

The effect of chatbot-based exercise interventions on physical activity, exercise habits, and sedentary behavior: A systematic review and meta-analysis of randomized controlled trials

DIGITAL HEALTH

Volume 11: 1–16

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/20552076251353731

journals.sagepub.com/home/dhj

Qian Wang¹ , Wenxian Yang² , Kai Xu³ and Lawrence T Lam^{1,4,5}

Abstract

Objectives: Chatbots, transcending the limitations of space and time while reducing user resistance, offer a personalized and autonomous approach potentially enhancing digital health interventions. This systematic review aims to quantitatively evaluate the effectiveness of chatbot-based exercise interventions (EIs) in randomized controlled trials (RCTs), emphasizing physical activity (PA), exercise habits (EH), and sedentary behavior (SB).

Methods: A comprehensive search was carried out across Embase, Web of Science, PubMed, and Cochrane databases for studies published from January 2010 to October 2024. The characteristics of the included studies and the employed chatbots were analyzed. A meta-analysis was performed to synthesize the impact of chatbot-based EIs on PA, EH, and SB. We carried out subgroup analyses (SAs) for investigating potential modifying effects.

Results: This review includes 12 studies conducted between 2013 and 2024, encompassing 2446 participants and 240 parent-child dyads, with participant ages ranging from 8 to 71 years. The chatbot-based EIs significantly enhanced PA compared to control groups (SMD = 0.20, 95% CI = 0.04–0.37, $p = 0.02$; $I^2 = 51\%$, $p = 0.04$). However, these interventions did not significantly impact EH (SMD = 0.29, 95% CI = –0.48–1.06, $p = 0.46$) or reduce SB (SMD = 0.32, 95% CI = –0.54–1.19, $p = 0.46$). Notably, SAs revealed a small but significant short-term increase in EH (SMD = 0.29, 95% CI = 0.08–0.50, $p = 0.006$).

Conclusion: Chatbot-based EIs appear promising in improving PA, particularly moderate to vigorous PA (MVPA). Nonetheless, further RCTs with diverse chatbot designs, larger sample sizes, low-resource settings, objective measures of SB (e.g., accelerometers), and extended follow-ups are essential to corroborate these findings. Future research should also explore the adaptation of chatbots to various socioeconomic and cultural contexts and the integration of technologies like ChatGPT into EIs.

Trial Registration: Meta-analysis PROSPERO: CRD42024609852

Keywords

Chatbot, exercise intervention, physical activity, exercise habit, sedentary behavior

Received: 30 December 2024; accepted: 4 June 2025

¹Faculty of Medicine, Macau University of Science and Technology, Macau SAR, China

²Jiangsu Research Institute of Sports Science, Nanjing, China

³Nanjing Sport Institute, Nanjing, China

⁴Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

⁵Faculty of Health, University of Technology Sydney, Sydney, Australia

Corresponding author:

Lawrence T Lam, Faculty of Medicine, Macau University of Science and Technology, Av. Wai Long, Macao SAR, China.

Emails: tmlam@must.edu.mo; lawrence.lam@sydney.edu.au;

lawrence.lam@uts.edu.au



Introduction

Non-communicable diseases (NCDs) are the predominant causes of mortality within the realm of public health globally, accounting for over 41 million deaths in 2019.¹ These diseases are particularly prevalent in regions with scarce resources and in low- and middle-income countries.^{2,3} The most prevalent NCDs include cardiovascular and chronic respiratory diseases, type 2 diabetes, and cancer. As the global population ages, the burden of NCDs is increasingly becoming a critical issue worldwide.⁴ Regrettably, projections indicate that the burden of NCDs is expected to escalate further in the future.⁵ In response, the World Health Organization (WHO) has spearheaded a global action plan aimed at the prevention and control of NCDs.⁶ A notable risk factor in the prevention of NCDs is physical inactivity, which is characterized by insufficient physical activity (PA) and excessive sedentary time. It has been estimated that approximately 3.2 million deaths annually can result from physical inactivity.⁷ PA involves any movement of the body's muscles that triggers the consumption of energy.⁸ It is essential throughout a person's life. It is crucial to enhance PA and reduce sedentary behavior (SB) from infancy to old age. However, merely increasing PA levels is insufficient; a balanced approach that addresses both activity and SB is optimal.

Several types of exercise interventions (EIs) have been developed to target overweight,⁹ obesity,¹⁰ and type 2 diabetes.¹¹ Most research indicates that these interventions positively affect overweight, obesity, and type 2 diabetes.^{12–14} EIs are an effective and economical approach to addressing the energy imbalance between consumption and expenditure. However, they present several challenges, including difficulty in maintaining sustainable engagement and in meeting the diverse needs of participants.^{15–18} Existing studies indicate that EI programs are more effective in the short term than in the long term.^{19,20} This may be because participants initially find the programs engaging, but their interest wanes over time as they become tired of the repetitive nature of the interventions. Consequently, sustainable engagement appears unattainable. Furthermore, individual differences are significant.²¹ Often, researchers implement the same intervention across diverse groups. Some participants may like the intervention, others may only slightly like it, and some may dislike it. Those who dislike the intervention often experience a low sense of self-efficacy and merely go through the motions without genuine engagement. The goal of behavioral intervention programs is to establish a habit of exercising. According to Fogg's behavior theory, motivation and competence are crucial factors in habit formation.²² However, targeted individuals exhibit varying levels of motivation and competence, and a single intervention style does not universally apply. It is widely recognized that a one-size-fits-all approach to interventions is ineffective.²³ Current EIs lack personalized designs meeting diverse participants' individual needs.

The WHO has emphasized the importance of digital health interventions in public health.²⁴ Chatbots have become a new trend in the digital environment due to their ability to sustain

engagement and meet diverse user requirements.²⁵ Chatbots, also known as conversational and virtual agents, engage in bilateral dialogue with users based on machine learning (ML),²⁶ deep learning (DL),²⁷ and natural language processing (NLP).²⁸ In these dialogue systems, chatbots can communicate with clients naturally and automatically in the form of text, images, and voice/speech by collecting users' information, thus meeting the needs of the current context.²⁹ Chatbots are applied in various settings, including business,³⁰ education,³¹ and healthcare.³² Behavioral theories such as the theory of planned behavior (TPB) and social cognitive theory (SCT) provide valuable frameworks for designing personalized chatbot interventions. TPB can guide the design of chatbots that target key factors influencing user intentions, such as perceived behavioral control, subjective norms, and attitudes.³³ By addressing specific factors, chatbots can guide users toward adopting more adaptive behaviors and making healthier choices. SCT underscores the significance of observational learning and self-regulation in effecting behavioral change.^{34,35} In the context of chatbot design, SCT can be operationalized by crafting interactive dialogues that enhance self-efficacy and encourage users to emulate positive behaviors. Furthermore, with the advancement of technology and the increasing number of e-device users, the application of chatbots in EIs attracts rising attentions. Chatbots, by overcoming the constraints of space and time and mitigating user resistance, have become a viable, personalized, and autonomous form of intervention.³⁶ Theoretically, integrating chatbots with EIs represents a promising synergy.

Numerous chatbot-based EIs have targeted mental health,^{37,38} and related health behaviors,^{39,40} encompassing total PA,⁴¹ moderate-to-vigorous PA (MVPA),⁴² daily step counts,⁴¹ exercise habits (EHs),⁴³ and SBs.⁴⁴ Despite these developments, a gap persists regarding the effectiveness of AI-driven chatbots in PA interventions. Some researchers contend that it is premature to conclusively determine the efficacy of chatbot interventions on PA,^{45–47} while other studies highlight the substantial impact of such interventions on improving PA.^{36,41,48} It is possible that the design of these interventions lacks robustness. Following the Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence, evidence quality is categorized into four grades, ranging from high to low: randomized controlled trials (RCTs), cohort studies and case-control studies, case series, and expert opinions without explicit critical appraisal.⁴⁹ A systematic review and meta-analysis should ideally include only high-quality articles, specifically RCTs. Additionally, although some chatbot-driven EIs have explored outcomes related to EHs and SB, they lack a systematic review and meta-analysis. This review aims at investigating the effectiveness of chatbot-based EIs on PA, EHs, and SBs.

Method

The protocol for this systematic review was registered with PROSPERO (ID: CRD42024609852) and conforms to the

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.⁵⁰ To maintain a high standard of quality, the review process rigorously adhered to established guidelines, encompassing steps involving the strategy for searching, selecting studies, and extracting data. No funding was received for this work.

Information sources

A comprehensive systematic search was conducted from January 2010 to October 2024 across multiple databases. All databases were last searched on 26 October 2024. The scientific literature databases searched included Web of Science, Embase, Cochrane, and PubMed.

