Received: 22 February 2022 | Revised: 7 November 2022 | Accepted: 30 November 2022

DOI: 10.1111/ijpo.12997

REVIEW ARTICLE

Pediatric OBESITY WILEY

The long-term effects of primary school-based obesity prevention interventions in children: A systematic review and meta-analysis

Michel S. Smit[1] | Mirte Boelens[1] | Famke J. M. Mölenberg[1] | Hein Raat[1] | Wilma Jansen[1,2]

[1]Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, the Netherlands

[2]Department of Social Development, City of Rotterdam, Rotterdam, the Netherlands

Correspondence
Wilma Jansen, Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, the Netherlands.
Email: w.jansen@rotterdam.nl

Funding information
City of Rotterdam, Department of Sport

Summary

Introduction: This systematic review and meta-analysis investigate the long-term effects of primary school-based obesity prevention interventions on body-mass index (and z-scores), waist circumference (and z-scores) and weight status.

Methods: Four databases were searched for studies from date of inception until June 8th, 2021. We included randomized controlled trials (RCT) and non-RCTs investigating effects ≥12 months post-intervention of primary school-based interventions with intervention duration ≥6 months and containing a diet and/or physical activity component on outcomes of interest. Articles were assessed on risk of bias and methodological quality by RoB2 and ROBINS-I. Meta-analysis was performed and results were narratively summarized. Evidence quality was assessed with GRADE.

Results: Nineteen studies were included, 9 were pooled in a meta-analysis. No long-term effects were found on body-mass index (+0.06 kg/m^2; CI95% = −0.38, 0.50; I^2 = 66%), body-mass index z-scores (−0.08; CI95% = −0.20, 0.04; I^2 = 36%), and waist circumference (+0.57 cm; CI95% = −0.62, 1.75; I^2 = 13%). Non-pooled studies reported mixed findings regarding long-term effects on body-mass index, body-mass index z-scores and weight status, and no effects on waist circumference and waist circumference z-scores. Evidence certainty was moderate to very low.

Discussion: No clear evidence regarding long-term effects of primary school-based interventions on obesity-related outcomes was found. Recommendations for further research and policy are discussed.

Prospero registration ID: CRD42021240446.

KEYWORDS
elementary school, prevention, program, lifestyle, sustained effect

Abbreviations: BMI, body-mass index; CI, confidence interval; GRADE, grading of recommendations assessment, development and evaluation; PRISMA-P, preferred reporting items for systematic reviews and meta-analysis protocol; PROSPERO, prospective register of systematic reviews; RCT, randomized controlled trial; RoB2, risk of bias 2; ROBINS-I, risk of bias in non-randomized studies of interventions; WC, waist circumference; WHO, World Health Organization; zBMI, BMI-for-age z-scores; zWC, WC-for-age z-scores.

1 | INTRODUCTION

Current estimates by the World Health Organization (WHO) are that almost one in five children worldwide are affected by overweight or obesity.[1] Although a recent stabilization has been confirmed regarding the overweight and obesity prevalence in higher-income countries, the prevalence overall remains high.[2,3] Studies have shown that effective interventions for obesity prevention accomplish more positive effects on body-mass index (BMI) for younger children.[4] A recent Cochrane review by Brown and colleagues (2019) counted 153 randomized controlled trials (RCT) investigating the effectiveness of interventions that include diet or physical activity components, or both, designed for preventing overweight and obesity in children and adolescents.[5] They have concluded that interventions that focus only on physical activity can reduce the risk of obesity for children aged 6 to 12 years old, whereas they found no evidence for the effectiveness of diet interventions and some evidence for the effectiveness of interventions with a combined diet component and physical activity component.[5] The authors could not entirely explain the differences in the effects of interventions by the setting or duration of the interventions.[5] Although obesity prevention interventions may be implemented in many settings, schools have been suggested to be the optimal place to deliver interventions due to continuous contact with children[6] and reaching children with a wide range of different backgrounds.[7] Additionally, the results of other systematic reviews support the promising role of school-based interventions,[8,9] specifically primary school-based interventions,[10] for the prevention of childhood obesity.

RCTs are often considered the golden standard for effectiveness evaluation studies.[11] However, only a limited number of RCTs have a long-term follow-up (≥12 months post-intervention). It might be challenging to conduct such long-term RCTs due to funding issues or because it is unethical to withhold children from effective interventions for more years. In their review, Brown and colleagues (2019) encourage researchers to perform more long-term follow-up studies into the sustained effects of interventions. If RCTs are not feasible, researchers might turn to other types of designs, like natural experiments, quasi-experimental and observational studies with comparison groups as these can also generate causal evidence with good external validity.[12–14] Insights in the long-term effects of interventions are especially relevant for the transition from childhood into adolescence as the WHO highlighted that adolescence is a period of excessive weight gain because adolescents gain more freedom in food choices, while their physical activity levels often decrease.[15]

To the best of our knowledge, a recent systematic review focusing on the long-term effects of primary school-based obesity prevention interventions in children is not available. Such a review aggregating the available evidence may either help policymakers in developing and implementing successful interventions of which the effects sustain into adolescence and adulthood or may highlight the need for supplementary interventions at later ages. Therefore the research question in this systematic review and meta-analysis is: What are the long-term effects (≥12 months follow-up post-intervention) of primary school-based obesity prevention interventions in children on multiple indicators of overweight and obesity? Intervention effects on BMI, BMI-for-age z-scores (zBMI), WC, WC-for-age z-scores (zWC) and weight status will be included as important measures of obesity and in predicting obesity-related health risk, as clinical guidelines suggest.[16]

2 | METHODS

2.1 | Protocol

A protocol for this systematic review and meta-analysis was developed in line with the Preferred Reporting Items for Systematic reviews and Meta-analysis Protocol (PRISMA-P).[17] This protocol was prospectively registered within Prospective Register of Systematic Reviews (PROSPERO) on 02-04-2021 with registration ID: CRD 42021240446.

2.2 | Inclusion and exclusion criteria

2.2.1 | Type of primary studies

We included RCTs and studies with other controlled experimental and observational designs in this review. We excluded studies with pre-post designs without control group.

2.2.2 | Intervention and setting

This review focusses on primary school-based obesity prevention interventions. Therefore, interventions had to be at least partially primary school-based to be included, but we allowed interventions in multiple settings. Following Brown and colleagues (2019), we included interventions containing at least a diet or physical activity component.[5] Interventions could also include both components or other components in addition to diet or physical activity.

Only interventions with a duration of ≥6 months were included in this review because, according to the theory of Prochaska and DiClemente, a duration of at least 6 months is needed for behaviour change to stabilize when it involves physical activity.[18] To study long-term effects, the intervention effects should have been studied following a period in which there was no exposure to the intervention. Thus, we only included primary studies in which intervention effects were studied ≥12 months post-intervention.

2.2.3 | Control

We included studies with a control or comparator condition without an intervention.

2.2.4 | Participants in intervention

As prospectively determined, we included studies in which at least 90% of the children were 6- to 12-year-olds at the start of the intervention. When we were not able to determine this criterion, all children in the study had to be within the age range for the study to be included. Furthermore, studies mentioning primary school-aged children, without the notion of an age range, were also considered eligible for inclusion.

2.2.5 | Participants at follow-up

Since this review focuses on long-term effects of interventions (≥12 months follow-up post-intervention), children studied at follow-up may have transcended the age of 12 years old. No criterion was set on age range at follow-up.

2.2.6 | Context

In this review primary studies regarding the general population were included. Therefore, children with overweight or obesity were allowed as a part of the participant sample. However, studies that were solely focusing on children with overweight or obesity were excluded since this review focusses on interventions preventing these conditions in the general population. Studies with participants with other clinical conditions were also not eligible for inclusion in this review.

We applied no further restrictions regarding subgroups of the population (e.g., country of origin or socioeconomic status). Furthermore, no restrictions were drawn upon the country of origin of the studies.

2.2.7 | Outcomes

Outcomes of interest for this review were BMI, zBMI, WC, zWC, and weight status. zBMI is a score in which BMI is adjusted to age and gender of the individual based on growth standards and zWC is a score in which WC is adjusted to age and gender.[19] Weight status can be categorized as having underweight, normal weight, overweight or obesity using common cut-off standards.

2.2.8 | Language

Studies in English were eligible for inclusion in this review. Studies in other languages were excluded due to language barriers of the authors.

2.3 | Search strategy

A systematic literature search without language restrictions or date limits was conducted to find relevant studies in four databases (Embase.com, Medline ALL Ovid, Web of Science SCI-EXPANDED & SSCI and Cochrane CENTRAL register of Trials) from date of inception until the 8th of June 2021. In addition, we hand-searched references of included studies and the Cochrane review by Brown and colleagues.[5]

Corresponding authors were contacted when the full texts were unavailable on the internet. Combined thesaurus terms and terms in title/abstract were used for three elements: intervention or program, body-mass index or weight status or overweight or obesity or waist circumference and children. The syntax and thesaurus terms within this search strategy were adapted to fit to each of the used databases. The full search is presented in supplemental file 1 (Appendix S1).

2.4 | Selection process

EndNote X9 Software was used to guide and organize the selection process of the studies that were found. Duplicates of records were retrieved and removed by guidance of Bramer and colleagues.[20] Hereafter title and abstracts were screened for eligibility by two independent reviewers (M.S. and M.B.). Then these studies were screened on full text to select the studies eligible for inclusion (M.S. and M.B.). A prospectively discussed order of reasons for exclusion was determined by the screening authors. Studies that were excluded based on full text screening, were categorized based on the reason of exclusion. When differences or disagreements between the two reviewers were encountered in the process, discussion with a third reviewer (W.J.) took place to reach consensus.

2.5 | Data extraction

All data was extracted to a standardized data extraction Excel sheet by two independent authors (M.S. and M.B.). Discrepancies were discussed until consensus was reached. If available, the following data was extracted for all included studies: (1) first author, (2) year of publication, (3) full reference, (4) study objective, (5) theoretical framework behind the intervention, (6) country of intervention, (7) funding sources, (8) conflict of interest statement, (9) study design, (10) intervention and control group characteristics (sample size, age, gender), (11) number of clusters, (12) numbers of participating primary schools, (13) intervention characteristics (duration, components and setting), (14) comparison or control condition, (15) follow-up period, (16) context, (17) outcomes at baseline, (18) outcomes at follow-up, (19) estimate of the effect sizes for the association between primary school-based intervention and outcome (coefficients, CI's, *p*-values) and (20) retention rates.

Specifically for the meta-analysis, mean and standard deviations were extracted for the outcome variables. When results from multiple models were available, results from the model with the most confounders added were extracted. When multiple follow-up time points were available for a study we extracted data for all follow-up time

points. For the purpose of the meta-analyses, we chose the time point of the longest follow-up within adolescence for inclusion in the meta-analysis, as suggested by the Cochrane Handbook.[21] The remaining earlier follow-up time points were included in an additional sensitivity meta-analysis.

2.6 | Risk of bias assessment

A risk of bias assessment was performed by two independent authors (M.S. and M.B.) to assess the quality of each included study. When differences or disagreements between the two reviewers were encountered in the process, discussion with a third reviewer (W.J.) took place to reach consensus. The revised Cochrane Collaboration's Risk of Bias 2 (RoB2) tool was used for assessing risk of bias for randomized trials[22] and the Risk of Bias in Non-randomized studies of interventions (ROBINS-I) tool was used for assessing risk of bias for non-randomized studies.[23] The RoB2 assesses bias (1) in the randomization process, (2) due to deviations from intended interventions, (3) due to missing outcome data, (4) due to measurement of outcomes and, (5) in selection of the reported result. All types of bias were rated as low risk, some concerns or high risk. The overall risk of bias score was determined by the highest risk score. The ROBINS-I contains items for bias (1) due to confounding, (2) in selection of participants into the study, (3) in classification of interventions, (4) due to deviations from intended interventions, (5) due to missing data, (6) in measurement of outcomes and, (7) in selection of the reported result. All domains were rated as having low, moderate, serious or critical risk of bias. Like in the ROB2, the overall risk of bias score was determined by the highest risk score.

2.7 | Data synthesis and analysis

The characteristics for all the included studies and interventions were described. We conducted a meta-analysis for intervention effects for which at least three studies were available with the same study design and outcome measure and if the outcome measures could be calculated into mean and standard deviation values per group at follow-up. We used the mean differences at follow-up in outcome measures between the intervention group and control group. We used random-effects models to account for the between-study variance. The meta-analyses were conducted in Review Manager (version 5.4). The overall pooled mean difference and 95% confidence intervals (CI) were calculated. Forest plots were created to present the meta-analysis results. I^2 measures have been presented as an indication for the proportion of total variability due to between study heterogeneity. Additional sensitivity analyses were performed including available earlier follow-up time points. The remaining extracted results that could not be pooled in the meta-analyses or sensitivity meta-analyses were narratively described. Therefore, studies could be considered as both pooled for some pooled outcome variables and non-pooled for its other non-pooled outcome variables. The narratively described results were divided into results derived from RCT studies and results derived from non-RCT studies.

Summary of findings tables were created with the Grading of Recommendations Assessment, Development and Evaluation (GRADE) pro software to present the evidence quality across studies and certainty of effects of the interventions.[24] The evidence quality was assessed according to the GRADE handbook.[25] The effects that were included in the meta-analyses were added in the summary of findings tables as pooled outcome variables. Results that could not be included in the meta-analysis were narratively described in the summary of findings tables as the non-pooled outcomes. For the narratively described non-pooled outcomes a subdivision was created for evidence resulting from RCT studies and evidence resulting from non-RCT studies. The GRADE approach for rating the evidence quality began with rating the study design. The evidence quality rating could then be reduced by rating the factors: risk of bias, inconsistency in results, indirectness of evidence, imprecision, and publication bias. The evidence quality rating could be increased by the factors: large magnitude of effect, plausible confounding and dose–response gradient. Possible final grades of evidence were high, moderate, low, or very low certainty.

3 | RESULTS

3.1 | Study selection

The process of inclusion and exclusion of articles during article selection is presented in a PRISMA flow chart in Figure 1.[17] A total of 15 887 records were identified through comprehensive database searches. After the removal of duplicates, 9188 records remained for screening. After screening on title and abstract, 211 articles remained for full text screening. From these 211 articles, a total of 192 articles were excluded. Most important reasons for exclusion were the study design ($n = 27$), follow-up length ($n = 113$) and age of the population ($n = 23$). Supplemental file 2 (Appendix S1) presents a complete overview of the excluded references during full text screening per exclusion category. Eventually, 19 articles were found eligible for inclusion into this systematic review.[26–44] Of these 19 articles, 9 could be included into either the meta-analysis or additional sensitivity meta-analysis on earlier time-points.[26,27,32,35–38,43,44]

3.2 | Study characteristics

The most important study characteristics are presented in Table 1. Study objective, theory behind the intervention, scale-up, funding, conflict of interest statement, intervention content and control content are presented in Table S1. Furthermore, Table S3 presents a complete detailed overview of the results that were extracted from the 19 included articles in this review. Of the 19 studies, 16 were published in or after 2010.[26–31,33–36,38–41,43,44] The 19 studies were performed in 11 different countries, subdivided over the continents Europe

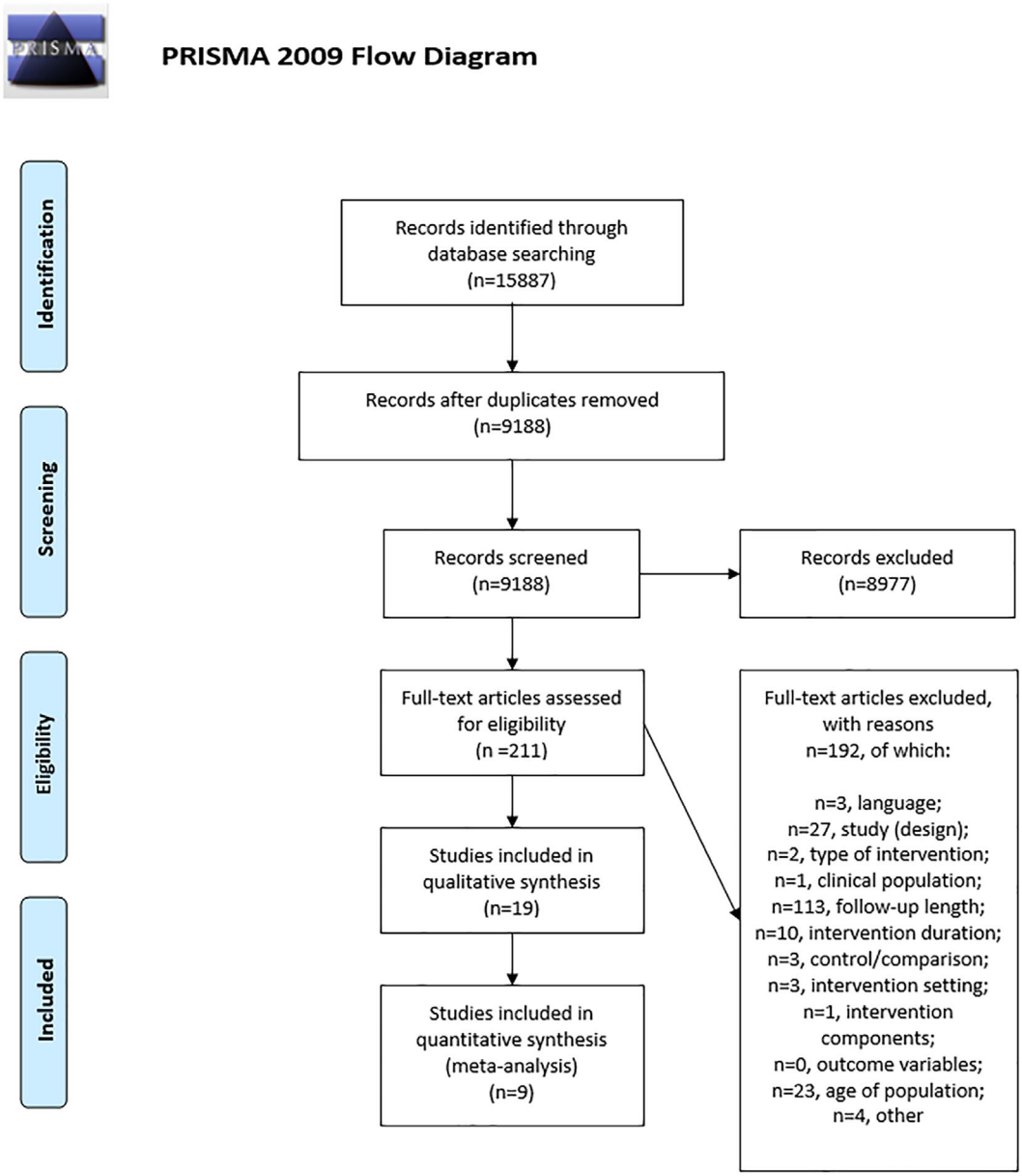

FIGURE 1 Flow diagram showing the selection process in this review

(10/19),[26,27,30,32,33,35,36,41,43,44] North-America (4/19),[29,37,39,40] Asia (3/19)[28,31,34] and Oceania (2/19).[38,42] Five studies reported to be based on the Social Cognitive Theory,[26,30,31,40,42] one reported the socio-ecological conceptual model,[36] one reported the Behaviour Change Theory,[37] one reported the concepts of the WHO Ottawa Charter for Health Promotion,[39] one reported Piaget's Cognitive Theory[40] and one reported the Behavioural Choice Theory.[42] Of the 19 studies, five interventions were considered a scale-up.[26,29,31,34,37] Most studies (13/19) were (cluster-)RCTs.[26–28,31,32,34–38,42–44] Other designs were a repeated measures crossover,[29] a non-randomized controlled trial,[30] two quasi-experimental,[33,40] a repeated cross-sectional[39] and a population-based prospective controlled[41] design. Four studies had a sample size between 100 and 250 participants,[29,38,40,41] six studies had a sample size between 251 and 500 participants,[27,30,32,33,36,42] five studies had a sample size between 501 and 1000 participants[28,31,35,39,44] and four studies had a sample size between 1001 and 5000 participants.[26,34,37,43]

