GoActive | Statistical Analysis Plan

A cluster-randomised controlled trial to evaluate the effectiveness and cost-effectiveness of the GoActive intervention to increase physical activity among 13-14 year-old adolescents**.**

Version 5.21 –April 2019

Contents

1 Introduction 3

2 Study outcomes 3

2.1 Additional socio-demographic information 6

2.1a Age and sex 6

2.1b Ethnicity and language spoken at home 6

2.1c Religious affiliation 6

2.1d Family socio-economic status 6

2.1e Family structure 7

3 Quality control, cleaning data and deriving variables 7

3.1 Accelerometer data 8

3.3 Questionnaire data: self-efficacy 9

3.4 Questionnaire data: social support for physical activity 9

3.5 Questionnaire data: friendship quality 9

3.6 Questionnaire data: well-being and self-esteem 9

3.7 Questionnaire data: shyness and sociability 10

3.8 Educational data 10

4 Trial analyses 10

4.1 Analysis populations 10

4.2 Descriptive characteristics 11

4.3 Analysis of primary and secondary outcomes 11

4.4 Effect modification analyses 12

5 Further considerations for analysis 12

5.1 Missing data 12

5.2 Implausible values of primary outcome 13

5.3 Multiplicity 13
6. Authors and Reviewers 14
7 References 15

1 Introduction

The aims and details of data collection of the GoActive cluster randomised controlled trial (CRCT) are provided in the main trial protocol paper, published in September 2017.[1] The purpose of this document is to outline the procedures for the main trial analysis, designed to assess the 10-month effectiveness of the GoActive intervention to increase average daily objectively measured moderate to vigorous physical activity (MVPA) among 13–14-year-old adolescents. This includes details on quality control of the data collected and the criteria used to define derived variables. In addition, it describes to how the intention-to-treat and per-protocol analyses will be defined and planned sensitivity analysis for missing data.

The analysis plan was approved by the Trial Steering Committee (TSC) on 11th April 2018 (i.e. before the long-term follow-up (T4) data are analysed).

Details of other analyses of data arising from the GoActive study will be the subject of later documents. This will include any cohort analyses, process evaluation, and cost-effectiveness analysis and long-term economic impact modelling (the latter two analysis plans are detailed in the GoActive Health Economics Analysis Plan, HEAP).

2 Study outcomes

Table 1 lists the primary and secondary outcomes of the GoActive study and describes the type of variable for each.

Table 1. GoActive study outcomes – all outcomes analysed separately at T3 and T4.

Variable	Type, Unit / Categories	Source, Comments
Primary outcome		
Average daily minutes of objectively measured moderate-to-vigorous physical activity (MVPA) at T4	Continuous; minutes	Continuous waveform data from wrist-worn Axivity monitor.[2]
Secondary outcomes: activity		
Average daily minutes of objectively measured moderate-to-vigorous physical activity (MVPA) at T3	Continuous; minutes	Continuous waveform data from wrist-worn Axivity monitor.[2]
Average daily minutes of MVPA during school time	Continuous; minutes	Continuous waveform data from wrist-worn Axivity monitor.[2]
Average daily minutes of MVPA during weekdays after school	Continuous; minutes	Continuous waveform data from wrist-worn Axivity monitor.[2]
Average daily minutes of MVPA at weekends	Continuous; minutes	Continuous waveform data from wrist-worn Axivity monitor.[2]
Average daily minutes of sedentary time, light intensity physical activity and overall physical activity (average acceleration)	Continuous; minutes	Continuous waveform data from wrist-worn Axivity monitor.[2]
Average daily minutes of sedentary time, light intensity physical activity and overall physical activity (average acceleration) during school time	Continuous; minutes	Continuous waveform data from wrist-worn Axivity monitor.[2]
Average daily minutes of sedentary time, light intensity physical activity and overall physical activity (average acceleration) during weekdays after school	Continuous; minutes	Continuous waveform data from wrist-worn Axivity monitor.[2]
Average daily minutes of sedentary time, light intensity physical activity and overall physical activity (average acceleration) at weekends	Continuous; minutes	Continuous waveform data from wrist-worn Axivity monitor.[2]
Change in intensity distribution	Continuous; minutes	Continuous waveform data from wrist-worn Axivity monitor.[2]

Physical activity type	Categorical; Never, Once, 2 to 3 times, 4+ times	Self-reported using items from the Youth Physical Activity Questionnaire (YPAQ).[3] Categories may be collapsed further depending on response distribution.
Secondary outcomes: anthropometry		
BMI	Continuous; score	Body mass index (BMI) z (SD) score calculated from height and weight data (i.e. weight / height^2 (kg/m^2)) collected onsite by trained research staff. Later categorised according to age and sex standardised IOTF thresholds.[4]
Body fat percentage	Continuous; %	Data collected onsite using bioelectrical impedance scales.
Waist circumference	Continuous; whole number (cm)	Data collected onsite by trained research staff.
Secondary outcomes: psychosocial		
Self-efficacy	Continuous; score in whole numbers (range 0-48)	8 self-reported items from Reynolds' Psychosocial Predictors of Physical Activity: Self-efficacy scale.[5]
Social support for physical activity	Continuous; score in whole numbers (range 9-36)	9 self-reported items from European Youth Heart Study.[6]
Friendship quality	Continuous; score in whole numbers (range 8-40)	8 self-reported items used in the ROOTS project (equally weighted).[7]
Well-being	Continuous; score in whole numbers (range 14-70)	Self-reported using 14-item Edinburgh-Warwick Wellbeing Scale.[8]
Self esteem	Continuous; score in whole numbers (range 10-40)	Self-reported using 10-item Rosenberg Self-Esteem Scale.[9]
Shyness and sociability	Continuous; score in whole numbers (range 10-50)	Self-reported using two 5-item measures from the EAS (Emotionality, Activity, Shyness and Sociability) temperament scale.[10]

2.1 Additional socio-demographic information

In addition to study outcomes identified in Table 1, socio-demographic information was collected as part of the GoActive study. These data have been collected at baseline only, unless otherwise stated. These data will primarily be used to describe the analysis sample, and also in future exploratory analyses (i.e. sex, socio-economic status, ethnicity, baseline physical activity, weight status).

2.1a Age and sex

Data on student age will be derived from self-report questionnaires (based on age in years and month of birth). Where possible, data self-reported by students will be used to identify sex (either Male, Female, or Prefer not to say). For students with missing data, or selecting multiple response options, sex will be imported from the electronic form used during anthropometric measurements.

2.1b Ethnicity and language spoken at home

Students self-reported their ethnicity from 20 response options (including free text for 'Any other background', 'Don't know' and 'Don't want to answer'). Multiple responses will be re-coded as follows:

- If a single ethnicity is selected (e.g. White British) in addition to 'Don't know', 'Don't want to answer', or 'Any other background' with no free text response, this will be recoded as the ethnicity that was identified by the student.
- If multiple backgrounds within one particular ethnicity category (e.g. Asian backgrounds, including Indian, Pakistan, Bangladeshi etc.) are selected, then the response will be recoded as 'Any other [e.g.] Asian background'.
- Where multiple 'Don't know' or 'Any other [background]' response options have been selected, the response will be recoded as the first distinct ethnicity (e.g. African) selected by the student.

Cleaned responses will be collapsed into 5 categories: White, Mixed/multiple ethnic background, Asian or Asian British, Black or Black British, or Other ethnic group.

2.1c Religious affiliation

Students were asked "What are your religious beliefs?". Religious affiliation will first be described using all options that were provided to students (no religion, Christian, Buddhist, Hindu, Jewish, Muslim, Sikh, other religion), but depending on the distribution of responses, may need to be further collapsed into a dichotomous variable (e.g. no religion vs. any religion selected).

2.1d Family socio-economic status

Family socio-economic status will be derived from student-reported data. At baseline, students were asked to indicate 1) their main caregivers, and 2) the highest level of education completed by the identified caregivers. Students also completed six items from the Family Affluence Scale[13] related to family car ownership, holidays, computers and availability of bathrooms, dishwasher and own bedroom. Where multiple response options are selected, items will be recoded with the most socio-economically deprived option (e.g. One bathroom, rather than Two or more bathrooms, in the family

home), in line with other conservative data processing approaches.[14] Family affluence, as a proxy for family socio-economic status, will be calculated by summing answers (possible range: 0 to 13), and dividing into pre-defined groups (i.e. affluence: low = 0-6, medium = 7-9, high = 10-13).[15] At least 4 out of 6 completed items will be required for inclusion in the summary score. At T2, students indicated whether or not they are eligible for free school meals, this will be used as a dichotomous variable with don't know and don't want to answer coded as not eligible (eligible for free school meals vs. not eligible).

2.1e Family structure

Family structure was collected by asking participants to select one or two main carers from a pre-defined list of 8 possible family members (e.g. father, step-mum, older brother). Family structure will first be described in 16 distinct categories (Mother and father, Parent and opposite sex step-parent, Same sex parents, Mother and other family member, Father and other family member, Mother only, Father only, Multiple family members, Other family member only, Other adult only, No carer selected, Mother and other adult, Father and other adult, Older brother or sister and other adult, Grandparent and other adult, Step-dad and other adult). Depending on the distribution of responses, these categories may be further collapsed (Mother and father vs. any other family structure selected, or other categories).

3 Quality control, cleaning data and deriving variables

The same quality control (QC), cleaning, and variable derivation procedures will be used for data that were collected at baseline (T1), interim (T2), immediate follow-up (T3), and long-term follow-up (T4). Timelines for data collection and subsequent quality control, cleaning, and derivation of variables are as follows:

- T1 data collection: September 2016 – January 2017, cleaned data release October 2017
- T2 data collection: February – May 2017, cleaned data release scheduled April 2018
- T3 data collection: April – July 2017, cleaned data release scheduled April 2018
- T4 data collection: scheduled for March – July 2018, cleaned data release November 2018

Apart from the process evaluation data, no follow-up data will be used in analysis until after the main analyses has been conducted.

All QC, data cleaning, and processing will be conducted on data sets not including intervention identifiers. Data cleaning will be conducted by Dr Kirsten Corder, supported by GoActive study assistants and members of the relevant function teams at the MRC Epidemiology Unit, University of Cambridge. Of this team, only Dr Kirsten Corder is unblinded to the intervention.

3.1 Accelerometer data

Information on activity will be derived from wrist-worn accelerometry (Axivity) data, with a 7-day wear protocol. Monitors will set up to record acceleration at 100Hz with a dynamic range of +-8*g*. Once returned, data from the monitors will be downloaded. The acceleration signal is composed of two main components; activity-related acceleration and gravity. Euclidean Norm Minus One (ENMO) can be interpreted as the magnitude of activity-related acceleration, accounting for (subtracting) gravitational acceleration, and represents a plausible approximation of human movement.[16,17] We suggest classifying time spent below 30mg of ENMO as sedentary (equivalent to 1.5 METs or ≤100 ActiGraph cpm), and time spent above 210mg ENMO will be classified as MVPA (equivalent to 4 METs or ≥2000 ActiGraph cpm).[18] Specific cut-points will be decided upon based on the best available evidence at the time of T4 data processing.

Given the 24-hour wear time protocol of the Axivity monitors, a diurnal adjustment will be used to reduce any bias caused by imbalances of non-wear.[19,20] Each day of possible wear will be divided into four time quadrants: morning (6am – 12pm), afternoon (12pm - 6pm), evening (6pm – midnight), and night (midnight – 6am). For students to be included in analyses, at least 12 hours of wear time will be required from the possible 42hrs in each quadrant (i.e. ≥12 hours from 7 possible mornings, ≥12 hours from 7 possible afternoons, and ≥12 hours from 7 possible evenings). The 'night' quadrant (i.e. midnight – 6am) will be classed sleep time for adolescents. For an individual hour to be included for analysis, at least 70% wear time will be required. These settings will be evaluated to ensure optimal use of the data. The final settings will be decided upon based on the best available evidence at the time of T4 data processing

Additional QC steps will include:

- Creating normal plots, histograms, and scatter plots to identify potentially implausible values.
- Scatter plots will compare baseline and follow-up values for each variable, and will be used to review different variables measured at the same time point that we would expect to be correlated (e.g. time spent in MVPA and overall physical activity).
- Outliers that deviate from the 'line' of the scatter plots by 2 SD or more on either axis will be considered implausible. For implausible values, the original data will be checked in the accelerometer software to make sure criteria have been applied correctly and any errors will be modified.

3.2 Anthropometric data

Anthropometric data will be cleaned using cleaning procedures similar to those conducted with accelerometry data. As the onsite data collection form used prevents the entry of pre-defined implausible values, visual review of histograms etc. will be only conducted as a secondary precaution. Scatter plots will however be used to compare each measure at baseline to the equivalent measure at follow-up, and will also match weight to height, weight to waist circumference, and weight to body fat percentage values at relevant time points.

Variables for weight status will be then derived. Body Mass Index (BMI) will be calculated from height and weight data, using weight / height^2 (kg/m^2).

Where:

- *oBMIag* is the observed BMI for a given sex and age (within 6 month categories)
- *mBMIag* is the mean BMI for participants of the same sex and same age as a given participant for whom the score is being derived
- *sdBMI*ag *is* the standard deviation of the mean BMI for participants of the same sex and same age as a given participant for whom the score is being derived

Data on BMI, waist circumference, and body fat percentage will be treated as continuous. The derived BMI z-scores will then be categorised in a new variable denoting Underweight, Normal weight, Overweight, or Obese, according to age and sex standardised International Obesity Task Force thresholds.[4]

3.3 Questionnaire data: self-efficacy

Self-efficacy is measured using 8 items from Reynolds' Psychosocial Predictors of Physical Activity: Self-efficacy scale,[5] covering self-efficacy for support seeking. A sum score across all items will be generated, with higher scores indicating greater self-efficacy for physical activity (range 0-48). During cleaning, for items for which multiple response options have been selected, data will be transformed to reflect the most conservative value.[14] To account for accidentally missed items, mean scores are derived including only those participants for whom at least 6 items have been completed. Normal plots, histograms, and scatter plots will be used, as for other variables, to review the normality of data and identify outliers.

3.4 Questionnaire data: social support for physical activity

Social support for physical activity was assessed using 9 self-reported items from European Youth Heart Study.[6] A mean score across all items will be generated, with higher scores indicating greater social support for physical activity (range 1-9). At least 5 out of 6 (support from family) and 2 out of 3 (support from friends) completed items will be required for inclusion in the summary score.

3.5 Questionnaire data: friendship quality

Friendship quality was measured using items previously included in the ROOTS prospective cohort study.[21] This 8-item scale rates friendship quality based on the availability, adequacy and intimacy of current friendships, and includes items related to number of friends, frequency of seeing friends, confiding in friends and episodes of teasing. Either 4 or 6 response options are provided for each item, and responses will be summed with higher scores indicating higher friendship quality. A minimum of 6 out of 8 item responses will be required for inclusion in the summary score.

3.6 Questionnaire data: well-being and self-esteem

Well-being was assessed using the 14-item Edinburgh-Warwick Wellbeing Scale.[8] Items (e.g. "I've been feeling useful") will be answered on a 5-point scale and responses summed, with higher scores

indicating lower levels of wellbeing. Self-esteem of participants will be assessed using the 10-item Rosenberg Self-Esteem Scale[9]. Each statement (e.g. “On the whole, I am happy with myself”) will be responded to on a four-point scale (Strongly agree to Strongly disagree), with higher scores denoting higher self-esteem. At least 12 responses to well-being items, and 8 to self-esteem, will be required for inclusion in the summary score.

3.7 Questionnaire data: shyness and sociability

Shyness and sociability was assessed with 10 items from the EAS (Emotionality, Activity, Shyness and Sociability) temperament scale, which identifies trait personality characteristics.[10] At least 3 out of 5 responses for shyness (and 3 out of 5 for sociability) will be required for inclusion in the summary score. Each item (e.g. “I make friends easily” (shyness) and “I like to be with people” (sociability)) will be ranked by participants from 1 ‘not typical’ to 5 ‘very typical’; reverse coding will be accounted for. Items will be summed, with higher scores indicating lower shyness and higher sociability.

3.8 Educational data

Anonymous, individual-level pupil attendance and academic performance will be collected at school-level from the National Pupil Database. Number of authorised (i.e. sickness or holiday) and unauthorised (i.e. truancy) days absent from school will be collected, and transformed into a percentage of each school year. Academic performance will be calculated as the sum of grade based points (A* = 58, A = 52…G = 16) and also as number of students gaining 5A*-C grades or the equivalent, consistent with the national reporting standard.[12]

4 Trial analyses

This section will describe the methods that are proposed to assess the effectiveness of the GoActive intervention on the primary and secondary outcomes. Analyses will be performed using Stata (version 14.2).[22] Trial analyses will be conducted on the clean data by Stephen Sharp (Senior Statistician at the MRC Epidemiology Unit, University of Cambridge).

4.1 Analysis populations

The primary analysis will use an Intention To Treat (ITT) population, which includes all participants in the group to which they were randomised, regardless of the intervention they received. A secondary analysis of the primary outcome only will use a Per Protocol (PP) population. Inclusion in the PP population for each participant was based on his/her degree of usage of the intervention website/submission of points and/or self-reported use of the intervention, and was defined jointly by key members of the GoActive project group (Stephen Sharp, Dr Stephanie Jong, Dr Esther van Sluijs, and Dr Kirsten Corder) after inspecting distributions. Participants who were active during tutor times at least two times during the last two weeks at T2, visited the website over 5 times and logged over 11 points were included in the per protocol population and are considered to have experienced the intervention as intended. This degree of usage of the intervention website and submission of points is equivalent to over the 75th percentile.

4.2 Descriptive characteristics

Baseline characteristics of the study population will be summarised separately within each randomised group. These characteristics will include age, sex, ethnicity, language spoken at home, religious affiliation, family socioeconomic status and structure, and weight status. For continuous variables, means and standard deviations will be reported, unless the variable has a highly skewed distribution, in which case the median, 25th and 75th percentiles will be presented. For categorical variables, the number and percentage of participants within each category will be presented. For each variable (continuous or categorical), the percentage of missing values will be reported. No statistical tests will be performed to compare randomised groups with respect to baseline characteristics.

4.3 Analysis of primary and secondary outcomes

Analyses will be reported according to the *Consort 2010: extension to cluster randomised trials* statement.[23] Table 2 provides an overview of the objectives, methods and proposed timelines for these analyses.

Table 2. Summary of GoActive analysis methods and timelines

Objective	Main methods	Planned timeline
Determine the effect of the GoActive intervention on average daily MVPA (the primary outcome)	ITT analysis. Linear regression (continuous outcome). Baseline values of the outcome included as a covariate (i.e. analysis of covariance, ANCOVA). Randomisation stratifiers, i.e. socio-economic status (low, high defined by pupil premium) and county (Cambridgeshire, Essex), will be included as covariates. Robust standard errors calculated to allow for the non-independence of individuals (clustering) within schools. Intervention effect will be the difference in mean change in MVPA (adjusted for baseline) between the intervention and control group, with a 95% confidence interval and p-value.	Analysis: Nov 2018 Submission: Jan 2019
Determine the effect of the GoActive intervention on secondary outcomes[a]	ITT analysis. Linear regression for continuous outcomes (e.g. self-efficacy) and multinomial regression for categorical outcomes (e.g. physical activity type). Baseline values of the outcome included as a covariate. Randomisation stratifiers, i.e. socio-economic status (low, high defined by pupil premium) and county (Cambridgeshire, Essex), will be included as covariates. Robust standard errors calculated to allow for the non-independence of individuals (clustering) within schools. For continuous outcomes, the intervention effect will be the difference in mean change in the outcome (adjusted for baseline) between the intervention and control group, with a 95% confidence interval. For categorical outcomes, intervention effects will be "relative risk ratios" (adjusted for baseline) comparing the intervention vs control group, for each category relative to a reference category, with 95% confidence intervals. No p-values will be calculated for the secondary outcomes.	Analysis: Nov 2018 Submission: Jan 2019

Secondary analysis: determine the effect of the GoActive intervention on the primary outcome in those who received intervention as intended	PP analysis.[b] Method as described above for the primary outcome.	Analysis: Nov 2018 Submission: Jan 2019

[a] For details of secondary outcomes, see Table 1.
[b] See 3.2a Analysis populations for details of PP analysis population.
Any continuous variables with skewed distributions will be log transformed prior to modelling.

4.4 Effect modification analyses

Effect modification (interaction) by pre-specified moderators will also be assessed in the above described regression models, for the primary outcome only (average daily minutes of MVPA at T4). These moderators will be:

- Sex (male vs. female)
- Socio-economic status (medium or lower vs. high according to FAS score[15])
- Ethnicity (White British background vs. any other ethnic background)
- Baseline physical activity (at least 60 average daily minutes of MVPA vs. less than 60 minutes if appropriate; otherwise, a median split will be used)
- Weight status (normal weight, overweight or obese)

A multiplicative interaction parameter between randomised group and each moderator in turn will be included in the regression model, and tested for significance using an F-test; if the p-value is <0.05, the intervention effect and 95% confidence interval will be estimated within each subgroup.

5 Further considerations for analysis

5.1 Missing data

Processing of individual variables for which data are missing is described in detail in Section 3. Data collection procedures in the GoActive study should minimise the volume of missing data, particularly for variables assessed using supervised questionnaires. If a participant has a missing value for an outcome at follow-up (e.g. waist circumference), he/she will be excluded from the analysis of that variable. This complete-case analysis is valid under the assumption that the outcome is missing at random (MAR) given randomised group and baseline.[24]

The pattern of missing data will be described. In the unlikely event that more than 10% of individuals have missing data for the primary outcome, the potential impact of deviations from the MAR assumption on the results for this outcome will be explored in sensitivity analyses using a pattern mixture model[25].

For continuous outcomes, those participants with a missing *baseline* value of the outcome variable will be included in the analysis using the missing indicator method[26], which is a valid method for pre-

randomisation measures in trials ensuring that no further participants are excluded while maintaining the advantage of improved precision.

5.2 Implausible values of primary outcome

A sensitivity analysis of the primary outcome will be performed in which any implausible values (as defined in section 3) are excluded. If the intervention effect estimate differs by more than 10% from that obtained when these values are included, both results will be reported in the paper.

5.3 Multiplicity

Given the large number of outcome variables, the focus of the results will be on estimated differences and 95% confidence intervals; p-values will only be reported for the primary outcome and for the interaction tests with respect to this outcome.

6. Authors and Reviewers

Authors:

Helen Elizabeth Brown, Kirsten Corder, Esther van Sluijs, Stephen J Sharp

MRC Epidemiology Unit, University of Cambridge

Reviewers:

Kate Hunt, Professor in Institute for Social Marketing, University of Stirling (previously Associate Director of MRC Social and Public Health Sciences Unit, University of Glasgow)

Russ Jago, Professor of Paediatric Physical Activity and Public Health, University of Glasgow

Ann Smith, Statistician, Astra Zeneca

Roz Fitches, Insight Coordinator, Living Sport

David Meads, Associate Professor, Faculty of Medicine and Health, University of Leeds

Jamie Millership, Teacher, Saffron Walden High School

Version 5.0 approved by reviewers on 11th April 2018

Clarifications in version 5.1 in July 2018

Clarifications in version 5.11 7th November 2018

Clarifications in version 5.13 18th March 2019

Final version 5.2 (with track changes removed) 18th March 2019

7 References

1 Brown HE, Whittle F, Jong ST, *et al.* A cluster randomised controlled trial to evaluate the effectiveness and cost-effectiveness of the GoActive intervention to increase physical activity among adolescents aged 13-14 years. *BMJ Open* 2017;**7**:e014419.

2 Doherty A, Jackson D, Hammerla N, *et al.* Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study. *PLoS One* 2017;**12**.

3 Corder K, van Sluijs EMF, Wright A, *et al.* Is it possible to assess free-living physical activity and energy expenditure in young people by self-report? *Am J Clin Nutr* 2009;**89**:862–70.

4 Cole TJ. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 2000;**320**:1240–1240.

5 Saunders RP, Pate RR, Felton G, *et al.* Development of Questionnaires to Measure Psychosocial Influences on Children's Physical Activity. *Prev Med (Baltim)* 1997;**26**:241–7.

6 Ommundsen Y, Page A, Ku PW, *et al.* Cross-cultural, age and gender validation of a computerised questionnaire measuring personal, social and environmental associations with children's physical activity: The European Youth Heart Study. *Int J Behav Nutr Phys Act* 2008;**5**:29.http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18489736

7 Goodyer IM, Herbert J, Tamplin a, *et al.* Short-term outcome of major depression: II. Life events, family dysfunction, and friendship difficulties as predictors of persistent disorder. *J Am Acad Child Adolesc Psychiatry* 1997;**36**:474–80.

8 Clarke A, Friede T, Putz R, *et al.* Warwick-Edinburgh Mental Well-being Scale (WEMWBS): validated for teenage school students in England and Scotland. A mixed methods assessment. *BMC Public Health* 2011;**11**:487.

9 Baumeister RF, Campbell JD, Krueger JI, *et al.* Rosenberg Self-Esteem Scale. 2003.

10 Buss AH, Plomin R. *Temperament (PLE: Emotion): Early Developing Personality Traits*. Psychology Press 2014. https://books.google.com/books?hl=en&lr=&id=zBSLBQAAQBAJ&pgis=1 (accessed 29 Sep2015).

11 National pupil database - GOV.UK. https://www.gov.uk/government/collections/national-pupil-database (accessed 10 Oct2017).

12 Education D for. Test and examination point scores used in the 2011 school and college performance tables. 2011;:1–8.

13 Currie C, Molcho M, Boyce W, *et al.* Researching health inequalities in adolescents: the development of the Health Behaviour in School-Aged Children (HBSC) family affluence scale. *Soc Sci Med* 2008;**66**:1429–36.

14 Corder K, van Sluijs EMF, McMinn AM, *et al.* Perception versus reality awareness of physical activity levels of British children. *Am J Prev Med* 2010;**38**:1–8.

15 Voráčová J, Sigmund E, Sigmundová D, *et al.* Family Affluence and the Eating Habits of 11- to 15-Year-Old Czech Adolescents: HBSC 2002 and 2014. *Int J Environ Res Public Health* 2016;**13**.

16 White T, Westgate K, Wareham NJ, *et al.* Estimation of physical activity energy expenditure during free-living from wrist accelerometry in UK adults. *PLoS One* 2016;**11**.

17 van Hees VT, Gorzelniak L, Dean León EC, *et al.* Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity. *PLoS One* 2013;**8**.

18 Trost SG, Loprinzi PD, Moore R, *et al.* Comparison of accelerometer cut points for predicting activity intensity in youth. *Med Sci Sports Exerc* 2011;**43**:1360–8.

19 Collings PJ, Wijndaele K, Corder K, *et al.* Levels and patterns of objectively-measured physical activity volume and intensity distribution in UK adolescents: the ROOTS study. *Int J Behav Nutr Phys Act* 2014;**11**:23.

20 Brage S, Westgate K, Wijndaele K, *et al.* Evaluation of a method for minimising diurnal information bias in objective sensor data. In: *ICAMPAM (Amherst)*. 2013.

21 Atkin AJ, Corder K, Goodyer I, *et al.* Perceived family functioning and friendship quality: cross-sectional associations with physical activity and sedentary behaviours. *Int J Behav Nutr Phys Act* 2015;**12**:23.

22 StataCorp. Stata Statistical Software: Release 15. *Coll Station TX StataCorp LLC* 2017.

23 Campbell MK, Piaggio G, Elbourne DR, *et al.* Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;**345**:e5661–e5661.

24 Sterne JAC, White IR, Carlin JB, *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;**338**:b2393.

25 White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. *Clin Trials* 2012;**9**:396–407.

26 White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. *Stat Med* 2005;**24**:993–1007.