ASPIRE-DNA

Assessment of the Impact of a Personalised Nutrition Intervention in Impaired Glucose Regulation

Statistical analysis

Author: Pierre Debeaudrap

Version 3 of October 19th, 2018

Version	Date	Summary of changes
1	July 6th, 2018	Reviewed by CG
2	September 13th, 2018	Additional information on the variables measurement HbA1c included as endpoint (following change in the protocol) Section exploratory analysis (leptin and comparison of the exploratory arm) completed Various minor correction including editing
3	October 19th, 2018	Final version

Table of contents

Abbreviations

BMI	Body Mass Index
CRF	Case report form
DBP	Diastolic Blood pressure
DNA	Deoxyribonucleic acid
FFQ	Food Frequency Questionnaire
FPG	Fasting Plasma Glucose
HDL	High-density lipoprotein
HOMA	Homeostasis model assessment of beta cell function
IFT	Impaired fasting glucose
IGR	Impaired Glucose regulation
IGT	Impaired glucose tolerance
IQR	Inter quartile range
IRI	Insulin resistance index
ITT	Intention to treat
LDL	Low-density lipoprotein
LMM	Linear mixed model
LTFU	Loss to follow up
OGTT	Oral glucose tolerance test
PG120	Plasma glucose at 12 minutes
QQ plot	Quantile – quantile plot
REML	Restricted Maximum Likelihood
SNP	single nucleotide polymorphisms
SBP	Systolic blood pressure

Investigating team

Chief Investigator	Prof. Nick Oliver
Co-Investigator	Prof. Chris Toumazou
Co-Investigator	Dr. Maria Karvela
Trial Manager	Dr. Caroline Golden
Statistician	Dr. Pierre de Beaudrap
Dietitian	Ms. Natalie Bosnic
Project Officer	Mrs. Maria Eze
Research Nurse	Ms. Judith Bedzo-Nutakor
Laboratory Scientist	Dr. Sara de Mateo Lopez
Laboratory Scientist	Mr. Khalid Mirza
Laboratory Scientist	Ms. Francesca Cavallo
Operations Director, DnaNudge	Dr. Tsz Kin Hon

Signature for approval

Name	Role	Signature	Date
Author			
Dr Pierre Debeaudrap	Statistician		19/10/2018
Approved by			
Prof. Nick Oliver	Chief investigator		22/10/2018
Dr Caroline Golden	Trial manager		22/10/2018

Background

Diabetes is amongst the most common long term conditions, with the number of people affected worldwide quadrupling from 108 million in 1980 to 422 million in 2014[1]. Its prevalence in people over 18 years of age has risen from 4.7% in 1980 to a staggering 8.5% in 2014. In 2012, there were 1.5 million deaths as a direct result of diabetes, making it the 8th leading cause of death amongst both sexes, and the 5th leading cause of death amongst women. There were a further 2.2 million deaths as a result of complications due to higher-than-optimal glucose levels[2]. In 2013, 6% of the UK adult population (2.7 million people) were diabetic[3], 90% of whom had type 2 diabetes[4]. A further 5 million people were estimated to be at high risk of developing type 2 diabetes (NHS n.d.). This has led to a cost of £8 billion per year to the NHS, 80% of which is due to diabetes-related complications such as cardiovascular disease, amputations, renal failure and sight loss[5].

The potential for lifestyle interventions to prevent type 2 diabetes in high-risk patients has been well demonstrated through various clinical trials[6,7]. Yet there is important variation in the response between individuals who receive lifestyle interventions because of behavioural factors such as personal motivation and adherence but also because of genetic factors[8]. When personalised, the lifestyle interventions could achieve greater effectiveness by optimising and adapting the intervention components to individuals' goals and characteristics as well as enhancing motivation to adhere to dietary advice. There is a need to determine the most-cost effective personalised interventions and how they could be applied.

The ASPIRE study aims to assess the use of genetic data to guide personalised nutrition in people with impaired glucose tolerance and to demonstrate the impact of personalised nutrition on glucose metabolism, anthropometry, non-glucose metabolism and lifestyle.

In this trial, it is assumed that DNA-based dietary advice is more effective in improving glucose metabolism in pre-diabetic participants compared to the advice received as part of standard care for pre-diabetic individuals. A DNA-based diet is also assumed to improve anthropometric measurements, lipid profile and blood pressure in pre-diabetic participants. If participants from the intervention arm are found to have lower fasting plasma glucose (FPG) compared to patients from the control arm, this will provide evidence that DNA-based dietary advice is an effective way to prevent diabetes.

This document details the planned analyses and endpoints derivation for the ASPIRE-DNA study.

Study objectives

1.1 Primary objectives

- To compare differences in the impact of a DNA-based diet and standard care in improving glucose regulation in pre-diabetic individuals

1.2 Secondary objectives

- To assess the ability of a DNA-based diet to improve the macro- and micro-nutrient profile of pre-diabetic participants.
- To assess the effect of a DNA-based diet on clinical markers including cholesterol, body composition and blood pressure, as they relate to diabetes risk.
- Where applicable, to assess changes in anthropometric measurements as a result of following a DNA-based diet, where they relate to diabetic risk reduction.

1.3 Exploratory objectives

- To assess the impact of providing DNA-based dietary guidelines via the DnaNudge App, on improving glucose regulation in pre-diabetic individuals
- To assess the impact of providing DNA-based dietary guidelines via the DnaNudge App, on the secondary outcomes outlined above.
- To explore the utility of an app as a delivery mechanism for DNA-based dietary advice.
- To examine the changes in leptin measurements via saliva samples during the study.

Study design

Open label randomised control trial. Participants will be randomised 1:1:1 to a) the control arm (n = 60) in which they will receive general health guidelines according to the NICE guidelines, b) the intervention arm (n = 60) in which they will receive DNA-based health guidelines via a genetic report, or to c) the exploratory arm (n = 60) in which they will receive DNA-based guidelines via the DnaNudge App.

The primary analysis will focus on the comparison between intervention arm and control arm.

Definitions

Impaired glucose regulation: impaired fasting glucose (IFG) or impaired glucose tolerance (IGT).

Impaired fasting glucose:
6.1mmol/L ≤ Fasting plasma glucose (FPG) ≤ 6.9mmol/L AND 2 hr plasma glucose: < 7.8 mmol/L

Impaired glucose tolerance:

- FPG ≥7mmol/L AND 7.8mmol/L ≤ 2-hour plasma glucose < 11mmol/L after glucose load
- HbA1c between 42 and 47 mmol/mol (6.0–6.4%)

Endpoints

1.4 Primary endpoint

The primary endpoint for the ASPIRE trial is the difference in FPG at week 6 between the intervention arm that receives DNA-based diet advice and the control arm. This outcome has been chosen because it is the most sensitive indicator of impaired glucose regulation. Endpoints assessment will not be blind to the study intervention arm allocated.

1.5 Secondary endpoints

Secondary outcomes include differences at weeks 6, 12 and 26 between arms in the cross sectional values and changes from baseline of the variables listed below.

- Variables related to glucose regulation:
  - FPG at other time points
  - Plasma glucose at 120 minutes (P120) following 75g OGTT
  - Insulin resistance index (IRI) computed using the Homeostasis model assessment (HOMA) of beta cell function
  - c-peptide at 120 minutes following 75g OGTT
  - HbA1c
- Variables related to anthropometric measurements
  - Body weight
  - Body Mass Index (BMI)
  - Lean mass
  - Fat mass
  - Waist circumference
- Variables related to lipid profile
  - Total cholesterol
  - Fasting triglycerides
  - Low-density lipoprotein (LDL) cholesterol
  - High-density lipoprotein (HDL) cholesterol
- Variables related to clinical markers
  - Systolic Blood pressure (SBP)
  - Diastolic blood pressure (DBP)
- Dietary intake
  - Total energy intake
  - Carbohydrate intake
  - Fat intake
  - Saturated fat intake
  - Salt intake

- Vitamin D
- Vitamin B6
- Vitamin B12

Study population

The study participants will be adults with impaired glucose regulation and fulfilling the following inclusion / exclusion criteria:

