Statistical Analysis Plan

A randomised controlled trial to investigate the effect of a structured education programme on women who have had gestational diabetes and are at risk of developing type 2 diabetes (The BABY STEPS study).

Trial registration

Sponsor Reference No: 0596
Ethics Ref No: 210608

SAP revision history

Date	Version	Justification for SAP version
05 September 2019	0.1	First draft for review
15 September 2019	0.2	Changes made for SAP reviewers' comments on v 0.1
21 September 2019	0.3	Changes made for SAP reviewers' comments on v 0.2
30 September 2019	1.0	Final version

SAP responsibilities

Role in SAP development	Name, affiliation	Role in trial
SAP author	Ghazala Waheed, University of Leicester	Study Statistician
SAP reviewer 1	Laura Gray, University of Leicester	Study Senior Statistician
SAP reviewer 2	Kamlesh Khunti, University of Leicester	Study chief Investigator
SAP reviewer 3	Ponnusamy Saravanan	Principal investigator

SAP signatures

Role	Name, affiliation	Date	Signature
Chief Investigator	Kamlesh Khunti, University of Leicester	01/10/2019	
SAP author Trial Statistician	Ghazala Waheed, University of Leicester	01/10/2019	
Supervising Statistician	Laura Gray, University of Leicester	01/10/2019	

1 Introduction

1.1 Trial background and rationale

Gestational diabetes is a type of diabetes that arises during pregnancy. Women with gestational diabetes have a two-fold increase in risk of pre-eclampsia and at least a seven-fold increase in risk of developing type 2 diabetes within 10 years [1, 2]. Gestational diabetes increases the risk of stillbirth and infants born to women with gestational diabetes are more likely to be large-for-gestational age, more likely to become obese during childhood, and more likely to become diabetic during adulthood [1, 3].

Observational evidence suggests that a healthy lifestyle is associated with reduced risk of gestational diabetes in women with no history of the condition (here, a healthy lifestyle is defined as not smoking, maintaining a healthy body weight, consuming a healthy diet, and exercising regularly) [4].

Randomised controlled trial and observational evidence also suggests an active lifestyle is associated with a reduced risk of progression from gestational diabetes to type 2 diabetes (here, an active lifestyle is defined as taking part in at least 150 minutes per week of moderate-intensity physical activity or at least 75 minutes per week of vigorous-intensity physical activity)[5].

1.2 Trial aim

The aim of this study is to assess the effectiveness of a structured education programme to improve physical activity in participants who have had gestational diabetes during any pregnancy up to 60 months (post birth) prior to recruitment, with the overall intention of reducing their risk of developing type 2 diabetes.

2 Methods

2.1 Trial design

The trial is a parallel group randomised controlled trial, in which participants are randomised to two groups (1:1). It will compare a structured education programme plus additional follow on support (a wrist worn activity monitor and interactive web based intervention) with usual care plus an information leaflet. The RCT will be conducted in two centres; University Hospitals of Leicester NHS Trust and George Eliot Hospital NHS Trust. Participants will be followed up for 12 months.

2.2 Randomisation

Randomisation will be stratified by age (<30years; ≥30years) and by ethnicity (White European; other) and participants will be randomised to either the control or the intervention arm.

2.3 Sample size

The primary outcome is based on increasing total physical activity as quantified by the Euclidean norm minus one (ENMO) method measured in milligravity units (mg). This is the

main measure of activity derived from the GENEActiv monitor. In order to detect a minimum clinically significant difference of 2.1 mg, which is equivalent to an overall increase in physical activity volume of approximately 30 minutes of brisk walking at 4km/h, assuming a standard deviation of 5.3 mg [6], a power of 80% and significance level of 5%, the sample size requires 202 participants. To allow for 20% loss to follow-up and 10% non-compliance of the GENEActiv monitor, we will therefore need to recruit 290 participants (145 in each arm).

2.4 Timing of final analysis

Analyses described in this SAP will be performed following completion of the trial and database lock.

2.5 Timing of outcome assessments

Outcomes will be assessed at 12 months, with intermediate assessments at 6 months which will include questionnaires and wearing the physical activity accelerometer.

3 Statistical principles

3.1 Confidence intervals and p-values

The estimates of effect will be reported with 95% confidence intervals, for both primary and secondary outcomes. All p-values shown will be two sided.

3.2 Missing values and outliers

Missing values represent a potential source of bias in clinical trials. Therefore, every effort will be made to ensure the data is fully explored for potential problems and have been correctly entered into the database. Participants who did not attend the baseline assessment after consent will be excluded from all analyses, but will be noted on the CONSORT flow chart. For the primary and secondary analyses, participants who do not have an outcome measurement will be excluded; as discussed in section 3.3.1. Sensitivity analyses will be performed to assess the robustness of the findings to missing data (detail in section 5.3.2).

Outliers will be identified by examining graph plots such as normality plots, box plots and histograms. Values that visually appear outside of the main distribution (i.e. outliers) will be further investigated and if this is the correct observation it will remain in the analysis, however further assumptions will be made to assess possible influences on the results using a sensitivity analysis; by removing the outlier or outliers to compare results. Where results from the two analyses are discrepant, this fact will be reported and discussed in the study publication.