Selection criteria

The selection criteria were stringently based on the PICOS framework (study types, outcome, comparison, intervention, population).⁵¹ Population: any demographic group; Intervention: any EI involving chatbots, conversational agents, or virtual assistants. Chatbots, such as Siri and ChatGPT, also known as virtual assistants or conversational agents, are increasingly being integrated into daily life. They can be categorized into three distinct types: (1) commercial voice assistants, encompassing Google Assistant, Alexa (Amazon), and Siri (Apple); (2) custom-designed text- or voice-based chatbots tailored for specific research or clinical interventions, developed using platforms like Dialogflow, Rasa, or Microsoft Bot Framework; and (3) AI-driven conversational agents based on large language models, such as ChatGPT or BERT, characterized by their dynamic, generative capabilities. These agents utilize advanced dialogue algorithms to enable natural language interactions through voice, text, or multimodal channels. The spectrum includes any chatbot form—text-based, voice-based, or both—and may operate as standalone software, within web browsers, on mobile devices, in applications, or as embodied conversational agents (ECAs). A constrained-input chatbot limits user input by various means such as length, format, language, or content. Conversely, unconstrained-input chatbots allow flexible, open-ended dialogues. Intervention settings (e.g., laboratory, hospital, home) and doses (i.e., duration, frequency, intensity) were not restricted. Comparison: studies involving a comparison condition (e.g., intervention group vs. control group or waitlist group or usual care group). Outcomes: eligible studies must measure PA outcomes, including total PA, moderate-to-vigorous PA, vigorous PA, moderate PA, light PA, or daily steps, or assess SB. Study type: only RCTs were included.

Exclusion criteria included duplicate records, article type, language, study design, and PICO relevance. Duplicate records encompassed identical publications from different databases, varying formats of the same study,

multiple publications of the same trial, and multilingual versions. Ineligible studies included literature reviews, commentaries, patent literature, protocols, and non-English publications. Studies not meeting the required design criteria—such as observational studies—were excluded. Additionally, studies were excluded if the intervention did not integrate both chatbot-related and exercise-based components. Studies that lacked sufficient statistical data on PA, EH, or SBs (e.g., means, standard deviations, effect sizes, or confidence intervals) and for which data could not be obtained from the authors were also excluded.

Data extraction and management

All studies were imported into Endnote X9, after which two authors eliminated duplicates to prevent data redundancy. The researchers meticulously reviewed the methodology sections of all included studies, focusing on sources of participants, data collection periods, study settings, and sample characteristics to ensure comprehensive understanding. Data extraction was collaboratively undertaken by two authors (WQ and YWX). Disagreements were resolved through discussion until everyone agreed, achieving 100% agreement. In cases where consensus could not be reached, an independent third-party expert was consulted to make the final decision regarding data extraction. Data were selected based on several criteria: (1) study characteristics, such as title, authors, and publication year; (2) characteristics of participants, including age and sample size; (3) types of RCT, e.g., two-arm, three-arm, cluster, stratified, and pilot studies; (4) duration of the intervention; (5) intervention programs; (6) type of chatbot utilized; (7) outcome measures; (8) the mean and standard deviation of outcomes. In instances where different studies utilized identical datasets, these were amalgamated into a singular study for analysis. If a study with incomplete data was identified, the authors endeavored to contact the corresponding author to acquire the necessary data or to calculate the mean and standard deviation using recommended formulas.

Quality of assessment

The risk of bias was independently assessed by two reviewers (YWX and WQ) using the Cochrane Risk of Bias Tool (version 2.0). Any unresolved disagreements were adjudicated by a third, qualified expert. The assessment encompassed seven domains: (1) selective reporting (reporting bias); (2) completeness of outcome data (attrition bias); (3) blinding of outcome assessment (detection bias); (4) blinding of participants and personnel (performance bias); (5) allocation concealment (selection bias); (6) random sequence generation (selection bias); and (7) other potential biases. Each domain was evaluated with three potential outcomes: high risk of bias (-), low risk of bias (+), or not applicable (NA). The reviewers' assessments

were based on these criteria, classifying the quality of each study as low, moderate, or high risk. The visualization of risk of bias was performed using RevMan 5.4 software.

Outcome of interest

The outcomes of interest included: (1) MVPA only; (2) total PA; (3) daily steps, expressed as steps per day or week; (4) EHs, assessed using the Self-Report Habit Index (SRHI) for exercise; (5) SB, quantified as duration of sedentary time, e.g., minutes per day or week.

Statistical analysis

The meta-analysis conducted in this review employed a random-effects model to estimate the degree of heterogeneity among the selected studies. This model posits that variations in the true effect size across studies can result from differences in intervention characteristics, participant demographics, and study methodologies. A random-effects model provides a more conservative estimate than a fixed-effect model and is deemed suitable in scenarios exhibiting moderate to substantial heterogeneity, as evidenced by an I^2 value exceeding 50%. The primary analysis, which included forest plots, heterogeneity tests, sensitivity analyses, and subgroup analyses (SAs), was facilitated using RevMan software version 5.4. Due to inconsistencies in measurement units, the standard mean difference (SMD) within a 95% confidence interval (CI) was employed to carry out statistical analysis. SAs were stratified following the type of chatbot and intervention duration (eight weeks or fewer versus more than eight weeks). These analyses assessed potential differences in effects across various subgroups. Differences between groups were deemed with statistical significance if the p -value fell below 0.05.

We used I^2 statistics for the quantitative assessment of the included studies' heterogeneity. The values of I^2 were categorized as follows: no heterogeneity ($0 \leq I^2 < 25\%$), moderate heterogeneity ($25\% \leq I^2 < 50\%$), substantial heterogeneity ($50\% \leq I^2 < 75\%$), and considerable heterogeneity ($75\% \leq I^2 \leq 100\%$).⁵² Egger's test was applied to evaluate the presence of publication bias. We carried out sensitivity analysis for ascertaining the robustness of the results, systematically excluding one study at a time to assess the impact of each study on the overall effect size.

Results

Study selection

Initially, a search of relevant keywords yielded 3508 records. Following a preliminary screening of titles and abstracts, we omitted 3281 studies. Further application of inclusion and exclusion criteria resulted in the elimination of 33 articles. Ultimately, 12 studies were selected for a systematic and quantitative review (Figure 1).

Study characteristics

This review encompasses 12 studies conducted between 2013 and 2024, involving a total of 2446 participants and 240 parent-child dyads (refer to Table 1).^{42–44,53–61} Among these studies, ten focused on individual participants, while two examined parent-child dyads. The studies targeted diverse populations, including children and the elderly. All articles were classified as high-quality given their design as RCTs. Specifically, eight studies implemented two-armed RCTs,^{43,44,54,55,57,59–61} including one pilot study.⁶⁰ Additionally, three studies were structured as three-armed RCTs,^{42,56,58} and one as a four-armed RCT.⁵³ The two-armed RCTs were more common, whereas the four-armed RCTs were less frequent. The bulk of the research selected has come out recently, with five articles emerging after 2020. Geographically, four studies originated from the USA,^{53,54,57,58} two from the Netherlands,^{42,56} and another two from the UK.^{55,59} The remaining studies, each conducted in a different country, came from Belgium,⁴⁴ Japan,⁶⁰ Canada,⁶¹ and Korea,⁴³ respectively. The most extensively studied intervention involving chatbot-based exercise was in the USA. Methodologically, five studies exclusively utilized objective monitoring devices, such as the Actigraph GT3X+ and Fitbit Flex 2HR. Four studies employed subjective questionnaires, including the Dutch Short Questionnaire to Assess Health-Enhancing PA and the International PA Questionnaire. Notably, three studies combined both objective measurements and subjective questionnaires.

The theoretical frameworks utilized across the studies were highly varied. Eleven articles encompassed multiple theories including the health action process approach, elaboration likelihood model, persuasive systems design model, tpb, self-regulation theory, precaution adoption process model, self-determination theory, self-monitoring theory, goal-setting theory, SCT, player experience and need satisfaction theory, agency, challenge, uncertainty, discovery, and outcomes framework, habit formation theory, health-behavior theory, and the transtheoretical model. Notably, the goal-setting theory and SCT were featured in three studies, while two studies concurrently utilized the self-determination theory and health action process approach. The time frame for the interventions extended from 4 weeks to a year. Each intervention featured a specific program involving a chatbot, such as #LIFEGOALS, Intelligent Personal Assistant Project, and Gabby. Outcomes varied; some studies focused on the MVPA of parent-child dyads. Additionally, beyond the MVPA indicator, outcomes included SB, daily step count, and EHs.