- Inclusion criteria
  - Provision of signed and dated informed consent form
  - Adults over 18 years of age
  - Impaired glucose regulation including IFG and IGT by FPG, OGTT or HbA1c criteria
  - Access to smartphone with an operating system of iOS 8.0 or above, or Android 4.0 or above.
- Exclusion criteria
  - Diabetic
  - Pregnant or planning pregnancy
  - Breastfeeding
  - Enrolled in other clinical trials
  - Have active malignancy or under investigation for malignancy
  - Severe visual impairment
  - Reduced manual dexterity
  - Use of psychiatric, anti-diabetic, and/or weight loss medication, and/or oral steroids
  - Bariatric surgery
  - History of illnesses that could interfere with the interpretation of the study results (e.g. HIV, Cushing syndrome, chronic kidney disease, chronic liver disease, hyperthyroidism, hereditary fructose intolerance, alcohol or substance abuse)
  - Unable to participate due to other factors, as assessed by the Chief Investigator

Statistical considerations

1.6 Analysis population

- Intention to treat population (ITT)

Primary analyses will be conducted following the intention to treat (ITT) principle. All participants who are randomised into the study will be included in the ITT analysis and the analysis is conducted according to the randomised treatment arm.

- Per protocol population

In addition, the primary endpoint will be analysed with the Per Protocol population (PPP), which consists of those subjects in the ITT population who complete the study with no significant deviations from the planned protocol procedures. The exclusion will be as follows:

- Participants without impaired glucose regulation before inclusion [see definition page 6]
- Significant protocol deviation such as non respect of inclusion / exclusion criteria
- Pregnancy during follow-up
- Participants loss to follow-up or who withdrew consent

1.7 Data description

Data will be collected as described in protocol version 0.18 of 16/07/2018 and managed as described in the data management system version 0.5 of 01/10/2018. Data will be collected in electronic case report form (eCRF) completed by investigators. All data will be on the virtual machine for the medical statistician to work with, the data will be in the form of data extractions (i.e. not the CRFs). If the original CRFs are needed, they will be provided in a read only format.

1.7.1 Recruitment and follow-up

Date of first and last study inclusion and of the last visit will be reported. A flow chart will be used to describe participants' flow. It will include the following information: number of persons screened, eligible, included, assessed at week 6 visit (see Figure 1). In addition, the number of participants screened, recruited, followed until study completion, lost to follow-up and who were withdrawn will be presented by month and intervention group in a table.

1.7.2 Data summary

Collected data will be summarised without consideration for treatment group as follows:

- Quantitative data: maximum and minimum values, median and interquartile range (IQR);
- Qualitative data: count (n) and proportion (%);

In addition, for each variable, the proportion of missing data will be reported. This information will be reported using the following output table (Table 1 and 2).

Figure 1. Example of flowchart

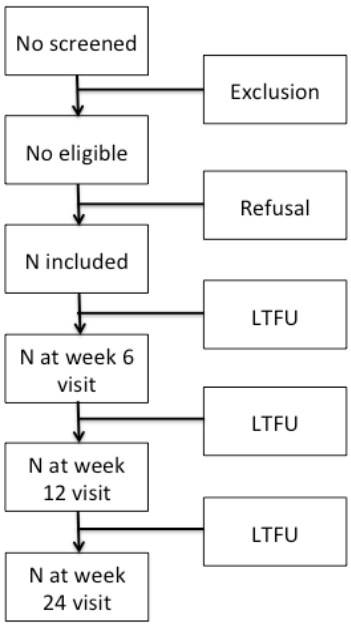

Table 1. Example of output table for main continuous variables

Variable	Label	N	Missing	Min	Max	P25	P75	median

Table 2. Example of output table for main categorical variables

Variable	Label	N	Missing	n	%	Cumulative	
						n	%

1.7.3 Data review and exploration of outliers and problematic distribution

The distribution of the main variables (e.g., FPG, triglyceride, cholesterol) will be assessed using published reference data. These distributions will be compared against a normal distribution using a quantile – quantile (QQ) plot. Skewness and kurtosis will be computed and additional assessment of normality may be conducted using the Kolmogorov-Smirnov test. Boxplot will be used to detect outliers.

Variables with outliers, missingness and/or problematic distribution will be reviewed with the some/all (where applicable) of the following members of the Trial Team:

- Ms. Natalie Bosnic (Dietitian)
- Ms. Judith Bedzo-Nutakor (Research Nurse)
- Mrs. Maria Eze (Project Officer)
- Dr. Caroline Golden (Trial Manager)
- Prof. Nick Oliver (Chief Investigator)

1.8 Baseline characteristics and description of the study population

Table 1 below lists the baseline variables of interest and presents the methods of derivation. The baseline variables will be summarised for each treatment arm using the mean with standard deviation (sd) for continuous data and count (n) with proportion (%) for categorical data. Statistical testing of between group imbalance in the baseline covariates will not be performed unless required for regulatory or publication reasons[9]. A tabular presentation will be used to display the results.

Table 1. Baseline variables

Description	**Definition**	**Type**	**Unit**	**Coding /derivation**	**CRF**
Age		Num	Year	Number of years from date of birth (months : year)	Anonymised Randomisation log
Sex			N (%)		Screening enrolment log
HbA1c		Num	Mmol/mol or %		Visit2-CRF
Randomisation group	Intervention group assigned at randomization				Anonymised Randomisation log
Height			m		Visit3-CRF
Weight			Kg		Visit3-CRF
Body mass index (BMI)			Kg/m^2		Visit3-CRF
BMI>30			N (%)		Visit3-CRF
Fat body mass			Kg		Visit3-CRF
Lean body mass			Kg		Visit3-CRF
Waist circumference			cm		Visit3-CRF
Systolic BP			mmHg		Visit3-CRF
Disatolic BP			mmHg		Visit3-CRF

Pulse			bpm		Visit3-CRF
Glucose t0	Plasma glucose at time 0 after 75g glucose load	Num	mM/L		Visit3-CRF
Glucose t120	Plasma glucose at 120' after 75g glucose load	Num	mM/L		Visit3-CRF
Insulin t120	Plasma Insulin at 120' after 75g glucose load		IU/L		Visit3-CRF
C-peptide t0	C-peptide at t0 after 75g glucose load		mM/L		Visit3-CRF
Total cholesterol			mM/L		Visit3-CRF
HDL-cholesterol			mM/L		Visit3-CRF
Non-HDL cholesterol			mM/L		Visit3-CRF
LDL cholesterol			mM/L		Visit3-CRF
Fasting TG			mM/L		Visit3-CRF

1.9 Analyses of quantitative data

Continuous outcomes will be compared between groups using the modified analysis of covariance (ANCOVA)[10]. This approach has been chosen because it is more powerful than other alternatives such as t-test[11].

A linear regression model will be used to analyse the association between post-intervention values of the outcomes with intervention arm adjusting for their baseline value, participant age and sex.

Let y_{it} denote the outcome value for individual i at time t, x_i her covariates (e.g. sex, age) and d_i her intervention arm. The ANCOVA model is given by:

$$y_{it} = \beta_0 + \beta_1 y_{i0} + \beta_2 x_i + \gamma d_i + \varepsilon_{it}$$

An extension of the basic ANCOVA model will consider an interaction between the intervention indicator and the baseline values to test the assumption that the treatment effect varies with baseline status.

In addition, changes in continuous outcomes from baseline will be analysed using a linear regression model of the difference between the outcome value at visit and its baseline value. Using the same notation, the change model is

$$\Delta y_{it} = y_{it} - y_{i0} = \beta_0 + \beta_1 x_i + \gamma d_i + \varepsilon_{it}$$

Covariates will be centered in the analyses. Robust linear regression *lmrob* from the R package robustbase will be used for the ANCOVA analysis[12] unless specified otherwise. It uses the MM-estimator to estimate the regression coefficient accounting for departure from normality of the outcomes.

Assessment of models' goodness of fit will include careful examination of residuals using graphical methods (scatter plot, boxplot and QQ plot) and detection of outliers. In addition, equality of the post-treatment variances will be tested using Levene's test.

1.10 Repeated measures analyses

Linear mixed model (LMM) will be used to analyse changes over time of continuous outcomes. Using the same notation (y_{it} for the outcome value for individual i at time t, x_i the covariates, and d_i for the intervention arm), the following model will be used to model the change in y_{it}

$$y_{it} = \beta_{0i} + \beta_1 t_{it} + \beta_2 X_i + \gamma_0 d_i + \eta_i + \varepsilon_{it}$$

LMM will be estimated using the REML method[13].

This model can be extended to account for non linearity in the time trend and for an interaction between time trend and intervention arm as follows:

$$y_{it} = \beta_{0i} + \beta_1 t_{it} + \beta_2 X_i + \gamma_0 d_i + \gamma_1 d_i t_{it} + \eta_i + \varepsilon_{it}$$

The assessment of the model's goodness of fit will be performed through graphical examination of the within-group residuals and random effects. It will include scatter plots of standardised residuals against fitted values as well as QQ plot and boxplot of standardised residuals and of random effects[14]. The homoscedastic normal errors assumptions will be tested using the Shapiro-Wilks test and White's test of homoscedasticity of error variances.

1.11 Non-parametric model

For variables showing strong departure from normality, the non-parametric Mann Whitney Wilcoxon rank sum test (two groups) or the Kruskal and Wallis (to compare the three arms) will be used.

1.12 Missing data

Missing data will be imputed using the multiple imputation with chained equation (MICE) method using the other variables including outcomes variables if there is no evidence against the assumption of data missing at random[15][17]. Missing at random refers to the fact that all factors related to the probability of a data being missing have been observed (they are missing completely at random meaning the missingness mechanism is completely stochastic). The cause for missing data will be carefully discussed with the investigator team to determine whether the missing at random assumption holds. The prediction of missing values will be performed after specifying a model from the variables distribution (e.g. linear regression model for normal distribution, logistic model for binary regression…). MICE will be performed using the R package mice with at least five imputed datasets[16].

Analysis of the primary endpoint

1.13 Derivation

Measurement. Under the protocol, participants are scheduled to have an OGTT at week 6 (visit 6, D42+/- 3 days). Plasma glucose at 0 minute will be measured after 12 hours fasting and before the 75g glucose load via an OGTT.

Data transformation. Available data from Arkadianos et al. suggests that PFG may not be normally distributed (see Annex)[17]. In the study by Arkadianos, post intervention PFG showed significant departure from normality with significant skewness (0.8, Agostinos test; p = 0.04). By contrast, no significant departure from normality was observed for PFG changes from baseline (Annex). Although the log transformation has been suggested in the literature, this transformation applied to Arkadianos data did not improve the normality of PFG distribution. Using the box-cox method, the best transformation was achieved with the boundary value of -2.

For this analysis, the preferred approach to handle departures from normality and outliers will be to use of robust statistical methods[18].

1.14 Planned analysis

The primary analysis will be conducted under the ITT principle using modified ANCOVA as described above. Raw and adjusted mean differences in PFG at week 6 between the intervention and control arm will be presented with 95% confidence interval (95% CI). The adjusted mean difference will be estimated using a linear regression model of week 6 FPG to baseline PFG and treatment arm adjusted for age and sex.

The primary null hypothesis is of no difference in week 6 FPG adjusted means between the intervention and the control arms. It will be tested against the alternative hypothesis of a difference between the two groups.

H_0: $\mu_0 = \mu_1$

H_A: $\mu_0 \neq \mu_1$

Where μ_0 and μ_1 are the week adjusted trimmed mean FPG in respectively the control and intervention group. The null hypothesis will be rejected if the robust adjusted 95% CI of the mean difference does not include 0.

- Power of the analysis

With 60 participants in each group and assuming that the average PFG changes from baseline to the week 6 will be -0.6 mmol/l in the intervention arm and -0.1 mmol/l in the control arm, that the baseline standard deviation (sd) of PFG is 0.7, the post-intervention sd is 0.6, and that the correlation between baseline and post-intervention measurement range between 0.6 and 0.4, the analysis will have a power greater than 0.9 to detect a difference between both arms.

- Sub-group and additional analysis of the primary outcome

An interaction between baseline FPG and treatment effect will be tested.

No sub-group analysis is planned for the primary endpoint.

- Sensitivity analysis

The robustness of the results will be tested using a non-parametric model as described above. Both results will be presented and discrepancies will be discussed.

Secondary endpoints

The same analyses (described in the method section and detailed for week 6 FPG above) will be computed for all of the following continuous variables. Therefore, only additional specific analyses are described in the section below.

1.15 Variables related to glucose regulation:

Participants are scheduled to have an OGTT at baseline (visit 3), weeks 6 (visit 6), week 12 (visit 8) and week 26 (visit 10).

- Fasting Plasma glucose
  - Derivation: plasma glucose after a 12 hours fast and before the 75g glucose load will be measured via an OGTT.
  - Analysis: modified ANCOVA and LMM; no additional analysis.
- Plasma glucose at 120 minutes following 75g OGTT (G120)
  - Derivation: Plasma G120 minutes after 75g glucose load will be measured via an OGTT.
  - Analysis: modified ANCOVA and LMM; no additional analysis.
- HbA1c will be measured by the Pathology Department in HammersmithHospital (Imperial College Healthcare NHS Trust).
  - Analysis: modified ANCOVA and LMM; no additional analysis.
- Insulin resistance index computed using the Homeostasis model assessment (HOMA) of beta cell function

- Derivation: the insulin resistance index (IRI) will be derived using the HOMA mathematical model[19]. It requires the FPG and fasting insulin concentration and is computed as follows:

$$IR = \frac{glucose \times insulin}{22.5}$$

where glucose stand for FPG in mmol/l and insulin for fasting insulin in mU/l.

- Analysis: modified ANCOVA and LMM; no additional analysis.

- C-peptide at 120 minutes following 75g OGTT will be measured by the Pathology Department in Hammersmith Hospital (Imperial College Healthcare NHS Trust).
  - Analysis: modified ANCOVA and LMM; no additional analysis.

1.16 Variables related to anthropometric measurements

Patients are scheduled to have anthropometric measurements at baseline and visits 4, 6, 8 and 10 (weeks 6, 12 and 26).

- Body weight
  - Derivation: weight will be recorded in light clothing without shoes to the nearest kilogram using the same scale during the trial
  - Analysis:
    - Main analysis: modified ANCOVA and LMM.
    - In addition, changes in weight over time will be analysed using a LMM model first with measurements taken at weeks 4, 6, 12 and 26.

- Body Mass Index (BMI)
  - Derivation: height will be measured bare foot to the nearest 1 cm using a stadiometer. BMI will be computed as

$$BMI = \frac{weight\,(kg)}{height^2\,(m)}$$

  - Analysis:
    - Main analysis: modified ANCOVA and LMM.
    - Changes in BMI over time will be analysed using a LMM model first with only weight measurements taken at weeks 6, 12 and 26 then with all data including weight reported by phones

- Lean mass

- Derivation: lean mass will be measured using Bioelectrical Impedance Analysis with the Body Composition Analyser Tanita BC-418
- Analysis: modified ANCOVA and LMM; no additional analysis.

- Fat mass
  - Derivation: fat mass will be measured using Bioelectrical Impedance Analysis with the Body Composition Analyser Tanita BC-418.
  - Analysis: modified ANCOVA and LMM; no additional analysis.

- Waist circumference
  - Derivation: waist circumference will be measured with to the nearest 0.5 cm at the midpoint of lower border of the costal margin and uppermost border of iliac crest using a non-stretchable measuring tape. Three consecutive measurements will be recorded.
  - Analysis: modified ANCOVA and LMM; no additional analysis.

1.17 Variables related to lipid profile

- Derivation: Total and HDL cholesterol and triglycerides will be measured by the Pathology Department in Hammersmith Hospital (Imperial College Healthcare NHS Trust). LDL- cholesterol will be computed using the Friedwald relation[20]:

$$LDL\ cholesterol = Total\ cholesterol - HDL\ cholesterol + \frac{triglyceride}{2.2}$$

- Analysis: modified ANCOVA and LMM; no additional analysis.

1.18 Variables related to clinical markers

- Systolic and diastolic blood pressure (BP)
  - Derivation: the mean of two measurements of the blood pressure with a digital sphygmomanometer. Blood pressure will be measured after at least 5' rest.
  - Analysis: the analysis will be conducted separately for systolic and diastolic BP using modified ANCOVA and LMM as described above; no additional analysis.

1.19 Variables related to nutrition

- Dietary intakes

- Derivation: dietary intakes (Total energy, carbohydrate, fat, saturated fat, salt and vitamin B6, B12, D) will be assessed using 24-hours recall questionnaire (food frequency questionnaire [FFQ]) at visits 4, 5, 7, 9 and 11. The FETA software will be used to derive average daily nutrition contents from the FFQ data. In addition, 25-hydroxyvitamin D and vitamin B6 will be measured directly by the Pathology Department in Hammersmith Hospital (Imperial College Healthcare NHS Trust).
- Analysis: the analysis will be conducted separately for each nutrient using modified ANCOVA and LMM as described above;
- Additional analysis: additional dietary intakes may be examined.

Exploratory analyses

The first pre-defined exploratory analysis will compare the exploratory arm with the two other arms. More specifically it will test the two following hypotheses: The exploratory arm confers similar or greater benefit compared to the interventional arm and greater benefit compared to the control arm.

The second hypothesis generating exploratory analysis will examine how the changes in leptin measurements correlate with changes in weight overall and by intervention arm. Leptin will be measured via saliva samples at weeks 0, 6, 12 and 26.

The last exploratory analysis will examine how outcomes in the exploratory arm vary across adherence levels. Adherence levels will be defined from the recorded activity of the application.

These analyses are exploratory in nature hence results will be interpreted cautiously.

Additional information

1.20 Significance level

All analyses will be conducted using a 5% significance level. No adjustment for multiplicity will be used.

1.21 Statistical software

Data management will be performed using the statistical software Stata version 12.1 (LP StataCorp). Statistical analyses will be performed with the statistical software R (R

Foundation for Statistical Computing). The R packages WRS2, robustbase, lme4 and robustlmm will be used.

References

1. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS medicine* 2006; **3**(11): e442.
2. WHO. Global report on diabetes. 2016. http://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf;sequence=1.
3. Gatineau M, Hancock C, Holman N, et al. Adult obesity and type 2 diabetes. 2014. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/338934/Adult_obesity_and_type_2_diabetes_.pdf (accessed June 14th, 2018.
4. Diabetes UK. Diabetes: facts and stats. 2014. https://www.diabetes.org.uk/resources-s3/2017-11/diabetes-key-stats-guidelines-april2014.pdf (accessed June 14th, 2018.
5. Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabetic medicine : a journal of the British Diabetic Association* 2012; **29**(7): 855-62.
6. Dunkley AJ, Bodicoat DH, Greaves CJ, et al. Diabetes prevention in the real world: effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes and of the impact of adherence to guideline recommendations: a systematic review and meta-analysis. *Diabetes care* 2014; **37**(4): 922-33.
7. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The New England journal of medicine* 2002; **346**(6): 393-403.
8. Franks PW, Poveda A. Lifestyle and precision diabetes medicine: will genomics help optimise the prediction, prevention and treatment of type 2 diabetes through lifestyle therapy? *Diabetologia* 2017; **60**(5): 784-92.
9. Senn S. Testing for baseline balance in clinical trials. *Statistics in Medicine* 1994; **13**(17): 1715-26.
10. Yang L, Tsiatis A. Efficiency study of estimators for a treatment effect in a pretest – posttest trial. *The American Statistician* 2001; **55**(5): 314-21.
11. O'Connell N, Dai L, Jiang Y, et al. Methods for analysis of pre-post data in clinical research: a comparison of five common methods. . *J Biol Biostat* 2017; **8**: 334.
12. Koller. Simulations for Sharpening Wald-type Inference in Robust Regression for Small Samples. 2018. https://cran.r-project.org/web/packages/robustbase/vignettes/lmrob_simulation.pdf.

13. Laird N, Ware J. Random-Effects Models for Longitudinal Data Biometrics 1982;38:963-974. *Biometrics* 1982; **38**(4): 963-74.

14. Bates JCPDM, Pinheiro JC, Pinheiro J, Bates D. Mixed-Effects Models in S and S-PLUS: Springer New York; 2000.

15. Rubin DB. Multiple Imputation for Nonresponse in Surveys: Wiley; 2009.

16. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011; **45**(3): 1-67.

17. Arkadianos I, Valdes AM, Marinos E, Florou A, Gill RD, Grimaldi KA. Improved weight management using genetic information to personalize a calorie controlled diet. *Nutrition journal* 2007; **6**: 29.

18. Wilcox RR. Introduction to Robust Estimation and Hypothesis Testing: Elsevier Science; 2016.

19. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 1985; **28**(7): 412-9.

20. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clinical chemistry* 1972; **18**(6): 499-502.