All interventions took place in a primary school setting,[26–44] although six interventions additionally contained a home setting component[28,35–37,42,44] and one intervention additionally contained a home setting, community setting and online setting component.[30] Classroom or school teachers (partially) delivered the interventions in most studies.[26,28–35,37,41] Three interventions solely contained a diet component[27,32,43] and seven interventions solely contained an physical activity component,[29,33,34,36,38,41,42] whereas nine interventions contained both a diet and an physical activity component.[26,28,30,31,35,37,39,40,44] Eight interventions contained an additional non-diet non-physical activity component,[26,28,35,37,39,40,42,44] which is

TABLE 1 Characteristics of included studies

Author and year	Intervention/ study name	Country	Study design	Settings	No of primary schools No of clusters in analysis	Intervention delivered by	Intervention components: Diet, Physical activity, Other	Intervention duration
Anderson, 2016[27]	The Active for Life Year 5 (AFLY5)	England	Cluster-randomized controlled trial	Primary school	60 primary schools 60 clusters	School teachers, Learning support assistants	Diet, Physical activity, Sedentary behaviour, Screen viewing	6–7 months
Bere, 2014[28]	The Fruits and Vegetables Make the Marks (FVMM) project	Norway	Cluster-randomized controlled trial	Primary school	38 primary schools 38 clusters	NR	Diet	9 months
Cao, 2019[29]	The Family-Individual-School (FIS)-based comprehensive childhood-obesity intervention	China	Cluster-randomized controlled trial	Primary school, Home	14 primary schools Clusters at class level; amount NR	Health teachers, Teachers	Diet, Physical activity	34 months
Gao, 2013[30]	The Dance Dance Revolution-based exercise intervention	United States of America	Repeated-measures crossover design	Primary school	1 primary school Clusters NR	Classroom teachers, Research assistants	Physical activity	7 months
Gorely, 2011[31]	The GreatFun2Run intervention	England	Non-randomized controlled trial	Primary school, Community, Home, Online	8 primary schools 8 clusters	School teachers	Diet, Physical activity	10 months
Habib-Mourad, 2020[32]	The Ajyal Salima ("Healthy Kids") intervention, originating from the "Health-E-PALS intervention".	Lebanon	Stratified cluster-randomized controlled trial	Primary school	36 primary schools 36 clusters	Teachers, Researchers (to train the teachers)	Diet, Physical activity	16 months (over two academic years)
James, 2007[33]	The Christchurch obesity prevention programme in schools (CHOPPS)	England	Cluster-randomized controlled trial	Primary school	6 primary schools 29 clusters (class level)	One investigator, School teachers	Diet	12 months
Jurak, 2013[34]	The enhanced PE programme	Slovenia	Quasi-experimental	Primary school	9 primary schools Clusters NR	General teacher, Specialist PE teacher	Physical activity	48 months
Li, 2010[35]	The Happy 10 program	China	Cluster-randomized controlled trial	Primary school	20 primary schools 132 clusters (class level)	Teachers, Health educators	Physical activity	1 school year
Llaurado, 2018[36]		Spain						28 months

TABLE 1 (Continued)

Author and year	Intervention/ study name	Country	Study design	Settings	No of primary schools No of clusters in analysis	Intervention delivered by	Intervention components: Diet, Physical activity, Other	Intervention duration
	The Educació en Alimentació (EDAl) program		Randomized controlled trial	Primary school, Home	38 primary schools Clusters NR	School teachers, university students who acted as health promoting agents	Diet, Physical activity, Sedentary behaviour	
Meyer, 2014[37]	The Kinder- und Jugendsportstudie (KISS)	Switzer- land	Cluster-randomized controlled trial	Primary school, Home	15 primary schools 28 clusters (class level)	PE teachers	Physical activity	12 months
Nader, 1999[38]	The Child and Adolescent Trial for Cardiovascular Health (CATCH) school and family intervention (CATCH III)	United States of America	Randomized controlled trial	Primary school, Home	96 primary schools Clusters NR	Classroom teachers, Food service personnel, PE specialists	Diet, Physical activity, Cigarette smoking	36 months
Nogueira, 2017[39]	The CAPO kids trial	Australia	Cluster-controlled randomized trial	Primary school	2 primary schools 2 clusters	Instructors	Physical activity	9 months
Ofosu, 2018[40]	APPLE schools	Canada	Repeated cross-sectional	Primary school	173 primary schools Clusters NR	School health facilitators	Diet, Physical activity, Mental health	36 months
Puma, 2013[41]	The Integrated Nutrition and Physical Activity Program (INPAP)	United States of America	Quasi-experimental	Primary school	Primary schools NR (1 school district) Clusters NR	Local resource teachers	Diet, Physical activity, Screen viewing	24 months
Rosengren, 2021[42]	The Paediatric Osteoporosis Prevention (POP) study	Sweden	Population-based prospective controlled intervention study	Primary school	4 primary schools Clusters NR	Regular school teachers	Physical activity	6–8 years
Salmon, 2008[43]	The Switch-Play intervention	Australia	Cluster randomized controlled trial	Primary school, Home	3 primary schools 17 clusters (class level)	PE teachers, Interventionists	Physical activity, Sedentary behaviour, Screen viewing	9 months
Stea, 2018[44]	The Fruits and Vegetables Make the Marks (FVMM) Project	Norway	Cluster-randomized controlled trial	Primary school	38 primary schools Clusters NR	NR	Diet	9 months
Tarro, 2014[45]		Spain	Cluster- randomized controlled trial		38 primary schools			28 months

(Continues)

TABLE 1 (Continued)

Author and year	Intervention/ study name	Country	Study design	Settings	No of primary schools No of clusters in analysis	Intervention delivered by	Intervention components: Diet, Physical activity, Other	Intervention duration
	The Educació en Alimentació (EDAl) program			Primary school, Home	2 clusters	University students who acted as health promoting agents	Diet, Physical activity, Sedentary behaviour	

Author and year	Follow-up length	No. of total group participants at follow-up (% female) No. of intervention group participants at follow-up (% female) No. of control group participants at follow-up (% female)	Baseline age total group (mean(SD)) Baseline age intervention group (mean(SD)) Baseline age control group (mean(SD))	Follow-up age total group (mean(SD)) Follow-up age intervention group (mean(SD)) Follow-up age control group (mean(SD))	Other specific population characteristics	Retention rate at follow-up
Anderson, 2016[27]	12 months	$n = 2221$ (50.8%) in total $n = 1064$ (48.9%) in intervention group $n = 1157$ (52.5%) in control group	9.5 (0.3) y in total group 9.5 (0.3) y in intervention group 9.5 (0.3) y in control group	NR in total group NR in intervention group NR in control group	NR	80.7%
Bere, 2014[28]	36 and 84 months	$n = 320$ (62%) in total $n = 112$ (59%) in intervention group $n = 208$ (64%) in control group	11.8 (NR) y in total group NR in intervention group NR in control group	NR in total group NR in intervention group NR in control group	At baseline, 45% of intervention group was in 7th grade. The remainder in 6th grade. At baseline, 48% of control group was in 7th grade. The remainder in 6th grade.	16.4%
Cao, 2019[29]	24 and 36 months	$n = 807$ (46.5%) in total $n = 406$ (42.9%) in intervention group $n = 401$ (50.1%) in control group	NR in total group 7.00 (0.41) y in intervention group 6.82 (0.24) y in control group	NR in total group 13.68 (0.40) y in intervention group 13.50 (0.24) y in control group	NR	33.0%
Gao, 2013[30]	12 months	$n = 156$ (41.2%) in total, including 53 participants with no long-term effects measured due to continued intervention in year 2. $n = 68$ (NR) in intervention group $n = 44$ (NR) in control group	10.32 (0.91) y in total group Between 10 and 12 y in intervention group Between 10 and 12 y in control group	10.28 (0.90) y in total group NR in intervention group NR in control group	The participating school had the highest obesity rate in the state of Utah, US. Vast majority of children were from economically disadvantaged, Latino immigrant families. Approximately 73% of children received free lunch.	79.3%

TABLE 1 (Continued)

Author and year	Follow-up length	No. of total group participants at follow-up (% female) No. of intervention group participants at follow-up (% female) No. of control group participants at follow-up (% female)	Baseline age total group (mean(SD)) Baseline age intervention group (mean(SD)) Baseline age control group (mean(SD))	Follow-up age total group (mean(SD)) Follow-up age intervention group (mean(SD)) Follow-up age control group (mean(SD))	Other specific population characteristics	Retention rate at follow-up
Gorely, 2011[31]	Between 18 and 20 months	$n = 421$ (NR) in total $n = 206$ (NR) in intervention group $n = 215$ (NR) in control group	NR in total group 8.8 (NR) y in intervention group 8.9 (NR) y in control group	NR in total group NR in intervention group NR in control group	The majority of participants were of white British ethnicity (intervention 94.8% and control 96.5%)	71.5%
Habib-Mourad, 2020[32]	12 months	$n = 806$ (NR) in total $n = 457$ (NR) in intervention group $n = 349$ (NR) in control group	9.95 (1.13) y in total group 9.81 (0.68) y in intervention group 10.13 (0.68) y in control group	NR in total group NR in intervention group NR in control group	The study included public schools (primarily low socioeconomic status) and private schools (primarily middle and high socioeconomic status).	65.1%
James, 2007[33]	24 months	$n = 434$ (48.2%) in total $n = 219$ (45.2%) in intervention group $n = 215$ (51.2%) in control group	8.6 (NR) y, with a range of 7.0–10.9 y in total group 8.7 (0.8) y for boys in intervention group 8.7 (1.0) y for girls in intervention group 8.6 (0.9) y for boys in control group 8.7 (0.9) y for girls in control group	11.6 (NR) y, with a range of 10.0–13.9 y in total group NR in intervention group NR in control group	NR	67.4%
Jurak, 2013[34]	84 months	$n = 324$ (48.8%) in total $n = 160$ (46.3%) in intervention group $n = 164$ (51.2%) in control group	NR in total group 7.76 (0.33) y in intervention group 7.71 (0.32) y in control group	NR in total group NR in intervention group NR in control group	Urban locations	72.8%
Li, 2010[35]	12 months	$n = 4120$ (48.6%) in total $n = 2028$ (46.4%) in intervention group $n = 2092$ (50.7%) in control group	9.3 (0.7) y in total group NR in intervention group NR in control group	NR in total group NR in intervention group NR in control group	NR	87.7%
Llaurado, 2018[36]	48 months	$n = 503$ (NR) in total $n = 154$ (NR) in intervention group	NR in total group (age range of 7–8 y). NR in intervention group	NR in total group 15.6 (0.53) y in intervention group	NR	21.4%

(Continues)

TABLE 1 (Continued)

Author and year	Follow-up length	No. of total group participants at follow-up (% female) No. of intervention group participants at follow-up (% female) No. of control group participants at follow-up (% female)	Baseline age total group (mean(SD)) Baseline age intervention group (mean(SD)) Baseline age control group (mean(SD))	Follow-up age total group (mean(SD)) Follow-up age intervention group (mean(SD)) Follow-up age control group (mean(SD))	Other specific population characteristics	Retention rate at follow-up
		$n = 349$ (NR) in control group	NR in control group	14.9 (0.36) y in control group		
Meyer, 2014[37]	36 months	$n = 173$ (50.3%) first graders in total $n = 116$ (61.2%) fifth graders in total $n = 108$ (47%) first graders in intervention group $n = 81$ (59%) fifth graders in intervention group $n = 65$ (55%) first graders in control group $n = 35$ (66%) fifth graders in control group	NR in total group 6.9 (0.3) y for first graders in intervention group 10.9 (0.5) y for fifth graders in intervention group 6.9 (0.3) y for first graders in control group 11.1 (0.6) y for fifth graders in control group	NR in total group 10.6 (0.3) y for first graders in intervention group 15.0 (0.5) y for fifth graders in intervention group 10.6 (0.4) y for first graders in control group 15.1 (0.6) y for fifth graders in control group	First graders: 31% overweight in intervention group and 14% overweight in control group at baseline. Fifth graders: 20% overweight in both groups at baseline. 14–31% migrants.	57.6%
Nader, 1999[38]	36 months	$n = 3660$ (NR) in total $n = 2164$ (NR) in intervention group $n = 1496$ (NR) in control group	8.74 (NR) y in total group NR in intervention group NR in control group	NR in total group NR in intervention group NR in control group	Ethnically diverse	72.7%
Nogueira, 2017[39]	12 months	$n = 240$ (43.8%) in total $n = 155$ (42.6%) in intervention group $n = 85$ (44.8%) in control group	10.6 (0.6) y in total group 10.5 (0.6) y in intervention group 10.7 (0.6) y in control group	12.3 (0.6) y in total group 12.2 (1.0) y in intervention group 12.3 (0.6) y in control group	NR	70.8%
Ofosu, 2018[40]	Between 48 and 60 months	$n = 540$ (NR) in total $n = 202$ (48.3%) in intervention group $n = 338$ (59.1%) in control group	NR in total group 10.8 (0.4) y in intervention group 10.9 (0.4) y in control group	NR in total group 13.8 (1.4) y in intervention group 14.0 (1.3) y in control group	From low socioeconomic neighbourhoods	NA
Puma, 2013[41]	48 months	$n = 191$ (NR) in total $n = 91$ (NR) in intervention group $n = 100$ (NR) in control group	NR in total group NR in intervention group NR in control group	NR in total group NR in intervention group NR in control group	2nd graders at baseline; 8th graders at follow-up. Rural area, many children (i.e., 62%) in free school lunch	NA

TABLE 1 (Continued)

Author and year	Follow-up length	No. of total group participants at follow-up (% female) No. of intervention group participants at follow-up (% female) No. of control group participants at follow-up (% female)	Baseline age total group (mean(SD)) Baseline age intervention group (mean(SD)) Baseline age control group (mean(SD))	Follow-up age total group (mean(SD)) Follow-up age intervention group (mean(SD)) Follow-up age control group (mean(SD))	Other specific population characteristics	Retention rate at follow-up
					program and 59.1% was Hispanic.	
Rosengren, 2021[42]	Between 36 and 60 months (mean 48 months)	$n = 124$ (46.8%) in total $n = 81$ (44.4%) in intervention group $n = 43$ (51.2%) in control group	NR in total group 7.5 (0.6) y for boys in intervention group 7.5 (0.6) y for girls in intervention group 8.0 (0.7) y for boys in control group 7.9 (0.6) y for girls in control group	NR in total group 18.8 (0.2) y for boys in intervention group 18.7 (0.4) y for girls in intervention group 18.8 (0.3) y for boys in control group 18.7 (0.3) y for girls in control group	Elementary schools with a uniform socioeconomic status	35.5%
Salmon, 2008[43]	12 months	$n = 268$ (NR) in total $n = 84$ (NR) in BM/FMS intervention group $n = 60$ (NR) in BM intervention group $n = 69$ (NR) in FMS intervention group $n = 55$ (NR) in control group	10.67 (0.42) y[a] for boys in total group 10.67 (0.33) y[a] for girls in total group NR in intervention group NR in control group	NR in total group NR in intervention group NR in control group	Low socioeconomic status	90.8%
Stea, 2018[44]	168 months	$n = 1081$ (53.5%) in total $n = 322$ (53.0%) in intervention group $n = 759$ (55.6%) in control group	NR in total group NR in intervention group NR in control group	26.5 (NR) y in total group 26.5 (0.53) y in intervention group 26.5 (0.51) y in control group	10 to 12 year olds at baseline.	52.8%
Tarro, 2014[45]	24 months	$n = 619$ (NR) in total $n = 198$ (NR) in intervention group $n = 421$ (NR) in control group	NR in total group NR in intervention group NR in control group	NR in total group 12.51 (0.52) y in intervention group 11.88 (0.35) y in control group	7 to 8 year olds at baseline.	26.3%

Abbreviations: BM, behavioural modification; FMS, fundamental movement skills; NA, not applicable; NR, not reported; PE, physical education; SD, standard deviation; US, United States; y, years.
[a]Reported in months, but calculated into years by us.

TABLE 2 Quality assessment for RCTs with the RoB2 tool

	Item 1 Randomization process	Item 2 Deviations from intended interventions	Item 3 Missing outcome data	Item 4 Measurement of the outcome	Item 5 Selection of the reported result	Overall score
Anderson, 2016[27]	Low risk	High risk	Low risk	Low risk	High risk	High risk
Bere, 2014[28]	High risk	Some concerns	Some concerns	High risk	Some concerns	High risk
Cao, 2019[29]	Some concerns	Some concerns	Some concerns	Low risk	Some concerns	Some concerns
Habib-Mourad, 2020[32]	Some concerns	High risk	Low risk	Low risk	Some concerns	High risk
James, 2007[33]	Low risk	Some concerns	High risk	Low risk	Some concerns	High risk
Li, 2010[35]	Low risk	High risk	Low risk	Low risk	Some concerns	High risk
Llaurado, 2018[36]	Low risk	Some concerns	Low risk	Low risk	Low risk	Some concerns
Meyer, 2014[37]	Low risk	Some concerns	Low risk	Low risk	Low risk	Some concerns
Nader, 1999[38]	Low risk	Some concerns	Low risk	Low risk	Low risk	Some concerns
Nogueira, 2017[39]	Low risk	Some concerns	High risk	Low risk	Some concerns	High risk
Salmon, 2008[43]	Low risk	Some concerns	Some concerns	Low risk	Some concerns	Some concerns
Stea, 2018[44]	Some concerns	High risk	High risk	High risk	Some concerns	High risk
Tarro, 2014[45]	Some concerns	Some concerns	Low risk	Some concerns	Low risk	Some concerns

further specified in Table 1. Ten interventions had a duration between 6 and 12 months,[26,27,29,30,32,34,36,38,42,43] seven had a duration between 13 and 36 months[28,31,35,37,39,40,44] and two had a duration of >36 months.[33,41] Seven studies investigated long-term effects at a follow-up of 12 months post-intervention,[26,28,29,31,34,38,42] seven studies investigated effects at a follow-up between 13 and 36 months post-intervention,[27,28,30,32,36,37,44] four studies investigated effects at a follow-up between 37 and 60 months post-intervention[35,39–41] and three studies investigated effects at a follow-up of >60 months post-intervention.[27,33,43]

3.3 | Risk of bias

The risk of bias assessment for the 13 RCTs with the RoB2 tool is reported in Table 2. The overall risk of bias was assessed as "high risk" in seven RCTs and as "some concerns" in six RCTs. The largest risk of bias in RCTs was due to deviations from intended interventions, being classified as high risk in 4/13 RCTs and as some concerns in 9/13 RCTs. The two main concerns with this bias were that deviations from the intended intervention could have affected outcomes and the absence of an intention to treat analysis. Risk of bias due to the randomization process mainly came forth from absence of information on both the randomization method and method used to hide the allocation sequence from participants. Risk of bias due to missing outcome data mainly came forth from missing outcome data not being equally distributed over groups. Finally risk of bias due to selection of the reported result came forth from studies lacking any pre-specified study protocol registration.

The risk of bias assessment for the six non-RCTs with the ROBINS-I tool is reported in Table 3. Among the six non-RCTs the overall risk of bias was assessed as "critical risk" in one non-RCT and as "serious risk" in five non-RCTs. The risk of bias in non-RCTs was largest with regard to confounding (mainly due to not controlling for baseline differences), classification of interventions (mainly due to lack of defining control group) and missing data (mainly due to the unequal distribution of missing data over groups). For all six non-RCTs no information was available to rate the domain of selection of the reported result, forthcoming from studies lacking a prospectively registered study protocol.

3.4 | Meta-analyses on long-term intervention effects

Comparisons were possible between intervention and control groups for BMI, zBMI and WC (Figure 2). Table 4 contains a summary of findings which presents the evidence quality and certainty for these pooled comparisons. The three funnel plots for these meta-analyses are in Figure S1, Figure S2 and Figure S3 respectively. In these meta-analyses the time points of the longest follow-up[27,35] within adolescence were included when there were multiple follow-up time points[27,35,44] of the same study available. A complete overview of the GRADE rating assessments per GRADE item is presented in Table S2.

Results from six studies ($n = 5453$ participants) that were included in the meta-analysis for the evaluation of the long-term effect of the interventions on BMI show that the interventions had no effect on BMI in primary school-aged children ≥12 months post-intervention.[27,32,35–38] The mean difference in BMI between intervention and control groups was 0.06 kg/m^2 (CI95% = −0.38, 0.50; $n = 5453$, $Z = 0.260$, $p = 0.79$). I^2 was indicated as 66% ($\tau^2 = 0.18$, $\chi^2 = 14.72$,

TABLE 3 Quality assessment for non-RCTs with the ROBINS-I tool

	Item 1 Confounding	Item 2 Selection of participants into the study	Item 3 Classification of interventions	Item 4 Deviations from intended interventions	Item 5 Missing data	Item 6 Measurement of outcomes	Item 7 Selection of the reported result	Overall score
Gao, 2013[30]	Serious risk	Low risk	Low risk	Low risk	NI	Low risk	No information	**Serious risk**
Gorely, 2011[31]	Moderate risk	Low risk	Low risk	Low risk	Serious risk	Serious risk	No information	**Serious risk**
Jurak, 2013[34]	Serious risk	Low risk	Low risk	Low risk	Serious risk	Low risk	No information	**Serious risk**
Ofosu, 2018[40]	Moderate risk	Serious risk	Serious risk	Low risk	NI	Low risk	No information	**Serious risk**
Puma, 2013[41]	Critical risk	Low risk	Serious risk	Low risk	NI	Low risk	No information	**Critical risk**
Rosengren, 2021[42]	Serious risk	Low risk	Low risk	Low risk	Moderate risk	Low risk	No information	**Serious risk**

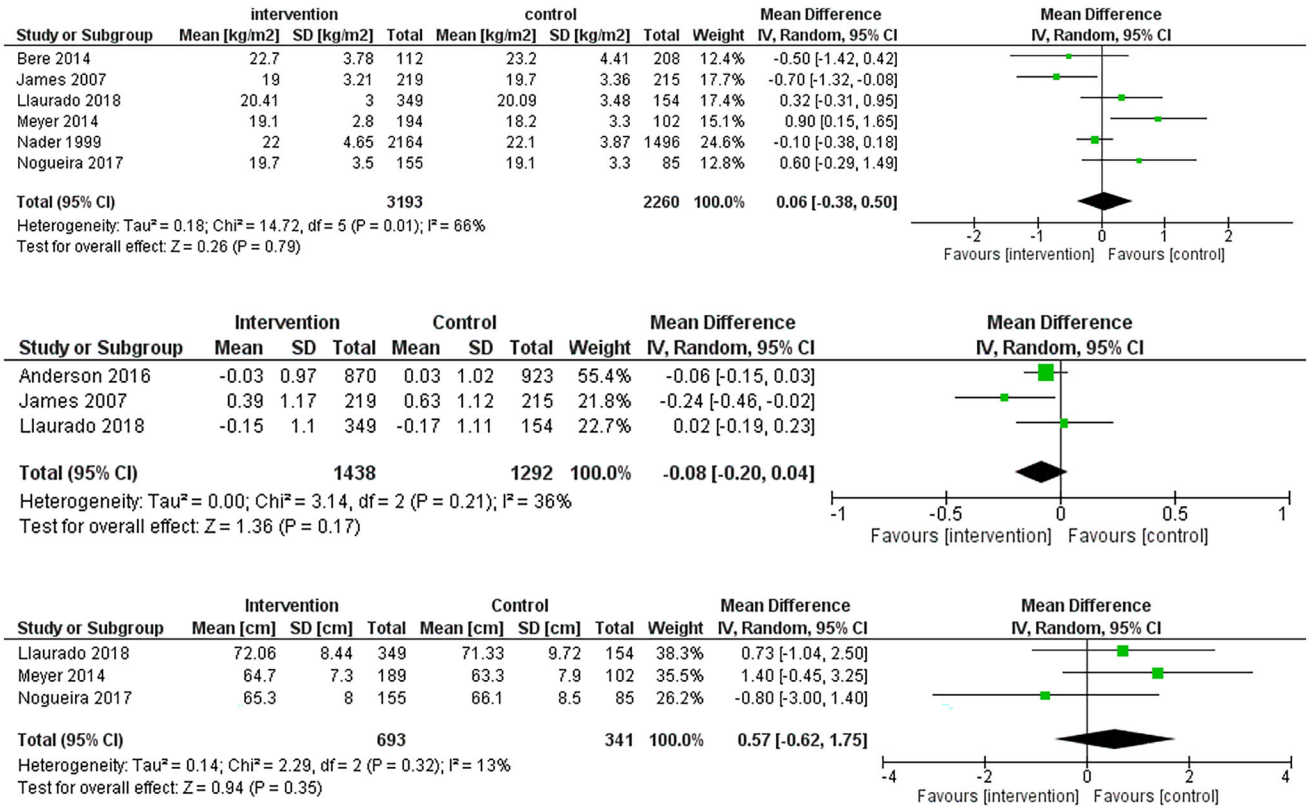

FIGURE 2 Three random-effects meta-analysis of the mean difference in baseline to follow-up change in respectively BMI, zBMI and waist circumference between intervention and control groups

df = 5, $p = 0.01$). The mean difference estimate for BMI has a very low evidence certainty according to the GRADE assessment due to risk of bias and imprecision. Results from three studies ($n = 2730$ participants) that were included in the meta-analysis on zBMI show that the interventions had no long-term effect on zBMI at ≥12 months of follow-up in primary school-aged children.[26,32,35] The mean difference in zBMI between intervention and control groups was −0.08 (CI95% = −0.20, 0.04; $n = 2730$, $Z = 1.360$, $p = 0.17$). I^2 was indicated as 36% ($\tau^2 = 0.00$, $\chi^2 = 3.14$, df = 2, $p = 0.21$). Certainty of the evidence for the long-term effect estimate of the interventions on zBMI is very low. Results from three studies ($n = 1034$ participants) that were included in the meta-analysis on WC show that the interventions had no long-term effect on WC at follow-up in primary school-aged children.[35,36,38] The mean difference in WC between intervention and control groups was 0.57 cm (CI95% = −0.62, 1.75; $n = 1034$, $Z = 0.940$, $p = 0.35$). I^2 was indicated as 13% ($\tau^2 = 0.14$, $\chi^2 = 2.29$, df = 2, $p = 0.32$). Certainty of the evidence for the long-term effect estimate of the interventions on WC is low.

3.5 | Sensitivity meta-analyses on long-term intervention effects

In the sensitivity analyses studies with shorter follow-up time-points[27,44] were used instead of longest follow-up time-points.[27,35] The forest plot and funnel plot of the additional sensitivity meta-analysis for BMI, in which the 84 months follow-up results[27] have been replaced by the 36 months follow-up results,[27] are presented in Figure S4 and Figure S7 respectively.

The forest plots and funnel plots of the additional sensitivity meta-analyses for zBMI and WC, in which the study by Llaurado and colleagues (2018)[35] has been replaced by the study by Tarro and colleagues (2014),[44] are presented in Figure S5, Figure S6, Figure S8 and Figure S9.

The results from the three additional sensitivity meta-analyses correspond to those of the three meta-analyses. The mean difference in BMI between intervention and control groups was 0.09 kg/m^2 (CI95% = −0.33, 0.50; $n = 5453$, $Z = 0.410$, $p = 0.68$).[27,32,35–38] The

TABLE 4 Summary of Findings for the sustained effect of primary school-based obesity prevention interventions in children aged 6–12 years

Patient or population: Primary school-aged children aged 6 to 12 years old
Setting: At least one component of the intervention was primary school-based
Intervention: Obesity prevention interventions with a diet component, physical activity component or both components.
Comparison: Control where no interventions took place

Pooled outcomes	No. of participants (studies) Follow-up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects: Risk with control group	Anticipated absolute effects: Risk difference with intervention group*
BMI	5453 (6 RCTs)	⊕○○○ Very low[a]	-	The mean BMI at follow-up was **20,55 kg/m^2**	MD **0.06 higher** (CI −0.38, 0.50)
zBMI	2730 (3 RCTs)	⊕○○○ Very low[a]	-	The mean zBMI at follow-up was **0.11**	MD **0.08 lower** (CI −0.20, 0.04)
WC	1034 (3 RCTs)	⊕⊕○○ Low[b]	-	The mean WC at follow-up was **67,62 cm**	MD **0.57 higher** (CI −0.62, 1.75)

Not-pooled outcomes	No. of participants (studies) Follow-up	Certainty of the evidence (GRADE)[1]	Effects narrative[#]
BMI	5820 (3 RCTs) ^^^^^^^^^^^^^^ 869 (3 non-RCTs)	⊕⊕○○ Low[c] ^^^^^^^^^^^^^^ ⊕○○○ Very low[a]	Mixed findings were reported by three RCT studies. Two studies found no association between group and BMI. The third study found that increase in BMI since baseline was lower in the intervention group compared with the control group. ^^ Mixed findings were reported by three non-RCT studies. Two studies found no association between group and BMI. The third study found a significant programme*year interaction effect indicating that the difference in BMI between the intervention group and control group decreased more and more toward the null during the follow-up.
BMI adjusted for the age-sex-population median	268 (1 RCT)	⊕⊕○○ Low[b]	One study reported that the change in BMI adjusted for the age-sex-population median was not significantly lower for the BM intervention group and FMS intervention group compared to the control group. However change in BMI minus sex-age population median was significantly lower for the combined BM/FMS intervention group compared to the control group.
zBMI	5733 (3 RCTs) ^^^^^^^^^^^^^^ 421 (1 non-RCT)	⊕⊕⊕○ Moderate[c] ^^^^^^^^^^^^^^ ⊕○○○ Very low[a]	Mixed findings were reported by three RCT studies. One study reported significant lower zBMI at 24 and 36 months follow-up in the intervention group compared with the control group for normal weighted children at baseline. A second study reported a significantly larger decrease in zBMI in the intervention group compared with the control group. A third study found no association between group and zBMI. ^^^ One non-RCT study found no associations between group and zBMI.
WC	421 (1 RCT)	⊕⊕○○ Low[d]	One RCT study found no association between group and WC.
zWC	2655 (2 RCTs)	⊕○○○ Very low[a]	No associations between group and zWC or change in zWC were found in two RCT studies.

TABLE 4 (Continued)

Not-pooled outcomes	No. of participants (studies) Follow-up	Certainty of the evidence (GRADE)[1]	Effects narrative[#]
Weight status	7059 (9 RCTs) ^^^^^^^^^^^^^^ 896 (3 non-RCT)	⨁◯◯◯ Very low[a] ^^^^^^^^^^^^^^ ⨁◯◯◯ Very low[a]	Nine RCT studies reported mixed findings. Six studies did not report significant associations between group and weight status. A seventh study reported a significantly lower percentage of overweight participants in the intervention group compared with the control group at 84 months follow-up, whereas no significant association was found at 36 months follow-up. An eighth study found a significant association between intervention group and lower overweight percentage at 36 months follow-up. The ninth study found a significant association between the combined BM/FMS intervention group and decrease in overweight percentage compared with the control group, but no significant association was found between the BM intervention group or FMS intervention group and change in overweight percentage compared with the control group. ^^ Three non-RCT studies found no associations between group and weight status.

Note : # = narrative effects for the not-pooled outcomes are based on the results in Table S4. **GRADE Working Group grades of evidence** High certainty: we are very confident that the true effect lies close to that of the estimate of the effect. Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect. Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.
Grading explanations : 1. Inconsistency component of GRADE is uncertain for the not-pooled outcomes. No available sufficient information on heterogeneity or confidence intervals. 2. Full explanations and reasons underlying downgrading are presented supplemental file 5 (Appendix S1). a. 2 levels downgraded for risk of bias and 1 level downgraded for imprecision. b. 1 level downgraded for risk of bias and 1 level downgraded for imprecision. c. 1 level downgraded for risk of bias. d. 2 levels downgraded for risk of bias.
Abbreviations: BM, behavioural modification; BMI, body-mass index; CI, confidence interval; FMS, fundamental movement skills; MD, mean difference; RCT, randomized controlled trial; WC, waist circumference.
* **The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

mean difference in zBMI between intervention and control groups was -0.08 kg/m^2 (CI95% $= -0.19, 0.02$; $n = 2846$, $Z = 1.550$, $p = 0.12$).[26,32,44] The mean difference in WC between intervention and control groups was 0.62 cm (CI95% $= -0.51, 1.75$; $n = 1150$, $Z = 1.080$, $p = 0.28$).[36,38,44]

3.6 | Long-term effectiveness of interventions on non-pooled outcomes

The long-term intervention effects on the non-pooled outcomes are narratively described in Table 4a subdivision is made between evidence originating from RCTs and evidence originating from non-RCTs. A complete overview of the GRADE rating assessments per GRADE item, and accompanied with our explanations for grading up and down, is presented in Table S2.

Six studies, of which three RCTs and three non-RCTs, studied intervention effects on BMI and findings were mixed. Of the three RCTs, two studies found no association between intervention and a long-term impact on BMI with follow-up periods of 24 and 168 months.[43,44] The third RCT found a lower BMI increase at 12 months of follow-up since baseline in the intervention group compared with the control group.[34] The certainty of the evidence across the RCTs following the GRADE assessment was rated as low. Of the three non-RCTs, two studies found no association between intervention and a long impact on BMI with follow-up periods of 20 and 48 months.[30,41] One study found a significant interaction effect between intervention and time. In this study the differences in BMI between groups decreased toward the null during the follow-up period, which suggests that the intervention effect did sustain up to 84 months of follow-up.[33] The certainty of the evidence across the non-RCTs following the GRADE assessment was rated as very low.

Only one non-RCT reported on the change since baseline in BMI adjusted for the age-sex-population median.[42] This study had three intervention arms and no significant effect on the outcome was found for the intervention group with solely a behavioural modification component and for the intervention group with solely a fundamental movement skills component in comparison with the control group. The intervention group with combined behavioural modification and functional movements skills components however was associated with a significant lower BMI adjusted for the age-sex-population median in comparison with the control group. Certainty of the evidence was rated as low.

Regarding zBMI, three RCTs reported mixed findings. One study found no association between intervention and zBMI at 12 months follow-up.[31] One study found significant lower zBMI values in the intervention group compared with the control group at 24 and 36 months of follow-up.[28] Another study reported a significantly larger decrease in zBMI from baseline to 12 months of follow-up in the intervention group compared with the control group.[34] For zBMI the certainty of the evidence across RCTs following the GRADE assessment was rated as moderate. Furthermore, one non-RCT found no association between intervention and zBMI at 20 months of follow-up.[30] Certainty of the evidence following the GRADE assessment was rated as very low.

One study examined the intervention effect on WC and found no association between obesity prevention interventions and WC at 20 months of follow-up.[30] The certainty of this evidence was rated as low. In addition, two studies reported on intervention effects on zWC at respectively 12 and 24 months of follow-up and both studies found no association between group and zWC or change in zWC.[26,32] The certainty of the evidence was rated as very low.

A total of 12 studies reported on the effect of obesity prevention interventions on weight status, of which nine RCTs and three non-RCTs. Of the nine RCTs, the majority of six studies found no association between the intervention and weight status.[26,31,32,35,43,44] In one RCT a lower percentage of overweight was found in the intervention group in comparison with the control group at 84 months of follow-up, whereas this association between group and percentage of overweight was not apparent at 36 months of follow-up.[27] Another RCT found a similar association at 24 and 36 months of follow-up.[28] Last, one RCT with 12 months follow-up found a significant association between the combined behavioural modification and functional movement skills intervention group and decrease in the percentage of children with overweight compared with the control group, but no association was found between the behavioural modification intervention group or functional movement intervention group and change in overweight percentage compared with the control group.[42] The certainty of the evidence across RCTs was rated as very low. Furthermore, three non-RCTs found no association between obesity prevention interventions and weight status ranging between 12 and 60 months of follow-up.[29,39,40]

4 | DISCUSSION

We included 19 studies in this review, of which 9 RCT studies were included in either the pooled meta-analyses or pooled sensitivity meta-analyses. Sixteen studies were used for non-pooled analyses. The results from the meta-analyses suggest no long-term effects of obesity prevention interventions on the pooled outcome WC, whereas the evidence is very uncertain about the long-term effects on BMI and zBMI. The additional sensitivity meta-analyses, including the earlier follow-up time points, revealed similar results. Narrative results have been subdivided in evidence originating from RCTs and evidence resulting from non-RCTs. Beneficial long-term effects on the non-pooled outcomes were found in only a limited number of studies, irrespective of study design. Summarizing, the results of this review do not show clear evidence regarding the long-term effects of primary school-based obesity prevention interventions in children on obesity-related outcomes.

In total five studies reported positive long-term intervention effects on non-pooled outcomes[27,28,33,34,42] and one study in the meta-analyses reported positive long-term intervention effects.[32] When having a closer look at these intervention studies (Table S3), they show much variation in characteristics such as intervention components, intervention duration and follow-up length. Likewise, the other 13 studies that found no evidence for a long-term effect ≥12 months post-intervention in any outcome were very diverse in terms of those characteristics. Furthermore, there is also much variability regarding additional settings on top of the primary school-based setting and regarding the theoretical framework underlying the intervention in both the studies for which long-term effects were found and the studies for which no long-term effects were found. As we could not statistically perform subgroup analysis due to lack of power, the role of intervention duration, intervention components, follow-up length, setting and theoretical framework on long-term intervention effectiveness remains uncertain. More targeted evidence in this area is needed and may be gathered using innovative study designs in future studies and by using predictive modelling combining intervention characteristics and individual participant data.[45]

The interpretation of studied long-term intervention effects above needs consideration of the immediate post-intervention effects as well. Some interventions might have been effective to start with and others not. We have been able to retrieve this information for 15 of the 19 intervention studies in both the included studies in this review[26,28–39,42,44] and in earlier published studies.[46–52] For 4 of these 15 studies, positive long-term effects were found.[28,32,34,42] Our findings indicate that the intervention effects sustained up to 12 months post-intervention in two studies[34,42] and have sustained and have been improving up to 36 and 24 months since immediately post-intervention in the other two studies.[28,32,46] When looking into the 11 studies for which no long-term effects on any outcome were found in this review,[26,29–31,33,35–39,44] three interventions were not effective immediately post-intervention as well on any outcome,[29,31,37,47] whereas eight interventions showed effectiveness on at least one outcome immediately post-intervention.[26,30,33,35,36,38,39,44,48–52]The lack of sustained and long-term intervention effects in most of the studies may have several causes, but the obesogenic environment that surrounds and shapes the health behaviour of children and adolescents certainly plays an important role,[53,54] opposing the effects that an ended intervention might have accomplished. However, methodological issues of the included studies might give alternative explanations for the lack of sustained and long-term intervention effects. Sample sizes in general were small, as 15 of the 19 studies had a sample size between 100 and 1000 participants. Attrition of participants at follow-up was apparent in all the included studies, and ranged between 16.4% and 87.7%. More specifically, the retention rates were even lower for the studies with a longer follow-

up length. Such participant attrition possibly results in power issues to adequately assess intervention effects at long-term follow-up.

This review has been aggregating the available evidence on the long-term effects of primary school-based obesity preventive interventions and intends to help the research field on long-term intervention effect studies forward. The findings in this review must be interpreted with caution due to a few limitations. Based on the GRADE assessment tool, the certainty of the evidence for the pooled and non-pooled outcomes was generally low to very low, mainly due to the relatively high risks of bias across studies. This is in line with our quality assessment score for the included studies. As mentioned above sample sizes and attrition in included studies might have influenced the lack of positive findings. Furthermore, the small sample sizes in combination with the low number of included studies influenced the possibilities for interpretation of heterogeneity in the meta-analyses.[55,56] As we included a small number of studies in each meta-analysis, the confidence intervals were wide. This causes uncertainty regarding the systemacy of the I^2-measure. A visual exploration of the forest plots suggests some additional undetected heterogeneity in the I^2-measure. Therefore, this review has only reported the I^2-measure, instead of interpreting and labelling the I^2-measure by using cut-off points. Furthermore, in this review we limited the scope of outcomes to BMI, zBMI, waist circumference and weight status. We acknowledge the fact that the use of other anthropometric indicators (such as fat percentage) and more in-depth considerations of methods to assess obesity (such as the 95th percentile) may lead to additional insights in the future. Final limitations are that trial registries were not searched and that non-English studies were not included in this systematic review.

In line with recommendations by Brown and colleagues (2019) we performed this review to aggregate the current evidence on the long-term effects of school-based obesity prevention interventions.[5] We cross-checked the included RCTs in this review against those in the Cochrane review by Brown, and substantially expanded the evidence base by our more recent search and by allowing other designs than RCTs (6 studies). In comparison, from the 13 included RCTs in this review, 3 RCTs were also included in the Cochrane review,[34,36,42] for 5 RCTs we included studies with a longer follow-up of included studies in the Cochrane review[26,28,31,32,37] and 5 RCTs were not included in the Cochrane review.[27,35,38,43,44]

Due to the limited number of available studies and the low quality of studies we cannot present clear evidence regarding the long-term effects of interventions on obesity-related outcomes. Based on the findings in this review, we have the following recommendations for future research. First of all, we emphasize the urgent need for higher quality follow-up studies to add to the knowledge base on the long-term effectiveness of primary school-based preventive interventions. Based on the shortcomings in quality of the reviewed studies, we recommend to prospectively register the studies, choosing appropriate analysis techniques, to adequately report on missing outcome data, to blind assessors and to adequately describe intervention and control group conditions. Clear information on control group content (attention control or usual practice) is important, because studies with attention control help researchers to disentangle the beneficial effects of attention and interpersonal interactions from the intervention itself.[57] RCTs are often considered the golden standard for these effectiveness evaluation studies,[11] but long-term follow-up of RCTs might be challenging due to funding issues or because it is unethical to withhold children from effective interventions for more years. Furthermore, the longer the follow-up period, the more likely it will become that power or attrition issues arise that make it more difficult to reach sound evidence on effectiveness. Therefore, to assess the long-term impact of interventions, other types of study designs may give important insights as well. Natural experiments, quasi-experimental and observational studies might be considered more feasible due to logistic or practical issues or can be considered more ethical than randomizing children for years in a control group withholding them from future interventions which are possibly effective.[12,13] In addition, using data from existing registries or ongoing cohort studies might be a way to derive insights on the long-term effects of interventions beyond their intervention period. Furthermore, we recommend the use of predictive modelling combining intervention characteristics and individual participant data in future studies to more adequately determine the effect of different intervention components and in population subgroups. If researchers choose for RCT designs, they should take into account the consequences of participant attrition at follow-up for their statistical power. We recommend researchers to take this into account for their power calculations and sample size determination when setting up new intervention effect evaluation studies. When reviewing the included interventions in this review, it becomes clear that interventions with a duration of multiple school years are underrepresented. More studies into the long-term effectiveness of interventions with longer durations are therefore recommended. Furthermore, when more evidence from high quality studies becomes available in the future, it will be valuable to review in more detail the differences in intervention effectiveness between subgroups of the population (such as differences in socioeconomic status or weight status) in an updated review. Finally, we recommend researchers to conduct more long-term effect studies in which both obesity-related outcomes and changeable lifestyle factors are measured. These studies are valuable to gain more knowledge on the pathways that underlie behaviour change in youth. It is likely that the association between obesity prevention interventions and their long-term effect on obesity-related outcomes such as BMI, zBMI, WC and weight status is mediated by intermediate factors related to a healthy lifestyle. Caloric overconsumption, poor food choices, sedentary behaviour and low levels of physical activity are changeable factors for the prevention of childhood obesity.[58–61] A recent systematic review and meta-analysis by Nally and colleagues observed that primary school-based interventions have no clear effect on healthy lifestyle behaviours such as sedentary behaviour, moderate-to-vigorous physical activity and diet behaviour.[62] Therefore, future research is warranted to explore the actual change in lifestyle that precedes outcomes on overweight on the short and longer term in primary school-based interventions.

Based on our findings, we have the following recommendations for practice. Policymakers and intervention developers should bear in mind that evidence of sustained effects of many included

interventions in this review is unclear. Besides methodological explanations, explanations can be found in the obesogenic environment that surrounds and shapes the health behaviour of children and adolescents,[53,54] opposing the accomplished effects immediately post-intervention. This may signify the need to implement prolonged interventions or additional interventions for children and adolescents to sustain and improve a healthy lifestyle. As a considerable number of the interventions showed immediate post-intervention effectiveness, more focus might also be needed on maintaining healthy behaviours. Whole system approaches with more focus on policies, systems and environments might be a promising way forward for policymakers in order to initiate and maintain behaviour among schoolchildren.[63] Therefore, a promising way for policy makers to intervene may therefore be policy, systems and environment changes supporting healthy behaviour in school and community settings.

5 | CONCLUSION

The results of this review contribute to knowledge on the long-term effects of obesity prevention interventions. It did not show clear evidence regarding the long-term effects of primary school-based obesity prevention interventions in children on obesity-related outcomes and certainty levels of the evidence were generally low. We recommend policy makers and intervention developers to implement prolonged interventions or additional interventions for children and adolescents to sustain a healthy lifestyle. Collaborating with researchers in early phases of intervention planning may be useful for this purpose Moreover, we emphasize the need for long-term effect evaluation studies using more innovative study designs. We specifically emphasize the urgent need for higher quality studies, to be able to reach stronger conclusions about the long-term effects of primary school-based obesity prevention interventions in children.

AUTHOR CONTRIBUTIONS

Michel S. Smit, Mirte Boelens, Famke J. M. Mölenberg, Hein Raat and Wilma Jansen contributed to the design of the study and the development of the protocol for this study. Michel S. Smit and Mirte Boelens performed the selection process of articles, data extraction and analysed the data. Michel S. Smit mainly wrote the manuscript and Mirte Boelens, Famke J. M. Mölenberg, Hein Raat and Wilma Jansen contributed to the manuscript by critical revisions and giving comprehensive feedback on multiple drafts. Michel S. Smit, Mirte Boelens, Famke J. M. Mölenberg, Hein Raat and Wilma Jansen read and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank Daan Nieboer for his advisory role in the execution of the meta-analysis. We thank Wichor Bramer for his role in the development of the systematic search strategy for this review.

FUNDING INFORMATION

Funding for this project was provided by the city of Rotterdam, the Netherlands.

CONFLICT OF INTEREST

Funding for this project was provided by the city of Rotterdam, the Netherlands. No conflict of interest was declared.

ORCID

Michel S. Smit https://orcid.org/0000-0003-4231-3972
Mirte Boelens https://orcid.org/0000-0003-4082-5237
Famke J. M. Mölenberg https://orcid.org/0000-0002-5305-9730
Hein Raat https://orcid.org/0000-0002-6000-7445
Wilma Jansen https://orcid.org/0000-0002-4453-9054

REFERENCES

1. World Health Organization. *Commission on Ending Childhood Obesity*. World Health Organization; 2019.
2. Garrido-Miguel M, Cavero-Redondo I, Alvarez-Bueno C, et al. Prevalence and trends of overweight and obesity in European children from 1999 to 2016: a systematic review and meta-analysis. *JAMA Pediatr*. 2019;173:e192430.
3. Abarca-Gómez L, Abdeen ZA, Hamid ZA, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128· 9 million children, adolescents, and adults. *Lancet*. 2017; 390:2627-2642.
4. Waters E, de Silva-Sanigorski A, Hall BJ, et al. Interventions for preventing obesity in children. *Cochrane Database Syst Rev*. 2011;12:CD001871.
5. Brown T, Moore TH, Hooper L, et al. Interventions for preventing obesity in children. *Cochrane Database Syst Rev*. 2019;7:CD001871.
6. Verjans-Janssen SRB, Gerards S, Kremers SPJ, Vos SB, Jansen MWJ, Van Kann DHH. Effects of the KEIGAAF intervention on the BMI z-score and energy balance-related behaviors of primary school-aged children. *Int J Behav Nutr Phys Act*. 2020;17:105.
7. Corder K, Schiff A, Kesten JM, van Sluijs EM. Development of a universal approach to increase physical activity among adolescents: the GoActive intervention. *BMJ Open*. 2015;5:e008610.
8. Bleich SN, Vercammen KA, Zatz LY, Frelier JM, Ebbeling CB, Peeters A. Interventions to prevent global childhood overweight and obesity: a systematic review. *Lancet Diabetes Endocrinol*. 2018;6:332-346.
9. Liu Z, Xu HM, Wen LM, et al. A systematic review and meta-analysis of the overall effects of school-based obesity prevention interventions and effect differences by intervention components. *Int J Behav Nutr Phys Act*. 2019;16:95.
10. Brown EC, Buchan DS, Baker JS, Wyatt FB, Bocalini DS, Kilgore L. A systematised review of primary school whole class child obesity interventions: effectiveness, characteristics, and strategies. *Biomed Res Int*. 2016;2016:4902714-4902715.
11. Hariton E, Locascio JJ. Randomised controlled trials—the gold standard for effectiveness research: study design: randomised controlled trials. *BJOG*. 2018;125:1716.
12. Smit MS, Raat H, Molenberg F, Wolfers MEG, Bannink R, Jansen W. Study protocol for the evaluation of long-term effects of the school-based obesity prevention program Lekker Fit! ('enjoy being fit'): a retrospective, controlled design. *BMJ Open*. 2021;11:e046940.
13. Gilmartin-Thomas JF, Liew D, Hopper I. Observational studies and their utility for practice. *Aust Prescr*. 2018;41:82-85.
14. Barnighausen T, Tugwell P, Rottingen JA, et al. Quasi-experimental study designs series-paper 4: uses and value. *J Clin Epidemiol*. 2017;89:21-29.
15. World Health Organization. *Consideration of the Evidence on Childhood Obesity for the Commission on Ending Childhood Obesity: Report of the Ad Hoc Working Group on Science and Evidence for Ending Childhood Obesity, Geneva, Switzerland*. World Health Organization; 2016.
16. Obesity: Guidance on the Prevention. *Identification, Assessment and Management of Overweight and Obesity in Adults and Children (NICE*

clinical guideline 43). National Institute for Health and Clinical Excellence; 2006.
17. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1.
18. Prochaska JO, DiClemente CC. Stages of change in the modification of problem behaviors. *Prog Behav Modif*. 1992;28:183-218.
19. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Org*. 2007;85:660-667.
20. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. *J Med Libr Assoc*. 2016;104:240-243.
21. Higgins J, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. Chichester: John Wiley & Sons; 2019.
22. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898.
23. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
24. GRADEpro GDT. *GRADEpro Guideline Development Tool [Software]*. McMaster University; 2020 (developed by Evidence Prime, Inc.). Available from gradepro.org
25. Schünemann H, Brozek J, Guyatt G, Oxman A. *GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations*. Updated October 2013. The GRADE Working Group; 2013.
26. Anderson EL, Howe LD, Kipping RR, et al. Long-term effects of the active for life year 5 (AFLY5) school-based cluster-randomised controlled trial. *BMJ Open*. 2016;6:e10957.
27. Bere E, Klepp KI, Overby NC. Free school fruit: can an extra piece of fruit every school day contribute to the prevention of future weight gain? A cluster randomized trial. *Food Nutr Res*. 2014;58.
28. Cao Z, Hua J, Zhang D, Thapa JR, Wang S. A cohort study assessing the sustainable long-term effectiveness of a childhood-obesity intervention in China. *Int J Epidemiol*. 2019;48:108-115.
29. Gao Z, Hannan P, Xiang P, Stodden DF, Valdez VE. Video game-based exercise, Latino children's physical health, and academic achievement. *Am J Prev Med*. 2013;44:S240-S246.
30. Gorely T, Morris JG, Musson H, Brown S, Nevill A, Nevill ME. Physical activity and body composition outcomes of the GreatFun2Run intervention at 20 month follow-up. *Int J Behav Nutr Phys Act*. 2011;8:74.
31. Habib-Mourad C, Ghandour LA, Maliha C, Dagher M, Kharroubi S, Hwalla N. Impact of a three-year obesity prevention study on healthy behaviors and bmi among Lebanese schoolchildren: findings from ajyal Salima program. *Nutrients*. 2020;12:1-14.
32. James J, Thomas P, Kerr D. Preventing childhood obesity: two year follow-up results from the Christchurch obesity prevention programme in schools (CHOPPS). *Br Med J*. 2007;335:762-764.
33. Jurak G, Cooper A, Leskosek B, Kovac M. Long-term effects of 4-year longitudinal school-based physical activity intervention on the physical fitness of children and youth during 7-year followup assessment. *Cent Eur J Public Health*. 2013;21:190-195.
34. Li YP, Hu XQ, Schouten EG, et al. Report on childhood obesity in China (8): effects and sustainability of physical activity intervention on body composition of Chinese youth. *Biomed Environ Sci*. 2010;23:180-187.
35. Llauradó E, Tarro L, Moriña D, Aceves-Martins M, Giralt M, Solà R. Follow-up of a healthy lifestyle education program (the EdAl study): four years after cessation of randomized controlled trial intervention. *BMC Public Health*. 2018;18:104.
36. Meyer U, Schindler C, Zahner L, et al. Long-term effect of a school-based physical activity program (KISS) on fitness and adiposity in children: a cluster-randomized controlled trial. *PLoS One*. 2014;9:e87929.
37. Nader PR, Stone EJ, Lytle LA, et al. Three-year maintenance of improved diet and physical activity: the CATCH cohort. *Arch Pediatr Adolesc Med*. 1999;153:695-704.
38. Nogueira RC, Weeks BK, Beck B. One-year follow-up of the CAPO kids trial: are physical benefits maintained? *Pediatr Exerc Sci*. 2017;29:486-495.
39. Ofosu NN, Ekwaru JP, Bastian KA, et al. Long-term effects of comprehensive school health on health-related knowledge, attitudes, self-efficacy, health behaviours and weight status of adolescents. *BMC Public Health*. 2018;18:515.
40. Puma J, Romaniello C, Crane L, Scarbro S, Belansky E, Marshall JA. Long-term student outcomes of the integrated nutrition and physical activity program. *J Nutr Educ Behav*. 2013;45:635-642.
41. Rosengren BE, Lindgren E, Jehpsson L, Dencker M, Karlsson MK. Musculoskeletal benefits from a physical activity program in primary school are retained 4 years after the program is terminated. *Calcif Tissue Int*. 2021;109:405-414.
42. Salmon J, Ball K, Hume C, Booth M, Crawford D. Outcomes of a group-randomized trial to prevent excess weight gain, reduce screen behaviours and promote physical activity in 10-year-old children: switch-play. *Int J Obes (Lond)*. 2008;32:601-612.
43. Stea TH, Tveter ET, Te Velde SJ, Vik FN, Klepp KI, Bere E. The effect of an extra piece of fruit or vegetables at school on weight status in two generations—14 years follow-up of the fruit and vegetables makes the Marks study. *PLoS One*. 2018;13:e0205498.
44. Tarro L, Llauradó E, Moriña D, Solà R, Giralt M. Follow-up of a healthy lifestyle education program (the Educació en Alimentació study): 2 years after cessation of intervention. *J Adolesc Health*. 2014;55:782-789.
45. Seidler AL, Johnson BJ, Golley RK, Hunter KE. The complex quest of preventing obesity in early childhood: describing challenges and solutions through collaboration and innovation. *Front Endocrinol (Lausanne)*. 2021;12:803545.
46. Cao ZJ, Wang SM, Chen Y. A randomized trial of multiple interventions for childhood obesity in China. *Am J Prev Med*. 2015;48:552-560.
47. Luepker RV, Perry CL, McKinlay SM, et al. Outcomes of a field trial to improve children's dietary patterns and physical activity. The Child and Adolescent Trial for Cardiovascular Health. CATCH collaborative group. *JAMA*. 1996;275:768-776.
48. Kipping RR, Howe LD, Jago R, et al. Effect of intervention aimed at increasing physical activity, reducing sedentary behaviour, and increasing fruit and vegetable consumption in children: active for life year 5 (AFLY5) school based cluster randomised controlled trial. *BMJ*. 2014;348:g3256.
49. Tarro L, Llaurado E, Albaladejo R, et al. A primary-school-based study to reduce the prevalence of childhood obesity—the EdAl (Educacio en Alimentacio) study: a randomized controlled trial. *Trials*. 2014;15:58.
50. Kriemler S, Zahner L, Schindler C, et al. Effect of school based physical activity programme (KISS) on fitness and adiposity in primary schoolchildren: cluster randomised controlled trial. *BMJ*. 2010;340:c785.
51. Nogueira RC, Weeks BK, Beck BR. Targeting bone and fat with novel exercise for peripubertal boys: the CAPO kids trial. *Pediatr Exerc Sci*. 2015;27:128-139.
52. Gorely T, Nevill ME, Morris JG, Stensel DJ, Nevill A. Effect of a school-based intervention to promote healthy lifestyles in 7-11 year old children. *Int J Behav Nutr Phys Act*. 2009;6:5.
53. Hill JO, Peters JC. Environmental contributions to the obesity epidemic. *Science*. 1998;280:1371-1374.

54. Swinburn B, Egger G. Preventive strategies against weight gain and obesity. *Obes Rev*. 2002;3:289-301.
55. von Hippel PT. The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Med Res Methodol*. 2015;15:35.
56. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I(2) in assessing heterogeneity may mislead. *BMC Med Res Methodol*. 2008;8:79.
57. LaFave SE, Granbom M, Cudjoe TKM, Gottsch A, Shorb G, Szanton SL. Attention control group activities and perceived benefit in a trial of a behavioral intervention for older adults. *Res Nurs Health*. 2019;42:476-482.
58. Kim J, Lim H. Nutritional management in childhood obesity. *J Obes Metab Syndr*. 2019;28:225-235.
59. Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014;384:766-781.
60. Luque V, Escribano J, Closa-Monasterolo R, et al. Unhealthy dietary patterns established in infancy track to mid-childhood: the EU childhood obesity project. *J Nutr*. 2018;148:752-759.
61. Thivel D, Aucouturier J, Doucet E, Saunders TJ, Chaput JP. Daily energy balance in children and adolescents. Does energy expenditure predict subsequent energy intake? *Appetite*. 2013;60:58-64.
62. Nally S, Carlin A, Blackburn NE, et al. The effectiveness of school-based interventions on obesity-related behaviours in primary school children: a systematic review and meta-analysis of randomised controlled trials. *Children (Basel)*. 2021;8:489.
63. Nobles JD, Radley D, Mytton OT. The action scales model: a conceptual tool to identify key points for action within complex adaptive systems. *Perspect Public Health*. 2022;142:328-337. 1757913921 1006747.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Smit MS, Boelens M, Mölenberg FJM, Raat H, Jansen W. The long-term effects of primary school-based obesity prevention interventions in children: A systematic review and meta-analysis. *Pediatric Obesity*. 2023; 18(3):e12997. doi:10.1111/ijpo.12997