3.3 Analysis populations

3.3.1 Complete Case Population

The level of missing data for each outcome will be assessed. If missing outcome data is present the initial analysis will be based on the complete cases as we have allowed 30% for loss to follow-up and non-compliance to GENEActiv monitor. All tests of the effect of treatment on outcomes will be conducted on a complete case (CC) population. That is, all the recruited participants will be included in the analysis, with the exception of those with

missing outcome data. This will be done on a 'by analysis' basis, i.e. only those participants with missing data for variables required for a specific analysis will be removed.

3.3.2 Intention-to-treat Population /Full analysis data

A sensitivity analysis using multiple imputation will be used to impute missing data, to assess the robustness of the conclusions to missing data. This will be a full intention to treat analysis. More detail is in section 5.3.2.2.

3.3.3 Per-protocol Population

Sensitivity analysis will also include a per-protocol analysis. The per protocol (PP) population are those who were compliant with the protocol and have complete data for the analysis concerned on 'by analysis' basis. In the control arm, the PP population will include all participants randomised to that arm with complete outcome data. In the intervention arm, the PP population will be defined as participants who have attended at least one of the two group sessions of the programme i.e. those randomised to intervention but not attending either group session will be excluded and have outcome with complete data. This will be a sensitivity analysis.

4 Trial population

4.1 Eligibility criteria

Eligibility criteria are described in The BABY STEPS Study Protocol v5; 25/05/2018 [7].

4.2 Recruitment

A CONSORT chart will summarise the flow of participants through the trial.

4.3 Baseline characteristics

Descriptive characteristics at baseline will be summarised by treatment arm. Numbers (with percentages) for binary and categorical variables, and means (and standard deviations) or medians (with lower and upper quartiles) as appropriate for continuous variables will be presented. The number of missing values will be reported in the footnote of the summary table. There will be no tests of statistical significance nor confidence intervals for differences between randomised groups on any of the baseline variables.

5 Outcome measures

5.1 Primary outcome measure

The primary outcome measure is change in overall volume of physical activity measured using accelerometer (GENEActiv monitor) between baseline and 12 months.

5.2 Secondary outcome measures

Below is a list of all of the secondary outcomes collected at baseline and follow-up.

(i) Ambulatory activity:

Participants will be asked to wear an accelerometer for eight days to record total volume of physical activity, which includes light, moderate, vigorous and moderate to vigorous physical activities (MVPA) and to complete an activity log to provide wear time information. To be included in an analysis a participant needs to provide at least one valid day of accelerometer data.

(ii) Anthropometric measures:

- Body weight (kg)
- Waist circumference (cm)
- Hip circumference (cm)
- Body Mass Index (kg/m2)

(iii) Clinical measures:

- Blood pressure (mmHg)
- Resting heart rate (bpm)
- Total cholesterol (mmol/l)
- HDL cholesterol (mmol/l)
- LDL cholesterol (mmol/l)
- HbA1c (mmol/mol, %)

(iv) Questionnaire measures

A number of different questionnaire measures will be completed by participants at baseline and 12 months. Details of these questionnaire measures are as follows:

- Recent Physical Activity Questionnaire (RPAQ)[8] – this will assess habitual physical activity in several domains including occupation, leisure, housework
- Health Related quality of life (EQ-5D-5L)[9]
- 5-A-Day Consumption and Evaluation Tool (FACET)[10] – this will assess the participants' fruit and vegetable intake.
- Hospital Anxiety and Depression Scale (HADS14)[11] – this will assess the severity of symptoms of anxiety and depression.
- Jenkins Self-efficacy for Exercise Expectations Scale[12] – this will assess the ability to exercise.

5.3 Analysis methods

5.3.1 Analysis of primary outcome

The primary outcome is change in overall volume of physical activity (average acceleration, ENMO) from baseline to 12 months. For the primary outcome treatment arm will be compared using linear regression modelling with a binary indicator for randomisation group as the explanatory variable, terms for stratification factors (age and ethnicity) as confounders, and adjustment for the change in accelerometer wear time from baseline and overall volume of physical activity at baseline.

The primary analysis at 12 months will be based on complete data.

5.3.2 Sensitivity analysis

5.3.2.1 Per Protocol Population

For the PP analysis, participants who have engaged with at least one of the two group session of the programme will be included. The PP will adhere to the same steps as the primary analysis.

5.3.2.2 Intention-to-treat Population

To allow for full analysis set, missing data will be imputed using a multiple imputation procedure which substitutes predicted values from a regression equation. The imputation will be carried out by the MI command in Stata 15. MI replaces missing values with multiple sets of simulated values to complete the data, performs standard analysis on each completed dataset, and adjusts the obtained parameter estimates for missing-data uncertainty using Rubin's rules to combine estimates[13].

