Statistical Analysis Plan

Generation Healthy Kids: A cluster-randomized controlled trial of a multi-component and multi-setting intervention to promote healthy weight and wellbeing in 6-11-year-old children in Denmark

This statistical analysis plan is reported according to the 2017 Guidelines for Content of Statistical Analysis Plans in Clinical Trials (1)

Clinical Trials registration no.: NCT05940675

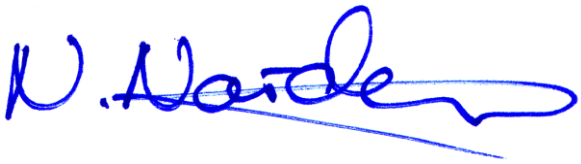
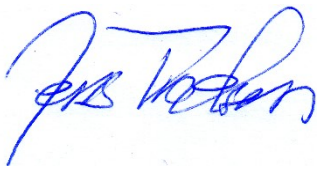
Table of Contents

Abbreviations

ADP	Air-displacement plethysmography
BIA	Bioimpedance analyzer
BMI	Body mass index
FFM	Fat free mass
FM	Fat mass
GHK	Generation Healthy Kids
IQR	Interquartile range
SAP	Statistical Analysis Plan

Section 1: Administrative information

Title *(Item 1A)*	Generation Healthy Kids: A cluster-randomized controlled trial of a multi-component and multi-setting intervention to promote healthy weight and wellbeing in 6-11-year-old children in Denmark
Trial registration *(Item 1B)*	Registered at ClinicalTrials.gov: NCT05940675 Registered at the Scientific Ethics Committee in the Region of Southern Denmark: S-20220094.
SAP version *(Item 2)*	Version 1.0, dated 4-June-2025 Version 2.0, dated 18 November 2025
Protocol version *(Item 3)*	This document has been written based on information in the study protocol version 6, dated 17-Jan-2025. All successive protocol versions are available at ClinicalTrials.gov. The study protocol is also published in PLOS One (2).
SAP revisions *(Item 4A-4B)*	**Revision 1: (changes highlighted with yellow in SAP):** • *Changes to section 5.5 Baseline characteristics:* ○ Item 25b: Description of baseline characteristics: Baseline values of body composition will be measured by bioimpedance (BIA). • *Changes to section 6.1 Outcome definitions:* ○ Item 26b: Measurement methods: In the primary analysis, BIA will be used to measure fat mass. • *Changes to section 6.2 Analysis methods* ○ Item 27a: Analysis methods: Adaptation to use of single measurement method. ○ Item 27b: Covariate adjustment: Change in strategy for co-variate selection. ○ Item 27e: Sensitivity analyses: Air-displacement plethysmography (ADP) will be used in two pre-specified sensitivity analyses: i) an ADP-only analysis (baseline to end-of-study) and ii) a bivariate model combining ADP and BIA. • *Changes to section 6.3 Missing data* ○ Item 28: Handling of missing data: BIA is used in the primary analysis of FM. The rationale for changes is elaborated in Appendix 1: SAP revision 1, dated 18 November 2025.
Names, affiliations, roles, and responsibilities of SAP contributors *(Item 5)*	**Natascha Holbæk Pedersen, PhD, postdoc** Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark. Role: Shared first author of successive SAP drafts. **Louise T. Thomsen, PhD, researcher** Department of Prevention, Health Promotion and Society, Steno Diabetes Center Copenhagen, Copenhagen, Denmark. Role: Shared first author of successive SAP drafts.

	Anders Grøntved, PhD, professor Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark. Role: Supervision, comments, and final approval of SAP. **Camilla Trab Damsgaard, PhD, associate professor** Department of Nutrition, Exercise and Sports, University of Copenhagen, Copenhagen, Denmark. Role: Supervision, comments, and final approval of SAP. **Peter Krustrup, PhD, professor** Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark. Role: Supervision, comments, and final approval of SAP. **Werner Vach, senior statistician** Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark; *and* Basel Academy for Quality and Research in Medicine, Basel, Switzerland. Role: Statistical supervision, comments, and final approval of SAP.
Signatures and dates *(Item 6)*	**Nicolai Baastrup Nordsborg (primary investigator), 18 November 2025**  **Jens Troelsen (co-primary investigator), 18 November 2025**  **Camilla Trab Damsgaard (investigator & work package lead), 18 November 2025**  **Peter Krustrup (investigator & work package lead), 18 November 2025** 

	Anders Grøntved (investigator & work package lead), 18 November 2025  **Werner Vach (senior statistician), 18 November 2025** 

Section 2: Introduction

2.1. Background and rationale

Item 7. Synopsis of trial background and rationale including brief description of research questions

Overweight and obesity among children are global problems affecting more than 340 million children and adolescents worldwide (3). The World Health Organization describes childhood obesity as one of the most serious public health challenges for the 21st century (4). In Denmark, the prevalence of overweight and obesity among primary school children is 12-13%, and this increases to 18-19% at the time of graduation (5).

Childhood overweight and obesity can have significant negative health consequences for children's wellbeing and long-term physical health. Children with overweight or obesity are more likely to experience bullying, stigmatization and suffer from low self-esteem, anxiety, and lower quality of life (6). In addition, children with overweight or obesity are more likely to have overweight or obesity as adults, thereby increasing their risk of several adverse health outcomes (7).

Prior interventions to prevent childhood obesity have generally shown no or only small effects on weight development in children, and the evidence of long-term effects is limited (8,9). Prior interventions have mainly been school-based and have focused primarily on physical activity and dietary interventions (8,9).

The "*Generation Healthy Kids (GHK)*" program is a multi-component, multi-setting intervention with the overall aim of promoting healthy weight development and wellbeing in children aged 6–11 years (2). The intervention targets multiple settings where children spend their daily lives, including families, schools, after-school clubs, and local communities. Within each setting, the intervention components target diet, physical activity, screen media use and sleep habits. The intervention components are described in detail in (2).

The GHK program is evaluated in a cluster-randomized trial running from August 2023 – June 2025 at 24 Danish schools (1rst-3rd grade). The overall aim of the trial is to investigate the effects of the multi-setting, multi-component GHK intervention on weight development, health, and wellbeing in Danish school children aged 6–11 years (2).

2.2. Scope of the present statistical analysis plan (SAP)

This SAP covers the analysis of the GHK trial's predefined *primary and secondary outcomes relating to body composition and anthropometric measures*, as described in the study protocol (2) and as further specified in the following sections.

The primary and secondary outcomes will be reported in the main effect paper from the study, including selected subgroup analyses, focusing on sex and adiposity, as specified in Section 2.3 *"Objectives"* and Section 6.2 *"Analysis Methods"*.

Sub-group analyses pertaining to differential effects of the intervention on body composition and anthropometry according to e.g. socioeconomic and demographic characteristics will be reported in a separate paper, which is not covered by the present SAP. Furthermore, analysis of other prespecified outcomes (relating to wellbeing and mental health, cognitive function, school performance and school absence, cardiometabolic health, physical fitness and motor functions, physical activity levels and physical literacy, dietary intake, nutrient status, food literacy, screen media practices, sleep time, and sleep quality (2)) are not covered by the present SAP.

2.3. Objectives

Item 8. Description of specific objectives or hypotheses

Primary objective

The primary objective covered by this SAP is to investigate the effect of the GHK intervention on healthy weight development, as measured by body composition and anthropometric measures, in the intervention group compared with the control group. By healthy weight development, we refer to patterns of growth and weight gain that are appropriate for the child's age, sex, and stage of maturation, and that support physical health and well-being.

The primary hypothesis is that the intervention will lead to healthier weight development (as further defined in section 6.1 *"Outcome definitions"*) in the intervention group compared to control group during the intervention period.

Secondary objectives

Secondary objectives are:

1. *To investigate the effect of the intervention over time,* i.e. the time required for the intervention to have an effect in the intervention group compared to control group.

2. *To investigate potential differential effects of the intervention by selected child characteristics:*

   A. Child adiposity level at baseline:
   We hypothesize that the effect of the intervention may differ depending on child adiposity levels at baseline. Due to the nature of the intervention, it is possible that children with high adiposity levels at baseline, or those at high risk of developing adiposity, may have a more pronounced benefit of the intervention than children with healthy baseline weight. However, conversely, it is also possible that children with normal weight at baseline (and their families) may be more likely to accept and participate in intervention activities, e.g. eat the school lunch, actively participate in FIT FIRST and adopt healthier screen/sleep habits. This could lead to stronger intervention effects in children with normal baseline adiposity levels.

   Such subgroup effects are important to investigate from a public health perspective, and they may be masked in the primary analyses of the overall population.

This will be addressed by investigation of a potential effect modification by children's baseline adiposity level (see further details in Section 6.2 "*Analysis Methods*", Item 27f). In case of demonstrating an effect in subgroups of children defined by adiposity level, this will be interpreted as having demonstrated a success of the intervention.

B. *Sex:*
Differential intervention effects will also be explored by sex. Investigating possible differential effects by sex is important to assess the potential public health impact of broader implementation of the GHK intervention, and to allow assessment of whether adjustments of the intervention may be necessary to increase effectiveness in either girls or boys. Furthermore, investigating potential differential effects by sex is considered relevant because growth and anthropometric development differ in girls and boys in the considered age spans.

Section 3: Study methods

3.1. Trial design

Item 9. Brief description of trial design, including type of trial (e.g., parallel group, multi-arm, crossover, factorial) and allocation ratio and may include brief description of interventions.

The GHK trial is a parallel group cluster-randomized superiority trial with two arms. A total of 24 schools were randomly allocated in a 1:1 ratio to intervention (the GHK multi-component intervention program) or control (no intervention). The intervention is described in detail in (2).

3.2. Randomization

Item 10. Randomization details, e.g., whether any minimization or stratification occurred (including stratifying factors used or the location of that information if it is not held within the SAP)

Randomization occurred at school level prior to inclusion of individual participants (i.e., children). A covariate-constrained randomization procedure was carried out to maximize school-level covariate balance between intervention and control group and to increase the probability that child baseline characteristics did not differ substantially between groups.

Allocation was conducted in two strata (1: Western Denmark, i.e. Region of Southern Denmark and 2: Eastern Denmark, i.e. Capital/Zealand Region) and was constrained by the following school-level covariates: Number of children in 1st and 2nd grade, proportion of children with parents of basic educational level (vocational education or lower), proportion of children with non-Danish background, rural or urban municipality, and number of planned weekly physical education lessons. The school-level covariates were obtained from the Ministry of Children and Education (10), and information on physical education lessons was provided by school principals. To ensure timely allocation of schools to their respective groups and allow schools sufficient time to plan the subsequent school year, randomization was conducted in two steps, with the first step based on the first 10 schools recruited, and the second step based on the last 14 schools recruited.

Randomization was performed using the 'cvcrand' command in STATA for covariate-constrained randomization in cluster-randomized trials (11). Randomization was carried out by a statistician not otherwise involved in the study, and the procedure was concealed from the investigators.

Due to the nature of the intervention, the study is open label, i.e. allocation is not blinded to participants, school staff, local communities, or research team members.

After randomization, one control school decided to withdraw from the study, leaving 12 intervention and 11 control schools in the study.

3.3. Sample size

Item 11. Full sample size calculation or reference to sample size calculation in protocol (instead of replication in SAP)

The sample size calculation for the study is described in the study protocol (2).

3.4. Framework

Item 12. Superiority, equivalence, or noninferiority hypothesis testing framework, including which comparisons will be presented on this basis.

The GHK trial is a superiority trial, i.e. the primary and secondary outcomes are intended to demonstrate superiority of the intervention over control.

The only exception to the superiority framework is the secondary outcome "prevalence of underweight". For this outcome, we expect equivalence of intervention vs control.

3.5. Statistical interim analyses and stopping guidance

Item 13a. Information on interim analyses specifying what interim analyses will be carried out and listing of time points

No interim analyses are planned.

Item 13b. Any planned adjustment of the significance level due to interim analysis

Not applicable.

Item 13c. Details of guidelines for stopping the trial early

No formal interim statistical analyses to inform potential early stopping of the trial are planned.

However, adverse events will be monitored via questionnaires to the parents at each measurement round, and all reported adverse events will be assessed by the clinically responsible physician. If a serious adverse event occurs, the principal investigator will notify the Regional Scientific Ethics Committee of Southern Denmark within seven days of becoming aware of such adverse event. The principal investigator may suspend or prematurely terminate parts of or the entire study for significant and documented reasons. If a suspicion relating to an unacceptable risk to study participants arises during the study, the principal investigator will suspend the study while the risk is assessed. The principal investigator will terminate the study if an unacceptable risk is confirmed.

3.5. Timing of final analysis

Item 14. Timing of final analysis, e.g., all outcomes analyzed collectively or timing stratified by planned length of follow-up

The analysis of the primary and secondary outcomes, as defined in the present SAP, will be performed collectively when all data have been collected, and data cleaning has been performed (expected analysis phase: September–November 2025).

3.6. Timing of outcome assessments

Item 15. Time points at which the outcomes are measured including visit "windows"

The study includes three measurement rounds:

- Baseline assessment (month 0): September–November 2023.
- Interim assessment (month 6–8, end of school year 1): May–June 2024.
- End-of-study assessment (month 18–20, end of school year 2): March–June 2025.

Description of the measurement procedures at each round can be found in the study protocol (2).

Section 4: Statistical principles

4.1. Confidence intervals and P values

Item 16. Level of statistical significance

All hypotheses testing will be based on two-sided tests at the alpha = 0.05 level.

Item 17. Description and rationale for any adjustment for multiplicity and, if so, detailing how the type 1 error is to be controlled

No adjustment for multiplicity will be performed.

Item 18. Confidence intervals to be reported

All confidence intervals presented will be 95% and two-sided.

4.2. Analysis populations

Item 20. Definition of analysis populations, e.g., intention to treat, per protocol, complete case, safety

The primary analysis will be conducted on an intention-to-treat basis including all children with consent for participation at baseline and at least one valid outcome measurement during the study.

4.3. Adherence and protocol deviations

Item 19a. Definition of adherence to the intervention and how this is assessed including extent of exposure

No assessment of adherence will be performed for the main effect paper. A thorough process evaluation assessing the degree of implementation and factors influencing implementation will be published separately as described in the study protocol (2) and in the separate process evaluation protocol (12).

Item 19b. Description of how adherence to the intervention will be presented

A process evaluation assessing degree of implementation and factors influencing implementation will be published separately as described in (2) and (12).

Item 19c. Definition of protocol deviations for the trial

Not applicable.

Item 19d. Description of which protocol deviations will be summarized

Not applicable.

Section 5: Trial population

5.1. Screening data

Item 21. Reporting of screening data (if collected) to describe representativeness of trial sample

Not relevant.

5.2. Eligibility

Item 22. Summary of eligibility criteria

All children enrolled in the participating school classes at baseline are eligible to participate. The number of participants eligible will be summarized in the CONSORT diagram, see Figure 1.

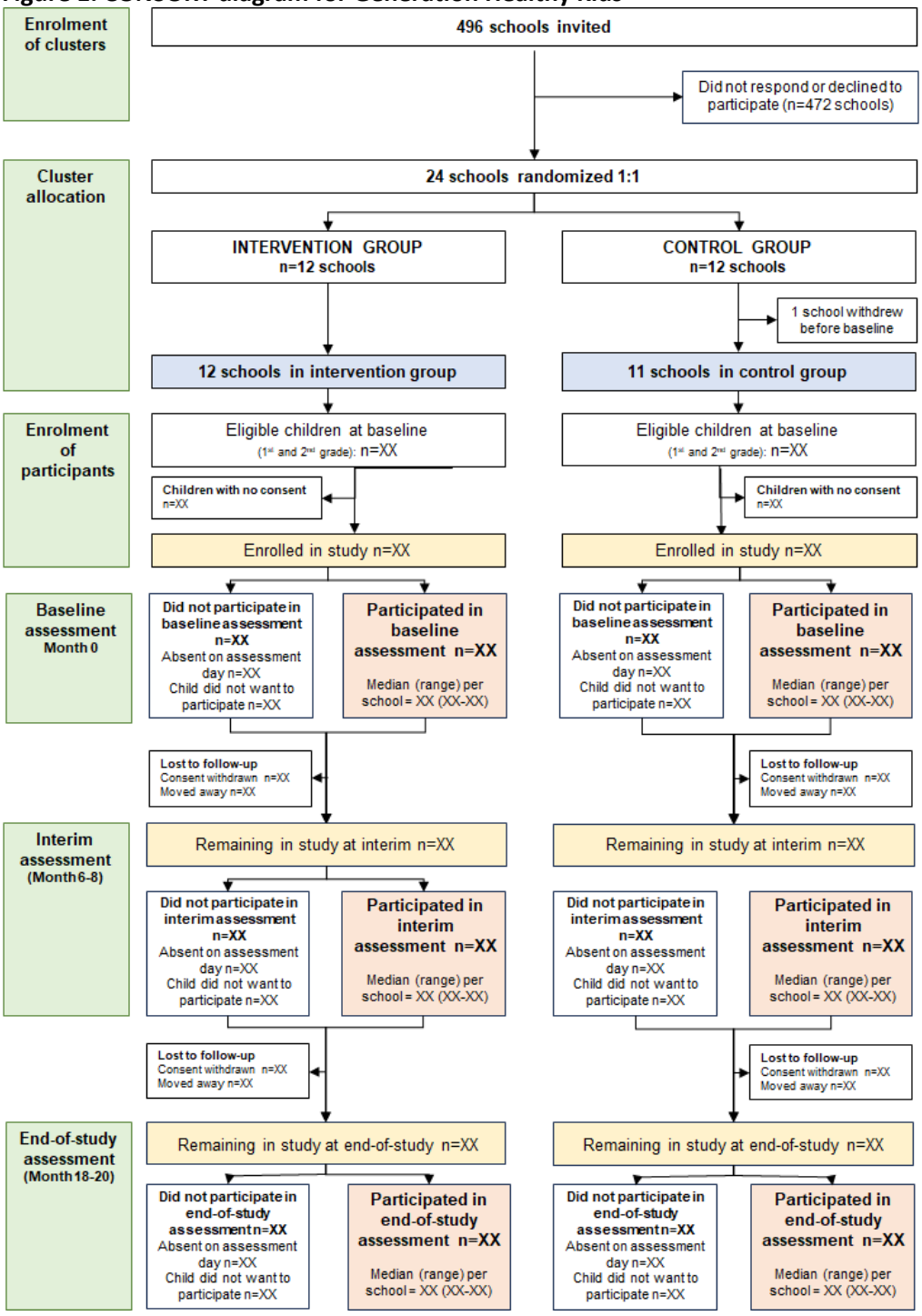
Children enrolled in the participating classes during the study period were also offered to participate in the study. Data from these children will not be included in the ITT analysis in the main effect paper, but may potentially be included in other descriptive or secondary analyses papers from the study (not covered by the present SAP). If so, this will be clearly stated in the papers.

5.3. Recruitment

Item 23. Information to be included in the CONSORT flow diagram

The flow of participants through the trial will be illustrated in accordance with the CONSORT extension for cluster-randomized trials (13) as shown in Figure 1.

Figure 1. CONSORT diagram for Generation Healthy Kids



5.4. Withdrawal/follow-up

Item 24a. Level of withdrawal, e.g. from intervention and/or from follow-up
Item 24b. Timing of withdrawal/lost to follow-up data
Item 24c. Reasons and details of how withdrawal/lost to follow-up data will be presented

Withdrawal of clusters (schools):
Before randomization, school management was informed that by signing up for the study, the school committed itself to participating in the study, regardless of the result of randomization. Nevertheless, one school randomized to control withdrew before study start, and no data were collected from children at this school. The withdrawn cluster (school) was not replaced.

Withdrawal of participants (children):
Parents were informed that they could withdraw their child from the study at any time and for any reason. In case of withdrawal, the child will be discontinued from the measurement schedule, and no further data will be collected, but children in the intervention group will still be offered the intervention activities implemented as part of the normal school day. Data collected before withdrawal will be included in the study.

The CONSORT diagram will show the number of children withdrawing their consent between baseline and interim, and between interim and end-of-study (Figure 1). Parents will be asked their reasons for withdrawal. If provided, these answers will be summarized descriptively in the main effect paper. The CONSORT diagram will also show the number of children who were lost to follow-up because they moved away from the school (see Figure 1).

5.5. Baseline characteristics

Item 25a. List of baseline characteristics to be summarized
Item 25b. Details of how baseline characteristics will be descriptively summarized

The following baseline characteristics will be summarized for the intervention versus control group:

- Cluster level (schools):
    - Number of participating school classes
    - Number of eligible children per school (median, IQR, range)
    - Mean proportion of parents with basic educational level in included year groups (%)
    - Mean proportion of parents with non-Danish background in included year groups (%)
    - Proportion of schools with urban vs rural location (%)

- Individual level (children):
    - Sex (male vs female) (n, %)
    - Age, years (median, IQR, range)
    - Grade level (1^{st}, 2^{nd}) (n, %)
    - Region (Eastern, Western) (n, %)
    - Sociodemographic characteristics:
        - Highest parental educational level (n, %)
        - Parental employment (n, %)
        - Family type (single vs two provider household) (n, %)
        - Country of origin (Danish or immigrant/descendant) (n, %)
    - Baseline body composition, measured by bioimpedance (BIA) (see further in Section 6.1., "Outcomes") ~~air-displacement plethysmography (ADP) with bioimpedance (BIA) values used for missing data (see further in Section 6.1, "Outcomes" and Section 6.3 "Missing data")~~:
        - FM, kg (mean, SD or median, IQR (depending on distribution))

        - FFM, kg (mean, SD or median, IQR (depending on distribution))
        - FM index, kg/m^2 (mean, SD or median, IQR (depending on distribution))
        - FFM index, kg/m^2 (mean, SD or median, IQR (depending on distribution))
        - FFM-to-FM ratio (mean, SD or median, IQR (depending on distribution))
        - Fat percentage, % (mean, SD or median, IQR (depending on distribution))
    - Baseline anthropometry:
        - Height, cm (mean, SD or median, IQR (depending on distribution))
        - Height z-score (mean, SD or median, IQR (depending on distribution))
        - Weight, kg (mean, SD or median, IQR (depending on distribution)))
        - BMI z-score based on WHO references (mean, SD or median, IQR (depending on distribution))
        - Prevalence of underweight, normal weight, overweight, obesity (n, %)
        - Waist circumference (mean, SD or median, IQR (depending on distribution))

Baseline characteristics of participants versus eligible non-participants:
Using register data from Statistics Denmark, we will obtain socioeconomic and demographic information on all eligible participants at baseline. This will enable us to compare participating children with eligible non-participants in the intervention and control groups.

We expect that the following baseline characteristics will be descriptively summarized for participants versus eligible non-participants:

- Age (median, IQR)
- Sex (n, %)
- Highest parental educational level (n, %)
- Parental employment (n, %)
- Family type (single vs two provider household) (n, %)
- Country of origin (Danish or immigrant/descendant) (n, %)

Section 6: Analysis

6.1. Outcome definitions

List and describe each primary and secondary outcome including details of:

- ***Item 26a. Specification of outcomes and timings. If applicable include the order of importance of primary or key secondary end points (eg, order in which they will be tested)***
- ***Item 26b. Specific measurement and units (eg, glucose control, hbA1c [mmol/mol or %])***
- ***Item 26c. Any calculation or transformation used to derive the outcome (eg, change from baseline, QoL score, time to event, logarithm, etc)***

To comprehensively evaluate the effect of the GHK intervention on healthy weight development, we have selected primary and secondary outcomes that together reflect different aspects of changes in body composition and anthropometric measures. The primary outcome is fat mass (FM), but we also include secondary outcomes allowing for differentiation between fat and lean mass development, assessment of abdominal fat distribution, prevalence of underweight, overweight and obesity, and evaluation of overall growth. **Table 1** provides an overview of the selected primary and secondary outcomes, and the rationale for each outcome is further detailed below.

Table 1. Overview of primary and secondary outcomes

Type of outcome	Category	Outcome	Description	Unit	Measurement method	Time frame
Primary outcome	Body composition	FM	Between-group difference in change in FM[1]	Kg	Bioimpedance analysis (BIA) (InBody270)[2] (primary analysis) *and* Air-displacement plethysmography (ADP) (BODPOD) (sensitivity analyses, cf. Item 27e). ~~Air-displacement plethysmography (ADP) (BODPOD).~~ ~~*and*~~ ~~bioimpedance analysis (BIA) (InBody270)[2]~~	Primary objective: From baseline to end-of-study Secondary objective: During the course of the intervention, i.e. from baseline to interim, and interim to end-of-study
Secondary outcomes	Body composition	FFM	Between-group difference in change[1] in FFM	Kg	BIA (InBody270)[2] (primary analysis) *and* ADP (BODPOD) (sensitivity analyses, cf. Item 27e). ~~ADP (BODPOD)~~ ~~*and*~~ ~~BIA (InBody270)[2]~~	Primary objective: From baseline to end-of-study Secondary objective: During the course of the intervention, i.e. from baseline to interim, and interim to end-of-study
		FM index	Between-group difference in change in FM index	Kg/m^2		
		FFM index	Between-group difference in change in FFM index	Kg/m^2		
		FFM-to-FM ratio	Between-group difference in change in FFM-to-FM ratio	N.A.		
		% FM	Between-group difference in change in % FM	%		
	Anthropometry	Height z-score	Between-group difference in change in standing height z-score based on WHO references.	cm	Portable stadiometer	Primary objective: From baseline to end-of-study Secondary objective: During the course of the intervention, i.e. from baseline to interim, and interim to end-of-study
		Weight status	Between-group difference in change[1] in prevalence of underweight, overweight and obesity, based on cutoffs by Cole et al. (14) and the International Task Force of Obesity.	%	InBody 270 and portable stadiometer	
		BMI z-score	Between-group difference in change in BMI z-score based on WHO references.	Kg/m^2	InBody 270 and portable stadiometer	
		Waist circumference	Between-group difference in change in waist circumference	mm	Non-elastic measuring tape at umbilicus level.	
Exploratory outcome	Anthropometry	Height z-score	Between-group difference in change in standing height z-score based on WHO references.	cm	Portable stadiometer	Primary objective: From baseline to end-of-study Secondary objective: During the course of the intervention, i.e. from baseline to interim, and interim to end-of-study

ADP, air-displacement plethysmography; BIA, bioimpedance analyzer; BMI, body mass index; FM, fat mass; FFM, fat free mass.

[1]As described below in section 6.2 (Item 27a), the outcomes of FM, FFM and waist circumference will be log-transformed in the analyses, i.e. changes will be expressed as mean percentage change. Supplementary analyses without log-transformation (expressing absolute changes) will also be conducted as secondary analyses.

[2]ADP was used at baseline and end-of-study. BIA was used at all three measurement rounds.

Primary outcome: Fat mass (FM)
The primary outcome is the difference in change in FM in the intervention group compared with the control group from baseline to end-of-study (Table 1).

We selected FM as our primary outcome because the intervention targets four behavioural components (diet, physical activity, screen media use, and sleep) which may all affect FM. In contrast, other aspects of body composition, in particular fat free mass (FFM), is only expected to be affected by the physical activity component. FM was considered a more appropriate primary outcome than weight-based measures (e.g. BMI z-score), because certain intervention components, particularly physical activity, may increase FFM (15), while e.g. the diet intervention may reduce FM. Such beneficial effects on body composition would not be reflected in changes in overall weight-based measures. Therefore, favourable intervention effects could be masked by using weight-based measures as the primary outcome.

We used FM in power calculations instead of FM index (2), because we anticipated that due to the cluster-randomized design, height would be equally distributed across intervention and control groups.

Measurement method for primary outcome:
The primary outcome is assessed by two measurement methods:

1) *Air-displacement-plethysmography (ADP) (BODPOD).* This method has demonstrated a level of agreement with the 4-compartment model comparable to that of dual-energy X-ray absorptiometry in children with obesity (16). For logistical and resource reasons, ADP was only applied at baseline and end-of-study
2) *Bioimpedance analyser (BIA) (InBody 270).* BIA is a highly feasible field method for estimating body composition, but it has shown lower agreement than ADP with the gold-standard 4-compartment model in children with obesity (16). The BIA method was applied at all three time points.

~~For the present study, we assume ADP to be the more accurate measurement (16), and the ADP results will therefore be used whenever available. BIA will only be used if no ADP measurement is available, i.e. for children with missing data on ADP at baseline or end-of-study, and for the measurements at the interim time point. Furthermore, as the ADP results are sensitive to the child's clothing, BIA results will be used if a child did not wear the same clothes (e.g. swimwear, light clothing) at baseline and end-of-study~~.

In the primary analysis of FM, we will use BIA data. ADP measures of FM will be used in pre-specified sensitivity analyses (See item 27e). Further details on reasons for this change are provided in Appendix 1 to this SAP.

Secondary outcomes:
The secondary outcomes have been selected to provide a more comprehensive assessment of the intervention's effect on different aspects of child healthy weight development and growth, as further detailed in the following:

a. Fat free mass (FFM): Including FFM as a secondary outcome is considered relevant because FFM can change independently of FM due to increased physical activity during the intervention. Thus, we expect that FFM may increase in the intervention group due to increased muscle mass and bone mineralization caused by the FIT FIRST intervention (15).

b. FM and FFM index: FM and FFM index will be considered relevant secondary outcomes in case the randomization procedure did not ensure equal distribution of height between intervention and control group.

c. FFM-to-FM ratio, FM percentage: These body composition measures are sensitive to changes in both FM and FFM and therefore offer relevant summary indicators which account for variations in both lean and adipose tissue.

d. Waist circumference: Waist circumference is included as a measure of abdominal fat distribution, because waist circumference is known to have distinct health implications and is an independent marker of cardiometabolic risk profile (17,18).

e. BMI z-score: Although weight-based measures have limitations in the context of evaluating the GHK intervention (as described above), BMI z-score is included as a secondary outcome to enable comparison with previous studies of which many used BMI z-score as the primary outcome (9).

f. Prevalence of underweight, overweight and obesity: The prevalence of weight categories (using the cutoffs defined by Cole et al. (14)) are included because these are widely used in clinical and public health practice, and they can elucidate the potential clinical impact of the intervention at the population level. Since the prevalence of obesity in this age group is only around 3% in the general Danish population (5,19), we expect to merge the categories of overweight and obesity to avoid small sample sizes that could violate assumptions underlying statistical analyses. The prevalence of underweight is explored to provide a comprehensive picture of the total effect of the intervention, and to investigate whether the intervention inadvertently leads to increased prevalence of underweight.

Measurement method for secondary outcomes:

- The secondary outcomes of FFM, FM index, FFM index, FFM-to-FM ratio and FM percentage are assessed ~~as described above for the primary outcome~~ using BIA. ADP measures will be used in pre-specified sensitivity analyses (see Item 27e).
- Weight is measured using the InBody 270 bioimpedance analyzer.
- Standing height is measured using a portable stadiometer to the nearest millimetre. Three measurements are obtained, and the mean is used.
- Waist circumference is measured non-elastic measuring tape at the level of the umbilicus. Three measurements are obtained, and the mean is used.

Exploratory outcome:

a. **Height z-scores:** This outcome is included as an exploratory outcome to investigate how the intervention influences normal growth.

Measurement method for exploratory outcome:

- Standing height is measured using a portable stadiometer to the nearest millimetre. Three measurements are obtained, and the mean is used.

Pre-specified hypotheses for the intervention's effect on primary, secondary, and exploratory outcomes:

- *FM, FM index, FM percentage, BMI z-score, waist circumference, and prevalence of overweight/obesity* are expected to change during the study period to a more favorable degree in

the intervention group than in the control group, with lower values being considered more favorable.

- *FFM and FFM index* are expected to change during the study period to a more favorable degree in the intervention group than in the control group, with higher values being considered more favorable.
- The *prevalence of underweight* is expected to be unaffected by the intervention (i.e. similar development in the intervention and control group).
- *Height z-score* (exploratory outcome) could be both unaffected by the intervention or could, theoretically, be higher in the intervention group than in the control group, e.g. due to healthier diet. Alternatively, a lower fat mass in the intervention group could lead to slightly later puberty and therefore lower height in the intervention group. Therefore, we have no specific hypotheses for this outcome.

Timing of outcome measurements:
The primary analysis will focus on the intervention's effect from baseline to end-of-study. However, to examine secondary objective no. 2 (see section 2.3 *"Objectives"*), we will also investigate patterns during the course of the intervention in the abovementioned outcomes, including the development from baseline to interim, and from interim to end-of study (Table 1).

6.2. Analysis methods

Item 27a. What analysis method will be used and how will the treatment effects be presented

Analysis of longitudinal ~~the primary~~ outcomes ~~(FM)~~ will be conducted using a mixed linear regression model. Fixed effects are the time point, the intervention group, the interaction between time point and intervention group~~, an indicator for the type of measurement used~~, and potential covariates (cf. Item 27b) ~~the covariates mentioned in Item 27b~~. Random effects are school, class, and child specific intercept and slope (with respect to time as a continuous covariate). The covariance matrix of the random intercept and the random slope is allowed to vary freely between the two intervention groups. The error variances are allowed to vary over time and between the intervention groups. ~~The error variance is allowed to vary between ADP and BIA measurements, and a constant difference in variance is assumed~~. The interaction between time point and intervention will be parametrized as time point specific intervention effects. The randomization will be taken into account by restricting the intervention effect at baseline to 0.~~, the difference of the variance of the random intercept between the intervention groups, and the difference in error variances at baseline between the intervention groups to 0.~~

Results of the ~~primary~~ analysis will be reported by the effect estimate together with a 95% confidence interval and a p-value of the null hypothesis of no intervention effect based on a Wald test. The treatment effect will be phrased as "difference in mean change between the intervention and control group". In case of using log-transformed outcomes, the treatment effect will be transformed as (exp(.)-1)*100 allowing an interpretation as the "difference in mean percentage change". The boundaries of the confidence interval will be transformed accordingly.

The structure of the model:

- Reflects the potential effects of school and classes on the outcome.
- Takes the correlation between measurements from the same child into account.
- ~~Allows using the known association of the primary outcome with height and a potential association with age or sex to increase the power by using them as covariates (if relevant).~~
- Allows to include all children from the primary analysis population in the ~~primary~~ analysis independent of missing values in the ~~primary~~ outcome at some time point (including baseline) – except for those missing the outcome at all time points.

- Requires only to assume missing at random (MAR) with respect to missing values in the primary outcome.
- Takes into account the use of two different measurement methods for the primary outcome.

Computations will be based on the default settings of the mixed command in Stata. In case of convergence problems, the complexity of the model will be reduced by the following steps until the convergence is reached: 1) Removal of variance components with estimates on the boundary. 21) Equal error variances across at the two follow up time points. 32) Equal error variances across the intervention groups. 43) Equal correlation between random intercept and random slope across the intervention groups.

The weights (estimated inverse probabilities of participation using register data, described in Item 27b) will be incorporated using the pweight option.

With respect to analyses using both BIA and BODPOD measurements, a corresponding bivariate model will be used (see sensitivity analysis, cf. section 27e). This will require modelling a potential systematic difference between the two measurements.

Item 27b. Any adjustment for covariates

The statistical power of RCTs can be improved by adjusting for (strong) prognostic factors. The approach described above implies already an adjustment for baseline variables. Consequently, there is only a need to adjust for factors which improve the prediction of the outcomes after adjustment for baseline values. Consequently, although FM and some of the secondary outcomes are known to be highly influenced by height, this does not necessarily imply a need to adjust for height.

The decision to adjust for potential prognostic factors will be based on an empirical investigation, whether the prediction of the outcome at the final time point can be improved by adding the factor to a model including the baseline values. Improvements of the adjusted R^2 by at least 0.1 will be regarded as an indicator for inclusion as covariate in the analysis model. The potential prognostic factors considered are height, weight, BMI, age and sex.

In case the use of inverse probability of participation weighting does not result in sufficient balance between the two intervention groups with respect to some factors, corresponding covariates may be added.

The primary outcome, FM, and some of the secondary outcomes are known to be highly influenced by the height of a child. It is hence intended to adjust for height. In the literature, an adjustment for the square of height is common. In an initial data analysis (cf. item 29), it will be checked whether this presents the optimal type of adjustment also in this population. Similarly, it is unclear whether there is an association of the outcomes (on the top of height) with age and sex to a degree, which implies an increase in the precision of the estimate for the intervention effect. Hence, the degree of association will also be investigated in an initial data analysis. The final decision about inclusion of these three variables and their functional form will be made based on the initial data analysis.

The following table represents the potential variables for an adjustment for the different outcomes:

Outcome	**Potential adjustment variables**
FM, FFM, % FM, FFM-to-FM ratio, waist circumference	Height, age, sex
FM index, FFM index	Age, sex
Height z-score, BMI z-score	–

As per CONSORT and since this is an RCT, besides adjustments listed above, results will be presented unadjusted.

Parental consent for participation was obtained after the schools' randomization results had been communicated. This has led to variation in participation rates between the intervention and control groups, suggesting a potential imbalance in participating characteristics. To adjust for this, background information available from Statistics Denmark, also for the children and parents who did not consent to participation, is used to model the probability of consenting as a function of background variables. This modelling is currently performed by Statistics Denmark and described in a separate document, which will be attached as supplement to the publication. The model will be used to derive inverse probability weights for each child participating as described by Li et al. (20,21). We will refer to these weights as *"inverse sampling probability weights".*

Item 27c. Methods used for assumptions to be checked for statistical methods

Check of univariate distributions

The distribution of any variable included in any later analysis will be depicted graphically in the overall population and stratified by age and sex. Time varying variables will be also stratified by time point of assessment, and the analysis will be extended to change scores, i.e. differences from baseline to follow-up.

Outliers will be identified visually, and data will be checked for potential explanations. In case of explanations indicating an undue effect of external factors, outliers might be removed. Any decision of this kind will be documented and included in the supplementary material of the paper.

Consistency checks of individual trajectories

For each longitudinal outcome, the individual longitudinal trajectories of each child will be inspected with respect to a uniform development. Suspicious measurements will be checked.

Check of expectations

The outcomes considered in this study are well-established. Population norms are available and associations with child characteristics are well-established. The distribution of these variables and their association with child characteristics in this study at baseline will be depicted and compared with data from the literature, for example national data on prevalence of childhood overweight and obesity (5), to establish the validity of the measurement methods and evaluate the representativeness of the study sample.

Item 27d. Details of alternative methods to be used if distributional assumptions do not hold, e.g., normality, proportional hazards, etc.

In case of skewed distributions or other anomalies, potential data transformations will be discussed, or additional sensitivity analyses will be defined.

Item 27e. Any planned sensitivity analyses for each outcome where applicable

To evaluate the robustness of our results to the method used to measure body composition, two sensitivity analyses will be conducted:

1. An ADP-only analysis: The analysis will be repeated using FM measured by ADP at baseline and end-of-study. This analysis will include only participants with a valid ADP measurement, i.e. those wearing standard clothing in the BODPOD.
2. A bivariate model combining ADP and BIA measurements at all time-points: This model will include up to five measurements of FM per child (ADP and BIA at baseline, BIA at interim, ADP and BIA at end-of-study).

~~To check imperfectness of inverse probability weighting in achieving distributional balance between the two intervention groups, we will try adding the same covariates, as in represented in weights, to the model analysing primary outcome.~~

The potential impact of violation of the missing at random assumption in the sense that children with higher or lower value of the outcome are more likely to drop out will be addressed by a tipping approach. Multiple imputations will be generated for the missing outcomes variables, and certain amounts delta will be added to the imputed values. The dependence of the effect estimates on the choice of delta will be reported.

To check whether the results are strongly affected by any single school, the main analysis will be repeated while leaving out one school at a time. This will help assess the robustness of the findings.

Complete case analysis: A complete case analysis will be conducted as a sensitivity analysis. This will include study participants without missing data on the considered outcome at all time points.

Item 27f. Any planned subgroup analyses for each outcome

Effect modification by baseline adiposity level

To explore potential effect modification by baseline adiposity, we will model the intervention effect as a function of the FM index at baseline. This will be approached by extending the model used in the analysis of the primary outcome by replacing the constant intervention effect with a restricted cubic spline function with knots at the 25%, 50%, and 75% percentile. The asymmetric choice reflects the expectation that the function mainly varies in the upper part of the value range. The estimated spline function will be plotted together with a pointwise 95% band, the p-value of an overall test of the null hypotheses of no intervention effect, and the p-value of a test of a non-constant intervention effect.

Effect modification by sex

Effect modification by sex will be investigated by adding an interaction between the intervention and sex to the model used in the primary analysis. From this model, we will report estimates of sex-specific intervention effect of their difference together with 95% confidence intervals.

Further investigations of effect modifications

Investigations of effect modification by other candidate variables will be reported in a subsequent paper.

6.3. Missing data

Item 28. Reporting and assumptions/statistical methods to handle missing data (e.g., multiple imputation)

Missing data on primary outcome: The primary analysis will use BIA as measurement method for which we have minimal missing data. ~~To maximize our effort in preventing missing data on the primary outcome (FM), if a child's measurements at baseline or end-of-study are not available from ADP, we will use the data from BIA.~~ The chosen approach for the primary analysis allows already efficient estimation of the intervention effect under the MAR assumption. Hence, there is no need to perform multiple imputation in the primary analysis. We do not expect any missing values in the variables we need for adjustment in the primary analysis.

6.4. Additional analyses

Item 29. Details of any additional statistical analyses required, e.g, complier-average causal effect analysis

Initial data analyses

The analyses described in this section will be performed prior to starting the primary analysis and on a version of the analysis data set including no information on the intervention group. Therefore, they are allowed to inform the primary analysis.

1) The volume measurements of the ADP method are known to be affected by temperature and humidity. The ADP measurement procedure includes a correction for this, but it is unclear whether this correction is sufficiently calibrated. For all ADP assessments, information on temperature and humidity will be available. Hence, it is possible to check the degree of calibration.

   This will be based on fitting a regression model with the volume measurement as outcome and temperature, humidity, age, sex and height as covariates. The model is assumed to be of the structure alpha + f(temperature,humidity) + g(age,sex,height). The optimal functional form of f and g will be determined by using fractional polynomials. In case of distinct differences of f from 0, the estimate of f will be used to derive a correction formula of the type V'= V + f(temperature, humidity) - f(t_0,h_0), such that V' can be interpreted as the volume to be expected if the measurement has been performed at temperature t_0 and at humidity h_0.

2) The investigation will be extended to investigate the effect of clothing and fasting time on the volume measurements. There will be no attempt to correct the measured volumes accordingly, as these variables are affected by the choice of the child. A potential impact of clothing will be taken into account in the primary analysis by fixing the measurement method in children who change clothing (i.e. BIA will be used if the child wears different clothing at baseline and end-of-study). No attempts are planned to take the potential influence of fasting time into account.

3) The influence of age, sex and height on the primary outcome will be investigated by using the model described for the primary analysis without using the treatment indicator. The optimal function form of height will be investigated by using fractioned polynomials. In case of a distinct difference from a quadratic function, an alternative transformation will be used in the primary analysis. For age and sex, the reduction in the error variance when using these variables will be determined. They will be included in the primary analysis if they imply a reduction of the error variance by more than 5%.

"Drop-out" analysis

~~In this study there are three different levels of data availability for the outcomes relating to body composition: ADP, BIA and missing. This level plays a role similar to the traditional "drop-out" indicator for longitudinal outcomes in RCTs, and understanding of the association of the levels with child characteristics and time points can contribute to interpreting the robustness of the study results.~~

The status "missing" with respect to BIA measurements can be ~~further~~ divided into "missing due to drop out from study" and "failure to participate in an assessment". The distribution of the resulting three-level variable will be described stratified by child characteristics, assessment time points, and intervention group.

Secondary outcomes

The intervention effects on secondary outcomes will be analysed by the same model as used for the primary analysis. The FFM-to-FM ratio will be log-transformed prior to the analysis. For the prevalence

of underweight and overweight/obesity, a logistic mixed model will be used. From these analyses no p-values will be reported.

Analyses of patterns in intervention effect over time (secondary objective no. 2):
The intervention effect at interim will be derived from the model fitted in the primary analysis.

6.5. Harms

Item 30. Sufficient detail on summarizing safety data, eg, information on severity, expectedness, and causality; details of how adverse events are coded or categorized; how adverse event data will be analyzed, i.e, grade 3/4 only, incidence case analysis, intervention emergent analysis
See above Item 13c.

6.6. Statistical software

Item 31. Details of statistical packages to be used to carry out analyses
All analyses are expected to be completed in StataNow 19 BE (StataCorp).

6.7. References to other relevant documents

Published study protocol: (2)
Clinical Trials registration: https://clinicaltrials.gov/study/NCT05940675

References

1. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C et al. Guidelines for the content of statistical analysis plans in clinical trials. *JAMA.* 2017;318):2337–43.
2. Thomsen LT, Schmidt-Persson J, Damsgaard CT, Krustrup P, Grøntved A, Krølner RF et al. Generation Healthy Kids: Protocol for a cluster-randomized controlled trial of a multi-component and multi-setting intervention to promote healthy weight and wellbeing in 6–11-year-old children in Denmark. Pontoppidan M, editor. *PLoS One*. 2024;19:e0308142.
3. World Health Organization. *Fact Sheet - obesity and overweight* [Internet]. 2021. Available from: https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight, cited 10 June 2025.
4. World Health Organization. *Report of the Commission on Ending Childhood Obesity.* Geneva: World Health Organization; 2016.
5. Biltoft-Jensen AP, Fagt S, Møller FS. Inequality in childhood overweight and obesity in Denmark [in Danish]. [Internet]. Copenhagen; 2021. Available from: https://www.dst.dk/Site/Dst/Udgivelser/nyt/GetAnalyse.aspx?cid=46225, cited 10 June 2025.
6. Buttitta M, Iliescu C, Rousseau A, Guerrien A, Rousseau A. Quality of life in overweight and obese children and adolescents: a literature review. *Quality of Life Research*. 2014;23:1117–39.
7. Reilly J, Kelly J. Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. *Int J Obes.* 2011;35:891–8.
8. Hodder RK, O'Brien KM, Lorien S, Wolfenden L, Moore THM, Hall A, et al. Interventions to prevent obesity in school-aged children 6-18 years: An update of a Cochrane systematic review and meta-analysis including studies from 2015–2021. *EClinicalMedicine*. 2022;54:101635.
9. Spiga F, Davies AL, Tomlinson E, Moore TH, Dawson S, Breheny K, et al. Interventions to prevent obesity in children aged 5 to 11 years old. *Cochrane Database of Systematic Reviews.* 2024;5:CD015328.
10. Danish Ministry of Children and Education. *Education Statistics* [in Danish] [Internet]. 2023 Available from: https://uddannelsesstatistik.dk/Pages/main.aspx, accessed 23 May 2023.
11. Gallis JA, Li F, Yu H, Turner EL. cvcrand and cptest: Commands for efficient design and analysis of cluster randomized trials using constrained randomization and permutation tests HHS Public Access. *Stata J.* 2018;18:357–78.
12. Krølner RF, Brautsch LAS, Vilhelmsen M, Hoeeg D, Pedersen NH, Thomsen LT, et al. Generation Healthy Kids: a protocol for the mixed methods process evaluation of a school- and community-based intervention targeting healthy weight and wellbeing among 6–11-year-olds in Denmark. Trials (submitted). 2025.
13. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: Extension to cluster randomised trials. *BMJ*. 2012;345(7881).
14. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatr Obes*. 2012;7:284–94.
15. Larsen MN, Nielsen CM, Helge EW, Madsen M, Manniche V, Hansen L, et al. Positive effects on bone mineralisation and muscular fitness after 10 months of intense school-based physical training for children aged 8-10 years: the FIT FIRST randomised controlled trial. *Br J Sports Med*. 2018;52:254–60.
16. Orsso CE, Silva MIB, Gonzalez MC, Rubin DA, Heymsfield SB, Prado CM, et al. Assessment of body composition in pediatric overweight and obesity: A systematic review of the reliability and validity of common techniques. *Obesity Reviews*. 2020;21(8): e13041.
17. Ross R, Neeland IJ, Yamashita S, Shai I, Seidell J, Magni P, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. *Nat Rev Endocrinol*. 2020;16:177–89.

18. van Dijk SB, Takken T, Prinsen EC, Wittink H. Different anthropometric adiposity measures and their association with cardiovascular disease risk factors: a meta-analysis. *Netherlands Heart Journal*. 2012;20:208–218.
19. Andersen MB, Thorsted A, Jezek AH, Due P, Sørensen TIA, Thygesen LC. *Overweight and obesity in Danish children and adolescents. Prevalence and sociodemographic distribution* [in Danish]. Copenhagen: Danish National Institute of Public Health; 2020.
20. Papadogeorgou G, Liu B, Li F, Li F. Addressing selection bias in cluster randomized experiments via weighting. *Biometrics.* 2025;81:ujaf013.
21. Li F, Tian Z, Bobb J, Papadogeorgou G, Li F. Clarifying selection bias in cluster randomized trials. *Clinical Trials*. 2022; 19:33–41.

Appendix 1: Rationale for SAP revision, version 2, 18 Nov 2025

1. Scope of changes to SAP

This revision applies to the following items in the SAP:

- 5.5 Baseline characteristics (Item 25b: Baseline characteristics)
- 6.1 Outcome definitions (Item 26b: Measurement methods)
- 6.2 Analysis methods (Item 27a: Analysis method, Item 27b: adjustment for covariates, Item 27e: Sensitivity analyses)
- 6.3 Missing data (Item 28: Handling of missing data)

2. Purpose of this Appendix

This Appendix explains and documents the revision to the primary outcome measurement method and describes the added prespecified sensitivity analyses. The changes are informed by results from the initial data inspection. Note: The initial data analyses were conducted on a dataset without information on allocation – i.e. the analyses were conducted irrespective of allocation to intervention or control group, and without knowledge of the intervention's potential effect on primary or secondary outcomes.

3. Summary of main changes to SAP

In the primary analysis, bioelectrical impedance (BIA) alone will be used to measure fat mass. This replaces the model described in the original SAP, where it was intended to use ADP as the primary measurement, with BIA used only for those with missing data on ADP.

ADP will be used in two pre-specified sensitivity analyses:

i. An ADP-only analysis: The primary analysis will be repeated using FM measured by ADP at baseline and end-of-study. This analysis will include only participants with a valid ADP measurement, i.e. those wearing standard clothing in the BODPOD.
ii. A bivariate model combining ADP and BIA at all time-points.

4. Rationale for changes

4.1. Measurement method for primary outcome:

In SAP version 1 (Item 29), we described that we would conduct initial analyses of the impact of clothing on ADP measurements, as it is well-established that non-standard clothing can affect volume estimates during ADP (1). Clothing was recorded during data collection, and clothing was subsequently categorized into three groups:

1. *Standard clothing*, i.e. tight-fitting underwear or swimwear (optimal),
2. *Project clothing*, i.e. clothes provided by the project, which fitted children variably,
3. *Other clothing*, e.g. children's own clothing.

After data cleaning and categorization of clothing, the number of measurements among the 1,359 participants in the analysis population were distributed as follows:

	Baseline		Interim		End-of-study	
	n	%	N	%	n	%
ADP available	**909**	**67**	**NA**	**-**	**1,039**	**76**
Clothing 1 (standard)	587	43	NA	-	708	52
Clothing 2 (project)	204	15	NA	-	231	17
Clothing 3 (other)	118	9	NA	-	100	7
BIA available	**1,274**	**94**	**1,225**	**90**	**1,197**	**88**
Total	**1,359**	**100**	**1,359**	**100**	**1,359**	**100**

Table S1. Number of measurements of primary outcome across measurement methods, measurement rounds and ADP clothing categories.

Table S1 shows that the total number of ADP measurements is substantially lower than the number of BIA measurements. A notable proportion of children declined the ADP assessment because they found the BODPOD intimidating.

In our efforts to investigate the impact of clothing, Bland–Altman plots stratified by clothing category and height were generated to assess agreement between BIA and ADP. These plots indicated that agreement varied systematically with both clothing category and child height (see Figure S1). These findings highlighted the impact of clothing, leading us to consider only clothing category 1 as valid. Consequently, less than half of participants had valid ADP measurements at baseline (43%) and approximately half at end-of-study (52%) (Table S1).

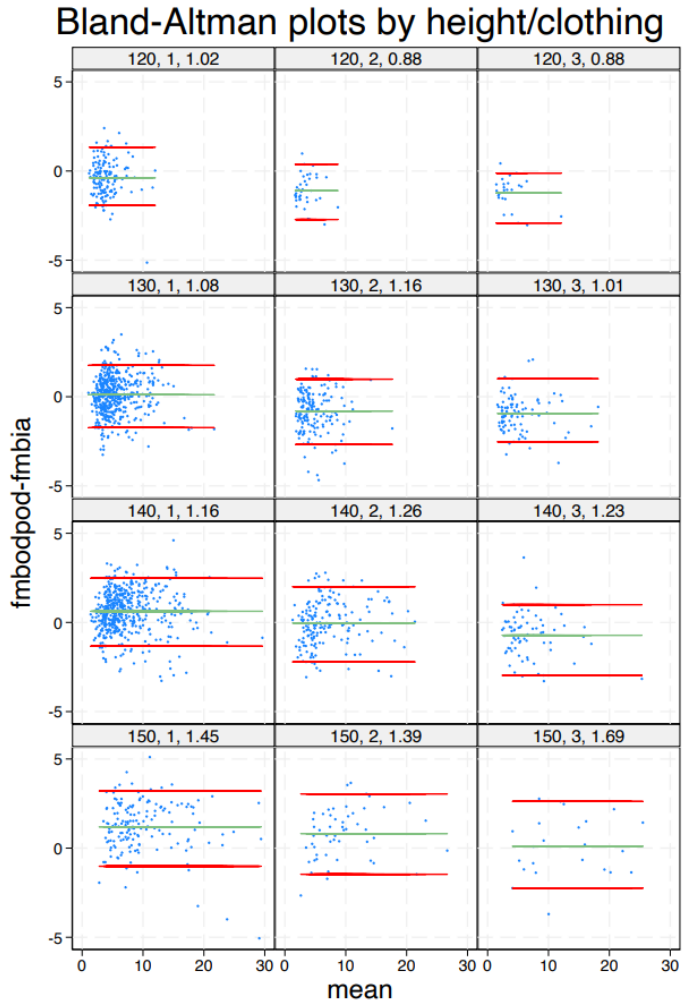


Figure S1: Bland–Altman plots showing agreement in fat mass between air-displacement plethysmography (ADP; fmbodpod) and bioelectrical impedance analysis (BIA; fmbia), stratified by clothing category (1, 2, 3) and height (120, 130, 140, 150 cm).

Given i) the considerable numbers of children with missing data on ADP under the validity restriction concerning clothing; and ii) the systematic differences between ADP and BIA related to clothing and height, we decided that the primary analysis of the intervention effect on FM will only include BIA measurements.

To be completely transparent and to examine the robustness of the study results on the selection of measurement method for the primary outcome, we will include two pre-specified sensitivity analyses:

i. An ADP-only analysis: The primary analysis will be repeated using FM measured by ADP at baseline and end-of-study. This analysis will only include the participants with a valid ADP measurement, i.e. those wearing standard clothing in the BODPOD.
ii. A bivariate model combining ADP and BIA measurements at all time-points: This model will include up to five measurements of FM per child (ADP and BIA at baseline, BIA at interim, ADP and BIA at end-of-study).

4.2. Rationale for other changes to SAP:

The changes to the analysis methods and adjustment for covariates (SAP Items 27a and 27b) reflect that:

- No need to take two measurement methods into account.
- All longitudinal continuous outcomes can be analysed by a (structurally) identical model.
- The decision to use partially log-transformed outcomes.
- A lack of specification how treatment effects should be phrased in the original SAP.
- Overstating the role of covariates in the original SAP.
- Limitations implied by the use of Stata's mixed command and the use of weights.

5. Timing of SAP changes and upload

The second version of the SAP was finalized and uploaded **before initiation of any analyses investigating the intervention's effect on primary or secondary outcomes**. The decision to change the measurement method for the primary outcome was thus unaffected by any analyses of intervention effectiveness. All changes were approved by all SAP contributors on 18 November 2025.

6. References

(1) COSMED (2020). *BOD POD® GS-X (model 2020) User Manual (6/2022). 210-3000 Rev E.* COSMED.