Characteristics of chatbot

The chatbots demonstrated considerable variation across the studies in terms of their names, delivery methods,

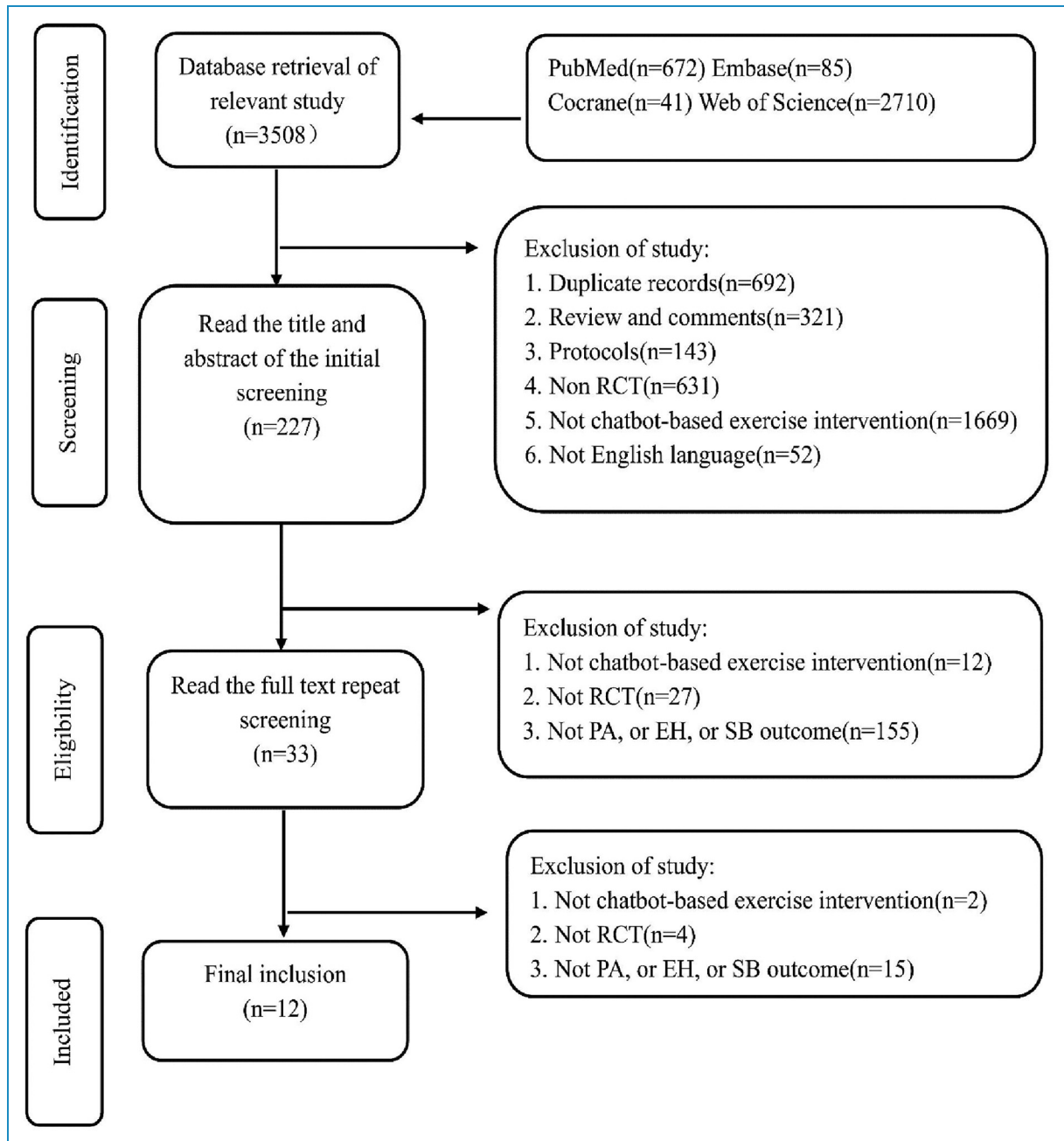


Figure 1. Article selection flow chart.

platforms, diversity, input and output modalities, functionalities, and password settings (refer to Table 2). Specifically, four studies employed chatbots with distinct names, such as Aimbot, Echo Dot, Ally, and Alexa. The primary delivery methods were computers and mobile phones: six studies utilized mobile phones, four employed computers, and two integrated both devices. Since 2020, interventions have predominantly utilized mobile phones or a combination of both mobile phones and computers. The delivery platforms encompassed software programs, websites, embodied intelligent personal assistants, and mobile

applications. Notably, software-based interventions, which first appeared in 2013, were discontinued subsequently, whereas mobile applications have been increasingly adopted since 2017. Half of the studies featured multi-component interventions, whereas the others relied on single-chatbot interventions. In terms of user interaction, seven studies featured constrained input, wherein participants communicated with the chatbots via pre-programmed routines. Conversely, two studies allowed for unconstrained input, facilitating more flexible and natural interactions. Three studies combined both types of input. All studies

Table 1. Characteristics results of a meta-analysis on the intervention.

publication	country	measurement	Participants	age (mean $\pm$ sd)	theory	program	duration	outcome	study
1 Peuters et al. (2024)	Belgium	Accelerometer Activity AX3, Euclidean Norm Minus One	279	13.6 $\pm$ 1.0	health action process approach; elaboration likelihood model; persuasive systems design model	#LIFE GOALS	14 weeks	PA, ST	quasi stratified two-armed RCT
2 Collombon et al. (2024)	Netherlands	GT3X-BT, SQUASH	248	59.4 $\pm$ 6.4	AP: theory of planned behavior; self-regulation theory; precaution adoption process model IM: self-determination theory	Computer-and-mobile-based PA interventions	6 months	MVPA	three-armed RCT
3 Sakane et al. (2023)	Japan	Pedometer	78	52 $\pm$ 6.5	self-monitoring and goal-setting theory	mHealth KENPO app intervention	12 weeks	daily step	two-armed pilot RCT
4 Tugault-Lafleur et al. (2023)	Canada	Fitbit Flex 2, PAQ(FOC), PAQ(SF), SBQ	214 parent-child dyads	Parents: 44 $\pm$ 6.2; children: 13 $\pm$ 2.2	social cognitive theory, the player experience and need satisfaction theory, the agency challenge, uncertainty, discovery, and outcomes framework	Aim2Be app intervention	6 months	MVPA, daily step, screen time	two-armed RCT
5 Carlin et al. (2021)	UK	ActiGraph GT3 accelerometer, FEAHQ-R	26 parent-child dyads	Phase 1: Parents: 40.5 $\pm$ 5.4; Children: 9.1 $\pm$ 2.0 Phase 2: Parents: 38.9 $\pm$ 5.2; Children: 7.9 $\pm$ 2.0	NA	Intelligent Personal Assistant Project	12 weeks	PA, ST, activity habit	two-armed RCT
6 Piao et al. (2020)	Korea	SRHI	121	NA/20–59	habit formation theory	Healthy Lifestyle Coaching Chatbot intervention	12 weeks	climbing habit	two-armed RCT
7 Kramer et al. (2020)	UK	Accelerometer	274	41.7 $\pm$ 13.5	health-behavior theory	Smartphone Walking App intervention	8 weeks	daily step	two-armed RCT
8 Gardiner et al. (2017)	USA	Self-Efficacy for Exercise Scale	61	35 $\pm$ 8.4	goal-setting theory	Gabby	12 weeks	PA	two-armed RCT
9 Hassoon et al. (2017)	USA	Fitbit Charge 2 HR	42	62.1 $\pm$ 9.8	Smarttext: supervised goal-based theory MyCoach: unsupervised goal-based theory with a reward condition	Physical Activity by Technology Help	4 weeks	daily step, PA	three-armed RCT
10 Friederichs et al. (2014)	Netherlands	SQUASH	958	42.9 $\pm$ 14.5	social cognitive theory, the trans-theoretical model	Web-based PA interventions based on MI	1 month	PA	three-armed RCT
11 Bickmore et al. (2013)	USA	Pedometer	263	71.3 $\pm$ 5.4	health Action Process, approach and self-determination theory	ECA-based physical activity	12 months	daily step	two-armed RCT
12 Bickmore et al. (2013)	USA	IPAQ	122	33.0 $\pm$ 12.6	social cognitive theory	ECA-based physical activity	2 months	PA, daily step	four-armed pilot RCT

ST: sedentary time; SQUASH: Dutch Short Questionnaire to Assess Health Enhancing Physical Activity; MVPA: moderate-to-vigorous physical activity; PAQ(FOC): Physical Activity Questionnaire for Older Children, PAQ(SF): Physical Activity Questionnaire Short Form; SBQ: Sedentary Behavior Questionnaire; FEAHQ-R: Family Eating and Activity Habits Questionnaire; SRHI: self-report habit index; IPAQ: International Physical Activity Questionnaire.

provided text-based output, with some also incorporating voice or speech and visual elements. The functionalities of the chatbots were categorized into three groups: delivering personalized, one-on-one behavioral counseling; providing automated goal-setting and feedback; and offering both tangible and psychological rewards. Password usage was infrequently mentioned, with only one study requiring participants to establish a password.

Risk of bias

In the twelve selected studies, the risk of bias was found to be optimal (refer to Table 3 and Figure 2). Owing to the RCT design, all studies involving randomization exhibited a low risk of selection bias. Furthermore, the majority of these trials provided detailed descriptions of the flow of participant enrolment in their documentation, suggesting a low risk of attrition bias. The researchers in these trials diligently adhered to methodological standards, particularly concerning allocation concealment and the blinding of participants or personnel. Importantly, since the outcomes were predominantly quantified using objective monitoring devices, the likelihood of detection bias was minimal. The researchers also devoted significant attention to mitigating other potential biases, which are elaborated upon subsequently.

Results of meta-analysis

Results of PA. The chatbot-based EI significantly improved PA compared to the control group, as indicated by a standardized mean difference (SMD = 0.20, 95% CI = 0.04–0.37, $p < 0.05$; $I^2 = 51\%$, $p = 0.04$; Figure 3). Notwithstanding, substantial heterogeneity was observed across studies. A pronounced effect was noted in the subgroup engaging in MVPA, where the intervention group demonstrated a significant increase compared to controls (SMD = 0.26, 95% CI = 0.05–0.46, $p < 0.05$). In contrast, no notable improvements were observed in the total PA subgroup (SMD = 0.28, 95% CI = –1.83–2.39, $p = 0.79$) or the daily step count subgroup (SMD = 0.16, 95% CI = –0.09–0.40, $p = 0.21$). Given the observed heterogeneity, further SAs were conducted based on intervention duration (≤ 8 weeks vs. > 8 weeks) and specific chatbot characteristics (delivery method and user input type).

Subgroup analysis by duration. In the analysis stratified by duration, the chat-based EI was linked with a modest increase in PA in the short term (SMD = 0.29, 95% CI = 0.08–0.50, $p < 0.05$), which had statistical significance ($p < 0.05$; see Figure 4). Conversely, in the long term, the intervention did not display a notable rise in PA (SMD = 0.06, 95% CI = –0.19–0.31, $p = 0.64$), indicating no statistical significance ($p > 0.05$). Notably, the long-term subgroup exhibited minimal heterogeneity ($I^2 = 30\%$, $p = 0.64$). The above results reveal that the duration of the

intervention contributes significantly to the heterogeneity observed in the primary analysis.

Subgroup analysis by type of chatbot. In the SA concerning the chatbot delivery method, children's PA was not notably improved in the mobile phone subgroup (SMD = 0.30, 95% CI = –0.22–0.82, $p = 0.25$; Figure 5). Conversely, a significant improvement was observed in the computer-delivered subgroup (SMD = 0.29, 95% CI = 0.14–0.44, $p < 0.05$; $I^2 = 26\%$, $p = 0.26$), with no notable heterogeneity detected. Furthermore, when comparing the effectiveness between mobile phone and computer groups directly, no notable effect on PA improvement was found (SMD = 0.02, 95% CI = –0.73–0.77, $p = 0.96$; $I^2 = 49\%$, $p = 0.14$). The above finding reveals that the chatbot delivery method is likely to affect the heterogeneity observed in the outcomes.

Analysis of chatbot input subgroups. In the SA concerning chatbot inputs, a notable improvement was documented in the group receiving both constrained and unconstrained inputs, with minimal heterogeneity observed (SMD = 0.37, 95% CI = 0.23–0.52, $p < 0.05$; $I^2 = 0\%$, $p = 0.58$). This result was statistically significant ($p < 0.05$; Figure 6). No notable heterogeneity was detected within this subgroup. Conversely, the group with constrained input alone did not exhibit a significant effect, also displaying low heterogeneity (SMD = 0.06, 95% CI = –0.11–0.23, $p = 0.46$; $I^2 = 0\%$, $p = 0.53$). In the subgroup receiving unconstrained input, PA was not notably improved (SMD = 0.28, 95% CI = –1.83–2.39, $p = 0.79$; $I^2 = 74\%$, $p = 0.05$), and substantial heterogeneity was noted, suggesting that input type may significantly influence the heterogeneity of outcomes.

EH results. Chatbot-based interventions did not demonstrate a significant enhancement in PA habits when compared to control groups. The difference observed was not statistically significant (SMD = 0.29, 95% CI = –0.48–1.06, $p = 0.46$; $I^2 = 54\%$, $p = 0.11$). Additionally, the analysis revealed no considerable heterogeneity among the studies ($I^2 = 54\%$, $p = 0.11$) (Figure 7).

Sedentary time results. The analysis indicated that chatbot-based interventions were not associated with a significant reduction in sedentary time relative to control groups (SMD = 0.32, 95% CI = –0.54–1.19, $p = 0.46$; $I^2 = 0\%$, $p = 0.91$). Furthermore, the studies displayed negligible heterogeneity ($I^2 = 0\%$, $p = 0.91$) (Figure 8).

Discussion

Overall effect of chatbot-based interventions

This study aims at evaluating the efficacy of chatbot-based interventions on PA, EHs, and SB. In assessing the overall impact, synthesizing and quantitatively analyzing twelve

Table 2. Characteristics results of a meta-analysis on the chatbot.

Author, year, country	Name	Delivery method	Delivery platform	Chatbot only or mutual-component	Input (constrained or unconstrained, or both)	Output	Function	Password
Peuters, 2024, Belgium	Chatbot	mobile phone	App	mutual-component	constrained	text	provide information and send supporting messages for individual support	NA
Collombon, 2024, Netherlands	Chatbot	computer, mobile phone	Website, embodied application	chatbot only	constrained and unconstrained	text	automatically provide personalized walking messages and feedback based on contextual data	NA
Sakane, 2023, Japan	Chatbot	mobile phone	App	mutual-component	constrained	text	give feedback of app data and sign of weight regain, chatbot-supported quiz on health and health behavior	NA
Tugault-Lafleur, 2023, Canada	Aimbot	mobile phone	App	mutual-component	constrained	text, image	involve a number of chat sessions between the user and a digital coach with preprogrammed prompts, questions, and answers	NA
Carlin, 2021, UK	Echo dot	computer, mobile phone	embodied intelligent personal assistants	mutual-component	unconstrained	text, voice	help users learn skills and complete tasks	Yes
Piao, 2020, Korea	ChatBot	mobile phone	App	chatbot only	unconstrained	text, image	remind them to perform stair-climbing activities, give feedback based on the behavior performed each day	NA
Kramer, 2020, UK	Ally	mobile phone	App	mutual-component	constrained	text	guide participants through physical activity interventions, set a personalized activity goal based on each participant's past activity data	NA
Gardiner, 2017, USA	NA	Computer	website	mutual-component	constrained	text, image, speech	greeting, behavior change counsel, remind inactive users	Yes
Hassoon, 2017, USA	Alexa	mobile phone	embodied intelligent personal assistants	chatbot only	constrained and unconstrained	voice, text	Smarttext: sent messages to the participant, MyCoach: provide coaching, learn about the participant behavior and general responses to maximize rewards	NA

(continued)

Table 2. Continued.

Author, year, country	Name	Delivery method	Delivery platform	Chatbot only or mutual-component	Input (constrained or unconstrained, or both)	Output	Function	Password
Friederichs, 2014, Netherlands	NA	Computer	website	chatbot only	constrained and unconstrained	text, avatar	ask questions and provide feedback like in a real-life situation	NA
Bickmore, 2013, USA	NA	Computer	software	chatbot only	constrained	text, image, speech	upload steps, provide positive reinforcement, have a problem-solving discussion for any barriers identified, set goal, give an exercise tip	NA
Bickmore, 2013, USA	NA	Computer	software	chatbot only	constrained	text	set goal, develop and review goals, help users solve problems, overcome barriers, and increase knowledge of health behaviour	NA

Table 3. Risk of bias summary table.

Risk of Bias Domain	Low Risk (%)	Unclear Risk (%)	High Risk (%)
Random sequence generation	100%	0%	0%
Allocation concealment	67%	%	33%
Blinding of participants/ personnel	58%	42%	0%
Blinding of outcome assessment	92%	0%	8%
Incomplete outcome data	50%	8%	42%
Selective reporting	75%	17%	13%

RCTs suggests that these interventions could be a promising method for enhancing PA. Notably, there was a significant improvement in MVPA, a finding consistent with other research.⁴⁸ However, these interventions did not significantly affect total PA or daily step count. There was a notable discrepancy in outcomes related to daily steps and total PA between this review and prior studies.⁴⁸ Furthermore, EHs or SB did not change notably. The absence of meta-analyses focusing on EHs and sedentary time renders this review potentially unique and pioneering. Surprisingly, even as interventions increased MVPA, they did not impact EHs or SB. This review found no correlation between PA, EHs, and sedentary time. Regarding the methodology, our findings build upon previous high-quality research. However, the blinding of participants and personnel was suboptimal. Seven out of the twelve reviewed trials did not implement blinding of personnel, which is considered undesirable, and one trial exhibited a notable likelihood of bias in performance. In terms of measurement, there was no uniform unit of measurement across the studies due to the use of various devices, which poses a challenge for standardization. Despite the powerful capabilities of advanced generative models like ChatGPT, no existing trials have employed such technologies in chatbot interventions. Nonetheless, these models offer substantial promise for enhancing the effectiveness, emotional engagement, and user experience of digital health interventions focused on exercise. The observed heterogeneity ($I^2 = 51\%$) in the results may result from changes in chatbot types, intervention settings, and participant demographics. This variability underscores the necessity to develop uniform methods for future research and reporting protocols to improve comparability. Due to the variability and limitations in the data, this analysis does not provide conclusive evidence regarding the efficacy of chatbot-based EIs.

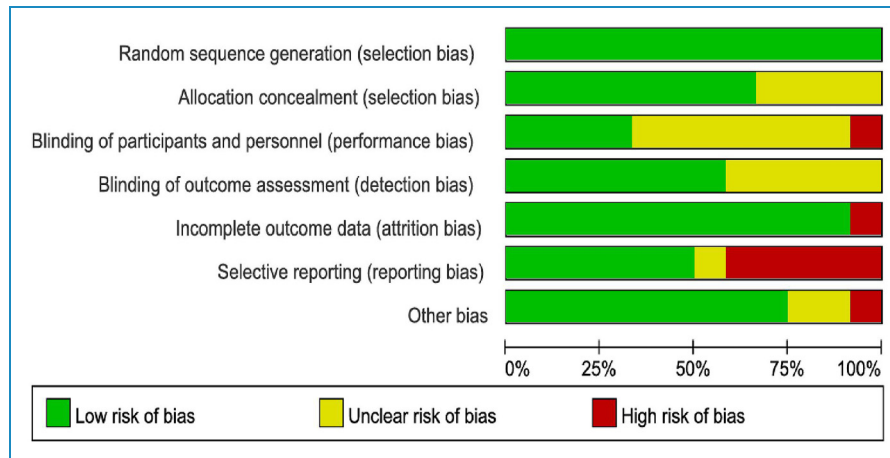


Figure 2. Risk of bias graph each risk of bias item presented as percentages.

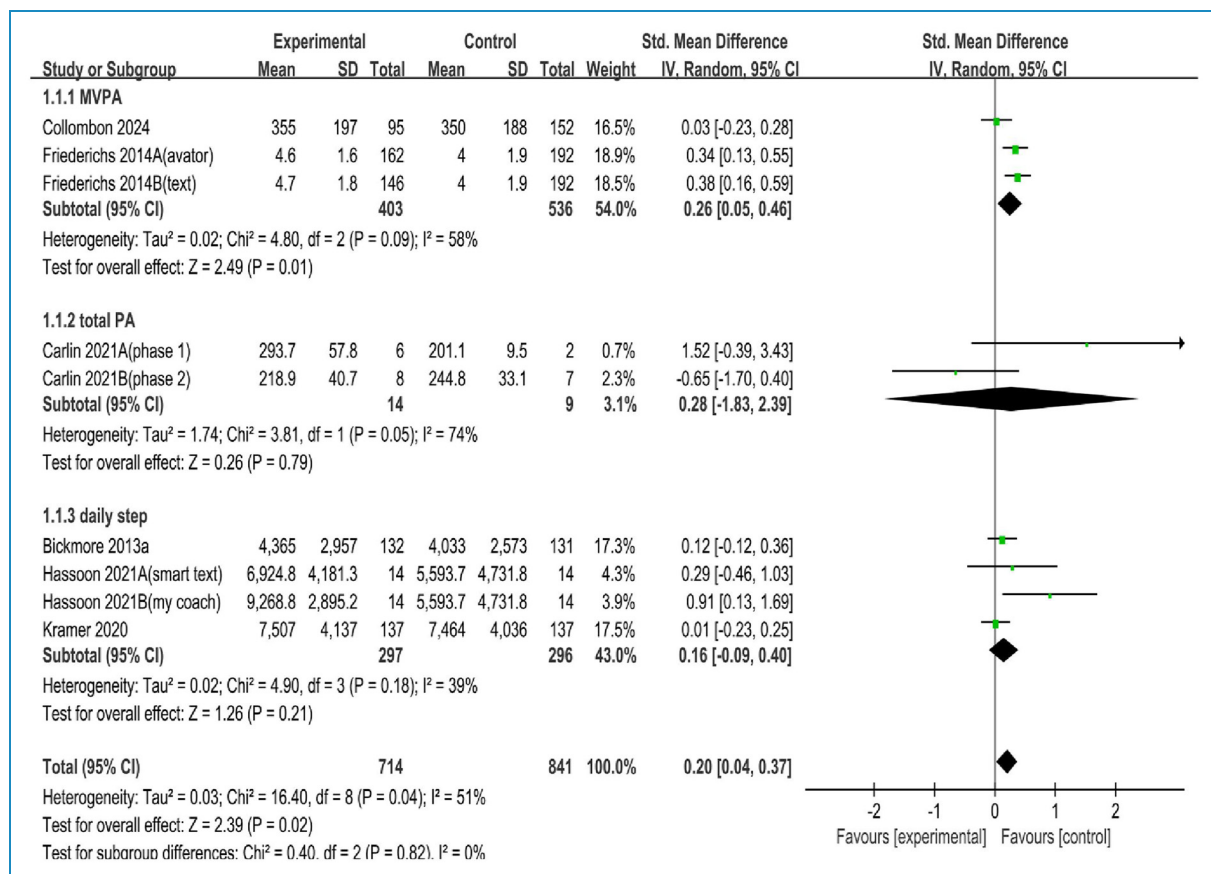


Figure 3. Forest plot of chatbot-based exercise intervention on PA.

Interpretations and suggestions

Interpretations of these review results could fall along a spectrum of possibilities, offering multiple perspectives for future development of chatbot-based EIs aimed at

increasing PA, fostering EHs, and reducing SB. SAs have indicated that these interventions are particularly effective in the short term, when implemented via desktop platforms, and when they incorporate both constrained and unconstrained chatbot input modalities. Three key factors

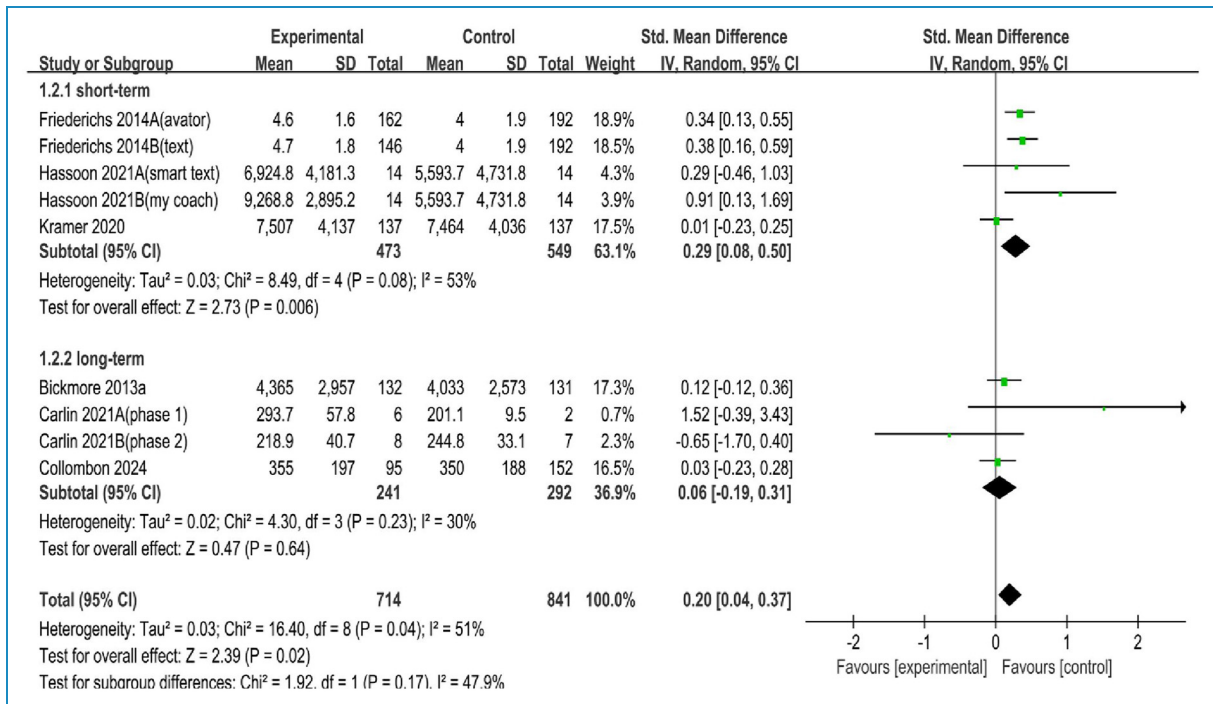


Figure 4. Forest plot of duration of intervention on PA.

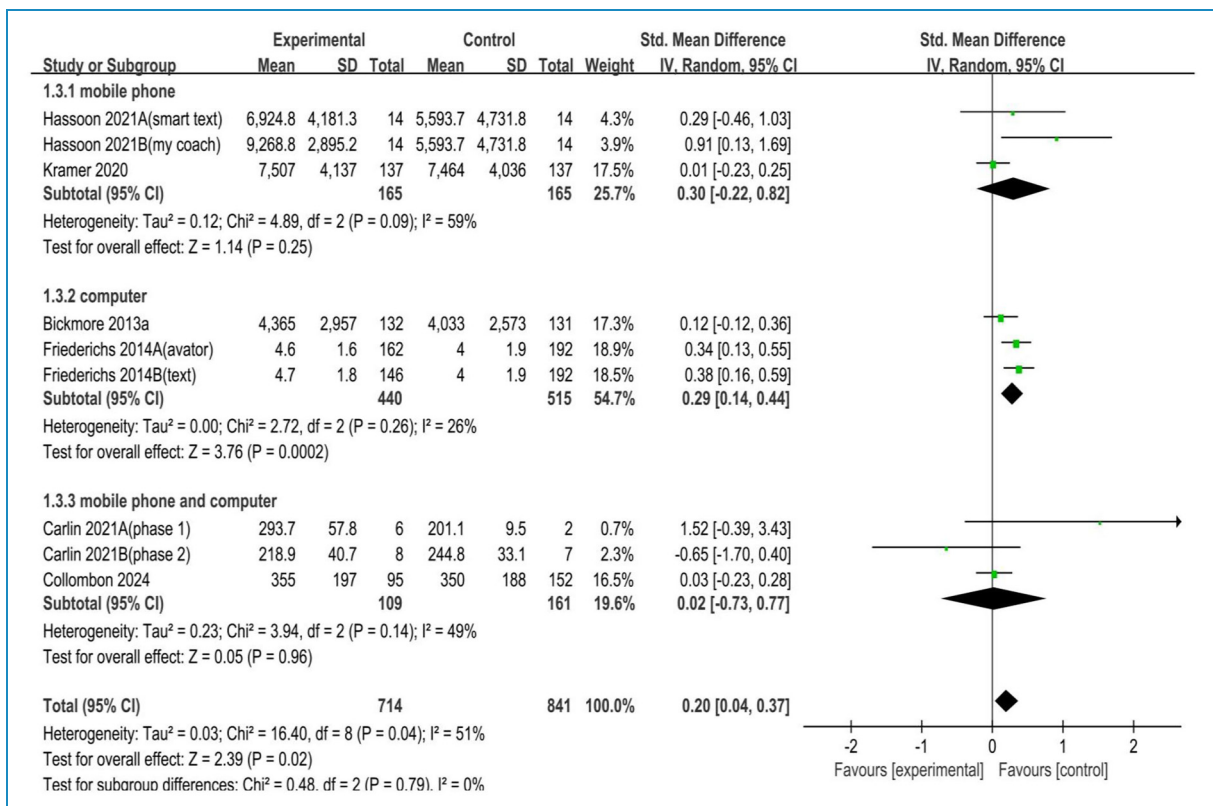


Figure 5. Forest plot of chatbot type of intervention on PA.

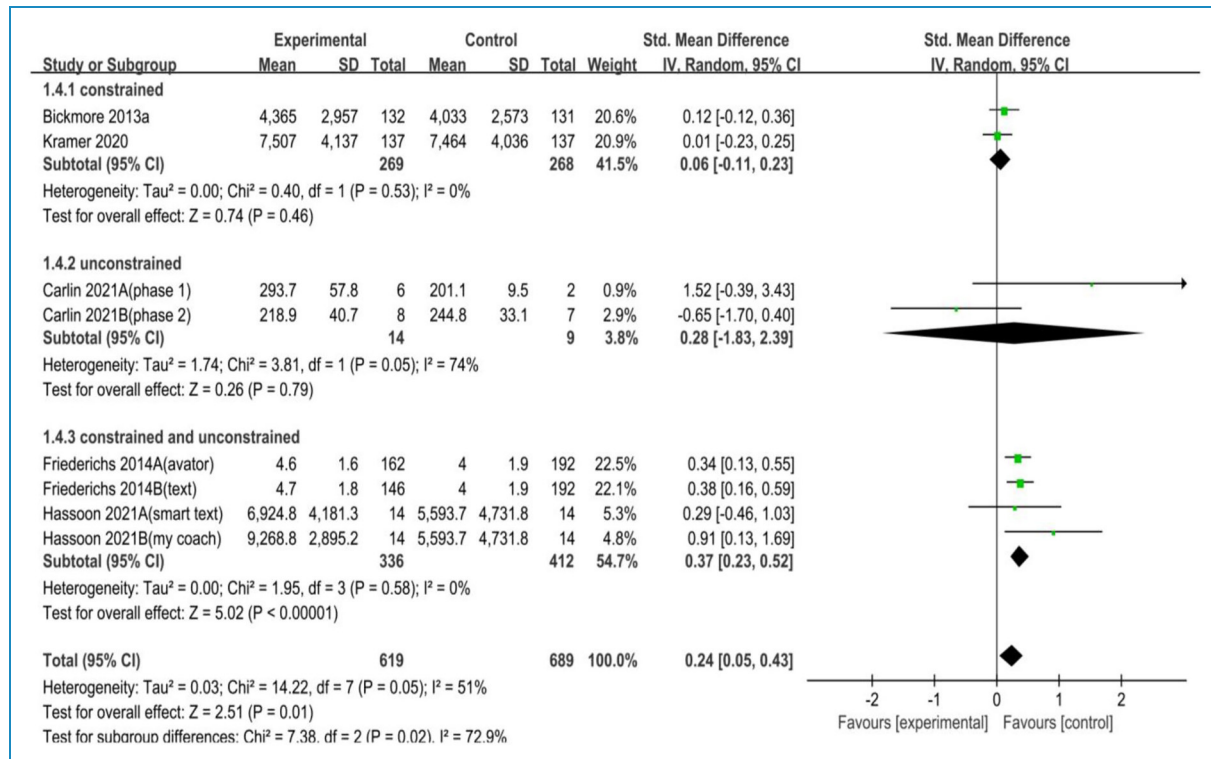


Figure 6. Forest plot of chatbot input of intervention on PA.

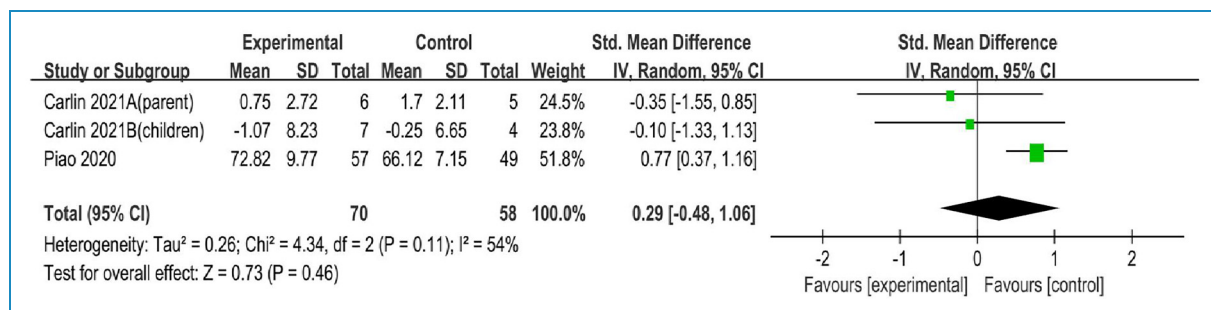


Figure 7. Forest plot of chatbot-based exercise intervention on exercise habits.

influencing the effectiveness of interventions were identified: duration, delivery method, and chatbot input modality (refer to Table 4).

Intervention duration appears to be a critical influencing factor. Short-term interventions tend to be more effective. Initial user engagement is frequently driven by the novelty or curiosity of the program; however, sustained participation often wanes due to the repetitive nature of the content. This decline in motivation, commonly termed as ‘intervention fatigue’, results from cognitive overload and emotional disconnection, particularly when the intervention lacks variability or fails to customize responses to individual user requirements. In the absence of personalized feedback or dynamic interactions,

participants may view the intervention as tedious or ineffectual, which can lead to disengagement or even burnout. Future research should focus on integrating a variety of engaging activities and periodically introducing new themes tailored to user preferences gathered via chatbot interactions. To maintain interest over time, chatbots should actively engage with users by providing personalized and attractive program suggestions. Furthermore, the establishment of consistent EHs is crucial. Embedding theories of goal-setting and habit formation within the chatbot algorithms may enhance long-term commitment. As users develop intrinsic motivation and autonomy in their exercise practices, the reliance on external interventions naturally decreases.

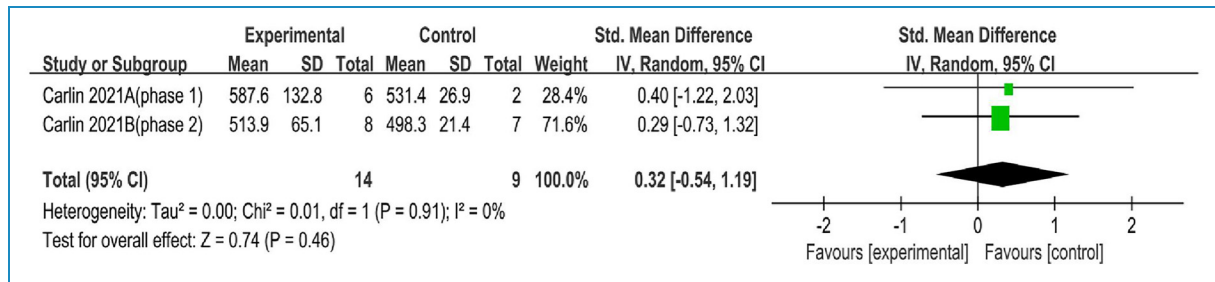


Figure 8. Forest plot of chatbot-based exercise intervention on sedentary time.

Table 4. The summary table of subgroup findings.

Key Factor	Findings	Interpretations	Future Recommendations
Duration	Short-term intervention was better.	Engagement tends to decline over time due to repetitive content and lack of personalization.	Incorporate varied and personalized content; use chatbot feedback to adapt themes and maintain interest.
Delivery	Computer-based chatbots were more effective than mobile-based ones.	Mobile phones can cause distractions (e.g., notifications), reducing adherence.	Improve mobile delivery with scheduled reminders, reduce distractions, and refine rather than abandon it.
Input	Combined constrained and unconstrained input yielded better outcomes.	Flexible input accommodates diverse user needs; rigid or too open formats can hinder some user groups.	Design age- and ability-appropriate input systems that are simplified for children and accessible for older adults.

The mode of delivery also emerged as a vital determinant of success. Notably, interventions delivered through computers proved more effective than those administered via mobile devices. Despite the ubiquity of mobile phones, their propensity for interruptions, such as notifications, can significantly detract from user engagement and adherence. Future interventions could address these challenges by incorporating features such as scheduled reminders, disabling notifications during sessions, or restricting access to other distracting applications. Nevertheless, given their convenience, mobile-based interventions warrant further refinement rather than complete abandonment.

Additionally, the modality of chatbot input is crucial. Interventions that combine constrained and unconstrained input modalities enable users to interact more flexibly, catering to a wide range of preferences. A standardized input mode may not suit all users, particularly vulnerable groups. For instance, children might misuse an entirely unrestricted input, whereas older adults may find overly rigid input formats challenging due to sensory impairments. Future studies should aim to develop input modalities that are age-appropriate and accessible, including simplified interfaces for children and enhanced usability features for older adults.

The importance of SB

The efficacy of chatbot-based interventions on SB remains suboptimal. A primary contributing factor is the prevalent research focus on enhancing PA while neglecting SB as a significant outcome. Furthermore, the development of sustained EHs generally requires long-term repetition within stable contexts and involves deep cognitive restructuring, as outlined in theories of habit formation. Chatbots, particularly those that do not track progress or provide long-term reinforcement, often lack the necessary persistence and contextual integration to support the development of these habits. Nevertheless, SB presents considerable risks to both physical and mental health.^{62–64} Although regular PA yields substantial health benefits, these do not completely mitigate the detrimental effects of extended periods of SB. Surprisingly, some studies included in this review have found that chatbot-based EIs showed a link to an increase in sedentary time. To overcome this issue, effectively reducing SB may require context-specific cues, such as sedentary alerts, which prompt users to take immediate action. To improve the effectiveness of chatbots in reducing SB, future interventions could incorporate passive monitoring tools (e.g., smartphone sensors or wearable devices) and deliver timely, personalized sedentary alerts.

Strengths and limitations

This review possesses some notable strengths. Primarily, it is the first to quantitatively evaluate the effectiveness of chatbot-based EIs on EHs and SB, offering a novel perspective for future research in this domain. Additionally, all included studies were RCTs, adhering to rigorous methodological standards. Consequently, the high quality of the selected articles lends significant credibility to the findings of this review. Furthermore, the studies encompassed participants from a broad age spectrum, ranging from young adults to the elderly. This demographic diversity enhances the generalizability of the results and minimizes potential biases. However, the review also presents certain limitations. Covered studies had small sample sizes, reflecting the nascent stage of research in chatbot-based EIs. Nonetheless, all studies met the minimum required sample sizes. Another limitation is that the included studies were predominantly carried out in high-income countries, such as the United States and the Netherlands, which may limit the applicability of the findings to low-resource settings. Future research should aim to evaluate chatbot-based EIs in a variety of socioeconomic and cultural contexts. Although the number of studies included was limited, this review integrated the most recent and comprehensive research available at the time.

Conclusion

The chatbot-based EI has been shown to significantly improve PA, particularly MVPA. However, it did not exert a significant impact on EH and SB. It is noteworthy that, in SAs, the duration and type of chatbot emerged as significant influencing factors. Subsequent studies can emphasize well-designed and tailored interventions that incorporate engaging and entertaining activities and apply theories related to the formation of EH. Attention should also be directed towards strategies for minimizing distractions during mobile phone use, and exploring the subdivision of diversified input modes for specific demographic groups. Since reducing sedentary time is a viable target in chatbot-based EIs, integration of passive monitoring tools (e.g., smartphone sensors or wearable devices) to deliver timely, personalized sedentary alerts is recommended. Furthermore, more RCTs involving diverse chatbots, conducted on a larger scale, in low-resource settings, and utilizing objective measures of SB (e.g., accelerometers), along with extended follow-up, are necessary to establish definitive conclusions. Future chatbot-based EIs should also aim to design chatbots that are adaptable to varied socioeconomic and cultural contexts and investigate the potential integration of technologies like ChatGPT into EIs.

Acknowledgements

The authors would like to acknowledge Hong Gu, and Yongwei, Li for their insightful advice on the concept analysis and strategic

direction of this study as well as their thoughtful consideration in the selection of terms.

ORCID iDs

Qian Wang  <https://orcid.org/0009-0005-2772-8127>

Wenxian Yang  <https://orcid.org/0009-0007-2451-4750>

Kai Xu  <https://orcid.org/0000-0002-5849-2664>

Lawrence T Lam  <https://orcid.org/0000-0001-6183-6854>

Ethical considerations

Ethics committee approval was not needed for this research because it did not engage in clinical trials involving animals or humans.

Author contributions

The team comprising WQ, Lawrence, YWX, and XK contributed to the review by designing its structure, searching for and screening literature, selecting applicable studies, extracting key data, compiling the results, and drafting and revising the manuscript. The final manuscript received the approval of all authors.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

The primary results of the study are recorded in both the article and supplementary materials. Contact the corresponding author for more information.

References

1. Global Burden of Disease Collaborative Network, Global Burden of Disease Study 2019 (GBD 2019) Results (2020, Institute for Health Metrics and Evaluation – IHME) <https://vizhub.healthdata.org/gbd-results/> (2019).
2. Kankeu HT, Saksena P, Xu K, et al. The financial burden from non-communicable diseases in low- and middle-income countries: a literature review. *Health Res Policy Syst* 2013; 11: 12.
3. Muka T, Imo D, Jaspers L, et al. The global impact of non-communicable diseases on healthcare spending and national income: a systematic review. *Eur J Epidemiol* 2015; 30: 251–277.
4. Fang EF, Xie CL, Schenkel JA, et al. A research agenda for ageing in China in the 21st century (2nd edition): focusing on basic and translational research, long-term care, policy and social networks. *Ageing Res Rev* 2020; 64: 26.
5. Bennett JE, Stevens GA, Mathers CD, et al. NCD Countdown 2030: worldwide trends in non-communicable

- disease mortality and progress towards sustainable development goal target 3.4. *Lancet* 2018; 392: 1072–1088.
6. World Health Organization. *Global action plan for the prevention and control of noncommunicable diseases 2013–2030*. Geneva: World Health Organization, 2013.
 7. World Health Organization. *Global status report on physical activity 2022*. Geneva: World Health Organization, 2022.
 8. Paxton RJ, Melanson EL, Stevens-Lapsley JE, et al. Physical activity after total knee arthroplasty: a critical review. *World J Orthop* 2015; 6: 614–622.
 9. Murphy ECS, Carson L, Neal W, et al. Effects of an exercise intervention using Dance Dance revolution on endothelial function and other risk factors in overweight children. *Int J Pediatr Obes* 2009; 4: 205–214.
 10. Wadden TA, Vogt RA, Kuehnel RH, et al. Exercise in the treatment of obesity: effects of four interventions on body composition, resting energy expenditure, appetite, and mood. *J Consult Clin Psychol* 1997; 65: 269–277.
 11. Hordern MD, Coombes JS, Cooney LM, et al. Effects of exercise intervention on myocardial function in type 2 diabetes. *Heart* 2009; 95: 1343–1349.
 12. Stoner L, Rowlands D, Morrison A, et al. Efficacy of exercise intervention for weight loss in overweight and obese adolescents: meta-analysis and implications. *Sports Med* 2016; 46: 1737–1751.
 13. Haghighi MM, Mavros Y and Singh MAF. The effects of structured exercise or lifestyle behavior interventions on long-term physical activity level and health outcomes in individuals with type 2 diabetes: a systematic review, meta-analysis, and meta-regression. *J Phys Act Health* 2018; 15: 697–707.
 14. Shawahna R, Batta A, Asa'ad M, et al. Exercise as a complementary medicine intervention in type 2 diabetes mellitus: a systematic review with narrative and qualitative synthesis of evidence. *Diabetes Metab Syndr-Clin Res Rev* 2021; 15: 273–286.
 15. Fiedler J, Eckert T, Wunsch K, et al. Key facets to build up eHealth and mHealth interventions to enhance physical activity, sedentary behavior and nutrition in healthy subjects - an umbrella review. *Bmc Public Health* 2020; 20: 1605–1621.
 16. Michie S, Yardley L, West R, et al. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *J Med Internet Res* 2017; 19: 13.
 17. Grady A, Yoong S, Sutherland R, et al. Improving the public health impact of eHealth and mHealth interventions. *Aust N Z Publ Health* 2018; 42: 118–119.
 18. McCarroll R, Eyles H and Ni Mhurchu C. Effectiveness of mobile health (mHealth) interventions for promoting healthy eating in adults: a systematic review. *Prev Med* 2017; 105: 156–168.
 19. Huang T, Zhao GG, Tan HY, et al. Effects of family intervention on physical activity and sedentary behavior in children aged 2.5–12 years: a meta-analysis. *Front Pediatr* 2021; 9: 10.
 20. Ridgers ND, Timperio A, Ball K, et al. Effect of commercial wearables and digital behaviour change resources on the physical activity of adolescents attending schools in socio-economically disadvantaged areas: the RAW-PA cluster-randomised controlled trial. *Int J Behav Nutr Phys Act* 2021; 18: 11.
 21. Williamson PJ, Atkinson G and Batterham AM. Inter-individual differences in weight change following exercise interventions: a systematic review and meta-analysis of randomized controlled trials. *Obes Rev* 2018; 19: 960–975.
 22. Fogg BJ and Euchner J. Designing for behavior change-new models and moral issues an interview with BJ fogg. *Res-Technol Manage* 2019; 62: 14–19.
 23. Cheng YC and Su CH. Evidence supports PA prescription for Parkinson's disease: motor symptoms and non-motor features: a scoping review. *Int J Environ Res Public Health* 2020; 17: 13.
 24. Who Guideline: Recommendations on Digital Interventions for Health System Strengthening. 2019.
 25. Nadarzynski T, Miles O, Cowie A, et al. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. *Digit Health* 2019; 5: 12.
 26. Usip PU, Udo EN, Asuquo DE, et al. A machine learning-based Mobile chatbot for crop farmers. In: *1st International Conference on Electronic Governance with Emerging Technologies (EGETC)*. 2022 Sep 12–14, Tampico, MEXICO, 2022, pp.192–211: Springer International Publishing Ag.
 27. Sperli G. A cultural heritage framework using a deep learning based chatbot for supporting tourist journey. *Expert Syst Appl* 2021; 183: 12.
 28. Aleedy M, Shaiba H and Bezbradica M. Generating and analyzing chatbot responses using natural language processing. *Int J Adv Comput Sci Appl* 2019; 10: 60–68.
 29. Abdul-Kader SA and Woods J. Survey on chatbot design techniques in speech conversation systems. *Int J Adv Comput Sci Appl* 2015; 6: 72–80.
 30. Selamat MA and Windasari NA. Chatbot for SMEs: integrating customer and business owner perspectives. *Technol Soc* 2021; 66: 12.
 31. Martins L, Fernández-Ferrer M and Puertas E. Analysing conversation pathways with a chatbot tutor to enhance self-regulation in higher education. *Educ Sci* 2024; 14: 21.
 32. Kim M, Oh J, Kim D, et al. Understanding user preferences in developing a mental healthcare AI chatbot: a conjoint analysis approach. *Int J Hum-Comput Interact* 2024; 41: 4813–4821.
 33. Ajzen I. The theory of planned behaviour: reactions and reflections. *Psychol Health* 2011; 26: 1113–1127.
 34. Bandura A. Health promotion from the perspective of social cognitive theory. *Psychol Health* 1998; 13: 623–649.
 35. Benight CC and Bandura A. Social cognitive theory of post-traumatic recovery: the role of perceived self-efficacy. *Behav Res Ther* 2004; 42: 1129–1148.
 36. Maher CA, Davis CR, Curtis RG, et al. A physical activity and diet program delivered by artificially intelligent virtual health coach: proof-of-concept study. *JMIR mHealth UHealth* 2020; 8: 11.

37. Booth F, Potts C, Bond R, et al. A mental health and well-being chatbot: user event log analysis. *JMIR mHealth UHealth* 2023; 11: e43052.
38. Schillings C, Meissner E, Erb B, et al. Effects of a chatbot-based intervention on stress and health-related parameters in a stressed sample: randomized controlled trial. *JMIR Ment Health* 2024; 11: 13.
39. Chew HSJ, Chew NW, Loong SSE, et al. Effectiveness of an artificial intelligence-assisted app for improving eating behaviors: mixed methods evaluation. *J Med Internet Res* 2024; 26: 14.
40. Wlasak W, Zwanenburg SP and Paton C. Supporting autonomous motivation for physical activity with chatbots during the COVID-19 pandemic: factorial experiment. *JMIR Form Res* 2023; 7: 15.
41. To QG, Green C and Vandelanotte C. Feasibility, usability, and effectiveness of a machine learning-based physical activity chatbot: quasi-experimental study. *JMIR mHealth UHealth* 2021; 9: e28577.
42. Collombon E, Bolman CAW, de Bruijn GJ, et al. The efficacy of online physical activity interventions with added mobile elements within adults aged 50 years and over: randomized controlled trial. *Appl Psychol-Health Well Being* 2024; 16: 1921–1943.
43. Piao M, Ryu H, Lee H, et al. Use of the healthy lifestyle coaching chatbot app to promote stair-climbing habits among office workers: exploratory randomized controlled trial. *JMIR mHealth UHealth* 2020; 8: e15085.
44. Peuters C, Maenhout L, Cardon G, et al. A mobile healthy lifestyle intervention to promote mental health in adolescence: a mixed-methods evaluation. *Bmc Public Health* 2024; 24: 19.
45. Oh YJ, Zhang JW, Fang ML, et al. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. *Int J Behav Nutr Phys Act* 2021; 18: 25.
46. Luo TC, Aguilera A, Lyles CR, et al. Promoting physical activity through conversational agents: Mixed methods systematic review. *J Med Internet Res* 2021; 23: 17.
47. Lyzwiniski LN, Elgendi M and Menon C. Conversational agents and avatars for cardiometabolic risk factors and lifestyle-related behaviors: scoping review. *JMIR mHealth UHealth* 2023; 11: 21.
48. Singh B, Olds T, Brinsley J, et al. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. *npj Digit Med* 2023; 6: 10.
49. Howick Jea. Oxford centre for evidence-based medicine levels of evidence, <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebml-levels-of-evidence> (2011).
50. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ-Br Med J* 2021; 372: 9.
51. Schardt C, Adams MB, Owens T, et al. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak* 2007; 7: 6.
52. Higgins Jea. Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022), www.training.cochrane.org/handbook (2022).
53. Bickmore TW, Schulman D and Sidner C. Automated interventions for multiple health behaviors using conversational agents. *Patient Educ Couns* 2013; 92: 142–148.
54. Bickmore TW, Silliman RA, Nelson K, et al. A randomized controlled trial of an automated exercise coach for older adults. *J Am Geriatr Soc* 2013; 61: 1676–1683.
55. Carlin A, Logue C, Flynn J, et al. Development and feasibility of a family-based health behavior intervention using intelligent personal assistants: randomized controlled trial. *JMIR Form Res* 2021; 5: 14.
56. Friederichs S, Bolman C, Oenema A, et al. Motivational interviewing in a web-based physical activity intervention with an avatar: randomized controlled trial. *J Med Internet Res* 2014; 16: 6.
57. Gardiner PM, McCue KD, Negash LM, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: a feasibility randomized control trial. *Patient Educ Couns* 2017; 100: 1720–1729.
58. Hassoon A, Baig Y, Naiman DQ, et al. Randomized trial of two artificial intelligence coaching interventions to increase physical activity in cancer survivors. *npj Digit Med* 2021; 4: 6.
59. Kramer JN, Künzler F, Mishra V, 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; 54: 518–528.
60. Sakane N, Suganuma A, Domichi M, et al. The effect of a mHealth app (KENPO-app) for specific health guidance on weight changes in adults with obesity and hypertension: pilot randomized controlled trial. *JMIR mHealth UHealth* 2023; 11: 11.
61. Tugault-Lafleur CN, Gonzalez OD, Macdonald J, et al. Efficacy of the Aim2Be intervention in changing lifestyle behaviors among adolescents with overweight and obesity: randomized controlled trial. *J Med Internet Res* 2023; 25: 20.
62. Rey-López JP, Bel-Serrat S, Santaliesra-Pasías A, et al. Sedentary behaviour and clustered metabolic risk in adolescents: the HELENA study. *Nutr Metab Cardiovasc Dis* 2013; 23: 1017–1024.
63. Same RV, Feldman DI, Shah N, et al. Relationship between sedentary behavior and cardiovascular risk. *Curr Cardiol Rep* 2016; 18: 7.
64. Teychenne M, Ball K and Salmon J. Sedentary behavior and depression among adults: a review. *Int J Behav Med* 2011; 17: 246–254.