The following procedure will be followed:

• The MI will be set as wide
• The MI will register imputation of the overall volume of physical activity at 12 months and at baseline
• A regression method will be used , where the registered variables will be individually adjusted for the randomisation and stratification factors (age and ethnicity), with 100 imputations to avoid biased estimates, rseed (2259) and the force options
• MI estimate will provide the final results using the same regression model as the primary analysis and covariate adjustments.

5.3.2.3 Effects on the number of valid days

Three separate models will be run; including participants who have worn accelerometer for the following:

- At least 2 valid days
- At least 3 valid days
- At least 4 valid days

5.3.3 Analysis of secondary outcomes

The analyses of the secondary outcomes will be conducted in a similar manner as the main analysis using the appropriate model type, logistic regression for binary outcomes, linear for continuous and ordinal for ordinal outcome.

The assumptions of each analysis will be assessed and alternate parameterisations will be considered where appropriate.

5.3.4 Subgroup analyses for primary outcome

We will conduct the following subgroup analyses to assess if the intervention effect is statistically different between these groups.

- Age (<> median)
- Number of episodes of GDM (1 and ≥2)
- BMI categories (Normal, overweight, obese) – calculated according to ethnicity threshold.
- BMI pre-pregnancy – as above but subject to obtaining the data
- Ethnicity (White, South Asian, other)
- Parity (1 and ≥2)
- Sessions attended – attended at least one session by used App and did not use App in categories:

  (1) did not register to use App,
  (2) registered but used App <10 times over 12 months,
  (3) registered and used the App fairly frequently

The main analysis for each subgroup will involve using the same analytic principles as the primary analysis, but stratifying by the subgroup to estimate the mean difference in the overall volume of physical activity when compared to the control group. Further subgroup analyses will investigate the interaction effects between the treatment and subgroup, this will include an additional interaction term to the model; 'subgroup × treatment'. Subgroup results will be graphically presented as forest plots, with the estimated difference of overall volume of physical activity between treatment groups with 95% confidence interval for the interaction.

5.4 Safety reporting

The number (%) of individuals experiencing either an adverse event or a serious adverse event, whether or not attributed to the study, will be summarised by randomised group.

5.5 Statistical software

Analyses will be performed using Stata version 15.0 [14].

6 References

1. Fadl HE, Östlund IK, Magnuson AF, Hanson US. Maternal and neonatal outcomes and time trends of gestational diabetes mellitus in Sweden from 1991 to 2003. Diabetic Medicine. 2010 Apr;27(4):436-41.
2. Bellamy L, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. The Lancet. 2009 May 23;373(9677):1773-9.
3. Ben-Haroush A, Hadar E, Chen R, Hod M, Yogev Y. Maternal obesity is a major risk factor for large-for-gestational-infants in pregnancies complicated by gestational diabetes. Archives of gynecology and obstetrics. 2009 Apr 1;279(4):539-43.
4. Zhang C, Tobias DK, Chavarro JE, Bao W, Wang D, Ley SH, Hu FB. Adherence to healthy lifestyle and risk of gestational diabetes mellitus: prospective cohort study. Bmj. 2014 Sep 30;349:g5450.
5. Bao W, Tobias DK, Bowers K, Chavarro J, Vaag A, Grunnet LG, Strøm M, Mills J, Liu A, Kiely M, Zhang C. Physical activity and sedentary behaviors associated with risk of progression from gestational diabetes mellitus to type 2 diabetes mellitus: a prospective cohort study. JAMA internal medicine. 2014 Jul 1;174(7):1047-55.
6. Bell JA, Hamer M, van Hees VT, Singh-Manoux A, Kivimäki M, Sabia S. Healthy obesity and objective physical activity–. The American journal of clinical nutrition. 2015 Jul 8;102(2):268-75.
7. Sukumar N, Dallosso H, Saravanan P, Yates T, Telling C, Shorthose K, Northern A, Schreder S, Brough C, Gray LJ, Davies MJ. Baby Steps–a structured group education programme with accompanying mobile web application designed to promote physical activity in women with a history of gestational diabetes: study protocol for a randomised controlled trial. Trials. 2018 Dec;19(1):682.
8. Golubic R, May AM, Borch KB, Overvad K, Charles MA, Diaz MJ, Amiano P, Palli D, Valanou E, Vigl M, Franks PW. Validity of electronically administered Recent Physical Activity Questionnaire (RPAQ) in ten European countries. PloS one. 2014 Mar 25;9(3):e92829.
9. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of life research. 2011 Dec 1;20(10):1727-36.
10. Ashfield-Watt PA, Welch AA, Godward S, Bingham SA. Effect of a pilot community intervention on fruit and vegetable intakes: use of FACET (Five-a-day Community Evaluation Tool). Public Health Nutrition. 2007 Jul;10(7):671-80.

11. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale-a review of validation data and clinical results. Journal of psychosomatic research. 1997 Jan 1;42(1):17-41.
12. Resnick B, Jenkins LS. Testing the reliability and validity of the self-efficacy for exercise scale. Nursing research. 2000 May 1;49(3):154-9.
13. Rubin DB. Inference and missing data. Biometrika. 1976 Dec 1;63(3):581-92.
14. StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC