0.60] |
| Kendzor, 2017 [134] | 76 | 67 | 17 | 46 | 43 | 15 | 0.8% | 0.51 [-0.19, 1.22] |
| Keyserling, 2008 [135] | 14 | 3.7 | 86 | 13 | 2.8 | 89 | 2.0% | 0.30 [0.01, 0.60] |
| Kim, 2018 [136] | 51.2 | 4.4 | 41 | 50.3 | 4.4 | 42 | 1.5% | 0.20 [-0.23, 0.63] |
| Koizumi, 2009 [138] | 27.2 | 14.7 | 34 | 19 | 9.8 | 34 | 1.3% | 0.65 [0.16, 1.14] |
| Kolt, 2012 [76] | 121.2 | 184.4 | 130 | 111.4 | 110.6 | 123 | 2.3% | 0.06 [-0.18, 0.31] |
| Li, 2017 [139] | 64.2 | 70.5 | 17 | 56 | 60.1 | 17 | 0.8% | 0.12 [-0.55, 0.80] |
| Li, 2020 [107] | 44.7 | 41.2 | 55 | 31.6 | 32.4 | 57 | 1.7% | 0.35 [-0.02, 0.73] |
| Lynch, 2019 [109] | 249.6 | 159.7 | 40 | 170.1 | 127.9 | 40 | 1.4% | 0.54 [0.10, 0.99] |
| Maxwell Smith, 2019 [140] | 312 | 229.3 | 34 | 240 | 177.7 | 33 | 1.3% | 0.35 [-0.14, 0.83] |
| Melville, 2015 [112] | 3 | 2.6 | 42 | 3.1 | 2.1 | 40 | 1.4% | -0.04 [-0.47, 0.39] |
| Pinto, 2013 [61] | 214 | 147.3 | 19 | 97 | 148 | 24 | 0.9% | 0.78 [0.15, 1.40] |
| Pinto, 2015 [62] | 70.3 | 65.9 | 36 | 16.5 | 31.9 | 32 | 1.2% | 1.01 [0.50, 1.52] |
| Pope, 2018 [117] | 34.2 | 18.7 | 12 | 37.8 | 20.4 | 8 | 0.5% | -0.18 [-1.07, 0.72] |
| Ribeiro, 2014 [63] | 473.5 | 913.2 | 101 | 58.5 | 484.3 | 94 | 2.1% | 0.56 [0.27, 0.85] |
| Samuels, 2011 [143] | 34.8 | 13.6 | 12 | 26.5 | 16.1 | 25 | 0.8% | 0.53 [-0.17, 1.23] |
| Ter Hoeve, 2018 [68] | 6.7 | 3 | 121 | 6.5 | 3.1 | 252 | 2.4% | 0.07 [-0.15, 0.28] |
| Unick, 2012 [145] | 133 | 217 | 11 | 44.8 | 124.6 | 12 | 0.6% | 0.49 [-0.35, 1.32] |
| Van der Weegen, 2015 [71] | 48.2 | 23.8 | 65 | 42.9 | 25.8 | 134 | 2.0% | 0.21 [-0.09, 0.51] |
| Van Hoye, 2018 [72] | 17.8 | 49.6 | 157 | 9.6 | 34 | 50 | 1.9% | 0.18 [-0.14, 0.49] |
| Wijsman, 2013 [147] | 11.1 | 21.8 | 108 | -0.1 | 15.4 | 105 | 2.1% | 0.59 [0.32, 0.86] |
| Yates, 2017 [78] | -4.1 | 19 | 287 | -5.9 | 17.7 | 272 | 2.7% | 0.10 [-0.07, 0.26] |
| **Subtotal (95% CI)** | | | **2717** | | | **2830** | **57.9%** | **0.33 [0.24, 0.43]** |
| Heterogeneity: Tau² = 0.04; Chi² = 85.05, df = 37 (P < 0.0001); I² = 56% | | | | | | | | |
| Test for overall effect: Z = 7.16 (P < 0.00001) | | | | | | | | |
| **4.2.3 Combined intervention** | | | | | | | | |
| Harris, 2018 [75] | 374 | 179 | 778 | 310 | 168 | 456 | 2.9% | 0.37 [0.25, 0.48] |
| **Subtotal (95% CI)** | | | **778** | | | **456** | **2.9%** | **0.37 [0.25, 0.48]** |
| Heterogeneity: Not applicable | | | | | | | | |
| Test for overall effect: Z = 6.15 (P < 0.00001) | | | | | | | | |
| **Total (95% CI)** | | | **5850** | | | **5133** | **100.0%** | **0.28 [0.21, 0.35]** |
| Heterogeneity: Tau² = 0.04; Chi² = 162.03, df = 61 (P < 0.00001); I² = 62% | | | | | | | | |
| Test for overall effect: Z = 7.82 (P < 0.00001) | | | | | | | | |
| Test for subgroup differences: Chi² = 4.23, df = 2 (P = 0.12), I² = 52.8% | | | | | | | | |



Subgroup Analysis by Intervention Design For the Outcome Total Physical Activity

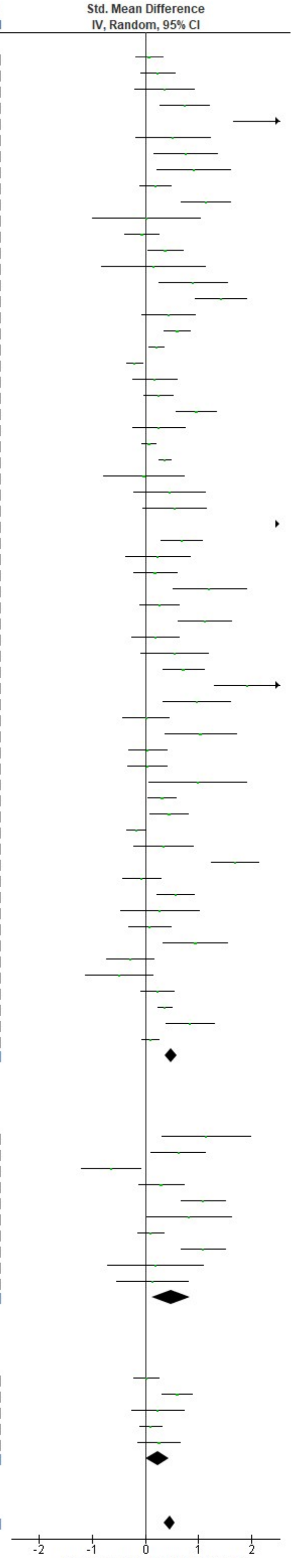
| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **4.3.1 Scalable (no human interaction)** | | | | | | | | |
| Aittasalo, 2006 [131] | 509.5 | 479.4 | 203 | 555 | 519.7 | 62 | 3.8% | -0.09 [-0.38, 0.19] |
| Carr, 2013 [148] | 171.4 | 234.5 | 25 | 121.8 | 174.6 | 28 | 2.6% | 0.24 [-0.30, 0.78] |
| Compernolle, 2015 [85] | 81.6 | 78 | 52 | 90.6 | 100.2 | 71 | 3.4% | -0.10 [-0.46, 0.26] |
| Kernot, 2019 [52] | 298.9 | 97.7 | 74 | 288.8 | 68.6 | 33 | 3.2% | 0.11 [-0.30, 0.52] |
| Mansi, 2015 [54] | 1,035 | 444 | 29 | 188 | 135 | 29 | 2.0% | 2.55 [1.84, 3.25] |
| Martin, 2015 [111] | 10.5 | 21.1 | 32 | -8 | 23 | 16 | 2.3% | 0.84 [0.21, 1.46] |
| Merom, 2007 [114] | 91 | 243.1 | 105 | 84.2 | 231.8 | 209 | 4.0% | 0.03 [-0.21, 0.26] |
| Prestwich, 2010 [118] | 3 | 1.79 | 88 | 2.3 | 14.3 | 46 | 3.5% | 0.08 [-0.27, 0.44] |
| Simons, 2018 [64] | 317 | 135 | 55 | 359.5 | 146 | 63 | 3.4% | -0.30 [-0.66, 0.06] |
| Suggs, 2013 [66] | 49.5 | 39.6 | 79 | 5.8 | 33.9 | 79 | 3.5% | 1.18 [0.84, 1.52] |
| Thorsteinsen, 2014 [153] | 574 | 435 | 12 | 502 | 323 | 8 | 1.5% | 0.17 [-0.72, 1.07] |
| Vandelanotte, 2018 [146] | 387.7 | 377.7 | 78 | 230 | 164.1 | 46 | 3.4% | 0.50 [0.13, 0.87] |
| **Subtotal (95% CI)** | | | **832** | | | **690** | **36.7%** | **0.39 [0.06, 0.73]** |
| Heterogeneity: Tau² = 0.30; Chi² = 97.59, df = 11 (P < 0.00001); I² = 89% | | | | | | | | |
| Test for overall effect: Z = 2.28 (P = 0.02) | | | | | | | | |
| **4.3.2 Non-scalable (mHealth combined with human interactions)** | | | | | | | | |
| Cadmus-Bertram, 2019 [84] | 1,463 | 489 | 24 | 1,343 | 395 | 23 | 2.5% | 0.26 [-0.31, 0.84] |
| Cruz, 2016 [44] | 279.5 | 74 | 13 | 212 | 53.9 | 13 | 1.7% | 1.01 [0.19, 1.83] |
| DeGreef, 2010 [73] | 301 | 106 | 20 | 260 | 104 | 21 | 2.3% | 0.38 [-0.24, 1.00] |
| DeGreef, 2011[89] | 93 | 66 | 60 | 40 | 56 | 32 | 3.0% | 0.84 [0.39, 1.28] |
| DeGreef, 2011a [74] | 176 | 108 | 43 | 65 | 68 | 24 | 2.6% | 1.15 [0.61, 1.68] |
| Dlugonski, 2012 [45] | 28.2 | 15.6 | 22 | 15.4 | 13.9 | 23 | 2.3% | 0.85 [0.24, 1.47] |
| Furber, 2010 [50] | 366.5 | 270.8 | 97 | 270.9 | 244.4 | 107 | 3.8% | 0.37 [0.09, 0.65] |
| Gill, 2019 [96] | 2.13 | 25.8 | 59 | 1.4 | 25.7 | 59 | 3.4% | 0.03 [-0.33, 0.39] |
| Keyserling, 2008 [135] | 29.8 | 6.5 | 86 | 28.6 | 5.7 | 90 | 3.7% | 0.20 [-0.10, 0.49] |
| Kolt, 2012 [76] | 168.9 | 190.5 | 130 | 146.3 | 163.5 | 123 | 4.0% | 0.13 [-0.12, 0.37] |
| Lane, 2015 [149] | 84.9 | 50.3 | 125 | 80.3 | 32.8 | 159 | 4.0% | 0.11 [-0.12, 0.35] |
| Lyons, 2017 [110] | 117 | 121 | 20 | 58 | 33 | 20 | 2.2% | 0.65 [0.01, 1.29] |
| Mailey, 2010 [150] | 299,792 | 102,800 | 24 | 251,625 | 83,081 | 23 | 2.5% | 0.51 [-0.08, 1.09] |
| Melville, 2015 [112] | 33.5 | 10 | 42 | 34 | 12 | 40 | 3.1% | -0.04 [-0.48, 0.39] |
| Motl, 2011 [151] | 24.7 | 18.8 | 23 | 12.4 | 14.2 | 25 | 2.4% | 0.73 [0.14, 1.32] |
| Müller, 2016 [56] | 383.4 | 843.4 | 18 | 377.4 | 842.6 | 21 | 2.3% | 0.01 [-0.62, 0.64] |
| Schwerdtfeger, 2012 [154] | 738.6 | 245.7 | 21 | 690.5 | 217.1 | 41 | 2.7% | 0.21 [-0.32, 0.74] |
| Sugden, 2008 [152] | 108.7 | 54.7 | 26 | 113.822 | 62.337 | 18 | 2.4% | -0.09 [-0.69, 0.51] |
| Van Hoye, 2018 [72] | 1.8 | 8.3 | 157 | 0.6 | 7 | 50 | 3.6% | 0.15 [-0.17, 0.47] |
| Wyke, 2019 [77] | 3,717 | 3,307 | 499 | 2,741 | 2,951 | 505 | 4.4% | 0.31 [0.19, 0.44] |
| Yates, 2017 [78] | -32.3 | 94.6 | 287 | -40.3 | 85 | 272 | 4.3% | 0.09 [-0.08, 0.25] |
| **Subtotal (95% CI)** | | | **1796** | | | **1689** | **63.3%** | **0.30 [0.18, 0.42]** |
| Heterogeneity: Tau² = 0.03; Chi² = 42.43, df = 20 (P = 0.002); I² = 53% | | | | | | | | |
| Test for overall effect: Z = 4.99 (P < 0.00001) | | | | | | | | |
| **Total (95% CI)** | | | **2628** | | | **2379** | **100.0%** | **0.34 [0.20, 0.47]** |
| Heterogeneity: Tau² = 0.10; Chi² = 140.11, df = 32 (P < 0.00001); I² = 77% | | | | | | | | |
| Test for overall effect: Z = 4.96 (P < 0.00001) | | | | | | | | |
| Test for subgroup differences: Chi² = 0.27, df = 1 (P = 0.60), I² = 0% | | | | | | | | |



## **Appendix 9:** Subgroup Analysis by Control Group Type

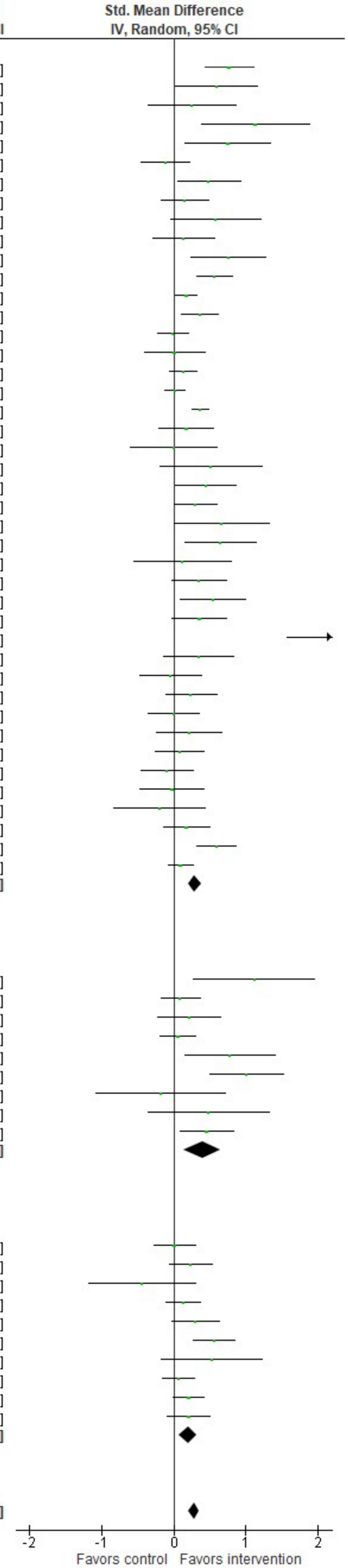
Subgroup Analysis by Control Group Type for The Outcome Walking

| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **3.1.1 No/minimal intervention control group** | | | | | | | | |
| Aittasalo, 2012 [41] | 457 | 306 | 123 | 431 | 403 | 118 | 1.7% | 0.07 [-0.18, 0.33] |
| Alsaleh, 2016 [79] | 36.7 | 109.8 | 66 | 16.5 | 58.7 | 79 | 1.5% | 0.23 [-0.09, 0.56] |
| Ashton, 2017 [80] | 1,588.2 | 2,740.6 | 26 | 575.4 | 2,735.1 | 24 | 1.2% | 0.36 [-0.20, 0.92] |
| Baker, 2008 [81] | 9,977 | 4,669 | 39 | 7,078 | 2,911 | 40 | 1.3% | 0.74 [0.28, 1.20] |
| Barwais, 2013 [82] | 1,625 | 553.8 | 18 | 483 | 175.9 | 15 | 0.7% | 2.61 [1.65, 3.57] |
| Butler, 2004 [83] | 56,396 | 3,315 | 17 | 53,666 | 6,543 | 16 | 1.0% | 0.52 [-0.18, 1.21] |
| Cadmus-Bertram, 2019 [84] | 6,697 | 2,878 | 24 | 4,853 | 1,703 | 23 | 1.1% | 0.76 [0.17, 1.36] |
| Coelho, 2018 [43] | 8,853 | 3,320 | 20 | 6,248 | 2,030 | 17 | 1.0% | 0.91 [0.23, 1.59] |
| Compernolle, 2015 [85] | 9,484 | 4,875 | 86 | 8,589 | 4,380 | 93 | 1.6% | 0.19 [-0.10, 0.49] |
| Creel, 2016 [86] | 6,084 | 873 | 52 | 5,253 | 421 | 35 | 1.3% | 1.13 [0.67, 1.60] |
| Croteau, 2004 [129] | 2,419 | 6,074.6 | 7 | 2,320 | 4,732 | 8 | 0.6% | 0.02 [-1.00, 1.03] |
| Croteau, 2007 [130] | 6,180 | 3,530 | 79 | 6,378 | 2,994 | 68 | 1.5% | -0.06 [-0.38, 0.26] |
| Dadaczynski, 2017 [87] | 526.5 | 240.6 | 80 | 442.9 | 200.4 | 64 | 1.5% | 0.37 [0.04, 0.70] |
| De Blok, 2006 [88] | 3,927 | 2,617 | 8 | 3,554 | 2,019.6 | 8 | 0.7% | 0.15 [-0.83, 1.13] |
| DeGreef, 2010 [73] | 9,601 | 5,002 | 20 | 5,538 | 3,877 | 21 | 1.0% | 0.89 [0.25, 1.54] |
| DeGreef, 2011[89] | 7,703 | 2,729 | 60 | 3,883 | 2,537 | 32 | 1.3% | 1.42 [0.94, 1.90] |
| DeGreef, 2011a [74] | 6,771 | 3,889 | 43 | 5,173 | 3,094 | 24 | 1.3% | 0.44 [-0.07, 0.94] |
| Demeyer, 2017 [90] | 7 | 32.5 | 129 | -10 | 23.5 | 132 | 1.7% | 0.60 [0.35, 0.85] |
| Dishman, 2009 [91] | 14.8 | 17.3 | 564 | 11.2 | 17.3 | 265 | 1.8% | 0.21 [0.06, 0.35] |
| Finkelstein, 2016 [48] | -480 | 2,568 | 201 | 43.7 | 2,640.7 | 599 | 1.8% | -0.20 [-0.36, -0.04] |
| Fjeldsoe, 2010 [94] | 16.7 | 89.4 | 45 | 0.3 | 89.4 | 43 | 1.4% | 0.18 [-0.24, 0.60] |
| Furber, 2010 [50] | 249.9 | 196 | 97 | 202.6 | 189.5 | 107 | 1.6% | 0.24 [-0.03, 0.52] |
| Gill, 2019 [96] | 1,646 | 3,302 | 59 | -1,485 | 3,171.5 | 59 | 1.5% | 0.96 [0.58, 1.34] |
| Glynn, 2014 [97] | 5,855 | 4,264 | 31 | 4,859 | 3,474 | 35 | 1.3% | 0.25 [-0.23, 0.74] |
| Hardeman, 2020 [98] | 8,419 | 3,224 | 417 | 8,191 | 3,003 | 442 | 1.8% | 0.07 [-0.06, 0.21] |
| Harris, 2018 [75] | 8,306 | 3,140 | 778 | 7,198 | 2,809 | 456 | 1.8% | 0.37 [0.25, 0.48] |
| Hornikx, 2015 [99] | 984 | 1,208 | 12 | 1,013 | 1,275 | 15 | 0.9% | -0.02 [-0.78, 0.74] |
| Hospes, 2009 [100] | 7,872 | 3,962 | 18 | 6,172 | 3,194 | 17 | 1.0% | 0.46 [-0.21, 1.13] |
| Houle, 2011 [101] | 9,850 | 3,282 | 23 | 7,970 | 3,433 | 22 | 1.1% | 0.55 [-0.05, 1.15] |
| Hultquist, 2005 [102] | 10,159 | 292 | 27 | 8,270 | 354 | 31 | 0.5% | 5.70 [4.51, 6.90] |
| James, 2015 [51] | 800.8 | 2,820.6 | 57 | -1,294 | 3,304.4 | 52 | 1.4% | 0.68 [0.29, 1.07] |
| Katzmarzyk, 2011[104] | 7,248 | 2,737 | 20 | 6,637 | 2,418 | 23 | 1.1% | 0.23 [-0.37, 0.83] |
| Kernot, 2019 [52] | 276.3 | 221.7 | 74 | 235 | 218.9 | 33 | 1.4% | 0.19 [-0.23, 0.60] |
| King, 2013 [106] | 253.5 | 248.7 | 20 | 26.8 | 67 | 19 | 1.0% | 1.21 [0.52, 1.89] |
| Li, 2020 [107] | 6,673 | 3,462 | 55 | 5,819 | 2,860 | 57 | 1.5% | 0.27 [-0.10, 0.64] |
| Long, 2013 [108] | 65,983 | 18,069 | 38 | 47,596 | 13,900 | 33 | 1.3% | 1.12 [0.61, 1.62] |
| Lynch, 2019 [109] | 8,193 | 3,301 | 37 | 7,539 | 3,404 | 40 | 1.3% | 0.19 [-0.26, 0.64] |
| Lyons, 2017 [110] | 6,193 | 3,183 | 20 | 4,586 | 2,476 | 20 | 1.1% | 0.55 [-0.08, 1.19] |
| Maher, 2015 [53] | 332 | 289 | 51 | 160 | 185 | 59 | 1.4% | 0.72 [0.33, 1.10] |
| Mansi, 2015 [54] | 9,792 | 2,053 | 29 | 6,551 | 1,154 | 29 | 1.1% | 1.92 [1.29, 2.55] |
| Martin, 2015 [111] | 1,067 | 2,098 | 32 | -1,042 | 2,202 | 16 | 1.1% | 0.97 [0.34, 1.61] |
| Melville, 2015 [112] | 4,823 | 2,059 | 42 | 4,784 | 2,613 | 40 | 1.4% | 0.02 [-0.42, 0.45] |
| Mutrie, 2012 [58] | 9,351 | 2,017 | 20 | 7,138 | 2,169 | 19 | 1.0% | 1.04 [0.36, 1.71] |
| Nolan, 2017 [59] | 272 | 2,275.9 | 63 | 155 | 2,557.1 | 59 | 1.5% | 0.05 [-0.31, 0.40] |
| Oliveira, 2019 [60] | 7,507 | 3,077 | 54 | 7,401 | 2,841 | 55 | 1.5% | 0.04 [-0.34, 0.41] |
| Paul, 2016 [115] | 5,791 | 2,952 | 15 | 2,947 | 2,399 | 8 | 0.7% | 0.99 [0.07, 1.90] |
| Poirier, 2016 [116] | 5,411 | 2,277 | 107 | 4,751 | 1,834 | 110 | 1.6% | 0.32 [0.05, 0.59] |
| Prestwich, 2010 [118] | 1.98 | 1.9 | 88 | 1.17 | 1.58 | 46 | 1.5% | 0.45 [0.09, 0.81] |
| Reijonsaari, 2012 [119] | 2,047 | 1,650 | 264 | 2,338 | 1,762 | 257 | 1.8% | -0.17 [-0.34, 0.00] |
| Roos, 2014 [120] | 3,810 | 10,993.5 | 29 | 817.3 | 2,349.2 | 22 | 1.2% | 0.35 [-0.21, 0.91] |
| Rowley, 2019 [121] | 9,015 | 2,842 | 97 | 4,654 | 1,447 | 32 | 1.3% | 1.69 [1.23, 2.14] |
| Simons, 2018 [64] | 7,741 | 4,553 | 55 | 8,061 | 5,112 | 63 | 1.5% | -0.07 [-0.43, 0.30] |
| Stacey, 2016 [65] | 10,849 | 5,127 | 75 | 8,014 | 4,568 | 58 | 1.5% | 0.58 [0.23, 0.93] |
| Tabak, 2014 [123] | 5,603 | 3,475.8 | 13 | 4,617 | 3,460 | 16 | 0.9% | 0.28 [-0.46, 1.01] |
| Thorndike, 2014 [124] | 7,886 | 3,622 | 50 | 7,600 | 3,492 | 49 | 1.4% | 0.08 [-0.31, 0.47] |
| Tudor-Locke, 2004 [69] | 9,123 | 4,539 | 24 | 5,622 | 2,405 | 23 | 1.1% | 0.94 [0.34, 1.55] |
| Vallance, 2016 [125] | 5,923 | 3,109 | 41 | 6,885 | 3,576 | 37 | 1.3% | -0.29 [-0.73, 0.16] |
| Van Blarigan, 2019 [126] | 10,047 | 4,461 | 20 | 12,541 | 5,535 | 19 | 1.0% | -0.49 [-1.13, 0.15] |
| Van Hoye, 2018 [72] | 13.2 | 29.6 | 157 | 6.7 | 26 | 50 | 1.6% | 0.23 [-0.09, 0.54] |
| Wyke, 2019 [77] | 9,801 | 3,730 | 464 | 8,518 | 3,254 | 471 | 1.8% | 0.37 [0.24, 0.50] |
| Yamada, 2012 [128] | 3,726 | 1,607 | 40 | 2,267 | 1,837 | 42 | 1.3% | 0.84 [0.38, 1.29] |
| Yates, 2017 [78] | -486 | 2,252 | 287 | -690 | 2,041 | 272 | 1.8% | 0.09 [-0.07, 0.26] |
| **Subtotal (95% CI)** | | | **5607** | | | **5037** | **80.7%** | **0.47 [0.36, 0.59]** |
| Heterogeneity: Tau² = 0.14; Chi² = 397.03, df = 61 (P < 0.00001); I² = 85% | | | | | | | | |
| Test for overall effect: Z = 8.29 (P < 0.00001) | | | | | | | | |
| **3.1.2 Alternative intervention control group** | | | | | | | | |
| Cruz, 2016 [44] | 10,440 | 4,012 | 13 | 6,430 | 2,613 | 13 | 0.8% | 1.15 [0.31, 1.99] |
| Duru, 2010 [92] | 9,883 | 14,014 | 34 | 2,426 | 8,429 | 28 | 1.2% | 0.62 [0.11, 1.14] |
| Engel, 2006 [93] | 150 | 117 | 24 | 216 | 87 | 30 | 1.2% | -0.64 [-1.19, -0.09] |
| Gell, 2015 [95] | 6,867.7 | 2,227 | 41 | 6,189 | 2,297 | 46 | 1.4% | 0.30 [-0.13, 0.72] |
| Izawa, 2012 [103] | 8,609.6 | 3,064.5 | 52 | 5,512.9 | 2,571.8 | 51 | 1.4% | 1.09 [0.67, 1.50] |
| Kawagoshi, 2015 [105] | 51.3 | 63.7 | 12 | 12.3 | 25.5 | 15 | 0.8% | 0.82 [0.02, 1.61] |
| Kolt, 2012 [76] | 107.4 | 164.9 | 130 | 92.2 | 120.5 | 123 | 1.7% | 0.10 [-0.14, 0.35] |
| Mendoza, 2015 [113] | 3,080 | 3,254 | 52 | 138 | 1,950 | 50 | 1.4% | 1.08 [0.67, 1.50] |
| Pope, 2018 [117] | 5,175 | 2,308 | 12 | 4,746 | 2,045 | 8 | 0.7% | 0.19 [-0.71, 1.08] |
| Talbot, 2003 [67] | 4,337 | 2,903 | 17 | 3,972 | 2,563 | 17 | 1.0% | 0.13 [-0.54, 0.80] |
| **Subtotal (95% CI)** | | | **387** | | | **381** | **11.6%** | **0.48 [0.12, 0.83]** |
| Heterogeneity: Tau² = 0.25; Chi² = 45.88, df = 9 (P < 0.00001); I² = 80% | | | | | | | | |
| Test for overall effect: Z = 2.60 (P = 0.009) | | | | | | | | |
| **3.1.3 Combined control group** | | | | | | | | |
| Merom, 2007 [114] | 52 | 107.17 | 105 | 48.9 | 126.83 | 209 | 1.7% | 0.03 [-0.21, 0.26] |
| Ribeiro, 2014 [63] | 969.7 | 1,464.1 | 101 | -181.5 | 2,284 | 94 | 1.6% | 0.60 [0.32, 0.89] |
| Spence, 2009 [122] | 690 | 450.2 | 32 | 592.5 | 368.9 | 31 | 1.3% | 0.23 [-0.26, 0.73] |
| Ter Hoeve, 2018 [68] | 8 | 2.6 | 121 | 7.7 | 3 | 252 | 1.7% | 0.10 [-0.11, 0.32] |
| Warren, 2014 [127] | 8,371 | 3,069 | 37 | 7,576 | 2,993 | 69 | 1.4% | 0.26 [-0.14, 0.66] |
| **Subtotal (95% CI)** | | | **396** | | | **655** | **7.7%** | **0.23 [0.01, 0.45]** |
| Heterogeneity: Tau² = 0.04; Chi² = 10.51, df = 4 (P = 0.03); I² = 62% | | | | | | | | |
| Test for overall effect: Z = 2.08 (P = 0.04) | | | | | | | | |
| **Total (95% CI)** | | | **6390** | | | **6073** | **100.0%** | **0.46 [0.36, 0.55]** |
| Heterogeneity: Tau² = 0.14; Chi² = 458.16, df = 76 (P < 0.00001); I² = 83% | | | | | | | | |
| Test for overall effect: Z = 8.98 (P < 0.00001) | | | | | | | | |
| Test for subgroup differences: Chi² = 3.75, df = 2 (P = 0.15), I² = 46.7% | | | | | | | | |



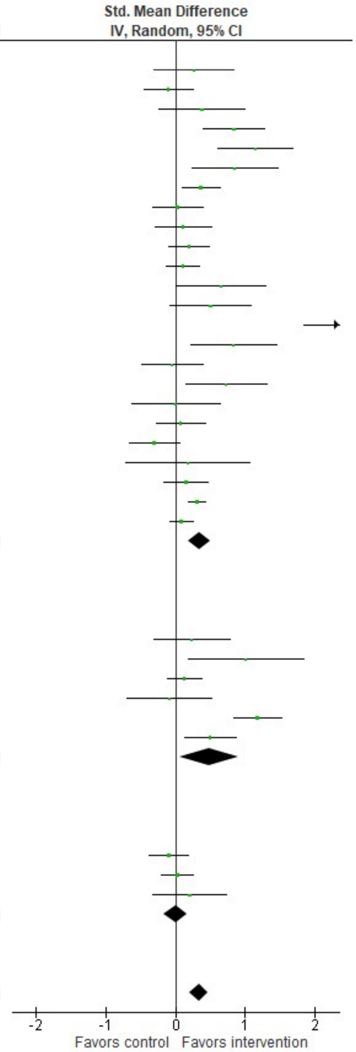
Subgroup Analysis by Control Group Type for The Outcome Moderate-to-vigorous Physical Activity

| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **3.2.1 No/minimal intervention control group** | | | | | | | | |
| Alsaleh, 2016 [79] | 237.9 | 121.1 | 66 | 108.9 | 198 | 79 | 1.8% | 0.77 [0.43, 1.11] |
| Ashton, 2017 [80] | 154.1 | 216.1 | 26 | 26.1 | 210.2 | 24 | 1.0% | 0.59 [0.02, 1.16] |
| Barnes, 2015 [42] | 0.2 | 1.6 | 23 | -0.2 | 1.5 | 19 | 0.9% | 0.25 [-0.36, 0.86] |
| Barwais, 2013 [82] | 649 | 494.9 | 18 | 192 | 221.3 | 15 | 0.7% | 1.13 [0.38, 1.87] |
| Cadmus-Bertram, 2019 [84] | 230 | 120 | 24 | 143 | 107 | 23 | 1.0% | 0.75 [0.16, 1.35] |
| Compernolle, 2015 [85] | 32.4 | 37.7 | 62 | 38.4 | 60 | 79 | 1.9% | -0.12 [-0.45, 0.22] |
| Creel, 2016 [86] | 37 | 41.6 | 52 | 20.4 | 17.1 | 35 | 1.4% | 0.48 [0.05, 0.92] |
| Dadaczynski, 2017 [87] | 310.9 | 129.4 | 80 | 291.8 | 114 | 64 | 1.9% | 0.15 [-0.17, 0.48] |
| DeGreef, 2010 [73] | 44 | 38 | 20 | 24 | 29 | 21 | 0.9% | 0.58 [-0.04, 1.21] |
| DeGreef, 2011[89] | 23 | 20 | 60 | 20 | 25 | 32 | 1.5% | 0.14 [-0.29, 0.57] |
| DeGreef, 2011a [74] | 82 | 84 | 43 | 27 | 39 | 24 | 1.2% | 0.76 [0.24, 1.28] |
| Demeyer, 2017 [90] | 8 | 21.1 | 140 | -3 | 17.5 | 140 | 2.3% | 0.57 [0.33, 0.80] |
| Dishman, 2009 [91] | 10.6 | 15.5 | 564 | 7.9 | 15.1 | 265 | 2.8% | 0.18 [0.03, 0.32] |
| Eakin, 2014 [46] | 163.6 | 155.6 | 119 | 114.6 | 108.6 | 127 | 2.3% | 0.37 [0.11, 0.62] |
| Edney, 2020 [47] | 108.1 | 74.1 | 272 | 108.9 | 52.6 | 130 | 2.5% | -0.01 [-0.22, 0.20] |
| Fjeldsoe, 2010 [94] | 18.3 | 167.3 | 45 | 16.4 | 167.4 | 43 | 1.5% | 0.01 [-0.41, 0.43] |
| Golsteijn, 2018 [133] | 331 | 234 | 208 | 301 | 219 | 211 | 2.6% | 0.13 [-0.06, 0.32] |
| Hardeman, 2020 [98] | 77.3 | 36.5 | 417 | 76.7 | 35.4 | 442 | 2.8% | 0.02 [-0.12, 0.15] |
| Harris, 2018 [75] | 374 | 179 | 778 | 310 | 168 | 456 | 2.9% | 0.37 [0.25, 0.48] |
| James, 2015 [51] | 33.9 | 145.3 | 57 | 9.6 | 128.6 | 52 | 1.7% | 0.18 [-0.20, 0.55] |
| Katzmarzyk, 2011[104] | 16.3 | 17.3 | 20 | 16.2 | 17.1 | 23 | 1.0% | 0.01 [-0.59, 0.60] |
| Kendzor, 2017 [134] | 76 | 67 | 17 | 46 | 43 | 15 | 0.8% | 0.51 [-0.19, 1.22] |
| Kernot, 2019 [52] | 204.1 | 131 | 74 | 150 | 96.9 | 33 | 1.5% | 0.44 [0.03, 0.86] |
| Keyserling, 2008 [135] | 14 | 3.71 | 86 | 13 | 2.8 | 89 | 2.0% | 0.30 [0.01, 0.60] |
| King, 2008 [137] | 310.6 | 298.3 | 19 | 135 | 208.2 | 18 | 0.8% | 0.66 [0.00, 1.33] |
| Koizumi, 2009 [138] | 27.2 | 14.7 | 34 | 19 | 9.8 | 34 | 1.3% | 0.65 [0.16, 1.14] |
| Li, 2017 [139] | 64.2 | 70.5 | 17 | 56 | 60.1 | 17 | 0.8% | 0.12 [-0.55, 0.80] |
| Li, 2020 [107] | 44.7 | 41.2 | 55 | 31.6 | 32.4 | 57 | 1.7% | 0.35 [-0.02, 0.73] |
| Lynch, 2019 [109] | 249.6 | 159.7 | 40 | 170.1 | 127.9 | 40 | 1.4% | 0.54 [0.10, 0.99] |
| Maher, 2015 [53] | 528 | 391 | 51 | 391 | 371 | 59 | 1.7% | 0.36 [-0.02, 0.74] |
| Mansi, 2015 [54] | 1,469 | 524 | 29 | 538 | 254 | 29 | 0.8% | 2.23 [1.57, 2.90] |
| Maxwell Smith, 2019 [140] | 312 | 229.3 | 34 | 240 | 177.7 | 33 | 1.3% | 0.35 [-0.14, 0.83] |
| Melville, 2015 [112] | 3 | 2.6 | 42 | 3.1 | 2.1 | 40 | 1.4% | -0.04 [-0.47, 0.39] |
| Murawski, 2019 [57] | 428.4 | 523.4 | 59 | 319.7 | 378.2 | 66 | 1.8% | 0.24 [-0.11, 0.59] |
| Nolan, 2017 [59] | 11 | 68.8 | 63 | 11 | 58.8 | 59 | 1.8% | 0.00 [-0.36, 0.36] |
| Pekmezi, 2017 [141] | 31.5 | 58.9 | 39 | 20.8 | 38.2 | 37 | 1.4% | 0.21 [-0.24, 0.66] |
| Sharp, 2016 [144] | 100 | 109 | 72 | 90 | 119 | 65 | 1.9% | 0.09 [-0.25, 0.42] |
| Simons, 2018 [64] | 27.3 | 26.7 | 55 | 30.1 | 31 | 63 | 1.7% | -0.10 [-0.46, 0.27] |
| Vallance, 2016 [125] | 143 | 132 | 41 | 146 | 155 | 37 | 1.4% | -0.02 [-0.47, 0.42] |
| Van Blarigan, 2019 [126] | 46.6 | 48.4 | 20 | 54.5 | 24.9 | 19 | 0.9% | -0.20 [-0.83, 0.43] |
| Van Hoye, 2018 [72] | 17.8 | 49.6 | 157 | 9.6 | 34 | 50 | 1.9% | 0.18 [-0.14, 0.49] |
| Wijsman, 2013 [147] | 11.1 | 21.8 | 108 | -0.1 | 15.4 | 105 | 2.1% | 0.59 [0.32, 0.86] |
| Yates, 2017 [78] | -4.1 | 19 | 287 | -5.9 | 17.7 | 272 | 2.7% | 0.10 [-0.07, 0.26] |
| **Subtotal (95% CI)** | | | **4493** | | | **3541** | **69.7%** | **0.29 [0.21, 0.38]** |
| Heterogeneity: Tau² = 0.05; Chi² = 125.93, df = 42 (P < 0.00001); I² = 67% | | | | | | | | |
| Test for overall effect: Z = 6.52 (P < 0.00001) | | | | | | | | |
| **3.2.2 Alternative intervention control group** | | | | | | | | |
| Cruz, 2016 [44] | 57.8 | 32.8 | 13 | 26.7 | 19.6 | 13 | 0.6% | 1.11 [0.28, 1.95] |
| Fjeldsoe, 2015 [49] | -8 | 262.4 | 104 | -29.2 | 196 | 107 | 2.2% | 0.09 [-0.18, 0.36] |
| Kim, 2018 [136] | 51.2 | 4.43 | 41 | 50.25 | 4.4 | 42 | 1.5% | 0.21 [-0.22, 0.64] |
| Kolt, 2012 [76] | 121.2 | 184.4 | 130 | 111.4 | 110.6 | 123 | 2.3% | 0.06 [-0.18, 0.31] |
| Pinto, 2013 [61] | 214 | 147.3 | 19 | 97 | 148 | 24 | 0.9% | 0.78 [0.15, 1.40] |
| Pinto, 2015 [62] | 70.3 | 65.9 | 36 | 16.5 | 31.9 | 32 | 1.2% | 1.01 [0.50, 1.52] |
| Pope, 2018 [117] | 34.2 | 18.7 | 12 | 37.8 | 20.4 | 8 | 0.5% | -0.18 [-1.07, 0.72] |
| Unick, 2012 [145] | 133 | 217 | 11 | 44.8 | 124.6 | 12 | 0.6% | 0.49 [-0.35, 1.32] |
| Vandelanotte, 2018 [146] | 148.8 | 181.1 | 78 | 79.8 | 77.1 | 46 | 1.7% | 0.45 [0.08, 0.82] |
| **Subtotal (95% CI)** | | | **444** | | | **407** | **11.4%** | **0.39 [0.14, 0.65]** |
| Heterogeneity: Tau² = 0.08; Chi² = 21.18, df = 8 (P = 0.007); I² = 62% | | | | | | | | |
| Test for overall effect: Z = 3.06 (P = 0.002) | | | | | | | | |
| **3.2.3 Combined control group** | | | | | | | | |
| Aittasalo, 2006 [131] | 95 | 83.7 | 203 | 94 | 70.9 | 62 | 2.1% | 0.01 [-0.27, 0.30] |
| Fischer, 2019 [132] | 228.4 | 157.8 | 63 | 194 | 136 | 128 | 2.0% | 0.24 [-0.06, 0.54] |
| Maselli, 2019 [55] | 134.4 | 83.4 | 11 | 185.5 | 126.3 | 21 | 0.7% | -0.44 [-1.18, 0.30] |
| Merom, 2007 [114] | 79 | 543.7 | 105 | 32.8 | 171.3 | 209 | 2.4% | 0.13 [-0.10, 0.37] |
| Prestwich, 2009 [142] | 1.3 | 1.2 | 60 | 0.97 | 1 | 94 | 1.9% | 0.30 [-0.02, 0.63] |
| Ribeiro, 2014 [63] | 473.5 | 913.2 | 101 | 58.5 | 484.3 | 94 | 2.1% | 0.56 [0.27, 0.85] |
| Samuels, 2011 [143] | 34.8 | 13.6 | 12 | 26.5 | 16.1 | 25 | 0.8% | 0.53 [-0.17, 1.23] |
| Ter Hoeve, 2018 [68] | 6.7 | 3 | 121 | 6.5 | 3.1 | 252 | 2.4% | 0.07 [-0.15, 0.28] |
| Vallance, 2008 [70] | 213 | 173 | 172 | 180 | 141 | 166 | 2.5% | 0.21 [-0.01, 0.42] |
| Van der Weegen, 2015 [71] | 48.2 | 23.8 | 65 | 42.9 | 25.8 | 134 | 2.0% | 0.21 [-0.09, 0.51] |
| **Subtotal (95% CI)** | | | **913** | | | **1185** | **18.9%** | **0.20 [0.08, 0.32]** |
| Heterogeneity: Tau² = 0.01; Chi² = 13.68, df = 9 (P = 0.13); I² = 34% | | | | | | | | |
| Test for overall effect: Z = 3.33 (P = 0.0009) | | | | | | | | |
| **Total (95% CI)** | | | **5850** | | | **5133** | **100.0%** | **0.28 [0.21, 0.35]** |
| Heterogeneity: Tau² = 0.04; Chi² = 162.02, df = 61 (P < 0.00001); I² = 62% | | | | | | | | |
| Test for overall effect: Z = 7.83 (P < 0.00001) | | | | | | | | |
| Test for subgroup differences: Chi² = 2.66, df = 2 (P = 0.26), I² = 24.7% | | | | | | | | |



Subgroup Analysis by Control Group Type for the Outcome Total Physical Activity

| Study or Subgroup | Intervention Mean | Intervention SD | Intervention Total | Control Mean | Control SD | Control Total | Weight | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|---|---|---|---|---|---|
| **3.3.1 No/minimal intervention control group** | | | | | | | | |
| Cadmus-Bertram, 2019 [84] | 1,463 | 489 | 24 | 1,343 | 395 | 23 | 2.5% | 0.26 [-0.31, 0.84] |
| Compernolle, 2015 [85] | 81.6 | 78 | 52 | 90.6 | 100.2 | 71 | 3.4% | -0.10 [-0.46, 0.26] |
| DeGreef, 2010 [73] | 301 | 106 | 20 | 260 | 104 | 21 | 2.3% | 0.38 [-0.24, 1.00] |
| DeGreef, 2011[89] | 93 | 66 | 60 | 40 | 56 | 32 | 3.0% | 0.84 [0.39, 1.28] |
| DeGreef, 2011a [74] | 176 | 108 | 43 | 65 | 68 | 24 | 2.6% | 1.15 [0.61, 1.68] |
| Dlugonski, 2012 [45] | 28.2 | 15.6 | 22 | 15.4 | 13.9 | 23 | 2.3% | 0.85 [0.24, 1.47] |
| Furber, 2010 [50] | 366.5 | 270.8 | 97 | 270.9 | 244.4 | 107 | 3.8% | 0.37 [0.09, 0.65] |
| Gill, 2019 [96] | 2.1 | 25.8 | 59 | 1.37 | 25.7 | 59 | 3.4% | 0.03 [-0.33, 0.39] |
| Kernot, 2019 [52] | 298.9 | 97.7 | 74 | 288.8 | 68.6 | 33 | 3.2% | 0.11 [-0.30, 0.52] |
| Keyserling, 2008 [135] | 29.8 | 6.5 | 86 | 28.6 | 5.7 | 90 | 3.7% | 0.20 [-0.10, 0.49] |
| Lane, 2015 [149] | 84.9 | 50.3 | 125 | 80.3 | 32.8 | 159 | 4.0% | 0.11 [-0.12, 0.35] |
| Lyons, 2017 [110] | 117 | 121 | 20 | 58 | 33 | 20 | 2.2% | 0.65 [0.01, 1.29] |
| Mailey, 2010 [150] | 299,792 | 102,800 | 24 | 251,625 | 83,081 | 23 | 2.5% | 0.51 [-0.08, 1.09] |
| Mansi, 2015 [54] | 1,035 | 444 | 29 | 188 | 135 | 29 | 2.0% | 2.55 [1.84, 3.25] |
| Martin, 2015 [111] | 10.5 | 21.1 | 32 | -8 | 23 | 16 | 2.3% | 0.84 [0.21, 1.46] |
| Melville, 2015 [112] | 33.5 | 10 | 42 | 34 | 12 | 40 | 3.1% | -0.04 [-0.48, 0.39] |
| Motl, 2011 [151] | 24.7 | 18.8 | 23 | 12.4 | 14.2 | 25 | 2.4% | 0.73 [0.14, 1.32] |
| Müller, 2016 [56] | 383.4 | 843.4 | 18 | 377.4 | 842.6 | 21 | 2.3% | 0.01 [-0.62, 0.64] |
| Prestwich, 2010 [118] | 2.96 | 1.8 | 88 | 2.3 | 14.3 | 46 | 3.5% | 0.08 [-0.28, 0.43] |
| Simons, 2018 [64] | 317 | 135 | 55 | 359.5 | 146 | 63 | 3.4% | -0.30 [-0.66, 0.06] |
| Thorsteinsen, 2014 [153] | 574 | 435 | 12 | 502 | 323 | 8 | 1.5% | 0.17 [-0.72, 1.07] |
| Van Hoye, 2018 [72] | 1.8 | 8.3 | 157 | 0.6 | 7 | 50 | 3.6% | 0.15 [-0.17, 0.47] |
| Wyke, 2019 [77] | 3,717 | 3,307 | 499 | 2,741 | 2,951 | 505 | 4.4% | 0.31 [0.19, 0.44] |
| Yates, 2017 [78] | -32.3 | 94.6 | 287 | -40.3 | 85 | 272 | 4.3% | 0.09 [-0.08, 0.25] |
| **Subtotal (95% CI)** | | | **1948** | | | **1760** | **72.0%** | **0.34 [0.19, 0.50]** |
| Heterogeneity: Tau² = 0.09; Chi² = 95.04, df = 23 (P < 0.00001); I² = 76% | | | | | | | | |
| Test for overall effect: Z = 4.40 (P < 0.0001) | | | | | | | | |
| **3.3.2 Alternative intervention control group** | | | | | | | | |
| Carr, 2013 [148] | 171.4 | 234.5 | 25 | 121.8 | 174.6 | 28 | 2.6% | 0.24 [-0.30, 0.78] |
| Cruz, 2016 [44] | 279.5 | 74 | 13 | 212 | 53.9 | 13 | 1.7% | 1.01 [0.19, 1.83] |
| Kolt, 2012 [76] | 168.9 | 190.5 | 130 | 146.3 | 163.5 | 123 | 4.0% | 0.13 [-0.12, 0.37] |
| Sugden, 2008 [152] | 108.7 | 54.7 | 26 | 113.8 | 62.3 | 18 | 2.4% | -0.09 [-0.69, 0.51] |
| Suggs, 2013 [66] | 49.5 | 39.6 | 79 | 5.8 | 33.9 | 79 | 3.5% | 1.18 [0.84, 1.52] |
| Vandelanotte, 2018 [146] | 387.7 | 377.7 | 78 | 230 | 164.1 | 46 | 3.4% | 0.50 [0.13, 0.87] |
| **Subtotal (95% CI)** | | | **351** | | | **307** | **17.6%** | **0.48 [0.06, 0.91]** |
| Heterogeneity: Tau² = 0.22; Chi² = 29.92, df = 5 (P < 0.0001); I² = 83% | | | | | | | | |
| Test for overall effect: Z = 2.23 (P = 0.03) | | | | | | | | |
| **3.3.3 Combined control group** | | | | | | | | |
| Aittasalo, 2006 [131] | 509.5 | 479.4 | 203 | 555 | 519.7 | 62 | 3.8% | -0.09 [-0.38, 0.19] |
| Merom, 2007 [114] | 91 | 243.1 | 105 | 84.2 | 231.8 | 209 | 4.0% | 0.03 [-0.21, 0.26] |
| Schwerdtfeger, 2012 [154] | 738.6 | 245.7 | 21 | 690.5 | 217.1 | 41 | 2.7% | 0.21 [-0.32, 0.74] |
| **Subtotal (95% CI)** | | | **329** | | | **312** | **10.5%** | **0.00 [-0.17, 0.17]** |
| Heterogeneity: Tau² = 0.00; Chi² = 1.07, df = 2 (P = 0.59); I² = 0% | | | | | | | | |
| Test for overall effect: Z = 0.04 (P = 0.97) | | | | | | | | |
| **Total (95% CI)** | | | **2628** | | | **2379** | **100.0%** | **0.34 [0.20, 0.47]** |
| Heterogeneity: Tau² = 0.10; Chi² = 140.16, df = 32 (P < 0.00001); I² = 77% | | | | | | | | |
| Test for overall effect: Z = 4.96 (P < 0.00001) | | | | | | | | |
| Test for subgroup differences: Chi² = 10.12, df = 2 (P = 0.006), I² = 80.2% | | | | | | | | |



**Appendix 10:** Study-specific Risk of Bias Judgments

| | Random sequence generation (selection bias) | Allocation concealment (selection bias) | Blinding of participants and personnel (performance bias) | Blinding of outcome assessment (detection bias) | Incomplete outcome data (attrition bias) | Selective outcome reporting | Other | Summary |
|---|---|---|---|---|---|---|---|---|
| Aittasalo 2006 [131] | 🔴 | 🟡 | 🟡 | 🔴 | 🔴 | 🟡 | 🟢 | High Risk |
| Aittasalo 2012 [41] | 🟢 | 🟡 | 🟡 | 🔴 | 🟡 | 🟢 | 🟢 | High Risk |
| Alsaleh 2016 [79] | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🟢 | 🟢 | High Risk |
| Ashton 2017 [80] | 🟢 | 🟢 | 🟡 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |
| Baker 2008 [81] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟡 | 🟢 | High Risk |
| Barnes 2015 [42] | 🟢 | 🟢 | 🟡 | 🟡 | 🟡 | 🟢 | 🟢 | Low Risk |
| Barwais 2013 [82] | 🟢 | 🟡 | 🔴 | 🟡 | 🟢 | 🟡 | 🟢 | Low Risk |
| Bennett 2008 [155] | 🟢 | 🟡 | 🔴 | 🟡 | 🟡 | 🟡 | 🟢 | Low Risk |
| Butler 2004 [83] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Cadmus-Bertram 2019 [84] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟡 | 🟢 | Low Risk |
| Carr 2013 [148] | 🟢 | 🟢 | 🟡 | 🟡 | 🟡 | 🟢 | 🔴 | High Risk |
| Coelho 2018 [43] | 🟢 | 🔴 | 🔴 | 🔴 | 🟢 | 🟡 | 🔴 | High Risk |
| Compernolle 2015 [85] | 🟢 | 🟡 | 🔴 | 🟢 | 🔴 | 🟢 | 🔴 | High Risk |
| Creel 2016 [86] | 🟡 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🔴 | High Risk |
| Croteau 2004 [129] | 🟢 | 🟡 | 🔴 | 🟢 | 🟢 | 🟢 | 🔴 | High Risk |
| Croteau 2007 [130] | 🟡 | 🟡 | 🟡 | 🟢 | 🟢 | 🟢 | 🔴 | High Risk |
| Cruz 2016 [44] | 🟢 | 🟢 | 🟡 | 🔴 | 🔴 | 🟢 | 🟢 | High Risk |
| Dadaczynski 2017 [87] | 🟢 | 🟡 | 🟡 | 🔴 | 🔴 | 🟢 | 🔴 | High Risk |
| De Blok 2006 [88] | 🟡 | 🟡 | 🟢 | 🟢 | 🔴 | 🟡 | 🔴 | High Risk |
| De Greef 2010 [73] | 🟡 | 🟢 | 🔴 | 🟢 | 🟢 | 🟡 | 🟢 | Low Risk |
| De Greef 2011 [89] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🔴 | 🟡 | High Risk |
| De Greef 2011 [74] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Demeyer 2017 [90] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🔴 | 🟢 | High Risk |
| Dishman 2009 [91] | 🔴 | 🟡 | 🟡 | 🔴 | 🔴 | 🟡 | 🔴 | High Risk |
| Dlugonski 2012 [45] | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🟡 | 🔴 | High Risk |
| Duru 2010 [92] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🔴 | 🟢 | High Risk |
| Eakin 2014 [46] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🟢 | High Risk |
| Edney 2020 [47] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟡 | 🟢 | Low Risk |
| Engel 2006 [93] | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🟡 | 🟡 | High Risk |
| Finkelstein 2016 [48] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🟢 | High Risk |
| Fischer 2019 [132] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🟢 | High Risk |
| Fjeldsoe 2010 [94] | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🟡 | 🔴 | High Risk |
| Fjeldsoe 2015 [49] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🟢 | High Risk |
| Furber 2010 [50] | 🟢 | 🟢 | 🟢 | 🔴 | 🟢 | 🟡 | 🟡 | Low Risk |
| Gell 2015 [95] | 🟡 | 🟡 | 🔴 | 🔴 | 🟢 | 🟡 | 🔴 | High Risk |
| Gill 2019 [96] | 🟡 | 🟡 | 🔴 | 🔴 | 🟢 | 🟢 | 🟢 | High Risk |
| Glasgow 2012 [156] | 🟢 | 🟡 | 🔴 | 🔴 | 🔴 | 🟢 | 🟡 | High Risk |
| Glynn 2014 [97] | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |
| Golsteijn 2018 [133] | 🟢 | 🟡 | 🔴 | 🟢 | 🔴 | 🟢 | 🟢 | High Risk |
| Hardeman 2020 [98] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |
| Harris 2018 [75] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🔴 | High Risk |
| Hornikx 2015 [99] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🟢 | 🔴 | High Risk |
| Hospes 2009 [100] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Houle 2011 [101] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟡 | 🔴 | High Risk |
| Hultquist 2005 [102] | 🟡 | 🟡 | 🟡 | 🔴 | 🟢 | 🟡 | 🔴 | High Risk |
| Izawa 2012 [103] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| James 2015 [51] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🟢 | High Risk |
| Kangasniemi 2015 [40] | 🟡 | 🟡 | 🔴 | 🟢 | 🔴 | 🔴 | 🔴 | High Risk |
| Katzmarzyk 2011 [104] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🔴 | High Risk |
| Kawagoshi 2015 [105] | 🟡 | 🟡 | 🔴 | 🟢 | 🔴 | 🟡 | 🟢 | High Risk |
| Kendzor 2017 [134] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🔴 | 🟢 | High Risk |
| Kernot 2019 [52] | 🟢 | 🟢 | 🔴 | 🔴 | 🔴 | 🟢 | 🟢 | High Risk |
| Keyserling 2008 [135] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🔴 | High Risk |
| Kim 2018 [136] | 🟡 | 🟡 | 🔴 | 🟢 | 🔴 | 🔴 | 🟢 | High Risk |
| King 2008 [137] | 🟢 | 🟡 | 🔴 | 🔴 | 🟢 | 🟡 | 🔴 | High Risk |
| King 2013 [106] | 🟢 | 🟡 | 🔴 | 🔴 | 🟢 | 🟢 | 🔴 | High Risk |
| Koizumi 2009 [138] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Kolt 2012 [76] | 🟢 | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🟢 | High Risk |
| Lane 2015 [149] | 🟡 | 🟡 | 🔴 | 🔴 | 🔴 | 🟡 | 🔴 | High Risk |
| Li 2017 [139] | 🟡 | 🔴 | 🔴 | 🟢 | 🟢 | 🟢 | 🔴 | High Risk |

| | Random sequence generation (selection bias) | Allocation concealment (selection bias) | Blinding of participants and personnel (performance bias) | Blinding of outcome assessment (detection bias) | Incomplete outcome data (attrition bias) | Selective outcome reporting | Other | Summary |
|---|---|---|---|---|---|---|---|---|
| Li 2020 [107] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |
| Long 2013 [108] | 🟡 | 🟡 | 🔴 | 🟢 | 🔴 | 🟡 | 🟢 | High Risk |
| Lynch 2019 [109] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🔴 | 🔴 | Low Risk |
| Lyons 2017 [110] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🔴 | 🟢 | High Risk |
| Maher 2015 [53] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |
| Mailey 2010 [150] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Mansi 2015 [54] | 🟡 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |
| Martin 2015 [111] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🟢 | 🔴 | High Risk |
| Maselli 2019 [55] | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🟡 | 🔴 | High Risk |
| Maxwell-Smith 2019 [140] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |
| Melville 2015 [112] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🔴 | 🟢 | High Risk |
| Mendoza 2015 [113] | 🟢 | 🔴 | 🟡 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |
| Merom 2007 [114] | 🟢 | 🟡 | 🔴 | 🔴 | 🟢 | 🟡 | 🔴 | High Risk |
| Motl 2011 [151] | 🟡 | 🟢 | 🔴 | 🔴 | 🟢 | 🟡 | 🟢 | High Risk |
| Müller 2016 [56] | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🔴 | 🟢 | High Risk |
| Murawski 2019 [57] | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🟡 | 🟢 | High Risk |
| Mutrie 2012 [58] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |
| Nolan 2017 [59] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🔴 | High Risk |
| Oliveira 2019 [60] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🟢 | High Risk |
| Paul 2016 [115] | 🔴 | 🔴 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Pekmezi 2017 [141] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🔴 | 🔴 | High Risk |
| Pinto 2013 [61] | 🟡 | 🟡 | 🟡 | 🔴 | 🟢 | 🔴 | 🟢 | High Risk |
| Pinto 2015 [62] | 🟡 | 🟢 | 🔴 | 🟢 | 🟢 | 🟡 | 🟢 | Low Risk |
| Poirier 2016 [116] | 🟢 | 🔴 | 🔴 | 🟢 | 🔴 | 🟢 | 🟡 | High Risk |
| Pope 2018 [117] | 🟢 | 🔴 | 🔴 | 🟢 | 🔴 | 🟡 | 🔴 | High Risk |
| Prestwich 2009 [142] | 🟢 | 🟡 | 🔴 | 🔴 | 🔴 | 🟡 | 🔴 | High Risk |
| Prestwich 2010 [118] | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🟡 | 🔴 | High Risk |
| Reijonsaari 2012 [119] | 🟢 | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🟢 | High Risk |
| Ribeiro 2014 [63] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🔴 | High Risk |
| Roos 2014 [120] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Rowley 2019 [121] | 🟡 | 🔴 | 🔴 | 🟢 | 🔴 | 🟡 | 🔴 | High Risk |
| Samuels 2011 [143] | 🟡 | 🟡 | 🔴 | 🟢 | 🔴 | 🟡 | 🔴 | High Risk |
| Schwerdtfeger 2012 [154] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Sharp 2016 [144] | 🟢 | 🟡 | 🔴 | 🔴 | 🔴 | 🟡 | 🔴 | High Risk |
| Simons 2018 [64] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🟢 | High Risk |
| Spence 2009 [122] | 🟡 | 🟡 | 🔴 | 🔴 | 🟢 | 🟡 | 🔴 | High Risk |
| Stacey 2016 [65] | 🟢 | 🔴 | 🔴 | 🟢 | 🔴 | 🔴 | 🔴 | High Risk |
| Sugden 2008 [152] | 🟢 | 🟡 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Suggs 2013 [66] | 🟢 | 🟡 | 🟢 | 🔴 | 🔴 | 🟡 | 🔴 | High Risk |
| Tabak 2014 [123] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |
| Talbot 2003 [67] | 🟢 | 🟡 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Ter Hoeve 2018 [68] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🔴 | 🔴 | High Risk |
| Thorndike 2014 [124] | 🟢 | 🟢 | 🟡 | 🟢 | 🟢 | 🟡 | 🟡 | Low Risk |
| Thorsteinsen 2014 [153] | 🟡 | 🟢 | 🔴 | 🔴 | 🟢 | 🟡 | 🟢 | High Risk |
| Tudor-Locke 2004 [69] | 🟡 | 🟡 | 🔴 | 🟢 | 🔴 | 🟡 | 🔴 | High Risk |
| Unick 2012 [145] | 🟡 | 🟡 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Vallance 2008 [70] | 🟢 | 🟢 | 🔴 | 🔴 | 🔴 | 🟢 | 🟢 | High Risk |
| Vallance 2016 [125] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟡 | 🔴 | High Risk |
| Van Blarigan 2019 [126] | 🟢 | 🔴 | 🔴 | 🟢 | 🟢 | 🟢 | 🟢 | High Risk |
| Vandelanotte 2018 [146] | 🟢 | 🟢 | 🟢 | 🔴 | 🔴 | 🟢 | 🟢 | High Risk |
| Van der Weegen 2015 [71] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🔴 | High Risk |
| Van Hoye 2018 [72] | 🟢 | 🟡 | 🔴 | 🟢 | 🟢 | 🟢 | 🔴 | High Risk |
| Warren 2014 [127] | 🟢 | 🟢 | 🔴 | 🟢 | 🔴 | 🟢 | 🔴 | High Risk |
| Wijsman 2013 [147] | 🟢 | 🟢 | 🔴 | 🟢 | 🟢 | 🟢 | 🟢 | Low Risk |

| | Random sequence generation (selection bias) | Allocation concealment (selection bias) | Blinding of participants and personnel (performance bias) | Blinding of outcome assessment (detection bias) | Incomplete outcome data (attrition bias) | Selective outcome reporting | Other | Summary |
|---|---|---|---|---|---|---|---|---|
| Wyke 2019 [77] | ● (green) | ● (green) | ● (red) | ● (red) | ● (green) | ● (green) | ● (red) | High Risk |
| Yamada 2012 [128] | ● (green) | ● (green) | ● (red) | ● (green) | ● (green) | ● (yellow) | ● (green) | Low Risk |
| Yates 2017 [78] | ● (green) | ● (green) | ● (red) | ● (green) | ● (green) | ● (green) | ● (green) | Low Risk |

*Note:* Table depicts risk of bias assessment per bias domain and an overall summary assessment. Summary assessment ranked as high if one domain other than "Blinding of participants and personnel" is ranked high.
●= low risk, ●= unclear risk, ●= high risk.

## **Appendix 11:** GRADE Profile

| Quality Assessment | | | | | | Summary of Findings | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Number of Participants | | | | |
| Number of Studies* | Risk of Bias | Inconsistency | Indirectness | Imprecision | Publication Bias | mHealth | Control | Effect Size as SMD (95% CI) | Signi-ficance | Quality |
| Walking (75 RCTs) | *Very Serious* 58/75 studies high risk of bias based on Cochrane criteria. Numerous studies with short intervention duration, small samples sizes and high attrition. As in most mHealth studies, blinding of participants and personnel is not possible. | *Not serious* Very high level of heterogenicity ($I^2$=84%) but range of results limited with 68/75 studies showing positive effects and CI [0.37-0.57] | *Not serious* Populations and interventions relevant | *Not serious* Low baseline risk (<5%), narrow confidence interval, sample size 12301) | *Undetected* Moderate sample size (mean=197, range=21-1442), funnel plot analysis does not suggest systematic publication bias. | 6307 | 5994 | 0.47 [0.37-0.57] | *P*<.001 | ●●○○ |
| MVPA (61 RCTs) | *Very Serious* 47/61 studies high risk of bias based on Cochrane criteria. Numerous studies with short intervention duration, small samples sizes and high attrition. As in most mHealth studies, blinding of participants and personell is not possible. | *Not serious* Moderate to high level of heterogenicity ($I^2$=63%) but range of results limited with 53/61 studies showing positive effects and CI [0.22-0.36] | *Not serious* Populations and interventions relevant | *Not serious* Low baseline risk (<5%), narrow confidence interval, sample size 10861) | *Undetected* Moderate sample size (mean=225, range=29-1442), funnel plot analysis does not suggest systematic publication bias. | 5787 | 5074 | 0.29 [0.22-0.36] | *P*<.001 | ●●○○ |
| Total PA (34 RCTs) | *Very Serious* 29/34 studies high risk of bias based on Cochrane criteria. Numerous studies with short intervention duration, small samples sizes and high attrition. As in most mHealth studies, blinding of participants and personnel is not possible. | *Not serious* Moderate to high level of heterogenicity ($I^2$=68%), moderate range of results [0.16-0.38] and 28/34 studies finding positive effects. | *Not serious* Populations and interventions relevant | *Not serious* Low baseline risk (<5%), narrow confidence interval, sample size 5144) | *Undetected* Moderate sample size (mean=189, range=15-1113), funnel plot analysis does not suggest systematic publication bias. | 2702 | 2442 | 0.27 [0.16-0.38] | *P*<.001 | ●●○○ |
| Energy Expenditure (8 RCTs) | *Very Serious* 7/8 studies high risk of bias based on Cochrane criteria. Numerous studies with short intervention duration, small samples sizes and high attrition. As in most mHealth studies, blinding of participants and personnel is not possible. | *Very serious* High level of heterogenicity ($I^2$=87%), large effect range CI [0.13-1.00], small sample. | *Not serious* Populations and interventions relevant | *Not serious* Low baseline risk (<5%), narrow confidence interval, sample size 834) | *Undetected* Small sample size (mean=128, range=37-463), funnel plot analysis does not suggest systematic publication bias. | 481 | 353 | 0.57 [0.13-1.00] | *P*=.01 | ○○○○ |

*Abbreviations*: GRADE, Grading of recommendations, assessment, development, and evaluation; RCT, randomized controlled trials; CI, confidence intervals

* Number of studies that report results at end of intervention.

# Article II

## Gamification 2.0: How Activity Tracking Apps Provide Long-term Motivation

Annette Mönninghoff[1,2]

[1]Institute for Customer Insight, University of St.Gallen, St.Gallen, Switzerland
[2]Institute for Mobility, University of St.Gallen, St.Gallen, Switzerland

## Abstract

One in four consumers is insufficiently physically active, resulting in substantial health risks. Conventional activity tracking apps motivate consumers but can promote motivational crowding out. Gamification can help overcome these challenges. This paper presents two case studies and introduces a new motivational framework for gamification.


**Keywords:** Gamification, mHealth, physical activity, crowding out, consumer motivation

## Introduction

Today, one in four consumers is insufficiently physically active (World Health Organization, 2018), resulting in substantial health risks. Insufficient physical activity (PA) increases the risk of chronic illnesses such as diabetes, cardiovascular diseases, and cancer (Lee et al., 2012).

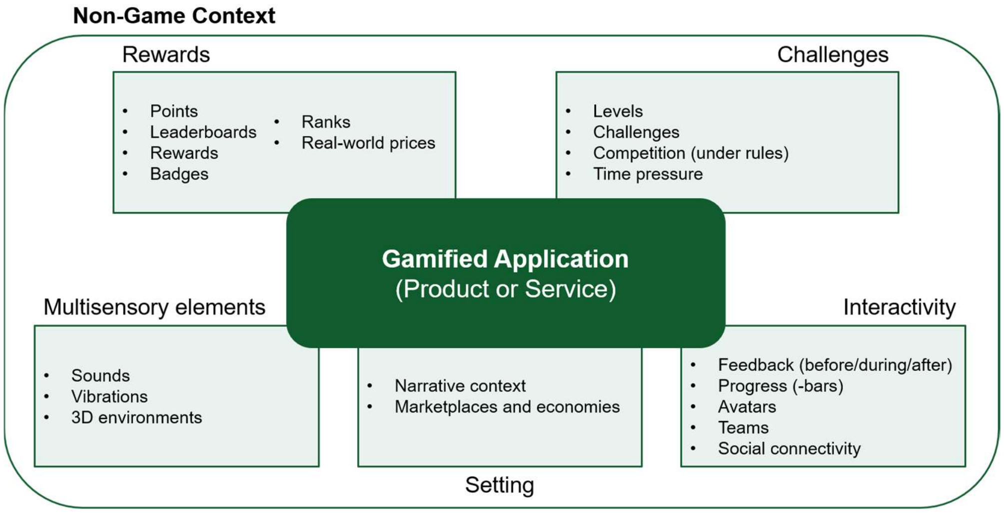
Activity tracking apps might help motivate consumers. Wearables and tracking apps have trended in recent years, with over 850 million sold to date (Statista, 2020). However, recent evidence has raised the question whether tracking apps can motivate consumers in the long term. A meta-analysis found no statistically significant effect of tracking apps on PA (Romeo et al., 2019). Surprisingly, when tracking apps were used for less than three months, they increased PA. However, in trials running for more than three months, no statistically significant increases were found (Romeo et al., 2019). In a series of experiments, Etkin (2016) found that wearable trackers promote motivational crowding out and theorized that activity trackers serve as an extrinsic incentive: They increase activity performance but decrease intrinsic walking motivation. This crowding out effect has also been documented in other fields (Koestner et al., 1999). Finally, a 2014 survey showed that one in three consumers already stops using their activity tracker six months after purchase (Endeavour Partners, 2014).

Could gamifying activity tracking apps help prevent motivational crowding out? This paper investigates how gamification promotes long-term motivation. Building on two case studies, it presents a *motivational framework for gamification* and makes various practical recommendations.

### Gamification and Motivation

Gamification has been defined as "the use of game design elements in non-game contexts" (Deterding et al., 2011, p. 9). Central to gamification are game elements. While common in many games, such elements are not part of every game. No single game element creates gameful experiences. Only integrating multiple game elements establishes intrinsically motivating environments in non-game contexts (Deterding et al., 2011). Game elements have been manifoldly modelled, leading scholars to criticize the lack of conceptual consensus (Johnson et al., 2016). Reeves and Read (2009) laid the foundations for a gamification framework by publishing ten "ingredients" of great games that can be used in non-game contexts to create consumer engagement. Building on this publication, scholars have developed various concepts ranging from 7 to 19 game elements (Hamari et al., 2014; Johnson et al., 2016; Lister et al., 2014). Some research has proposed classifying game elements into subgroups (Schlager et al., 2017). Figure 1 summarizes existing conceptualizations.

**Figure 1:** Existing Conceptual Framework for Gamification




*Source:* Author's illustration based on Hamari et al. (2014); Johnson et al. (2016); Lister et al. (2014); Reeves & Read (2009); Schlager et al. (2017)

## Impact of gamification on motivation

Understanding whether gamification might drive long-term behavior change toward a more active lifestyle requires exploring its impact on consumer motivation. Using game elements in non-game contexts (e.g., fitness tracking) potentially makes these products more enjoyable and engaging (Flatla et al., 2011). Playing games is fun and absorbing. It may create flow, i.e., being fully involved in and adequately challenged by an activity (Csikszentmihalyi, 2000; Mathwick & Rigdon, 2004).

Scholars differentiate two types of activity motivation: Consumers may be motivated to perform an activity either for its own sake (i.e., intrinsic motivation) or for an ulterior end (i.e., extrinsic motivation; cf. goal system theory, Kruglanski et al., 2018). Gamification per se is associated with intrinsic motivation (Hamari et al., 2014). However, this approach falls short when multiple, very different game elements are employed in non-game contexts. Not all game elements promote intrinsic motivation. Some elements (e.g., rewards, points, and leaderboards) are extrinsic incentives and thus motivate users by providing ulterior consumption motives. Extrinsically motivating game elements do not gamify PA itself, instead, they gamify the means used to motivate PA and thus often occur only after the activity has been performed. Other elements, however, have the potential to motivate intrinsically, by making the activity itself more enjoyable. Intrinsic activity motivation leads to more sustainable behavior change than extrinsically motivated behavior and is thus crucial in establishing long-term motivation (Teixeira et al., 2012). Building on self-determination theory, the satisfaction of three basic psychological needs can help explain how gamification promotes intrinsic motivation (Ryan & Deci, 2017): first, *challenges* support users' striving for *competence*, which positively impacts intrinsic motivation. *Levels* within gamified applications help ensure that users are optimally challenged. Second, *avatars* enable customizing a virtual counterpart, while *3D environments* promote exploration, with both elements satisfying users' need for *autonomy* (Peng et al., 2012). Third, *social connectivity* may satisfy users' need for *relatedness* and, thus, their intrinsic motivation.

*Motivational crowding out*

Motivational crowding out occurs when extrinsic incentives diminish intrinsic motivation over time (Etkin, 2016; Kruglanski et al., 1975). Extrinsically motivating game elements (e.g., points, leaderboards) may be perceived as controlling, thus diminishing users' sense of autonomy. Also, competition may negatively impact intrinsic motivation by reducing users' sense of competence if their performance is inferior to that of others (Ryan & Deci, 2017).

The context in which extrinsic incentives are employed moderates their impact on intrinsic motivation (Kruglanski et al., 1975). If incentives are perceived as integral (i.e., blend with the activity itself), they do not evoke motivational crowding out. Within gamified activity tracking apps, points, leaderboards, and competition may be integral and hence prevent motivational crowding out (Kruglanski et al., 1975; Schlager et al., 2017).

**Gamification in Activity Tracking Apps**

Existing research shows that over half of all health and fitness apps are gamified (Lister et al., 2014). A review of 132 health and fitness apps revealed that social or peer pressure was the game element used most (45.2%), followed by digital rewards (24.1%), competition (18.4%), and leaderboards (14.2%; Lister et al., 2014). A more recent review of 19 digital interventions in the health and wellness domain identified rewards (84.2%), leaderboards (31.6%), and avatars (31.6%) as the most common game elements (Johnson et al., 2016).

These reviews reveal the broad variety of gamified tracking apps. There is no single way of gamifying trackers, merely countless possible combinations of game elements. The reviewers criticize the lack of theoretical grounding in how game elements are selected. They also highlight that the most commonly used elements tend to be extrinsic motivators. This is surprising given the original goal of gamification to harness the motivational power of games to intrinsically encourage long-term behavior change (Deterding et al., 2011).

Many studies have proven the short-term effectiveness of gamified applications in increasing PA (Hamari et al., 2014; Johnson et al., 2016; Thorsteinsen et al., 2014). However, the longer-term motivational impact of applications remains under-researched (Johnson et al., 2016; Schlager et al., 2017; Zuckerman & Gal-Oz, 2014). Some authors have questioned that gamification is motivating per se, claiming there is only "anecdotal evidence" for this effect (Schlager et al., 2017, p. 31). Others have found that gamified health applications often have no or even negative effects on cognitive constructs such as motivation (Johnson et al., 2016). Scholars consider the mismatch between the most common game elements (i.e., rewards, leaderboards, avatars) and their motivational power the main reason why some studies report negative user reactions to gamified apps.

## Research Objectives & Methodology

The literature review revealed that most apps gamify post-activity feedback by employing extrinsically motivating game elements. However, existing research has evaluated apps as a single product without considering the various game elements used. Moreover, few studies have reviewed the long-term motivational impact of gamified tracking apps. Addressing these research gaps, this study presents two cases and seeks to answer two research questions:

1. How can gamifying activity tracking apps motivate consumers in the long term?
2. Which game elements should be used to promote long-term behavior change toward more active lifestyles?

*Zombies, Run!* and *Jump Jump Froggy 2* were selected for case studies. Both apps use intrinsically motivating game elements, were created by developers with a scientific background in motivation theory, and received positive media coverage (Howard, 2016). The walkthrough method (Light et al., 2018) was used to develop the case studies. Telephone interviews with the app developers were conducted and relevant public documents (e.g., press releases, newspaper articles, websites) and consumer reviews (e.g., YouTube, App stores) were analyzed.

## Findings

### Case study 1: Zombies, Run!

Launched in 2012, Zombies, Run! has become the world's most popular gamified activity tracking app with more than 5 million downloads and 200,000 active users (Six to Start & Naomi Alderman, 2020). The app leverages several game elements and aims to motivate users to run or walk more. Unlike conventional gamified trackers, Zombies, Run! gamifies the run itself by leveraging intrinsically motivating game elements instead of focusing solely on post-run rewards. Its most important game element is narrative context: Users assume the role of Runner 5, who has survived the zombie apocalypse. By running, users complete missions and receive important supplies which they can use after the run to build Abel Township, the last surviving town populated by humans. The storytelling uses audio (i.e., game element sound), so users need not look at the screen while running.

CEO Adrian Hon describes Zombies, Run! as a combination of game, tracking app, and audio storytelling that transforms the experience of running. Long-term motivation is created by a captivating narrative context that makes consumers immediately want to take up the next mission. The more users run, the more missions (i.e., game element levels) they get to complete, thus creating the optimal challenge. In the storytelling context, users randomly encounter zombies and are asked to speed up (i.e., game elements time pressure and challenge). The app supports users' sense of autonomy by offering various options: First, users can self-select how long their mission should be. The audio storytelling is then spaced across the run. Second, users can choose their music playlist or connect to music streaming services. The app also includes post-run rewards; these, however, are deeply integrated into the narrative, preventing motivational crowding out. The rewards consist of supplies collected while running. After the run, users can use these supplies to develop Abel Township, a 3D environment on a separate tab of the tracking app (see figure 2). The app automatically tells consumers their average running pace and summarizes core statistics (pace, kilometers, steps). The app deliberately includes no competitive elements as the developers believe that leaderboards, direct competition, and ranks demotivate many users, particularly if they are not proficient runners. The app is designed to appeal to a broad audience: CEO Adrian Hon describes users as non-runners or occasional runners who would like to run more often than they currently do.

*If you are like me and you don't really care much for running or exercising in general, then Zombies, Run! is definitely for you. You come for the great story, stay for the endorphins!*
*(Mack, app user)*

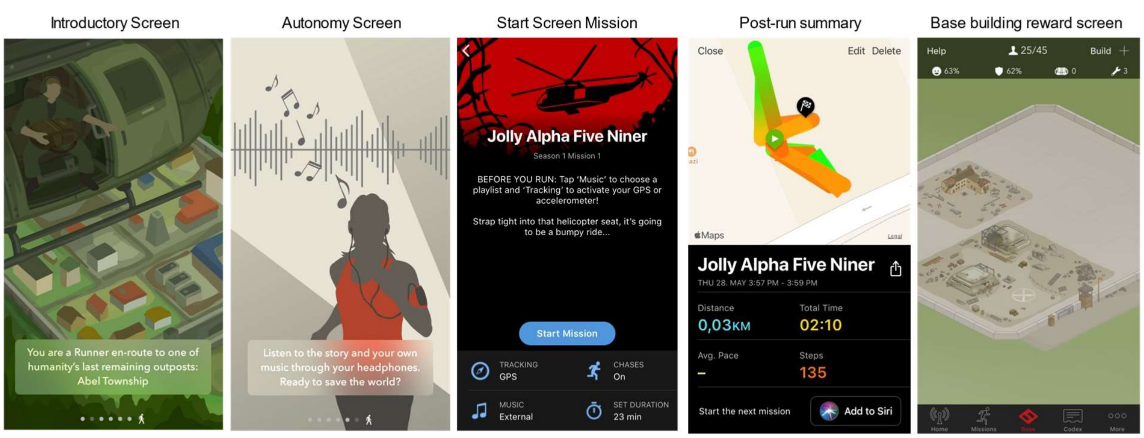
*I enjoy running, but it gets boring. Having the motivation to go on longer runs, to get out every day and run is hard. With Zombies, Run! I want to get out there and run. The story is so compelling that I need to get out there and find out what happens next.*
*(C587, app user)*

*After a few months of bouncing around different workout styles I found this app and let me tell you it works wonders. As soon as I start to space out and get bored, the story comes on again and I'm able to focus on that and get excited for the next piece of the story.*
*(Dogs_Luvr, app user)*

*This running app has gotten me through more tough runs and long weeks than I can tell you. I wanted a fun way to start back to running - and now I am fully hooked. […] I am about a year in and still completely loving the app.*
*(Yogi 2.0, app user)*

Evidence of the long-term motivational power of *Zombies, Run!* is manifold. A comparison of churn rates supports the long-term motivational power of *Zombies, Run!*: Over 50% of paying users stay with *Zombies, Run!* for more than a year, while research shows that regular activity tracking apps have a retention rate of just 6% 30 days after installment (Haslam, 2019). Also, over 200 missions have been taped since the launch, speaking to the need for ever new content for long-term users. Further, a randomized controlled trial confirmed that *Zombies, Run!* increases intrinsic motivation (Cowdery et al., 2015), consequently favoring long-term motivation.

**Figure 2:** Zombies, Run! Overview




*Source:* Six to Start & Naomi Alderman, Zombies, Run! Version 8.72

## Case study 2: Jump Jump Froggy 2

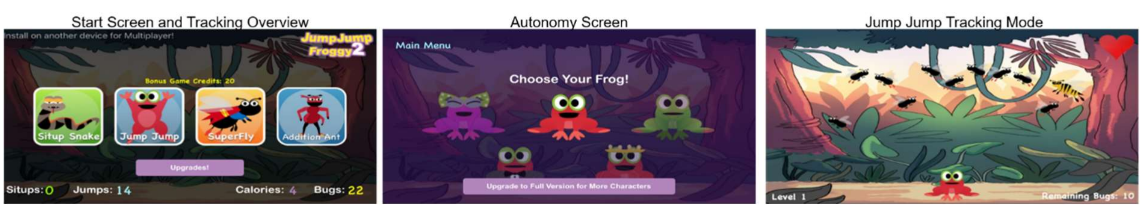
Launched in 2014, *Jump Jump Froggy 2* has been downloaded over 40,000 times. This app leverages multiple game elements and aims to motivate primary schoolchildren to be physically active. It was created by Timothy Charoenying, a cognitive scientist and software developer. A former teacher, he noticed that young children lacked opportunities to be active at school. His observation reflects the fact that, globally, 80% of kids do not meet the WHO’s physical activity recommendations. Like *Zombies, Run!*, *Jump Jump Froggy 2* also gamifies the activity instead of using post-activity game elements. Kids choose from different activity options (sit-ups or jumping). Next, they select their virtual avatar (among different types of frogs). Then tracking begins, and the children are asked to jump up and down to catch flies, to do sit-ups to move a snake around the phone screen, or to jump up and turn their body to collect fruits faster than an evil bee can catch them (i.e., game element *narrative context*). Fun music (i.e., game element *sound*) and an engaging comic-like 2D environment (figure 3) motivate the kids to perform the respective activity. To keep the app challenging, there are multiple game levels within each tracking option and task pace increases (i.e., game element *time pressure*). The app automatically recognizes movements through accelerometer data. Progress bars and points at the bottom of the screen inform the kids about their performance.

*I love that it gets my kids off the couch while still appealing to their love for video games (Sorrento76, app user)*

*Love watching my 7- and 10-year-old son and daughter choosing to do actual sit-ups and push-ups! So much better than just exercising their thumbs!” (HTC2011, app user)*

Positive consumer reviews, highlighting how much fun and how engaging *Jump Jump Froggy 2* is, attest to its intrinsic (long-term) activity motivation. Using the app in school contexts, Timothy Charoenying has observed high sustained student engagement. An analysis of consumer reviews also provides initial evidence for the long-term motivational power of the app. However, further quantitative evidence is needed to confirm the app’s lasting positive motivational impact.

**Figure 3:** Jump, Jump Froggy 2 Overview

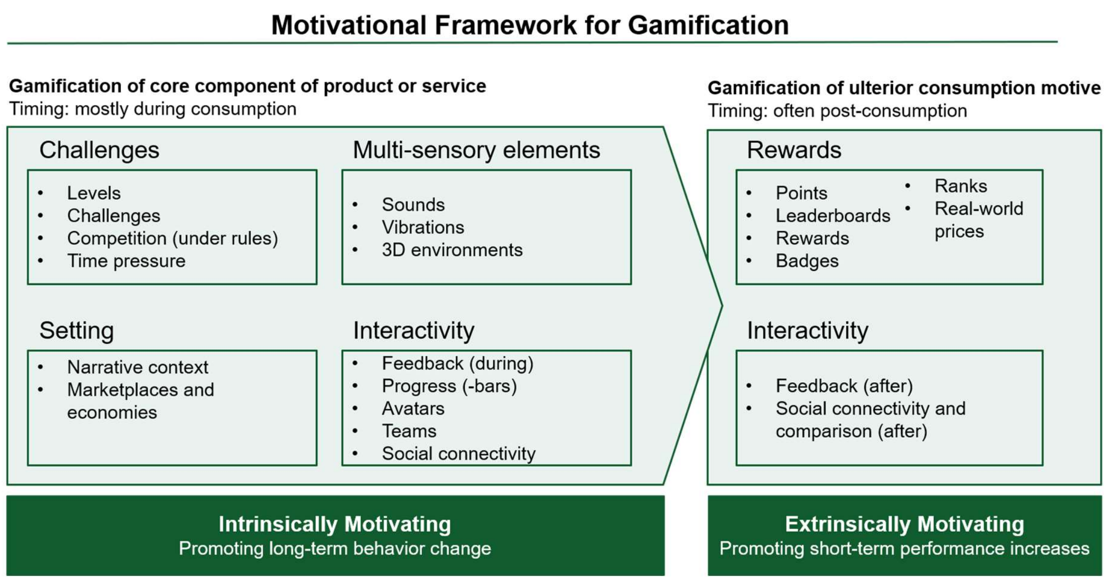



Source: Human Factored Design LLC, Jump Jump Froggy 2 Version 2.12

## A Motivational Framework for Gamification

The case studies revealed the need for a more granular assessment of gamification. As the motivational impact of game elements varies, gamification should not be evaluated as a single, unified phenomenon but broken down into individual game elements. Some game elements, especially those that gamify an integral component of a service or product, motivate intrinsically by enhancing consumers' sense of competence, autonomy, or relatedness; other elements motivate extrinsically, by providing a gamified ulterior end that motivates consumption. Building on the existing conceptual framework of gamification (figure 1), this paper proposes a new *motivational framework for gamification* (figure 4).

**Figure 4:** Motivational framework for gamification




*Source:* Author's illustration based on case study analyses (*Zombies, Run!* and *Jump Jump Froggy*) and on Hamari et al., 2014; Johnson et al., 2016; Lister et al., 2014; Reeves & Read, 2009; Schlager et al., 2017.

The proposed framework distinguishes two types of game elements: those gamifying a core component of a product or service, and those that gamify an ulterior consumption motive. To promote long-term motivation, this novel framework suggests that tracking app developers should focus on gamifying the activity itself. In contrast, gamifying post-tracking feedback (i.e., ulterior motive to be active) provides only short-term performance increases and potential motivational crowding out.

The two case studies powerfully illustrate how extrinsically motivating game elements can be employed without diminishing intrinsic motivation. *Zombies, Run!*

enables users to collect rewards while running. Instead of traditional rewards (e.g., points or badges), runners gather construction materials or food items to build Abel Township after the run. Rewards, then, are not just an extrinsic incentive, but are meaningful within the game context and integral to the game. Thus, the framework suggests using extrinsically motivating game elements only if they are employed for gamifying the consumption experience and are meaningful within the game context.

## Discussion

Prevalence of insufficient PA is high and involves substantial health risks. Recent evidence has called into question whether conventional activity tracking apps can motivate long-term behavior change and suggests that they may cause motivational crowding out. Although gamification is believed to help overcome these challenges, evidence of its motivational long-term impact is limited.

Based on a literature review and two case studies, this paper proposes a new *motivational framework for gamification*. The framework provides guidance to app developers and public health officials on how best to gamify activity tracking apps to promote long-term motivation and behavior change. The framework suggests developers should focus on gamifying core components of the activity tracking process with intrinsically motivating game elements (e.g., sound, narrative context, and avatars). This marks a fundamental shift for leading tracking apps (e.g., Strava, Runtastic), which currently mostly rely on gamifying post-activity feedback with leaderboards, badges, and rewards.

However, consumers might react differently to gamified activity tracking apps. The case studies show that moderately active consumers require intrinsically motivating game elements to drive behavior change. In contrast, highly active consumers – as currently targeted by Strava or Runtastic – possess high intrinsic motivation and thus might not require intrinsically motivating game elements. Thus, the findings provide initial evidence that companies might increase their consumer base by marketing intrinsically motivating gamified tracking solutions to moderately active consumers. The innovative fitness equipment company *Peloton* is an example of a firm which successfully markets to moderately active consumers by offering an intrinsically motivating, gamified bike.

Further research is needed to test the new conceptual framework given the limitations of this study. The framework was developed based on a literature review and the limited evidence of two case studies on innovative activity tracking apps. The latter were developed for English-speaking, high-income countries, limiting the generalizability of the findings. Future research should focus on validating the framework with quantitative evidence to further evaluate the long-term motivational power of gamified activity tracking apps.

**Management Summary**

This article reviews gamification literature and presents two case studies on activity tracking apps to evaluate the motivational impact of gamification. It finds that most tracking apps gamify post-activity feedback with extrinsically motivating game elements (e.g., leaderboards, points, and rewards). The article presents a new motivational framework for gamification and concludes that the activity tracking process itself should be gamified (e.g., with narrative context, sound, levels and challenges) to motivate intrinsically and promote long-term motivation within inactive consumer groups.

**Main Propositions**

1. Gamifying activity tracking apps potentially motivates consumers to adopt more active lifestyles.
2. Conventional gamified activity tracking apps mostly leverage extrinsically motivating game elements and focus on already active target groups.
3. To create long-term motivation and to reach moderately active consumer groups, fitness apps should gamify the tracking experience itself.
4. Assessing the various game elements employed and their motivational impact leads to a more differentiated evaluation of gamification.

**Lessons Learned**

1. Leading fitness apps should tap into the underserved but promising target group of insufficiently active consumers (>25% of the total population) by expanding their product offerings to include intrinsically motivating game elements such as narrative context, sound, and time pressure.
2. Future research should take a more granular perspective when evaluating gamified applications and quantitatively validate the proposed motivational framework for gamification.
3. Policymakers should leverage gamified fitness apps to address the growing physical activity crisis as such devices can potentially motivate long-term behavior change and reach large population groups.

## References

Cowdery, J., Majeske, P., Frank, R., & Brown, D. (2015). Exergame apps and physical activity: The results of the ZOMBIE trial. *American Journal of Health Education*, *46*(4), 216–222. https://doi.org/10.1080/19325037.2015.1043063.

Csikszentmihalyi, M. (2000). Beyond boredom and anxiety. Jossey-Bass Publishers.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining 'gamification'. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, 9–15. https://doi.org/10.1145/2181037.2181040.

Endeavour Partners (2017, April 27). Inside Wearables Part 2: Key developments for adoption and engagement. https://medium.com/@endeavourprtnrs/inside-wearables-part-2-july-2014-ef301d425cdd.

Etkin, J. (2016). The hidden cost of personal quantification. *Journal of Consumer Research*, *42*(6), 967–984. https://doi.org/10.1093/jcr/ucv095.

Flatla, D. R., Gutwin, C., Nacke, L. E., Bateman, S., & Mandryk, R. L. (2011). Calibration games: Making calibration tasks enjoyable by adding motivating game elements. *UIST'11 - Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 403–412. https://doi.org/10.1145/2047196.2047248.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 3025–3034. https://doi.org/10.1109/HICSS.2014.377.

Haslam, J. (2019, November 1). Fitness trackers see a boost in January — but do users keep their resolutions? Adjust, Blog. https://www.adjust.com/blog/health-tracker-installs-and-retention-data/.

Howard, J. (2016, July 14). Pokemon 'no': 7 other games to get you moving instead of Pokemon Go. *CNN Health*. https://edition.cnn.com/2016/07/14/health/pokemon-go-exercise-games-apps/index.html.

Johnson, D., Deterding, S., Kuhn, K. A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions*, *6*, 89–106. https://doi.org/10.1016/j.invent.2016.10.002.

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. https://doi.org/10.1037/0033-2909.125.6.627

Kruglanski, A. W., Fishbach, A., Woolley, K., Bélanger, J. J., Chernikova, M., Molinario, E., & Pierro, A. (2018). A structural model of intrinsic motivation: On the psychology of means-ends fusion. Psychological Review, 125(2), 165–182. https://doi.org/10.1037/rev0000095.

Kruglanski, A. W., Riter, A., Amitai, A., Margolin, Bath-Shevah Shabtai, L., & Zaksh, D. (1975). Can money enhance intrinsic motivation? A test of the content-consequence hypothesis. Journal of Personality and Social Psychology, 31(4), 744–750. https://doi.org/10.1037/0022-3514.31.4.744.

Lee, I.-M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., & Katzmarzyk, P. T. (2012). Impact of physical inactivity on the world's major non-communicable diseases. The Lancet, 380(9838), 219–229. https://doi.org/10.1016/S0140-6736(12)61031-9.Impact.

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. New Media and Society, 20(3), 881–900. https://doi.org/10.1177/1461444816675438.

Lister, C., West, J. H., Cannon, B., Sax, T., & Brodegard, D. (2014). Just a fad? Gamification in health and fitness apps. Journal of Medical Internet Research, 16(8), e9. https://doi.org/10.2196/games.3413.

Mathwick, C., & Rigdon, E. (2004). Play, flow, and the online search experience. Journal of Consumer Research, 31(2), 324-332.

Peng, W., Lin, J., Pfeiffer, K. A., & Winn, B. (2012). Need satisfaction supportive game features as motivational determinants: An experimental study of a self-determination theory guided exergame. Media Psychology, 15(2), 175–196. https://doi.org/10.1080/15213269.2012.673850.

Reeves, B., & Read, J. L. (2009). Total Engagement: Using Games and Virtual Worlds to Change the Way People Work and Businesses Compete. Harvard Business Review Press.

Romeo, A., Edney, S., Plotnikoff, R., Curtis, R., Ryan, J., Sanders, I., Crozier, A., & Maher, C. (2019). Can smartphone apps increase physical activity? Systematic review and meta-analysis. Journal of Medical Internet Research, 21(3), e12053. https://doi.org/10.2196/12053.

Ryan, R. M., & Deci, E. L. (2017). Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. Guilford Press.

Schlager, T., Müller-Stewens, J., Sprott, D., Häubl, G., & Herrmann, A. (2017). Gamification Paradigm (Working paper). Université de Lausanne. https://www.unil.ch/dmk/en/home/menuinst/research-1/working-papers.html

Six to Start & Naomi Alderman. (2020). Zombies, Run! Presskit Factsheet. https://zombiesrungame.com/presskit/, (accessed: May 29, 2020).

Statista. (2020, November 23). Wearables unit shipments worldwide by vendor from 2014 to 2019. https://www.statista.com/statistics/515634/wearables-shipments-worldwide-by-vendor/.

Teixeira, P. J., Silva, M. N., Mata, J., Palmeira, A. L., & Markland, D. (2012). Motivation, self-determination, and long-term weight control. International Journal of Behavioral Nutrition and Physical Activity, 9(21). https://doi.org/10.1186/1479-5868-9-22.

Thorsteinsen, K., Vittersø, J., & Svendsen, G. B. (2014). Increasing physical activity efficiently: An experimental pilot study of a website and mobile phone intervention. International Journal of Telemedicine and Applications, 2014. https://doi.org/10.1155/2014/746232.

World Health Organization. (2018). Global action plan on physical activity 2018–2030: more active people for a healthier world. https://doi.org/10.1016/j.jpolmod.2006.06.007.

Zuckerman, O., & Gal-Oz, A. (2014). Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Personal and Ubiquitous Computing, 18(7), 1705–1719. https://doi.org/10.1007/s00779-014-0783-2.

# Article III

## The Effect of a Future-Self Avatar mHealth Intervention on Physical Activity and Food Purchases: The FutureMe Randomized Controlled Trial

Annette Mönninghoff[1,2], Klaus Fuchs[3], Jing Wu[4], Jan Albert[4], Simon Mayer[4]

[1]Institute for Mobility, University of St.Gallen, St.Gallen, Switzerland
[2]Institute for Customer Insight, University of St.Gallen, St.Gallen, Switzerland
[3]Center for Artificial Intelligence, ETH Zurich, Switzerland
[4]Institute of Computer Science, University of St.Gallen, Switzerland

## Abstract

***Background*:** Insufficient physical activity and unhealthy diets are contributing to the rise in non-communicable diseases. Preventative mobile health (mHealth) interventions may enable reversing this trend, but present bias might reduce their effectiveness. Future-self avatar interventions have resulted in behavior change in related fields, yet evidence whether such interventions can change health behavior is lacking.

***Objective*:** Our primary objectives are to investigate the impact of a future-self avatar mHealth intervention on physical activity and food purchasing behavior, and to examine the feasibility of a novel automated nutrition tracking system. We also aim to understand how this intervention impacts related attitudinal and motivational constructs.

***Methods*:** We conducted a 12-week parallel randomized-controlled trial (RCT), followed by semi-structured interviews. German-speaking smartphone users aged ≥18 years living in Switzerland, and using at least one of the two leading Swiss grocery loyalty cards, were recruited for the trial. Data were collected from November 2020 to April 2021. The intervention group received the FutureMe intervention—a physical activity and food purchase tracking mobile phone application that uses a future-self avatar as the primary interface and provides participants with personalized food basket analysis and shopping tips. The control group received a conventional, text- and graphic-based primary interface intervention. We pioneered a novel system to track nutrition leveraging digital receipts from loyalty card data analyzing food purchases in a fully automated way. Data were consolidated in 4-week intervals and non-parametric tests were conducted to test for within- and between-group differences.

***Results*:** We recruited 167 participants; 95 eligible participants were randomized into either the intervention (n=42) or control group (n=53). The median age was 44.00 years (IQR 19.00), and the gender ratio was balanced (female 52/95, 55%). Attrition was unexpectedly high with only 30 participants completing the intervention, negatively impacting the statistical power of our study. The FutureMe intervention led to directional, small increases in physical activity (median +242 steps/day) and to directional improvements in the nutritional quality of food purchases (median –1.28 British Food Standards Agency Nutrient Profiling System Dietary Index points) at the end of the intervention. Intrinsic motivation significantly increased ($P$=.03) in the FutureMe group, but decreased in the control group. Outcome expectancy directionally increased for the FutureMe group, but decreased for the control group. Leveraging loyalty card data to track the nutritional quality of food purchases was found to be a feasible and an accepted fully automated nutrition tracking system.

***Conclusions*:** Preventative future-self avatar mHealth interventions promise to encourage improvements in physical activity and food purchasing behavior in healthy population groups. A full-powered RCT is needed to confirm this preliminary evidence and to investigate how future-self avatars might be modified to reduce attrition, overcome present bias, and promote sustainable behavior change.

***Keywords*:** mHealth, preventative medicine, avatar, present bias, nutrition tracking, physical activity, randomized controlled trial.

# Introduction

### Non-Communicable Diseases and Risk Factors

Non-communicable diseases (NCDs) are today's leading cause of death. They are responsible for over half of the years lived in disability and drive 59-80% of total health care costs in affluent, developed countries like New Zealand or Switzerland [1–3]. Among others, insufficient physical activity and an unhealthy, sodium-rich diet are recognized as key risk factors for NCDs by the World Health Organization (WHO) [4]. Increased salt consumption has been shown to increase blood pressure, while sodium-rich, low-fruit and low-whole-grain diets may increase the risk of death from cardiovascular diseases, cancer, and diabetes [5–7]. Insufficient physical activity has been linked to an increased risk of cancer, cardiovascular diseases, diabetes, dementia, and depression [8–11].

To combat the rise in NCDs, the WHO has released a global action plan providing clear recommendation on the required physical activity levels and maximum salt consumption [4]. However, the latest studies show that globally, 28% of adults do not meet the WHO's physical activity guidelines [12]. Moreover, salt intake was approximately twice the recommended amount in 2010 [13], and has not improved significantly since [14].

### Physical Activity and Salt Reduction mHealth Interventions

Wide-scale prevention programs are needed to make significant improvements toward realizing the WHO's NCD action plan. Mobile health (mHealth) interventions can make an important contribution toward efficiently reaching the wider population. While numerous physical activity mHealth interventions have proven effective in a wide range of populations [15–18], their effectiveness in decreasing salt consumption is less clear [19]. Preliminary results suggest that while mHealth interventions might lower participants' salt consumption by raising awareness, the quality of the evidence gathered in such studies is very low. Scholars have also criticized the lack of rigorous research designs [19]. In addition, mHealth nutrition interventions, such as those aiming to reduce sodium intake, regularly experience low user adoption and high attrition [20–22].

A review of commercially available nutrition-tracking mHealth application shows that most such interventions rely on manual or semi-automated tracking technologies [23], and require significant time investment from users, favoring underreporting [24] and high attrition rates as a result. There is a lack of studies employing improved technological means such as automated digital receipt processing [25,26], image analysis, or sensor-based tracking. Thus, there is a need for high-quality research that evaluates automated tracking systems in rigorous research settings [23,27].

### Challenges of Preventative mHealth Programs

Preventative mHealth interventions (ie, mHealth interventions targeting healthy population groups) have shown to be less effective than curative mHealth interventions (ie, mHealth interventions targeting sick or at-risk population groups) [15,18]. The reasons behind this moderating effect are manifold and remain little explored. The Health Action Process Approach (HAPA) is a widely used and validated theoretical framework that can serve as a starting point for explaining differences in intervention effectiveness [28]. According to the associated model, health behavior change is best described as a two-phase process consisting of a pre-intentional motivational phase and a post-intentional volitional phase [28]. Within the first phase, risk awareness (ie, the awareness that one’s behavior may cause a health risk), positive outcome expectancies (ie, the belief that a change in behavior will lead to reduced health risk), and action self-efficacy (ie, the belief in one’s ability to perform preventative health behavior) are fundamental for intention formation [28].

However, present bias may cause low risk awareness and decrease positive outcome expectancies in healthy population groups and thus limit the effectiveness of preventative mHealth interventions [29]. Present bias describes a situation where people take sub-optimal decisions as they over-discount future gains [29–32]. When facing preventative health decisions, consumers are asked to make unpleasant behavior changes today (ie, eating more vegetables instead of meat, exercising instead of watching movies), while the benefits of their behavior occur in the future (ie, better future health, longer life expectancy). In reverse, the negative consequences of a risky lifestyle are not immediately evident but only occur with a time lag, hindering the formation of positive outcome expectancies (ie, low symptom salience) [33].

### Future-Self Avatars to Overcome Present Bias and Increase Risk Awareness

Existing research argues that immediate monetary incentives can help overcome present bias in health decision making and improve adaptation of preventative mHealth interventions [34–36]. However, extrinsic monetary incentives might lead to motivational crowding out and diminish intrinsic motivation for behavior change over time [37,38].

Alternative strategies to overcome present bias and to increase risk awareness have been developed in related fields, such as pension planning [31] and taxation [39]. One little-explored possible solution are future-self avatars (ie, avatars that confront users with an aged version of themselves). Such avatars have been shown to increase users' tendency to accept later gains over immediate ones [31]. Evidence on the impact of avatars on health decision making is still scarce but preliminary results indicate that avatars can be leveraged to increase physical activity in obese populations [40], to promote smoking cessation in young adults [41], to reduce sodium intake [42], and to help manage depression [43]. To date, however, no study has investigated the impact of a preventative future-self avatar mHealth intervention in a randomized controlled trial (RCT) setting.

### Objectives

Accordingly, the primary aim of this study is twofold: first, to understand the impact of a preventative future-self avatar mHealth intervention on physical activity and on the nutritional quality of food purchases; second, to test the acceptability of a fully-automated nutrition tracking system leveraging grocery receipts in an RCT. Our secondary aim is to understand the impact of a future-self avatar preventative mHealth intervention on motivational and recovery self-efficacy, outcome expectancy, and types of motivation.

## Methods

### Study Design

We conducted our study in cooperation with a large Swiss health insurance company that facilitated participant recruitment. A 12-week parallel RCT design was employed, followed by semi-structured interviews with a selection of 15 participants who completed the trial. Recruitment began in November 2020 and ended in December 2020. Data were collected from November 2020 to April 2021. Potential participants were contacted via the insurance company's email newsletter and informed about the study, eligibility criteria, and data privacy policies. Additionally, social media networks were leveraged to recruit trial participants. Potential participants received a direct link to the iOS or Android app store where they could download the app-based intervention.

After downloading the mobile app, answering a survey on eligibility criteria, and consenting to the data privacy statement, participants were randomized into an intervention or a control group (1:1 ratio) via a random allocation algorithm programmed into the app. Participants did not receive any information about the existence of two different app versions (intervention and control).

Apart from the in-depth interviews, data collection was fully-automated, thus preventing reporting, detection, and performance bias. The semi-structured interviews were conducted by trained researchers who were neither the primary researchers nor the authors of this study. Selection of interview participants was blinded, meaning that interviewers had no knowledge of the interviewees' group allocation nor the outcome data collected in the RCT. The trial has been registered on ClinicalTrials.gov (NCT04505124), and the study is reported in accordance with the CONSORT guidelines [44].

**Study Population**

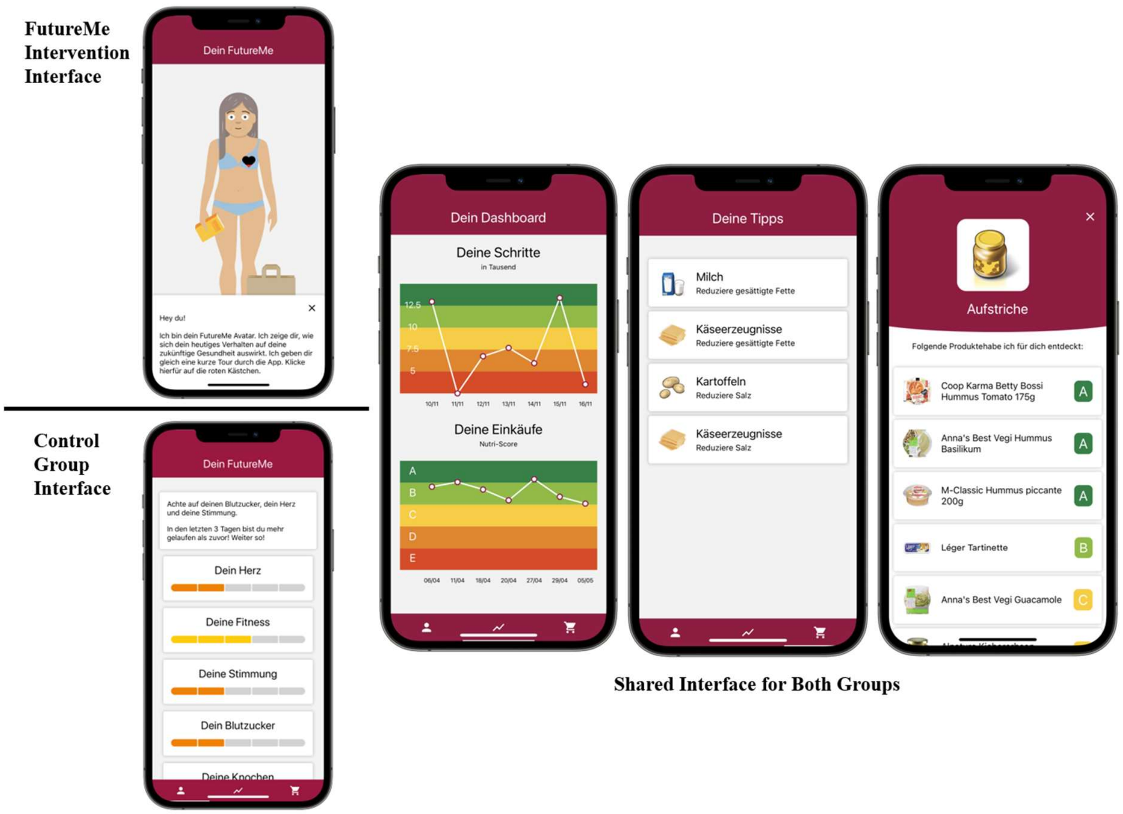
Eligible participants were smartphone users aged ≥ 18 years, living in Switzerland, and German-speaking. Participants had to be participating in at least one of the two leading grocery loyalty card programs in Switzerland. They were asked to confirm that they had no medical condition that prevented increased levels of physical activity or changes in daily nutrition.

**FutureMe Intervention**

The FutureMe app aims to promote the overall nutritional quality of participants' food purchases, and to encourage increases in physical activity to ≥7,500 steps per day, based on prior findings that health benefits can be achieved at this level [45–47]. Participants allocated to the FutureMe intervention received the FutureMe app. This comprised a future-self avatar, a dashboard displaying the number of daily steps and the nutritional quality of food purchases, a personalized basket analysis showing food categories with the highest improvement potentials, and personalized product recommendations of healthier product alternatives within the food categories purchased by the participant (Figure 1).

Product recommendations were meant to help users make healthier choices by substituting frequently consumed and rather unhealthy food items with alternatives of higher nutritional quality, without requiring users to modify the structure of their diets. All app components were updated daily, ensuring that participants received the most recent data. The intervention was fully automated and required no human-to-human interaction, making it highly scalable, tailored, and cost-effective [15].

**Figure 1:** Overview of The FutureMe and Control Intervention Mobile Application




*Future-self Avatar*

The future-self avatar was designed based on the HAPA behavior change model and encouraging findings of a prior pilot study by Fuchs et al [42]. The future-self avatar aims to increase participants' awareness of the future health consequences of their current nutritional and activity behaviors in a fun and engaging way (cf, risk perception, outcome expectancy, and self-efficacy [28]). It thus seeks to minimize the negative impact of present bias [31].

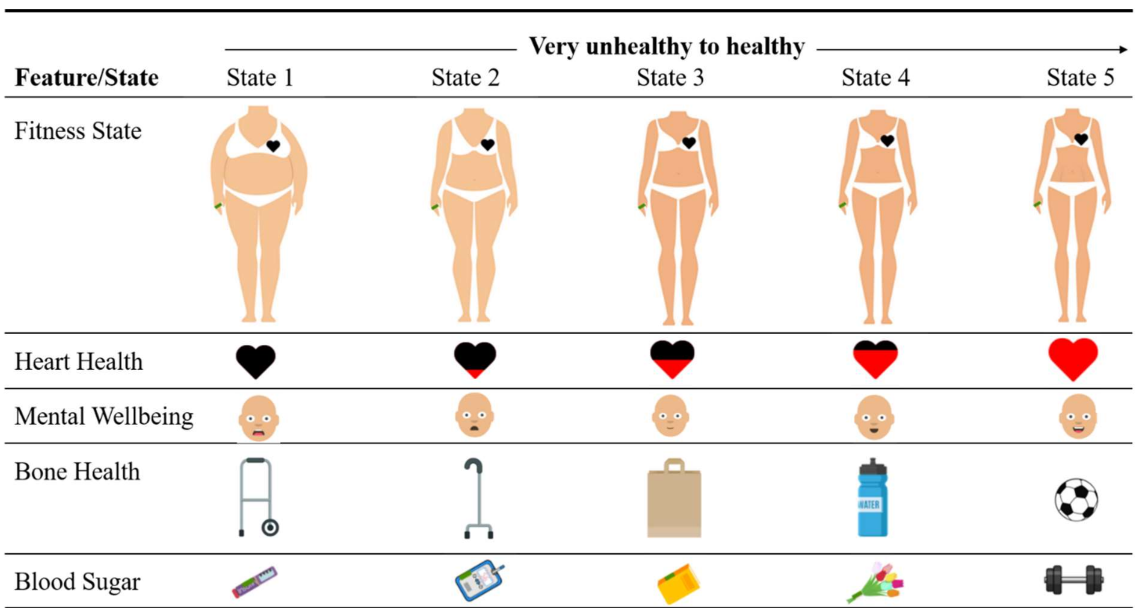
Participants could personalize their avatar during the sign-up process to ensure that the avatar best resembled themselves, in order to increase personal relevance and to positively impact health behavior change [48]. The avatar was depicted using a cartoon-style, unanimated 2D design to avoid feelings of eeriness and repulsion as described by the uncanny valley effect [49]. When first opening the app, participants were exposed to an aging simulation in which their avatar aged +20 years compared to their current age. The future-self avatar had 5 features that consumers could affect with their food purchases and physical activity behavior, with each feature having 5 different states (see Table 1 and Figure 2).

**Table 1:** System Rules Defining the Feature States of The Future-self Avatar

| Feature | Behavioral Influencer | Weighting | Calculation Scheme | Unit |
|---|---|---|---|---|
| Fitness State | Physical Activity | 100% | Average past 7 days | Steps/day |
| Heart Health | Sodium | 50% | Average past 12 baskets | FSA-NPS-DI points[a] |
| | Saturated fatty acids | 50% | Average past 12 baskets | |
| Mental Well-being | Fruit & Vegetable | 50% | Average past 12 baskets | FSA-NPS-DI points[a] |
| | Fiber | 50% | Average past 12 baskets | |
| Bone Health | Proteins | 100% | Average past 12 baskets | FSA-NPS-DI points[a] |
| Blood Sugar | Sugars | 100% | Average past 12 baskets | FSA-NPS-DI points[a] |

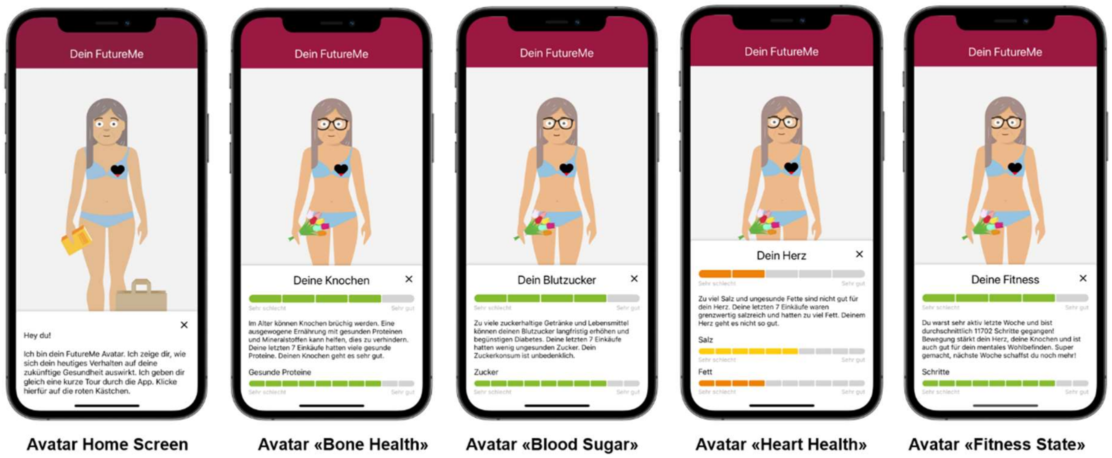
[a] British Food Standards Agency Nutrient Profiling System Dietary Index points (FSA-NPS-DI points) [50].

**Figure 2:** Avatar Feature States




The 5 features were clickable. If clicked, a pop-up screen opened, providing users with consequential health feedback and how their individual past behavior had impacted the feature. Participants were provided with an encouraging message to improve the respective health behavior (eg, *Feature Fitness State* "You were very active last week and took an average 11,702 steps/day! Physical activity improves your heart health, strengthens your bones, and also benefits your mental well-being! Well done, next week you can do even more"; Figure 3).

**Figure 3:** Overview of Avatar Features




*Fully Automated Nutrition and Physical Activity Tracking*

Nutrition and physical activity tracking within the FutureMe intervention app pioneered a novel and fully automated tracking system. Food purchases recorded via loyalty cards were used as a proxy for nutrition, as South African studies leveraging credit card data have shown this to be a promising method for changing everyday health behavior [25,26]. We gathered food purchase data using the loyalty cards from Switzerland's two leading grocery retailers. Data were analyzed using a comprehensive product database comprising over 55,000 products and their nutritional information. Participants were requested to connect their loyalty cards once while signing up. Step data were collected by the accelerometer on participants' mobile phone.

## Control Intervention

Apart from the future-self avatar, the control intervention was identical to the FutureMe intervention (see Figure 1). Instead of the future-self avatar, the control app used a more traditional text- and graphics-based interface to provide consequential health feedback to participants (see Figure 1). The control intervention used the same automated data tracking system as the FutureMe intervention.

## Outcomes and Measurement

Demographic information as well as attitudinal and motivational constructs were collected at baseline (T0), and after completing the 12-week intervention (T3), using an online survey that was programmed into the mHealth app. Step and food purchase data were collected at baseline (T0) and continuously throughout the 12-week intervention period using the fully automated tracking system previously described. At baseline, we collected steps per day for the past 6 days prior to sign-up and for all grocery baskets of the past 2 years that were registered on the loyalty cards. User engagement was recorded continuously for 12 weeks after completing the sign-up process. Continuously collected data were aggregated into three 4-week periods (T1, T2, T3). Qualitative data were collected from a group of randomly selected participants in semi-structured video interviews after the 12-week intervention period (T3).

The primary outcomes were as follows:

1. Physical activity (in steps/day) was measured objectively to avoid reporting biases [51]. Step information was collected through the built-in accelerometer in iOS or Android mobile phones as these devices are widely owned and require no additional investment [52], and thus offer a scalable solution to track physical activity. Mobile phones have been shown to reliably measure steps if carried continuously by participants, and when steps are taken at modest to fast pace [52–54].
2. The nutritional quality of food purchases was measured based on the British Food Standards Agency Nutrient Profiling System [50,55] (in FSA-NPS-DI points; –15 most healthy to +40 least healthy) [56]. Nutritional quality was calculated by separating solid foods and beverages, in accordance with the calculation method suggested by Julia et al [57]. We only report results for solid foods.

The secondary outcomes were the nutritional sub-categories comprising the Nutri-Score framework (Outcomes 3-8) [58], user engagement (Outcome 9), and attitudinal and motivational constructs (Outcomes 10-14):

1. Sugars, excluding fructose and lactose, in grams (g) per 100g food purchases (in FSA-NPS-DI points; 0 most healthy to +10 least healthy).
2. Saturated fatty acids in g per 100g food purchases (in FSA-NPS-DI points; 0 most healthy to +10 least healthy).
3. Sodium in mg per 100g food purchases (in FSA-NPS-DI points; 0 most healthy to +10 least healthy).

4. Fruit, vegetables, legumes, and nuts in % per 100g food purchases (in FSA-NPS-DI points; 0 least healthy to +5 most healthy).
5. Fibers in g per 100g food purchases (in FSA-NPS-DI points; 0 least healthy to +5 most healthy).
6. Proteins in g per 100g food purchases (in FSA-NPS-DI points; 0 least healthy to +5 most healthy).
7. User engagement (in app logins/week).
8. Motivational self-efficacy, measured with an adjusted four-item scale adopted from Schwarzer et al [59] on a 7-point Likert scale (1 completely disagree to 7 completely agree). The scale measured participants' confidence in their capability to be physically active and shop healthily in general (eg, "In my everyday life, I know how to shop healthily," Appendix 1). Internal consistency was Cronbach's α=.728.
9. Recovery self-efficacy, measured with an adjusted two-item scale adopted from Schwarzer et al [59] on a 7-point Likert scale (1 completely disagree to 7 completely agree). The scale measured participants' confidence in their capability to be physically active and shop healthily after a setback (eg, "After my vacation, I'm sure that I'll go back to balanced shopping, even if I have to get used to it again," Appendix 1). Internal consistency was Cronbach's α=.652.
10. Outcome expectancy, measured with an adjusted six-item outcome expectancy scale adopted from Renner and Schwarzer [60] on a 7-point Likert scale (1 completely disagree to 7 completely agree) that measured participants' expectations about how their present behavior will impact their future health (eg, "I believe that I can positively influence my health in old age with my current exercise and shopping behavior," Appendix 1). Internal consistency was Cronbach's α=.934.
11. Intrinsic motivation, measured with an adjusted three-item autonomous motivation scale adopted from Williams et al [61] on a 7-point Likert scale (1 completely disagree to 7 completely agree). The scale measured the degree to which participants were intrinsically motivated to be physically active and shop

healthily (eg, “I exercise and shop healthily because I personally believe it’s best for my health,” Appendix 1). Internal consistency was Cronbach’s α=.838.

12. Extrinsic motivation, measured with an adjusted three-item controlled motivation scale adopted from Williams et al [61] on a 7-point Likert scale (1 completely disagree to 7 completely agree). The scale measured the degree to which participants were extrinsically motivated to be physically active and shop healthily (eg, “I exercise and shop healthily because I want to see positive metrics on my activity tracker and health app,” Appendix 1). Internal consistency was Cronbach’s α=.690.

In addition to the quantitative outcomes, we collected qualitative data in 15 semi-structured video interviews. Data were gathered on app usefulness and ease of use as these factors have been shown to impact technology acceptance [62]. Furthermore, data were collected on the relationship that participants formed with the avatar and on areas for improving the FutureMe intervention in order to guide future research.

**Sample Size**

No other studies have so far compared a future-self avatar interface with a more conventional self-monitoring interface regarding their ability to motivate both physical activity increases and nutrition improvements in a healthy population sample. Given the limited availability of mHealth studies reporting Nutri-Score improvements based on shopping data, we focused on results from comparable mHealth physical activity studies to calculate sample size.

A systematic review of mHealth physical activity interventions found the step increase at the end of the intervention in a mixed population sample at 1566 steps [15] compared to both active and passive control groups. Baseline activity measurements for healthy population samples have been reported as ranging from 6745 to 6994 (2422 to 2620 SD) steps/day [63,64]. Using an α of 0.05 with a power of 80%, we estimated the minimum sample size for our trial to be 74-88 participants [65]. Dropout rates in mHealth studies vary significantly, depending on intervention duration and intervention components. A comparable 12-week mHealth physical activity study using scalable intervention components without human-to-human interaction reported a dropout-rate of 22.4% [66]. Anticipating a 20% dropout rate, we inflated the sample size to 100.

## Statistical Analysis

Nonparametric tests were conducted to examine intervention effects. A Wilcoxon signed-rank test was used to compare the outcome variables before and after the intervention to test for within-group effects. To test for between-group differences between the FutureMe and the control group, we conducted Mann-Whitney-U tests or Pearson Chi-Square tests. The level of significance was .05. Analyses were performed with SPSS Statistics 27 and Python 3.8.5.

Continuously measured data (physical activity, nutritional quality of food purchases, nutritional sub-categories, and user engagement) were analyzed in 4-week periods with T1 representing the mean value summarizing weeks 1-4, T2 summarizing weeks 5-8, and T3 summarizing weeks 9-12. Baseline values (T0) were also summarized depending on the outcome variable and data availability. For physical activity, T0 represents the mean steps per day of the 6 days prior to trial commencement. For the nutritional quality of food purchases and the nutritional sub-categories, T0 represents all foods purchased within 4 weeks prior to trial commencement.

Sensitivity analysis was conducted for demographic variables and primary outcomes to test for randomness of missing data and attrition bias [67,68]. Missing data were not replaced given the limitations of imputation methods in samples with high attrition where data are missing at random [67]. To test for intervention effects and statistical differences, we used complete case analysis.

## Qualitative Data Analysis

All 15 semi-structured video interviews were audio-recorded, fully transcribed, coded, and analyzed following the principles of thematic content analysis [69]. Additionally, demographic information on all 15 interview candidates was quantitatively summarized.

## Ethics

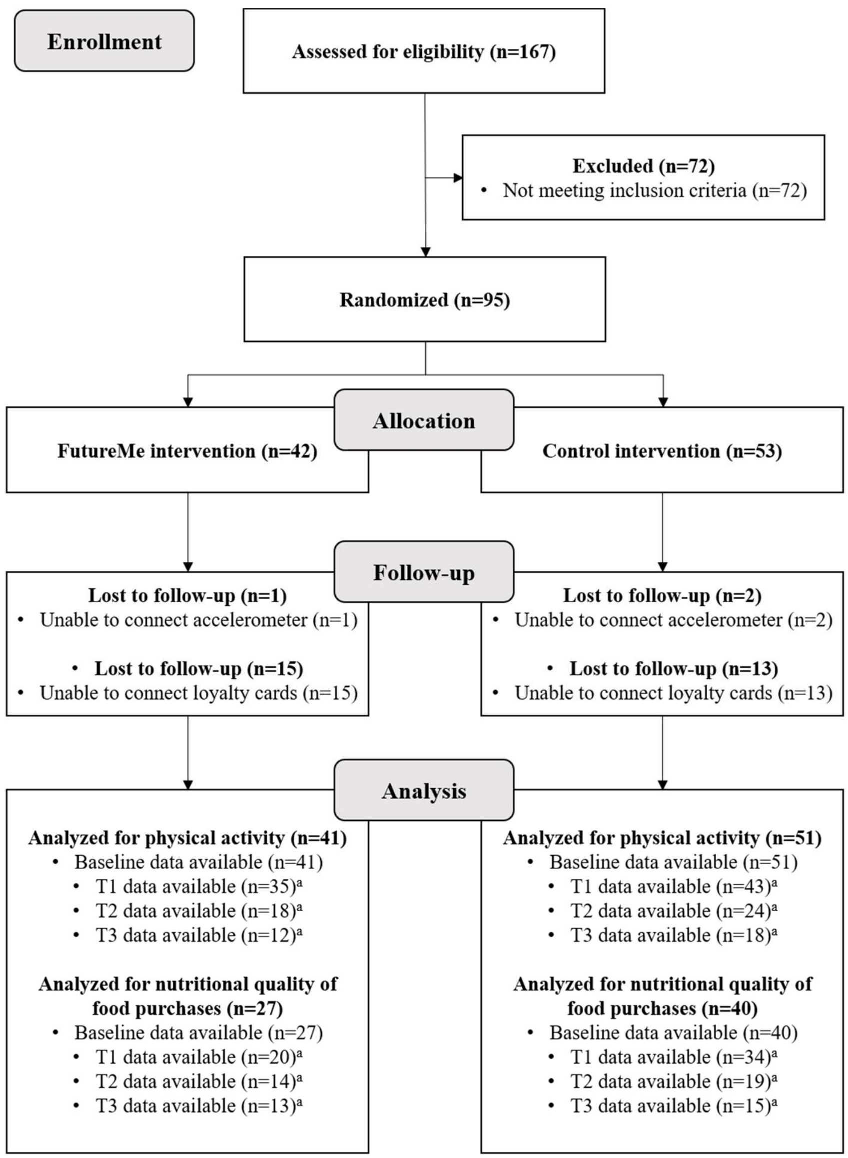
Participants gave their electronic informed consent to participate in the study prior to commencement. The Ethics Committee of the University of St.Gallen approved the protocol, recruitment strategies, and data privacy policies of this study (HSG-EC-2020-06-12-A).

# Results

## Demographic Data and Baseline Characteristics

As shown in the CONSORT flow chart (Figure 4), 167 participants downloaded the FutureMe/Control application from November 11 to December 15, 2020. Seventy-two participants were excluded as they did not meet the inclusion criteria. Ninety-five participants were randomized to either the FutureMe intervention (n=42) or the control intervention (n=53).

**Figure 4:** Consort Flow Chart




[a] Limited data available at T1-3 as participants of both groups stopped using the FutureMe/control app. Data was only automatically pulled when users logged into the app.

One participant in the intervention group and two participants in the control group were lost as they were unable to connect their mobile phone accelerometer to the intervention/control application. Fifteen participants in the intervention group and 13 participants in the control group were lost as they were unable to connect their loyalty cards to the FutureMe app. We saw unexpectedly high attrition during the 12-week intervention period (68.42%, n=95), with only 13 participants of the intervention group and 18 participants of the control group using the app during the last 4 weeks of the trial. As shown in Table 2, the median age of participants was 44 years (IQR 19.00). The gender ratio was balanced with 55% female participants (52/95). Median household size was 2 (IQR 2.00) and most participants had normal body weight (59%, 56/95). The majority of participants were recruited into the trial through the online panel of a Swiss health insurance company (80%, 76/95). No baseline differences existed between the intervention and the control group in regard to demographic data.

**Table 2:** Demographic Data at Baseline

| Demographic data | Total population (N=95) | Intervention group (n=42) | Control group (n=53) | *P* Value |
|---|---|---|---|---|
| **Gender, n (%)** | | | | .997[b] |
| Male | 43 (45%) | 19 (45%) | 24 (45%) | |
| Female | 52 (55%) | 23 (55%) | 29 (55%) | |
| **Age in years, median (IQR)** | 44.00 (19.00) | 47.00 (15.00) | 42.00 (25.00) | .46[c] |
| **BMI[a], n (%)** | | | | .96[b] |
| >30 | 11 (12%) | 5 (12%) | 6 (11%) | |
| 25 - 30 | 28 (29%) | 12 (29%) | 16 (30%) | |
| <25 | 56 (59%) | 25 (60%) | 31 (58%) | |
| **HH Size, median (IQR)** | 2.00 (2.00) | 2.00 (2.00) | 3.00 (2.00) | .61[c] |
| **Recruiting source, n (%)** | | | | .50[b] |
| Helsana | 76 (80%) | 33 (79%) | 43 (81%) | |
| University | 5 (5%) | 1 (2%) | 4 (8%) | |
| Social Media | 6 (6%) | 3 (7%) | 3 (6%) | |
| Other | 8 (8%) | 5 (12%) | 3 (6%) | |

[a]BMI: Body Mass Index

[b] Pearson Chi-Square test was performed to detect statistical differences.

[c] Mann-Whitney U Test was performed to detect statistical differences.

At baseline, the median steps per day were 4624.00 (IQR 4497.79), that is, roughly half the number of steps per day compared to a representative Swiss sample [70]. The median nutritional quality of food purchases was 6.13 FSA-NPS-DI points (IQR 4.02), which is in line with the values reported in a large-scale study [71]. Food purchases were low-sugar (median 0.88 FSA-NPS-DI points (IQR 1.20)) and contained moderate amounts of saturated fatty acids (median 3.01 FSA-NPS-DI points (IQR 1.66)), sodium (median 2.50 FSA-NPS-DI points (IQR 1.40)), and protein (median 3.16 FSA-NPS-DI points (IQR 0.93)). The share of fruits, vegetables, legumes, nuts, and fibers was rather low with a median of 0.30 (IQR 0.48) and 1.53 FSA-NPS-DI points (IQR 0.77), respectively.

No baseline differences were found between the intervention and control groups in regard to physical activity or the total nutritional quality of food purchases. Participants in the FutureMe intervention had significantly higher amounts of saturated fatty acids in their food purchases (*P*=.03) than the control group. Overall, participants showed high levels of motivational and recovery self-efficacy (median 6.0 (IQR 1.25); median 6.0 (IQR 1.50)), as well as outcome expectancy (median 6.2 (IQR 1.60)). Participants in the intervention group scored significantly higher on recovery self-efficacy than those in the control group (*P*=.03). Participants exhibited high intrinsic motivation to live a healthy lifestyle (median 6.0 (IQR 1.33)) and were less extrinsically motivated (median 3.0 (IQR 2.33)). Baseline characteristics in regard to all outcomes are presented in Table 3.

**Table 3:** Primary and Secondary Outcomes at Baseline

| Primary and secondary outcomes | Total population | n | Intervention group | n | Control group | n | *P* Value[e] |
|---|---|---|---|---|---|---|---|
| **Physical activity, median (IQR)** | | | | | | | |
| Walking (in steps/day) | 4624.00 (4497.79) | 92 | 4314.83 (3620.33) | 41 | 5084.83 (5220.33) | 51 | .53 |
| **Nutritional quality of food purchases, median (IQR)** | | | | | | | |
| Total Nutritional quality (in FSA-NPS-DI points)[a] | 6.13 (4.02) | 67 | 6.73 (5.03) | 27 | 5.83 (4.11) | 40 | .14 |
| **Nutritional sub-categories, median (IQR)** | | | | | | | |
| Sugars (in FSA-NPS-DI points)[b] | 0.88 (1.20) | 67 | 0.83 (1.08) | 27 | 0.91 (1.31) | 40 | .73 |
| Saturated fatty acids (in FSA-NPS-DI points)[b] | 3.01 (1.66) | 67 | 3.82 (2.83) | 27 | 2.93 (1.37) | 40 | .03 |
| Sodium (in FSA-NPS-DI points)[b] | 2.50 (1.40) | 67 | 2.77 (1.56) | 27 | 2.42 (1.37) | 40 | .42 |
| Fruit, vegetables, legumes and nuts (in FSA-NPS-DI points)[c] | 0.30 (0.48) | 67 | 0.28 (0.53) | 27 | 0.31 (0.47) | 40 | .64 |
| Fibers (in FSA-NPS-DI points)[c] | 1.53 (0.77) | 67 | 1.49 (0.70) | 27 | 1.56 (0.83) | 40 | .85 |
| Protein (in FSA-NPS-DI points)[c] | 3.16 (0.93) | 67 | 3.20 (0.70) | 27 | 3.15(1.03) | 40 | .31 |
| **Attitudinal and motiational constructs, median (IQR)** | | | | | | | |
| Motivational self-efficacy[d] | 6.00 (1.25) | 95 | 6.00 (1.31) | 42 | 6.00 (1.50) | 53 | .65 |
| Recovery self-efficacy[d] | 6.00 (1.50) | 95 | 6.00 (1.50) | 42 | 6.00 (1.50) | 53 | .03 |
| Outcome expectancy[d] | 6.20 (1.60) | 95 | 6.10 (1.65) | 42 | 6.20 (1.60) | 53 | .31 |
| Intrinsic motivation[d] | 6.00 (1.33) | 95 | 5.83 (1.75) | 42 | 6.00 (1.50) | 53 | .13 |
| Extrinsic motivation[d] | 3.00 (2.33) | 95 | 2.67 (2.00) | 42 | 3.33 (2.00) | 53 | .65 |

[a] FSA-NPS-DI point scale: -15 most healthy to +40 least healthy.
[b] FSA-NPS-DI point scale: 0 most healthy to +10 least healthy.
[c] FSA-NPS-DI point scale: 0 least healthy to +5 most healthy.
[d] Likert scale: 1 completely disagree to 7 completely agree.
[e] Mann-Whitney U test was performed to detect statistical differences.

**Physical Activity**

As shown in Table 4, participants of both groups increased their physical activity levels within the first 4 weeks of the intervention (FutureMe T1 median 4577.85 steps/day (IQR 4620.26); Control T1 median 5361.50 steps/day (IQR 5960.88)). Participants in the intervention group consistently increased their physical activity over the course of the intervention from median 4314.81 steps/day (IQR 3620.33) at baseline to median 6042.31 steps/day (IQR 6242.31) at T2, and to median 4556.94 (IQR 5226.94) at the end of intervention. In contrast, in the control group, the number of steps per day decreased after a short-term increase in T1, from median 5084.83 (IQR 5220.33) steps/day at baseline to median 2909.32 steps/day (IQR 5128.09) at T2 and to median 4821.64 steps/day (IQR 5228.79) at the end of the intervention. However, for both groups, the changes in the number of steps over the course of the intervention were not statistically significant. Nor were any significant differences in steps/day evident between the intervention and control groups (Table 5).

## Nutritional Quality of Food Purchases

Participants in both groups improved the nutritional quality of their food purchases over the course of the intervention (Table 4), albeit insignificantly. Participants in the intervention group shopped less healthily within the first 4 weeks of the intervention (median 6.79 FSA-NPS-DI points (IQR 3.67)) compared to the baseline (median 6.73 FSA-NPS-DI points (IQR 5.03)). However, they improved their shopping behavior in T2 (median 6.14 FSA-NPS-DI points (IQR 2.63)) and at the end of the intervention (median 5.45 FSA-NPS-DI points (IQR 3.19)). Participants in the control group consistently improved their shopping behavior over the course of the intervention from median 5.83 FSA-NPS-DI points (IQR 4.11) at baseline to median 5.70 FSA-NPS-DI points (IQR 5.31) at T1, to median 4.42 FSA-NPS-DI points (IQR 5.42) at T3, and to median 4.86 FSA-NPS-DI points (IQR 3.75) at the end of the intervention. At the end of the intervention, the control group purchased significantly healthier food than the FutureMe group (*P*=.02). However, the control group's food baskets were already healthier at baseline, although not at statistically significant levels.

**Table 4:** Primary Outcomes by Timepoint and Group

| Primary outcome variables | 4-week period actuals | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Baseline (T0)[b] | n | T1[c] | n | T2[d] | n | EoS (T3)[e] | n |
| **Physical Activity, median (IQR)** | | | | | | | | |
| Walking (in steps/day) | | | | | | | | |
| Avatar | 4314.83 (3620.33) | 41 | 4577.85 (4620.26) | 35 | 6042.45 (6242.31) | 18 | 4556.94 (5226.94) | 12 |
| Control | 5084.83 (5220.33) | 51 | 5361.50 (5960.88) | 43 | 4909.32 (5128.09) | 24 | 4821.64 (5228.79) | 18 |
| | | | | | | | | |
| **Nutritional quality of food purchases, median (IQR)** | | | | | | | | |
| Total nutritional quality (in FSA-NPS-DI-points)[a] | | | | | | | | |
| Avatar | 6.73 (5.03) | 27 | 6.79 (3.67) | 20 | 6.14 (2.63) | 14 | 5.45 (3.19) | 13 |
| Control | 5.83 (4.11) | 40 | 5.79 (5.41) | 34 | 4.42 (5.42) | 19 | 4.86 (3.75) | 14 |

[a] FSA-NPS-DI point scale: -15 most healthy to +40 least healthy.
[b] Baseline: For steps baseline defined as average steps/day 6 days prior to enrolling in the trial;
for shopping-related outcome variables baseline is defined as nutritional value of all foods purchased within the 4 weeks before the trial.
[c] T1: Week 1-4 average values
[d] T2: Week 5-8 average values
[e] EoS: End of Study, defined as mean value week 9-12.

**Table 5:** Within- and Between-group Comparisons of Primary Outcomes by Timepoint

| Outcome variables | Within-group comparison vs. baseline | | | | | | Between-group comparison | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Baseline vs. T1 | | Baseline vs. T2 | | Baseline vs. EoS | | Baseline[b] | | T1[c] | | T2[d] | | EoS[e] | |
| | Z | *P* value | Z | *P* value | Z | *P* value | U | *P* value | U | *P* value | U | *P* value | U | *P* value |
| **Physical Activity, median (IQR)** | | | | | | | | | | | | | | |
| Walking (in steps/day) | | | | | | | 965 | .53 | 656 | .34 | 212 | .93 | 105 | .92 |
| Avatar | -0.3 | .79 | -0.283 | .80 | -0.39 | .73 | | | | | | | | |
| Control | -0.9 | .38 | -0.438 | .68 | -0.57 | .60 | | | | | | | | |
| | | | | | | | | | | | | | | |
| **Nutritional quality of food purchases, median (IQR)** | | | | | | | | | | | | | | |
| Total nutritional quality (in FSA-NPS-DI-points)[a] | | | | | | | 470 | .14 | 505 | .29 | 660 | .44 | 796 | .02 |
| Avatar | -1.8 | .07 | -0.299 | .77 | -0.56 | .58 | | | | | | | | |
| Control | -0.4 | .70 | -0.213 | .83 | -1.26 | .21 | | | | | | | | |

[a] FSA-NPS-DI point scale: -15 most healthy to +40 least healthy.
[b] Baseline: For steps baseline defined as average steps/day 6 days prior to enrolling in the trial;
for shopping-related outcome variables baseline is defined as nutritional value of all foods purchased within the 4 weeks before the trial.
[c] T1: Week 1-4 average values
[d] T2: Week 5-8 average values
[e] EoS: End of Study, defined as mean value week 9-12.
Comment: *P* value reported as exact significances due to small sample sizes.

## Nutritional Sub-Categories

Appendix 2 provides a summary of all secondary outcomes for all relevant timepoints. It also presents the results of the Wilcoxon Sign-Rank tests and the Mann-Whitney U tests. Across both groups, no statistically significant improvements were found in nutritional sub-categories during the intervention. As for directional developments, for the control group, improvements in nutritional quality of food purchases were driven by reductions in sugar per 100g food purchases between baseline (median 0.91 FSA-NPS-DI points (IQR 1.31)) and the end of the study (median 0.55 FSA-NPS-DI points (IQR 1.07)). For the intervention group, improvements in nutritional quality of food purchases were driven by reductions in saturated fatty acids and sodium between baseline (saturated fatty acids median 3.82 FSA-NPS-DI points (IQR 2.83); sodium median 2.77 FSA-NPS-DI points (IQR 1.56)) and the end of the study (saturated fatty acids median 3.82 FSA-NPS-DI points (IQR 2.83); sodium median 2.77 FSA-NPS-DI points (IQR 1.56)). For both groups, the amount of protein, fruit, vegetables, and fibers per 100g in food purchases was relatively stable throughout the trial or even decreased slightly across both groups. At the end of the intervention, the control group purchases significantly healthier food than the FutureMe group in regard to sugar (*P*=.01) and saturated fatty acids (*P*=.01). However, the control group's food baskets already contained significantly lower amounts of saturated fatty acids at the baseline (*P*=.03).

### User Engagement

User engagement was similar across the FutureMe and control interventions (see Appendix 2). Engagement was highest during the first 4 weeks of the intervention (T1) where the FutureMe and control groups recorded a median of five logins per week (IQR 10.00) and four logins per week (IQR 10.00), respectively. User engagement decreased to one login per week for the FutureMe group for T2 (IQR 3.00) and remained flat until the end of the study (IQR 1.00). For the control group, user engagement also dropped to a median of one login per week at T2 (IQR 3.00) but increased to two logins per week (IQR 2.00) at the end of the study.

### Attitudinal and Motivational Constructs

Appendix 2 provides a summary of all attitudinal and motivational constructs pre- and post-intervention. It also presents the results of the Wilcoxon Sign-Rank tests and the Mann-Whitney U tests. Intrinsic motivation to lead a healthy lifestyle significantly (*P*=.03) increased from a median value of 5.83 (IQR 1.75) at baseline to 6.0 (IQR 1.67) at the end of the study for participants in the FutureMe intervention. In contrast, intrinsic motivation decreased in the control group throughout the intervention. The extrinsic motivation to lead a healthy lifestyle increased in both groups from baseline (FutureMe median 2.67 (IQR 2.00); control median 3.33 (IQR 2.00)) to the end of the study (FutureMe median 3.67 (IQR 2.50); control median 4.00 (IQR 2.08)). However, the changes were not statistically significant. At the end of the study, the FutureMe group showed significantly (*P*=.03) higher levels of recovery self-efficacy (median 7.00; (IQR 1.00)) compared to the control group (median 5.67; (IQR 1.25)).

Motivational self-efficacy did not change significantly in either group. Outcome expectancy increased for the FutureMe group between baseline (median 6.10; (IQR 1.65)) and the end of the study (median 7.00; (IQR 2.08)). However, the change was not statistically significant. In contrast, outcome expectancy decreased in the control group between baseline (median 6.20; (IQR 1.60)) and the end of the study (median 5.67; (IQR 1.62)), although not at statistically significant levels.

## Sensitivity Analysis

To test for attrition bias and randomness of missing data, we compared demographic and primary outcome variables at baseline (T0) between participants who did not finish the trial (ie, dropouts) and participants who completed the trial (ie, trial completers). The results are presented in Table 6. We found no statistical differences between dropouts and trial completers with respect to our primary outcome variables, indicating a low risk of attrition bias. Comparing demographic variables revealed no statistical differences in regard to gender ratio, body mass index, or household size. Trial completers, however, were found to be significantly younger (*P*=.02) than dropouts. Overall, our sensitivity analysis confirmed that missing data can be considered completely random in our trial and that no systematic error can thus be expected [68]. The large amount of missing data, however, reduces the statistical power of our analysis [67,68].

**Table 6:** Results of The Sensitivity Analysis

| Demographics and primary outcomes | Dropouts | n | Trial completers | n | *P* Value[e] |
|---|---|---|---|---|---|
| **Gender, n (%)** | | 65 | | 30 | .66 |
| Male | 28 (43%) | | 15 (50%) | | |
| Female | 37 (57%) | | 15 (50%) | | |
| **Age in years, median (IQR)** | 48.00 (20.00) | 65 | 42.00 (18.00) | 30 | .02 |
| **BMI[a], n (%)** | | | | | .92[f] |
| >30 | 9 (14%) | 65 | 1 (3%) | 30 | |
| 25 - 30 | 18 (28%) | 65 | 11 (37%) | 30 | |
| <25 | 38 (58%) | 65 | 18 (60%) | 30 | |
| **HH Size, median (IQR)** | 2.00 (2.00) | 65 | 3.00 (2.00) | 30 | .54 |
| **Physical activity, median (IQR)** | | | | | |
| Walking (in steps/day) | 4897.83 (4286.13) | 64 | 4077.50 (5499.05) | 28 | .95 |
| **Nutritional quality of food purchases, median (IQR)** | 5.64 (3.87) | 42 | 6.38 (4.26) | 28 | .24 |
| Total nutritional quality (in FSA-NPS-DI points)[b] | | | | | |

[a] BMI: Body Mass Index.

[a] FSA-NPS-DI point scale: -15 most healthy to +40 least healthy.

[e] Mann-Whitney U test was performed to detect statistical differences.

[f] Mann-Whitney U test was performed comparing absolute BMI scores.

## Results of Qualitative Study

Out of all participants who finished the trial, 15 participants were randomly selected to participate in semi-structured video interviews. The selected participant demographics were similar to the overall trial demographics. Forty-seven percent of participants (7/15) were female. The median age was 34.00 (IQR 24.00) and the median household size was 2 (IQR 1.00). Fifty-three percent of participants (8/15) had received the FutureMe intervention, 47% (7/15) the control intervention.

### Ease of Use and Usefulness

While most participants experienced some challenges while signing up, particularly with connecting their loyalty cards to the application, they found using the app intuitive and easy. App functionalities were said to be simple and easily understandable.

> *The app is very simple, easy to read and intuitive to use. Really without a lot of bells and whistles or hidden features.* [User #13, female]

Overall, many users appreciated how easy the app was to use—after the initial setup—and that data synchronized automatically and effortlessly into the app.

> *Personally, I was motivated by the fact that the effort is relatively low, it is automatically synchronized, whether it is sports or shopping*. [User #1, male]

The app's observed usefulness depended largely on how well participants perceived that tracking reflected their actual physical activity and food purchasing behavior. Participants who shopped mainly at the two participating grocery retailers, and whose physical activity behavior mostly involved walking or running, rated app usefulness as very high. They felt that the app provided helpful insights, liked the color-guidance on the overview tracking screen and were inspired by the food tips.

> *Above all, to bring the bar up to green. As far as possible everywhere, but I haven't achieved that yet.* [User #8, male]
> *I just found it very exciting with the purchases. […] I had the feeling that I buy very consciously and healthily, and sometimes I was in the orange and red area, where I thought just because I bought a pizza once, it drags the whole curve*

*down. That is then nevertheless exciting and stimulates you to think about it and to consider whether it would not make sense to buy the wholewheat pizza dough or a vegan alternative instead of the normal pizza.* [User #9, male]

Participants who purchased their food at different grocery retailers or local markets, who regularly ate takeaways, or who shared the loyalty card with other family members found the app less useful. They felt that it did not reflect their actual behavior. They also stated that they would rate usefulness higher if tracking were more accurate.

*It does not reflect things as they are. Because I don't buy a lot of vegetables and fruit at these two retailers.* [User #14, male]
*I was annoyed that my son just once again bought chips and peanuts.* [User #5, female]

**Participant-Avatar Relationship**

Users who rated the app's tracking accuracy high reported that the avatar motivated behavior change and that checking-in on the avatar was a key reason for logging into the app.

*The avatar was super cool.* [User #6, male]
*One app feature that motivated me was the avatar. I always clicked on the different items and checked to make sure that I improve on all categories*. [User #4, female]

Users liked that they could personalize their avatar and mentioned that the avatar somewhat resembled them. In contrast, most participants did not much like the avatar's Mood feature (ie, the facial expression depending on fruit, vegetable, and fiber consumption). In particular when the avatar expressed sadness, users reported that this did not reflect how they felt. Also, more generally, users with less healthy food purchases or low physical activity behavior were more critical, with some users feeling discouraged by the avatar's challenging feedback. They felt that some of the avatar's reactions to their behavior were exaggerated and did not reflect their state of health. This suggests that users struggled to fully understand that their avatar did not reflect their current health but rather their future health.

*I could only partially identify with the avatar. I found especially that he gained weight very quickly. I couldn't understand why it was so fast. I had the feeling that he represented a worse image than my actual condition*. [User #11, male]

Users who rated the app's tracking accuracy as low identified less with the avatar.

*I mostly buy drinks and non-food items at the two participating retailers, so nothing that affects the avatar. Therefore, I can't identify with it.* [User #15, female]

**Areas of Improvement**

Users suggested improvements for the app's tracking functionality, its analysis, insights, and feedback, and incentive schemes. The most frequently mentioned area of improvement concerned broadening synchronized data sources and improving tracking accuracy. Although users appreciated that the automatic synchronization of food purchases required no personal effort, they wanted to be able to manually add data from other retailers, to include restaurant consumption, or to exclude foods purchased by other household members or unconsumed (food waste). Receipt scanning was mentioned multiple times as a preferred option for manually adding data, followed by drop-down functionalities. Further, users indicated that breaking down food purchases over a consumption period would improve app usefulness as certain items (eg, oil, salt, or condiments) are usually consumed over a longer period of time.

*It would be cool if you could actually control the food analysis in the app via consumption and not the purchase, by dividing a product according to days or months in which I consume it*. [User #1, male]

Regarding activity tracking, users felt that including physical activity beyond the number of steps and improving app compatibility with leading fitness watches would be valuable improvements. Another related improvement often mentioned was improving data synchronization speed. In the FutureMe trial, food and physical activity data were updated daily. However, as an app login was required to synchronize loyalty card data with the app, it took approximately two days for food purchases to appear in the app. Users mentioned that they would have preferred to have food purchases available immediately after shopping.

Regarding the app's analysis options, users liked the food tips, but felt that these were not updated regularly enough. Some users indicated that they would have liked to receive push notifications once new data were available in the app, or motivating push messages if their shopping or physical activity behavior deteriorated or improved. Furthermore, various participants mentioned that they would have liked to see the least healthy items in their shopping basket on a separate screen, as they felt that this would help to improve their shopping behavior, instead of merely receiving shopping tips for healthier alternatives.

> *You only see alternative product suggestions in the app. But it would be cool to see which products you purchased that you should consume less of. That would be more helpful.* [User #4, female]

Regarding feedback, some participants were discouraged by the avatar's sometimes harsh language or appearance. They felt that empathetic, positive language—even if they shopped unhealthily—would be more encouraging, which aligns with motivational interviewing theory [72]. Furthermore, participants mentioned that animating the avatar or being able to adjust its clothing during the trial would improve user engagement, also with the overall app.

> *If you could dress the avatar however you want, I'd love that*. [User #15, female]

Lastly, some users suggested implementing an extrinsic incentive scheme in the application. Various users mentioned that an integrated financial bonus system or discounts for healthier food options would increase their motivation to use the app.

## Discussion

### Principal Findings

Insufficient physical activity and unhealthy diets are contributing to the rise in NCDs. Scalable, preventative interventions are needed to combat this trend. To date, however, evidence of scalable mHealth interventions targeting both physical activity and nutritional improvements is scarce.

This is the first study to examine the impact of a future-self avatar mHealth intervention on physical activity and on the nutritional quality of food purchases. Using a 12-week RCT design, we found that a future-self avatar mHealth intervention leads to directional, small increases in physical activity (median +242 steps/day) and to directional improvements in the nutritional quality of food purchases (median -1.28 FSA-NPS-DI points). Low nutritional quality of food purchases is associated with a higher risk of developing chronic diseases and higher levels of obesity [58]. Moreover, increases in physical activity can lower the risk of metabolic syndromes and reduce all-cause mortality [73,74]. Our results thus provide preliminary evidence that scalable, preventative mHealth interventions that use future-self avatars and automatic tracking of food choices via digital receipts from loyalty cards can contribute to reducing NCDs.

However, our study found no statistically significant changes in either of its primary outcomes over the course of the 12-week intervention. One underlying reason why we found only directional yet no statistically significant improvements might be the reduced statistical power of our research due to the unexpectedly high attrition rate (68.42%, n=95) and the resulting small sample at the end of the intervention. A meta-review of weight-loss interventions found that attrition rates vary significantly between studies (9-90%) but found no consistent predictors [75]. Also, prior research has noted that high dropout rates are a “natural feature” of mHealth interventions [20]. Compared to traditional randomized-controlled drug trials, adherence to an mHealth intervention is largely at the participant’s discretion [20]. This applies in particular when an intervention is not critical to participant well-being, as was the case with our preventative mHealth study. Further research with larger sample sizes is thus needed to understand whether a future-self avatar mHealth intervention can effectively reduce present bias and drive health behavior change.

To our knowledge, our study is the first mHealth intervention to implement fully automated food purchase tracking through loyalty cards in an RCT design outside of South Africa [25,26]. Based on a 12-week RCT and 15 semi-structured interviews, we found that our technical solution is feasible and reliably tracks food purchases without requiring any participant effort after the initial setup of the FutureMe app. Prior work has found that individuals from northern-European countries consume between 71.5% and 79.8% of their mean daily energy intake at home [76]. Analyzing food purchases through automized receipt tracking and a comprehensive nutritional product database is thus a promising method to guide improvements in overall nutrition, and also contributes to the need for automated and scalable nutrition tracking solutions [25,26]. Our study found that some participants were open to further improving the comprehensiveness of the analyzed food data by manually adding out-of-home food consumption and by increasing the number of included retailers. Future research should thus evaluate how to further optimize automated and semi-automated nutrition tracking systems so as to increase user retention.

Present bias offers one possible explanation as to why preventative mHealth interventions are less effective than curative mHealth interventions [15,32]. Our study thus examined whether a future-self avatar intervention can reduce present bias by increasing outcome expectancy. In our 12-week RCT, we found directional improvements in outcome expectancy for the FutureMe group while this decreased in the control group. The results of our qualitative interviews show that some users did not understand that the FutureMe avatar reflected the future consequences of their current health behavior. They assumed instead that the FutureMe avatar mirrors today's health consequences. This effect could have undermined significant increases in outcome expectancy. However, we found that the future-self avatar intervention statistically significantly increased intrinsic motivation ($P$=.03) while no increase occurred in the control intervention. Intrinsic motivation is associated with more sustainable behavior change than extrinsically motivated behavior [77]. Future-self avatars, as an element of gamification, might thus meaningfully contribute to improving the longer-term effectiveness of mHealth interventions, compared to more traditional text-based and numerical interfaces. Future research should explore how the interfaces and functionalities of a future-self avatar mHealth intervention might be improved to more clearly communicate future health consequences, given the promising results this technology has produced in overcoming present bias in related research fields [31].

**Limitations**

The strengths of our study are its use of objectively measured data for all primary and secondary outcomes, its rigorous research design, as well as its innovative future-self avatar interface and fully automated tracking system. However, it is not without limitations. First, the unexpectedly high attrition rate limited the sample size and statistical power of our study. We conducted a sensitivity analysis to test for attrition bias and found that participants who discontinued the app were not different from participants who finished the trial, based on their baseline demographics (with the exception of age), physical activity and food purchasing behavior.

Second, the timing of our study may have impacted our results. Prior research has found that physical activity behavior is seasonally dependent [78]. Our study was conducted from November 2020 to April 2021 and results thus may have been influenced by decreases in daylight and temperatures, as well as by unusual physical activity or food purchasing behavior due to the end-of-year holiday season and New Year's resolutions. Also, the rise of the Covid-19 pandemic during the trial period may have impacted physical activity patterns and grocery purchases.

Lastly, we only tracked and analyzed food purchased at Switzerland's two leading grocery retailers. These represent 69% of the grocery market in Switzerland [79]. While this share is considerable, non-inclusion of other food purchases (eg, local farmers' markets, butcher stores, or discounters) may have impacted our results. Also, physical activity tracking was limited to measuring steps and did not include other forms of physical activity, which may have constrained our findings. Additionally, for iOS users, we only captured the number of steps recorded by their mobile phone. For Android users, we captured both the steps recorded by their mobile phone and those recorded by other devices connected to Google Fit.

## Conclusions

Our RCT found that a future-self avatar mHealth intervention leads to directional improvements in physical activity levels and food purchasing behavior, and significantly increases intrinsic motivation to lead a healthy lifestyle. Leveraging loyalty card data to track the nutritional quality of food purchases was found to be a feasible and accepted nutrition tracking technology and thus contributes to the search for scalable and automated tracking solutions. However, the high attrition rate and resulting small sample size of our study require further research to evaluate whether future-self avatar mHealth interventions can improve physical activity and food purchases at statistically significant levels. The use of future-self avatars may need to be modified in order to prevent attrition, reduce present bias, and promote sustainable behavior change.

**Acknowledgments**

We thank the Swiss National Science Foundation (Project #188402) and Helsana Zusatzversicherungen AG for financially supporting this study.

Annette Mönninghoff, Klaus Fuchs, and Simon Mayer conceived and designed the study. Michael Zeltner developed the front-end application, BitsaboutMe supported the back-end integration of loyalty card data, and Klaus Fuchs oversaw the development of the nutritional product database. Annette Mönninghoff led the research project, wrote the trial protocol, performed the data analysis, and wrote the manuscript. Jing Wu performed the food purchase data analysis. Jan Albert provided technical support for data updates. All authors provided critical reviews on earlier drafts of the manuscript and approved the final version of the manuscript before submission.

**Conflicts of Interest**

This study was funded partly by Helsana Zusatzversicherungen AG. Helsana supported participant recruitment but was not involved in the study design, data analysis and interpretation, or in reviewing and approving the manuscript for publication.

**Abbreviations**

**FSA-NPS-DI:** British Food Standards Agency Nutrient Profiling System Dietary Index
**IQR**: interquartile range
**mHealth:** mobile health
**NCDs:** Non-communicable diseases
**RCT:** randomized-controlled trial
**WHO:** World Health Organization

## References


1. Benziger CP, Roth GA, Moran AE. The Global Burden of Disease Study and the Preventable Burden of NCD. Glob Heart; 2016;11(4):393–397 [doi: 10.1016/j.gheart.2016.10.024]
2. Bundesamt für Gesundheit. Herausforderung nichtübertragbare Krankheiten. Natl Strateg zur Prävention nichtübertragbarer Krankheiten 2017-2024. 2016.
3. Blakely T, Kvizhinadze G, Atkinson J, Dieleman J, Clarke P. Health system costs for individual and comorbid noncommunicable diseases: An analysis of publicly funded health events from New Zealand. PLoS Med 2019;16(1):1–21 [doi: 10.1371/journal.pmed.1002716]
4. World Health Organization. Global action plan for the prevention and control of non communicable diseases 2013-2020. Geneva: World Health Organization; 2013. URL: https://apps.who.int/iris/handle/10665/94384ISBN:978 92 4 150623 6 [accessed 2021-06-01]
5. Aburto NJ, Ziolkovska A, Hooper L, Elliott P, Cappuccio FP, Meerpohl JJ. Effect of lower sodium intake on health: Systematic review and meta-analyses. BMJ 2013;346(7903):1–20 [doi: 10.1136/bmj.f1326]
6. Mozaffarian D, Fahimi S, Singh GM, Micha R, Khatibzadeh S, Engell RE, Lim S, Danaei G, Ezzati M, Powles J. Global Sodium Consumption and Death from Cardiovascular Causes. N Engl J Med 2014;371(7):624–634 [doi: 10.1056/NEJMoa1304127]
7. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, Mullany EC, Abate KH, Abbafati C, Abebe Z, Afarideh M, Aggarwal A, Agrawal S, Akinyemiju T, Alahdab F, Bacha U, Bachman VF, Badali H, Badawi A, Bensenor IM, Bernabe E, Biadgilign SKK, Biryukov SH, Cahill LE, Carrero JJ, Cercy KM, Dandona L, Dandona R, Dang AK, Degefa MG, El Sayed Zaki M, Esteghamati A, Esteghamati S, Fanzo J, Farinha CS e. S, Farvid MS, Farzadfar F, Feigin VL, Fernandes JC, Flor LS, Foigt NA, Forouzanfar MH, Ganji M, Geleijnse JM, Gillum RF, Goulart AC, Grosso G, Guessous I, Hamidi S, Hankey GJ, Harikrishnan S, Hassen HY, Hay SI, Hoang CL, Horino M, Islami F, Jackson MD, James SL, Johansson L, Jonas JB, Kasaeian A, Khader YS, Khalil IA, Khang YH, Kimokoti RW, Kokubo Y, Kumar GA, Lallukka T, Lopez AD, Lorkowski S, Lotufo PA, Lozano R, Malekzadeh R, März W, Meier T, Melaku YA, Mendoza W, Mensink GBM, Micha R, Miller TR, Mirarefin M, Mohan V, Mokdad AH, Mozaffarian D, Nagel G, Naghavi M, Nguyen CT, Nixon MR, Ong

KL, Pereira DM, Poustchi H, Qorbani M, Rai RK, Razo-García C, Rehm CD, Rivera JA, Rodríguez-Ramírez S, Roshandel G, Roth GA, Sanabria J, Sánchez-Pimienta TG, Sartorius B, Schmidhuber J, Schutte AE, Sepanlou SG, Shin MJ, Sorensen RJD, Springmann M, Szponar L, Thorne-Lyman AL, Thrift AG, Touvier M, Tran BX, Tyrovolas S, Ukwaja KN, Ullah I, Uthman OA, Vaezghasemi M, Vasankari TJ, Vollset SE, Vos T, Vu GT, Vu LG, Weiderpass E, Werdecker A, Wijeratne T, Willett WC, Wu JH, Xu G, Yonemoto N, Yu C, Murray CJL. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019;393(10184):1958–1972 [doi: 10.1016/S0140-6736(19)30041-8]

8. Lee I-M, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Impact of Physical Inactivity on the World's Major Non-Communicable Diseases. Lancet 2012;380(9838):219–229 [doi: 10.1016/S0140-6736(12)61031-9]
9. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, Oyeyemi AL, Perez LG, Richards J, Hallal PC. Progress in physical activity over the Olympic quadrennium. Lancet; 2016;388(10051):1325–1336 [doi: 10.1016/S0140-6736(16)30581-5]
10. Bianchini F, Kaaks R, Vainio H. Weight control and physical activity in cancer prevention. 2002;3:5–8 [doi: 10.1046/j.1467-789x.2002.00046.x.]
11. Choi KW, Chen CY, Stein MB, Klimentidis YC, Wang MJ, Koenen KC, Smoller JW. Assessment of Bidirectional Relationships between Physical Activity and Depression among Adults: A 2-Sample Mendelian Randomization Study. JAMA Psychiatry American Medical Association; 2019;76(4):399 [doi: 10.1001/jamapsychiatry.2018.4175]
12. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016 : a pooled analysis of 358 population-based surveys with 19 million participants. Lancet Glob Heal; 2018;6(10):e1077–e1086 [doi: 10.1016/S2214-109X(18)30357-7]
13. Powles J, Fahimi S, Micha R, Khatibzadeh S, Shi P, Ezzati M, Engell RE, Lim SS, Danaei G, Mozaffarian D. Global, regional and national sodium intakes in 1990 and 2010: A systematic analysis of 24 h urinary sodium excretion and dietary surveys worldwide. BMJ Open 2013;3(12) [doi: 10.1136/bmjopen-2013-003733]
14. Thout SR, Santos JA, McKenzie B, Trieu K, Johnson C, McLean R, Arcand JA, Campbell NRC, Webster J. The Science of Salt: Updating the evidence on global

estimates of salt intake. J Clin Hypertens 2019;21(6):710–721 [doi: 10.1111/jch.13546]

15. Mönninghoff A, Kramer J-N, Hess AJ, Ismailova K, Teepe G, Tudor Car L, Müller-Riemenschneider F, Kowatsch T. Long-term effectiveness of mHealth physical activity interventions: a systematic review and meta-analysis of randomized-controlled trials. J Med Internet Res 2021;23(4):1–26 [doi: 10.2196/26699]
16. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The Effect of Physical Activity Interventions Comprising Wearables and Smartphone Applications on Physical Activity: a Systematic Review and Meta-analysis. Sport Med - Open; 2018;4(1):42 [doi: 10.1186/s40798-018-0157-9]
17. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, Crozier A, Maher C. Can smartphone apps increase physical activity? a systematic review and meta-analysis (Preprint). J Med Internet Res 2018;21(3):1–14 [doi: 10.2196/12053]
18. Smith DM, Duque L, Huffman JC, Healy BC, Celano CM. Text Message Interventions for Physical Activity: A Systematic Review and Meta-Analysis. Am J Prev Med; 2020;58(1):142–151[doi: 10.1016/j.amepre.2019.08.014.]
19. Ali SH, Luo R, Li Y, Liu X, Tang C, Zhang P. Application of Mobile Health Technologies Aimed at Salt Reduction: Systematic Review. JMIR mHealth uHealth 2019 Apr 17;7(4):e13250 [doi: 10.2196/13250]
20. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):1–9 [doi: 10.2196/jmir.7.1.e11]
21. Brindal E, Freyne J, Saunders I, Berkovsky S, Smith G, Noakes M. Features predicting weight loss in overweight or obese participants in a web-based intervention: Randomized trial. J Med Internet Res 2012;14(6) [doi: 10.2196/jmir.2156]
22. Laing BY, Mangione CM, Tseng CH, Leng M, Vaisberg E, Mahida M, Bholat M, Glazier E, Morisky DE, Bell DS. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients. Ann Intern Med 2014;161(10 0):S5–S12 [doi: 10.7326/M13-3005]
23. Fuchs KL, Haldimann M, Vuckovac D, Ilic A. Automation of Data Collection Techniques for Recording Food Intake: A Review of Publicly Available and Well-Adopted Diet Apps. 9th Int Conf Inf Commun Technol Converg ICT Converg Powered by Smart Intell ICTC 2018 IEEE; 2018;58–65 [doi: 10.1109/ICTC.2018.8539468]

24. Svensson Å, Larsson C. A mobile phone app for dietary intake assessment in adolescents: An evaluation study. JMIR mHealth uHealth 2015;3(4) [doi: 10.2196/mhealth.4804]
25. Sturm R, An R, Segal D, Patel D. A cash-back rebate program for healthy food purchases in South Africa: Results from scanner data. Am J Prev Med 2013;44(6):567–572 [doi: 10.1016/j.amepre.2013.02.011]
26. An R, Patel D, Segal D, Sturm R. Eating better for less: A national discount program for healthy food purchases in South Africa. Am J Health Behav 2013;37(1):56–61 [doi: 10.5993/AJHB.37.1.6]
27. Hassannejad H, Matrella G, Ciampolini P, De Munari I, Mordonini M, Cagnoni S. Automatic diet monitoring: a review of computer vision and wearable sensor-based methods. Int J Food Sci Nutr; 2017;68(6):656–670 [doi: 10.1080/09637486.2017.1283683]
28. Schwarzer R. Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008;57(1):1–29. [doi: 10.1111/j.1464-0597.2007.00325.x]
29. Wang Y, Sloan FA. Present bias and health. J Risk Uncertain 2018;57(2):177–198 [doi: 10.1016/j.physbeh.2017.03.040]
30. Ersner-Hershfield H, Garton MT, Ballard K, Samanez-Larkin GR, Knutson B. Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. Judgm Decis Mak 2009;4(4):280–286 [PMID:19774230]
31. Hershfield HE, Goldstein DG, Sharpe WF, Fox J, Yeykelis L, Carstensen LL, Bailenson JN. Increasing saving behavior through age-progressed renderings of the future self. J Mark Res 2011;48 [doi: 10.1509/jmkr.48.SPL.S23]
32. Hunter RF, Tang J, Hutchinson G, Chilton S, Holmes D, Kee F. Association between time preference, present-bias and physical activity: implications for designing behavior change interventions. BMC Public Health; 2018 Dec 19;18(1):1388 [doi: 10.1186/s12889-018-6305-9]
33. Baicker K, Mullainathan S, Schwartzstein J. Behavioral Hazard in health insurance. Q J Econ 2015;130(4):1623–1667 [doi: 10.1093/qje/qjv029]
34. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, Faulkner GE. Financial incentives for exercise adherence in adults: Systematic review and meta-analysis. Am J Prev Med; 2013;45(5):658–667 [doi: 10.1016/j.amepre.2013.06.017]

35. Kramer J-N, Tinschert P, Scholz U, Fleisch E, Kowatsch T. A Cluster-Randomized Trial on Small Incentives to Promote Physical Activity. Am J Prev Med [Internet] 2019 Feb;56(2):e45–e54 [doi: 10.1016/j.amepre.2018.09.018]

36. Wall J, Ni Mhurchu C, Blakely T, Rodgers A, Wilton J. Effectiveness of monetary incentives in modifying dietary behavior: A review of randomized, controlled trials. Nutr Rev 2006;64(12):518–531 [doi: 10.1111/j.1753-4887.2006.tb00185.x.]

37. Etkin J. The Hidden Cost of Personal Quantification. J Consum Res 2016;42(6):967–984 [doi: 10.1093/jcr/ucv095]

38. Kruglanski AW, Riter A, Amitai A, Margolin, Bath-Shevah Shabtai L, Zaksh D. Can money enhance intrinsic motivation? A test of the content-consequence hypothesis. J Pers Soc Psychol 1975;31(4):744–750 [doi: 10.1037/0022-3514.31.4.744]

39. Lockwood BB. Optimal Income Taxation with Present Bias. Am Econ J Econ Policy 2020;12(4):298–327 [doi: 10.1257/pol.20180762]

40. Navarro J, Cebolla A, Llorens R, Borrego A, Baños RM. Manipulating self-avatar body dimensions in virtual worlds to complement an internet-delivered intervention to increase physical activity in overweight women. Int J Environ Res Public Health 2020;17(11):1–14 [doi: 10.3390/ijerph17114045]

41. Karekla M, Savvides SN, Gloster A. An Avatar-Led Intervention Promotes Smoking Cessation in Young Adults: A Pilot Randomized Clinical Trial. Ann Behav Med 2020;54(10):747–760 [doi: https://doi.org/10.1093/abm/kaaa013]

42. Fuchs K, Meusburger D, Haldimann M, Ilic A. NutritionAvatar: Designing a Future-Self Avatar for Promotion of Balanced, Low-Sodium Diet Intention. Proc Biannu Conf Ital SIGCHI Chapter Padova, Italy: ACM; 2019; 1–10 [doi: 10.1145/3351995.3352044]

43. Pinto MD, Hickman RL, Clochesy J, Buchner M. Avatar-based depression self-management technology: Promising approach to improve depressive symptoms among young adults. Appl Nurs Res; 2013;26(1):45–48 [doi: 10.1016/j.apnr.2012.08.003]

44. Schulz K, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. BMC Med 2010;8(18) [doi: 10.1186/1741-7015-8-18]

45. Tudor-locke C, Hatano Y, Pangrazi RP, Kang M. How Many Steps Are Enough? Med Sci Sport Exerc [Internet] 2008 [cited 2020 Mar 4];40(7S):537–543 [doi: 10.1249/MSS.0b013e31817c7133]

46. Tudor-Locke CE, Myers AM, Bell RC, Harris SB, Wilson Rodger N. Preliminary outcome evaluation of the First Step Program: A daily physical activity intervention for individuals with type 2 diabetes. Patient Educ Couns 2002;47(1):23–28 [doi: 10.1016/s0738-3991(01)00169-0]
47. Sugiura H, Sugiura H, Kajima K, Mirbod SM, Iwata H, Matsuoka T. Effects of long-term moderate exercise and increase in number of daily steps on serum lipids in women: Randomised controlled trial. BMC Womens Health 2002;2(3) [doi: 10.1186/1472-6874-2-3]
48. Ahn SJ, Fox J, Hahm JM. Using Virtual Doppelgängers to Increase Personal Relevance of Health Risk Communication. Lecture Notes in Computer Science 2014; 8637:1-12 [doi: 10.1007/978-3-319-09767-1_1]
49. Moore RK. A Bayesian explanation of the “Uncanny Valley” effect and related psychological phenomena. Sci Rep 2012;2:1–5 [doi: 10.1038/srep00864]
50. Julia C, Méjean C, Touvier M, Péneau S, Lassale C, Ducrot P, Hercberg S, Kesse-Guyot E. Validation of the FSA nutrient profiling system dietary index in French adults—findings from SUVIMAX study. Eur J Nutr; 2016;55(5):1901–1910 [doi: 10.1007/s00394-015-1006-y]
51. Cerin E, Cain KL, Oyeyemi AL, Owen N, Conway TL, Cochrane T, VAN Dyck D, Schipperijn J, Mitáš J, Toftager M, Aguinaga-Ontoso I, Sallis JF. Correlates of Agreement between Accelerometry and Self-reported Physical Activity. Med Sci Sports Exerc 2016;48(6):1075–84. PMID:26784274
52. Duncan MJ, Wunderlich K, Zhao Y, Faulkner G. Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. J Sports Sci; 2018;36(15):1695–1704 [doi: 10.1080/02640414.2017.1409855]
53. Nolan M, Mitchell JR, Doyle-Baker PK. Validity of the Apple iPhone®/iPod Touch® as an Accelerometer-Based Physical Activity Monitor: A Proof-of-Concept Study. J Phys Act Heal 2014 May;11(4):759–769 [doi: 10.1123/jpah.2011-0336]
54. Major MJ, Alford M. Validity of the iPhone M7 motion co-processor as a pedometer for able-bodied ambulation. J Sports Sci 2016 Dec;34(23):2160–2164 [doi: 10.1080/02640414.2016.1189086]
55. Wu J, Fuchs K, Lian J, Haldimann M, Schneider T, Mayer S, Byun J, Gassmann R, Brombach C, Fleisch E. Estimating Dietary Intake from Grocery Purchase Data - A Comparative Validation of Relevant Indicators . Preprints 2021;1(0):1–23. [doi: 10.20944/preprints202107.0055.v1]

56. Julia C, Péneau S, Buscail C, Gonzalez R, Touvier M, Hercberg S, Kesse-Guyot E. Perception of different formats of front-of-pack nutrition labels according to sociodemographic, lifestyle and dietary factors in a French population: Cross-sectional study among the NutriNet-Santé cohort participants. BMJ Open 2017;7(6):1–11 [doi: 10.1136/bmjopen-2017-016108]
57. Julia C, Touvier M, Méjean C, Ducrot P, Péneau S, Hercberg S, Kesse-Guyot E. Development and validation of an individual dietary index based on the British Food Standard Agency nutrient profiling system in a French context. J Nutr 2014 Dec;144(12):2009–17 [doi: 10.3945/jn.114.199679]
58. Julia C, Hercberg S, World Health Organization. Development of a new front-of-pack nutrition label in France: the five-colour Nutri-Score. Public Heal Panor 2017;03(04):712–725.
59. Schwarzer R, Schüz B, Ziegelmann J, Lippke S, Luszczynska A, Scholz U. Adoption and Maintenance of Four Health Behaviors: Theory-Guided Longitudinal Studies on Dental Flossing, Seat Belt Use, Dietary Behavior, and Physical Activity. Ann Behav Med 2007;33(2):156–166 [doi: 10.1007/BF02879897]
60. Renner B, Schwarzer R. Documentation of the Scales of the Research Project: "Risk Appraisal and Consequences in Korea" (RACK). Risk Heal Behav. 2005. p. 1–49.
61. Williams GC, Ryan RM, Deci EL. Health-Care, Self-Determination Theory Questionnaire Packet; 2016. URL: https://selfdeterminationtheory.org/health-care-self-determination-theory/ [accessed 2021-06-01]
62. King WR, He J. A meta-analysis of the technology acceptance model. Inf Manag 2006;43(6):740–755 [doi: 10.1016/j.im.2006.05.003]
63. Ashton LM, Morgan PJ, Hutchesson MJ, Rollo ME, Collins CE. Feasibility and preliminary efficacy of the "HEYMAN" healthy lifestyle program for young men: a pilot randomised controlled trial. Nutr J; 2017;16(1):1–17 [doi: 10.1186/s12937-017-0227-8]
64. Gell NM, Wadsworth DD. The use of text messaging to promote physical activity in working women: A randomized controlled trial. J Phys Act Heal 2015;12(6):756–763 [doi: 10.1123/jpah.2013-0144]
65. Rosner B. Fundamentals of biostatistics. 7th editio. Boston: Brooks/Cole, Cengage Learning; 2011. ISBN:978-0-538-73349-6
66. Compernolle S, Vandelanotte C, Cardon G, De Bourdeaudhuij I, De Cocker K. Effectiveness of a web-based, computer-tailored, pedometer-based physical

activity intervention for adults: A cluster randomized controlled trial. J Med Internet Res; 2015 Feb 1;17(2):e38 [doi: 10.2196/jmir.3402]

67. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts. BMC Med Res Methodol; 2017;17(1):1–10 [doi: 10.1186/s12874-017-0442-1]
68. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009 Jun 29;338:b2393 [doi: 10.1136/bmj.b2393]
69. Bryman A. Social research methods. 5th edition. Oxford: Oxford University Press; 2016. ISBN:978-0-19-968945-3
70. Sequeira MM, Rickenbach M, Wietlisbach V, Tullen B, Schutz Y. Physical activity assessment using a pedometer and its comparison with a questionnaire in a large population survey. Am J Epidemiol 1995 Nov 1;142(9):989–99 [doi: 10.1093/oxfordjournals.aje.a117748]
71. Murakami K. Nutritional quality of meals and snacks assessed by the Food Standards Agency nutrient profiling system in relation to overall diet quality, body mass index, and waist circumference in British adults. Nutr J 2017 Sep 13;16(1):57 [doi: 10.1186/s12937-017-0283-0]
72. Lisetti C, Yasavur U, De Leon C, Amini R, Rishe N, Visser U. Building an on-demand avatar-based health intervention for behavior change. Proc 25th Int Florida Artif Intell Res Soc Conf FLAIRS-25 2012;175–180.
73. Sisson SB, Camhi SM, Church TS, Tudor-Locke C, Johnson WD, Katzmarzyk PT. Accelerometer-Determined Steps/Day and Metabolic Syndrome. Am J Prev Med; 2010;38(6):575–582 [doi: 10.1016/j.amepre.2010.02.015]
74. Oftedal S, Holliday EG, Attia J, Brown WJ, Collins CE, Ewald B, Glozier N, McEvoy M, Morgan PJ, Plotnikoff RC, Stamatakis E, Vandelanotte C, Duncan MJ. Daily steps and diet, but not sleep, are related to mortality in older Australians. J Sci Med Sport; 2020;23(3):276–282 [doi: 10.1016/j.jsams.2019.09.018]
75. Moroshko I, Brennan L, O'Brien P. Predictors of dropout in weight loss interventions: A systematic review of the literature. Obes Rev 2011;12(11):912–934 [doi: 10.1111/j.1467-789X.2011.00915.x]
76. Orfanos P, Naska A, Trichopoulou A, Grioni S, Boer JMA, Van Bakel MME, Ericson U, Rohrmann S, Boeing H, Rodríguez L, Ardanaz E, Sacerdote C,

Giurdanella MC, Niekerk EM, Peeters PHM, Manjer J, Van Guelpen B, Deharveng G, Skeie G, Engeset D, Halkjær J, Jensen AM, McTaggart A, Crowe F, Stratigakou V, Oikonomou E, Touvier M, Niravong M, Riboli E, Bingham S, Slimani N. Eating out of home: Energy, macro-and micronutrient intakes in 10 european countries. The european prospective investigation into cancer and nutrition. Eur J Clin Nutr 2009;63:S239–S262 [doi: 10.1038/ejcn.2009.84]

77. Teixeira PJ, Palmeira AL, Vansteenkiste M. The role of self-determination theory and motivational interviewing in behavioral nutrition, physical activity, and health: An introduction to the IJBNPA special series. Int J Behav Nutr Phys Act 2012;9:2–4 [doi: 10.1186/1479-5868-9-17]

78. Alahmari AD, Mackay AJ, Patel ARC, Kowlessar BS, Singh R, Brill SE, Allinson JP, Wedzicha JA, Donaldson GC. Influence of weather and atmospheric pollution on physical activity in patients with COPD. Respir Res; 2015;16(1):1–9 [doi: 10.1186/s12931-015-0229-z]

79. GfK. Marktanteile der führenden Detailhändler in der Schweiz im Jahr 2018. 2018 URL:https://de.statista.com/statistik/daten/studie/368538/umfrage/marktanteile-der-fuehrenden-unternehmen-im-lebensmittelhandel-in-der-schweiz/ [accessed 2021-06-01]

**Appendix 1:** Scales for Attitudinal and Motivational Constructs Used in The FutureMe Randomized-controlled Trial

Legend: 1=completely disagree, 7=completely agree

***Motivational self-efficacy (adapted from Schwarzer et al. 2007)*** [59]

| A | **In my everyday life, …** | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| A1 | … I manage to exercise three times a week for 30 min. | | | | | | | |
| A2 | … I can make sure that I exercise at least once a week | | | | | | | |
| A3 | … I know how to shop healthily | | | | | | | |
| A4 | … I make sure I have a balanced shopping cart when I shop. | | | | | | | |

***Recovery self-efficacy (adapted from Schwarzer et al. 2007)*** [59]

| B | **Imagine you come back from vacation and need to get back into your daily routine.**<br>**After my vacation, I'm sure that…** | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| B1 | … I'll go back to balanced shopping, even if I have to get used to it again. | | | | | | | |
| B2 | … I'll get back to regular exercise, even if I don't see immediate results. | | | | | | | |

***Outcome expectancy (adapted from Renner and Schwarzer, 2005)*** [60]

| C | **Do you think your behavior today can affect your health as you age?**<br>**I believe that…** | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| C1 | … I can positively influence my health in old age with my current exercise and shopping behavior. | | | | | | | |
| C2 | … I am reducing my risk of obesity in old age with my current exercise and shopping behavior. | | | | | | | |
| C3 | … I am reducing my risk of developing diabetes in old age with my current exercise and shopping behavior. | | | | | | | |
| C4 | … I am lowering my risk of suffering from cardiovascular disease in old age with my current exercise and shopping behavior. | | | | | | | |
| C5 | … I can improve my overall mental well-being in old age with my current exercise and shopping behavior. | | | | | | | |
| C6 | … I can improve my mobility in old age with my current exercise and shopping behavior. | | | | | | | |

***Intrinsic & extrinsic motivation (adapted from Williams, Ryan and Deci, 2016)*** [61]

| **D** | **Why is it important for you to exercise and shop healthily?**<br>**I exercise and shop healthily because…** | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| D1 | … I personally believe it's best for my health. | | | | | | | |
| D2 | … it's consistent with my life goals. | | | | | | | |
| D3 | … it is very important to be as healthy as possible. | | | | | | | |
| D4 | … I want to see positive metrics on my activity tracker and health app. | | | | | | | |
| D5 | … I want others to see that I am being healthy. | | | | | | | |
| D6 | … I want praise from others. | | | | | | | |

# **Appendix 2:** Secondary Outcomes by Timepoint and Group, Within- and Between-group Comparisons

| Outcome variables | 4-week period actuals | | | | | | | | Within-group comparison vs. baseline[h] | | | | | | Between-group comparison[i] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Baseline vs. T1 | | Baseline vs. T2 | | Baseline vs. EoS | | Baseline | | T1 | | T2 | | EoS | |
| | Baseline[d] | n | T1[e] | n | T2[f] | n | EoS[g] | n | Z | *P* value | Z | *P* value | Z | *P* value | U | *P* value | U | *P* value | U | *P* value | U | *P* value |
| **Nutritional sub-categories, median (IQR)** | | | | | | | | | | | | | | | | | | | | | | |
| Sugars (in FSA-NPS-DI points)[a] | | | | | | | | | | | | | | | 565 | .73 | 484 | .19 | 592.5 | .97 | 823.5 | .006 |
| Avatar | 0.83 (1.08) | 27 | 0.95 (0.90) | 20 | 1.06 (0.39) | 14 | 0.89 (0.47) | 13 | -0.561 | .58 | -1.542 | .12 | -0.56 | .58 | | | | | | | | |
| Control | 0.91 (1.31) | 40 | 0.65 (0.86) | 34 | 0.74 (1.07) | 19 | 0.55 (1.07) | 14 | -0.994 | .32 | -0.471 | .64 | -0.3 | .76 | | | | | | | | |
| Saturated fatty acids (in FSA-NPS-DI points)[a] | | | | | | | | | | | | | | | 412 | .03 | 493.5 | .23 | 542 | .54 | 824.5 | .006 |
| Avatar | 3.82 (2.83) | 27 | 3.44 (1.34) | 20 | 3.53 (0.78) | 14 | 3.47 (1.36) | 13 | -1.478 | .14 | 0.00 | 1.00 | -0.13 | .90 | | | | | | | | |
| Control | 2.93 (1.37) | 40 | 3.06 (2.30) | 34 | 2.07 (2.23) | 19 | 3.10 (1.66) | 14 | -0.128 | .90 | -0.731 | .47 | -1.45 | .15 | | | | | | | | |
| Sodium (in FSA-NPS-DI points)[a] | | | | | | | | | | | | | | | 527 | .42 | 458 | .11 | 614 | .82 | 732 | .10 |
| Avatar | 2.77 (1.56) | 27 | 2.76 (1.82) | 20 | 2.43 (1.50) | 14 | 2.21 (1.46) | 13 | -0.818 | .41 | -0.128 | .90 | -0.21 | .83 | | | | | | | | |
| Control | 2.42 (1.37) | 40 | 2.40 (1.69) | 34 | 2.31 (1.18) | 19 | 2.86 (2.13) | 14 | 0.00 | 1.00 | -0.471 | .64 | -0.13 | .90 | | | | | | | | |
| Fruit, vegetables, legumes and nuts (in FSA-NPS-DI points)[b] | | | | | | | | | | | | | | | 555 | .64 | 449 | .08 | 722 | .13 | 681 | .31 |
| Avatar | 0.28 (0.53) | 27 | 0.22 (0.32) | 20 | 0.15 (0.49) | 14 | 0.20 (0.28) | 13 | -0.561 | .58 | -0.561 | .58 | -0.56 | .58 | | | | | | | | |
| Control | 0.31 (0.47) | 40 | 0.12 (0.22) | 34 | 0.20 (0.52) | 19 | 0.10 (0.17) | 14 | -1.274 | .20 | -0.77 | .44 | -1.36 | .17 | | | | | | | | |
| Fibers (in FSA-NPS-DI points)[b] | | | | | | | | | | | | | | | 611 | .85 | 558.5 | .67 | 722 | .13 | 709 | .17 |
| Avatar | 1.49 (0.70) | 27 | 1.44 (0.95) | 20 | 1.60 (0.54) | 14 | 1.69 (0.98) | 13 | -1.172 | .24 | -0.471 | .64 | -0.64 | .52 | | | | | | | | |
| Control | 1.56 (0.83) | 40 | 1.84 (1.17) | 34 | 1.91 (1.05) | 19 | 1.75 (0.94) | 14 | -0.644 | .52 | 1.733 | .08 | -0.13 | .90 | | | | | | | | |
| Protein (in FSA-NPS-DI points)[b] | | | | | | | | | | | | | | | 509.5 | .31 | 579.5 | .86 | 543 | .54 | 711 | .17 |
| Avatar | 3.20 (0.70) | 27 | 2.84 (1.18) | 20 | 3.28 (0.62) | 14 | 3.26 (0.60) | 13 | -0.459 | .65 | -0.644 | .52 | -0.39 | .70 | | | | | | | | |
| Control | 3.15 (1.03) | 40 | 3.15 (0.91) | 34 | 2.88 (0.92) | 19 | 3.42 (1.17) | 14 | -0.731 | .47 | -1.174 | .24 | -0.82 | .41 | | | | | | | | |
| **User engagement, median (IQR)** | | | | | | | | | | | | | | | | | | | | | | |
| Logins (logins/week) | | | | | | | | | | | | | | | - | - | 729.00 | .28 | 132.50 | .93 | 50.00 | .47 |
| Avatar | - | - | 5.00 (10.00) | 36 | 1.00 (3.00) | 15 | 1.00 (1.00) | 11 | - | - | - | - | - | - | | | | | | | | |
| Control | - | - | 4.00 (10.00) | 47 | 1.00 (3.00) | 18 | 2.00 (2.00) | 11 | - | - | - | - | - | - | | | | | | | | |
| **Attitudes & motivation, median (IQR)** | | | | | | | | | | | | | | | | | | | | | | |
| Motivational self-efficacy[c] | | | | | | | | | | | | | | | 1080.00 | .81 | - | - | - | - | 47.00 | .65 |
| Avatar | 6.00 (1.31) | 42 | - | - | - | - | 6.00 (1.38) | 9 | - | - | - | - | 0.000 | 1.00 | - | - | - | - | - | - | - | - |
| Control | 6.00 (1.50) | 53 | - | - | - | - | 6.13 (1.00) | 12 | - | - | - | - | -0.240 | .83 | - | - | - | - | - | - | - | - |
| Recovery self-efficacy[c] | | | | | | | | | | | | | | | 1061.00 | .69 | - | - | - | - | 24.00 | .03 |
| Avatar | 6.00 (1.50) | 42 | - | - | - | - | 7.00 (1.00) | 9 | - | - | - | - | -0.431 | .67 | - | - | - | - | - | - | - | - |
| Control | 6.00 (1.50) | 53 | - | - | - | - | 5.75 (1.25) | 12 | - | - | - | - | -1.725 | .13 | - | - | - | - | - | - | - | - |
| Outcome expectancy[c] | | | | | | | | | | | | | | | 1110.50 | .99 | - | - | - | - | 39.00 | .31 |
| Avatar | 6.10 (1.65) | 42 | - | - | - | - | 7.00 (2.08) | 9 | - | - | - | - | -0.338 | .81 | - | - | - | - | - | - | - | - |
| Control | 6.20 (1.60) | 53 | - | - | - | - | 5.67 (1.62) | 12 | - | - | - | - | -0.196 | .87 | - | - | - | - | - | - | - | - |
| Intrinsic motivation[c] | | | | | | | | | | | | | | | 1073.50 | .77 | - | - | - | - | 32.00 | .13 |
| Avatar | 5.83 (1.75) | 42 | - | - | - | - | 6.00 (1.67) | 9 | - | - | - | - | -2.217 | .03 | - | - | - | - | - | - | - | - |
| Control | 6.00 (1.50) | 53 | - | - | - | - | 5.50 (1.50) | 12 | - | - | - | - | -1.491 | .16 | - | - | - | - | - | - | - | - |
| Extrinsic motivation[c] | | | | | | | | | | | | | | | 871.00 | .07 | - | - | - | - | 47.50 | .65 |
| Avatar | 2.67 (2.00) | 42 | - | - | - | - | 3.67 (2.50) | 9 | - | - | - | - | -1.612 | .12 | - | - | - | - | - | - | - | - |
| Control | 3.33 (2.00) | 53 | - | - | - | - | 4.00 (2.08) | 12 | - | - | - | - | -0.060 | 1.00 | - | - | - | - | - | - | - | - |

[a] FSA-NPS-DI point scale: 0 most healthy to +10 least healthy.
[b] FSA-NPS-DI point scale: 0 least healthy to +5 most healthy.
[c] Likert scale: 1 completely disagree to 7 coompletely agree.
[d] Baseline: for steps defined as average steps/day 6 weeks prior to enrolling in the trial; for shopping-related outcome variables baseline is defined as nutritional value of all foods purchased within the 4 weeks before the trial.
[e] T1: Week 1-4 average values.
[f] T2: Week 5-8 average values.
[g] EoS: End of Study, defined as mean value week 9-12.
[h] Wilcoxon sign-rank test was performed to detect statistical differences.
[i] Mann-Whitney U test was performed to detect statistical differences.

# Article IV

## The Future of Medicine: Accelerating Digitization in Healthcare. Why We Need It and Where We Struggle

Christoph Franz[1], Annette Mönninghoff[2,3]

[1]F. Hoffmann-La Roche AG, Basel, Switzerland
[2]Institute for Customer Insight, University of St.Gallen, St.Gallen, Switzerland
[3]Institute for Mobility, University of St.Gallen, St.Gallen, Switzerland

## Abstract

The Covid-19 pandemic has put the spotlight on the lack of digital infrastructure in today’s healthcare systems and demonstrated why digitized healthcare is a competitive advantage.

**Keywords:** Digitization, digital health, Covid-19, health policy.

## Introduction

The Covid-19 pandemic has hit us by surprise. When news about an unknown virus started spreading in January 2020 and an increasing number of new cases were detected daily around the globe (Taylor, 2021), we were not ready to manage a global pandemic.

### You can't manage what you can't measure

Particularly in Europe, healthcare systems were not built to efficiently consolidate infection numbers at a national scale. Instead, labs or clinics in Germany, for example, were communicating test results to regional public health offices via FAX, telephone or non-standardized emails which were then hand-typed into the public health offices' systems (Kinkartz, 2021). It took Germany six months to introduce a digitized detection system (DEMIS) that allowed labs to electronically communicate Covid test results to the national institute for disease control and prevention, ultimately speeding up the information flow and significantly improving data quality (Robert Koch Institut, 2021). The lack of digitized information flows is just one example for the underdevelopment of healthcare systems with regard to digital capabilities. Today, more than one year into the pandemic, many of the national contract tracing apps in EU countries are yet to be fully accepted and used by the population (Strempel, 2021). Also, Germany is still struggling to transfer the millions of originally analog vaccination records into the EU's digital vaccination pass (Deutsches Ärzteblatt, 2021).

### Digitized healthcare brings competitive advantage

In contrast, the backbone of Israel's healthcare system is digital, which comes as a competitive advantage, not only in times of a pandemic. The highly centralized Israeli system builds on a history of over 25 years of electronic medical records (EMR) for every citizen, allowing the country to coordinate their vaccination campaign seamlessly and effectively (Ministry of Economy and Industry State of Israel, 2020; Phlio, 2021). Because of Israel's comprehensive EMR infrastructure that systematically consolidates medical information in machine-readable, anonymized databases, the country could provide immediate feedback on vaccination success rates and side effects, along with demographic and medical record data for all vaccinated citizens (Regev, 2021). Taking advantage of this dataset, vaccine producer BioNtech/Pfizer offered prioritized distribution to Israel, leveraging the country as a pilot case for the real-life effect of a national vaccination campaign on pandemic dynamics (Regev, 2021). Without its digital health infrastructure, Israel would not have been able to re-open the country as fast as it did.

**Why we need to accelerate digitization beyond the pandemic**

We argue that beyond the Covid-pandemic, we need to accelerate digitization in healthcare in order to future-proof healthcare systems and to capitalize on countless opportunities:

1. *Digitization accelerates medical progress*

   Medical progress in recent decades has led to an exponential increase in the complexity of medical research. Increasingly, real-world data (RWD) complements findings from traditional clinical trials (Ramagopalan et al., (2020); Swift et al., (2018)). Together with ever smarter artificial-intelligence (AI) based tools, RWD opens the door to predictive medicine at scale. This is particularly relevant in researching conditions where preventive care has the greatest effect. According to what we know today, Alzheimer’s disease is one such condition. Every year, 10 million people develop Alzheimer’s disease, and the number of cases rises sharply as the population ages (World Health Organization, 2020). RWD allow us to systematically search for clues that help us better understand diseases and their development, and to diagnose them at an earlier stage. In the future, the success of pharmaceutical companies will not only depend on the best natural scientists, but also the best software engineers and access to the best anonymized health data. For societies, a digitized healthcare system and quality-controlled, machine-readable health record databases will become a competitive advantage. China, the United States, and Israel are heavily investing in these capabilities (Cyranoski, (2016); Ministry of Economy and Industry State of Israel, 2020), Europe needs to catch-up.

2. *Digitization improves the quality of healthcare*

   Until the 1950s, the rule of thumb was that it takes half a century for medical knowledge to double. Today, experts estimate that medical knowledge doubles every 70 days (Densen, 2011). This leads to complexity and a flood of innovations that cannot be mastered by humans alone. To further improve the quality of healthcare, doctors will increasingly be augmented by digital assistance systems. Such systems rely on AI and will support image-diagnosis, suggest treatment paths and ensure that even less experienced doctors follow treatment guidelines, hereby avoiding mistakes and improving the quality of care. To future-proof healthcare systems, we must invest in digital assistance tools and encourage research at the intersection of medicine and computer science.

### 3. *Digitization creates efficiencies*

Growing medical costs are a major challenge for healthcare systems especially in demographically aging societies (Fleisch et al., 2021). While some cost drivers like demographic change are unavoidable, there are others that are caused by inefficiencies and waste within the system. Digitalization can play a decisive role in reducing or preventing inefficiencies, which in turn can alleviate some of the cost pressure being experienced by healthcare systems. There are many areas where treatment is inefficient. Emergency rooms are one such example. Statistics show that when people come to emergency rooms, this is medically unnecessary in 71% of cases (Truven Health Analytics, 2013). These visits could have been avoided. Digital triage apps can help guide patients to an appropriate and more cost-effective treatment option. Digital health companies like Oscar Health have pioneered such apps with proven success, herby reducing healthcare costs without compromising on quality of care (Fleisch et al., 2021). In order to manage rising medical costs in the future, we need to implement connected networks of insurers, providers and patients that increase cost transparency and push performance-based remuneration schemes.

### 4. *Digitization improves access to healthcare*

Digitization is the key, if not the only way, to provide the 3.4 billion low-income people globally, those who earn less than $5.50 a day, with access to quality healthcare (World Bank, 2018). The World Health Organization estimates that by 2030, we will be facing a shortage of 12.9 million healthcare workers globally (World Health Organization, 2013). The situation is particularly acute in Sub-Saharan Africa and Asia (Ho, 2018; World Health Organization, 2013). But also, in Europe, remote areas have been faced with thinned out healthcare networks (Kassenärztliche Bundesvereinigung, 2016). Telemedicine can extend access to medical care in remote places, remote diagnostic tools efficiently bring expert knowledge to lower-skilled healthcare centers when needed, smart digital assistants can relieve doctors from administrative tasks, thereby freeing up much needed medical capacities. To improve access to healthcare, we need to improve mobile-phone and data networks, and to integrate telemedicine and remote diagnostic services into regular care networks.

### *5. Digitization supports preventative care*

The most fatal diseases are not infectious diseases, but non-communicable diseases such as diabetes, cardiovascular illnesses, or cancer which cause the majority of deaths worldwide (World Health Organization, 2021). To a large percentage, the risk of getting these diseases is driven by modifiable lifestyle-related factors. Smoking causes 7 million deaths per year, insufficient physical activity 2 million, and excess salt consumption 4 million (World Health Organization, 2021). Instead of treating a growing number of patients with chronic illnesses, the focus must shift towards prevention. Mobile digital health solutions (mHealth) can effectively support behavioral change towards healthier lifestyles. Because of their ubiquitous nature, they provide scalable and cost-effective solutions to support health behavior change in large populations. mHealth solutions also expand the reach of doctors beyond their clinic and offer continuous patient support, which improves therapy outcomes for chronic illnesses. The diabetes app MySugr showcases how supplementing traditional care with mHealth solutions improves quality of care (Fleisch et al., 2021). To future-proof healthcare systems, we need to leverage mHealth solutions to supplement current treatment regimes, and to transform healthcare systems towards prevention-oriented systems that safeguards health instead of treating diseases.

## References

BBC. (2021, March 26). *Covid passports: What are different countries planning?* https://www.bbc.com/news/world-europe-56522408.

Cyranoski, D. (2016, January 6). *China embraces precision medicine on a massive scale.* Nature. https://www.nature.com/articles/529009a

Densen, P. (2011). Challenges and opportunities facing medical education. *Transactions of the American Clinical and Climatological Association*, 122, 48.

Deutsches Ärzteblatt. (2021). Impfdokumentation: Die Arztpraxis ist kein zweites Bürgeramt. *Dtsch Arztebl. 118*(21), A-1026.

Fleisch, E., Franz, C., Herrmann, A., Mönninghoff, A. (2021). *Die Digitale Pille: Eine Reise in die Zukunft unseres Gesundheitssystems.* Campus Verlag.

Ho, A. (2018, September 17). *AI can solve China's doctor shortage. Here's how*. World Economic Forum. https://www.weforum.org/agenda/2018/09/ai-can-solve-china-s-doctor-shortage-here-s-how/

Kassenärztliche Bundesvereinigung. (2016, October 5). *Deutschlandweite Projektion 2030 – Arztzahlenentwicklung in Deutschland*. https://www.kbv.de/media/sp/2016_10_05_Projektion_2030_Arztzahlentwicklung.pdf

Kinkartz, S. (2021, January 26). *Gesundheitsämter: Mit Papier, Stift und Fax gegen Corona.* Deutsche Welle. https://www.dw.com/de/gesundheits%C3%A4mter-mit-papier-stift-und-fax-gegen-corona/a-56347106.

Ministry of Economy and Industry State of Israel. (2020). *Digital Health: The Israeli Promise.*

Philio, H. (2021). *Israel's COVID vaccination success: How a digital infrastructure and other factors have made Israel a world vaccination leader.* KPMG. Retrieved April, 20, 2021, from https://home.kpmg/xx/en/home/insights/2021/04/israel-covid-vaccination-success.html.

Ramagopalan, S. V., Simpson, A., & Sammon, C. (2020). Can real-world data really replace randomised clinical trials?. *BMC medicine, 18*(1), 1-2.

Regev, D. (2021, January 11). *The secret to Israel's coronavirus vaccination success.* Politico. https://www.politico.eu/article/israel-coronavirus-vaccine-success-secret/.

Robert Koch Institut. (2021, January 21). DEMIS – *Deutsches Elektronisches Melde- und Informationssystem für den Infektionsschutz.* https://www.rki.de/DE/Content/Infekt/IfSG/DEMIS/DEMIS_node.html;jsessionid=EF1DC376645DF985BC776E230B8828F4.internet082

Strempel, P. (2021, February 9). *The European struggle with Covid-19 contact tracing app continues.* Eit Digital. https://www.eitdigital.eu/newsroom/news/2021/the-european-struggle-with-covid-19-contact-tracing-app-continues/

Swift, B., Jain, L., White, C., Chandrasekaran, V., Bhandari, A., Hughes, D. A., & Jadhav, P. R. (2018). Innovation at the intersection of clinical trials and real-world data science to advance patient care. *Clinical and translational science*, *11*(5), 450-460.

Taylor, D. (2021, March 17). A Timeline of the Coronavirus Pandemic. *New York Times*. https://www.nytimes.com/article/coronavirus-timeline.html

Truven Health Analytics. (2013, April 25). *Truven Health Analytics Study Finds Most Emergency Room Visits Made by Privately-Insured Patients Are Avoidable*. Business Wire. https://www.businesswire.com/news/home/20130425005100/en/Truven-Health-Analytics-Study-Finds-Most-Emergency-Room-Visits-Made-by-Privately-Insured-Patients-Are-Avoidable

World Bank. (2018, October 17). *Nearly Half the World Lives on Less than $5.50 a Day*. https://www.worldbank.org/en/news/press-release/2018/10/17/nearly-half-the-world-lives-on-less-than-550-a-day

World Health Organization (2013, November 11). *Global health workforce shortage to reach 12.9 million in coming decades.* https://apps.who.int/mediacentre/news/releases/2013/health-workforce-shortage/en/index.html

World Health Organization. (2020, September 21). *Dementia*. https://www.who.int/news-room/fact-sheets/detail/dementia.

World Health Organization. (2021, April 13). *Noncommunicable diseases.* https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

# Curriculum Vitae

## Personal information

| | |
|---|---|
| Name | Annette Barbara Isabell Mönninghoff |
| Date of birth | August 14, 1987 |
| Place of birth | Herrenberg, Germany |

## Education

| | |
|---|---|
| 02/2019 – 08/2021 | University of St.Gallen, Switzerland<br>*Ph.D. in Management* |
| 09/2009 – 01/2012 | University of St.Gallen, Switzerland<br>*M.A. Strategy and International Management*<br>*Double Degree in Business Education* |
| 01/2011 – 05/2011 | Cornell University, United States<br>*Exchange semester* |
| 09/2006 – 07/2009 | WHU - Otto Beisheim School of Management, Germany<br>*B.Sc. Business Administration* |
| 07/2008 – 11/2008 | University of Sydney, Australia<br>*Exchange semester* |
| 09/1997 – 07/2006 | Otto Hahn Gymnasium Nagold, Germany<br>*Abitur* |

## Scholarships

| | |
|---|---|
| 2019 | Behavioral Lab, University of St.Gallen, Switzerland<br>*Research Grant* |
| 2007 - 2011 | Studienstiftung des deutschen Volkes, Germany<br>*Full Scholarship for Undergraduate & Graduate Studies* |

## Work experience

| | |
|---|---|
| Since 02/2019 | Institute for Customer Insight, University of St.Gallen, Switzerland<br>*Project Leader and Research Associate* |
| 08/2013 – 12/2018 | Beiersdorf Inc, United States<br>*Multiple Marketing Positions (last: Senior Brand Manager)* |
| 09/2011 – 07/2013 | Beiersdorf AG, Germany<br>*Junior Strategy Consultant CEO Office, Marketing Trainee* |