Daniel Stahl

Department of Biostatistics, King's College London

Analysis plan of MOVE-IT

TABLE OF CONTENTS

1. Brief description of the trial 3
1.1 Research Objectives 3
1.2 What is the trial design? 4
1.3 Randomization 4
1.4. Setting 5
1.5 Target population 5
1.6 Eligibility screening 5
1.7. Baseline measures 6
1.7. Duration of treatment period 8
1.8. Sample size and power calculations 9
1.9. The original proposed statistical analyses of main outcome variables 10
2. Data Analysis Plan: Data description 11
2.1 **Description of available data** 11
2.2 **Recruitment: GP and Participant Flow** 11
2.3. Baseline comparability of randomised groups 13
2.4. Adherence to allocated treatment 13
2.5. Adherence to the protocol 13
2.6 **Loss to follow up and other missing data** 14
2.7 Adverse event reporting 16
3. Statistical Data Analyses Plan: Comparison of body weight and physical activity between intervention groups (Primary outcome analysis) 17
3.1. General comments 17
3.2. Primary outcome "Physical Activity" 18
3.3. Overview of analyses of primary outcome: 19
3.3.1. Primary Intention to treat analysis 19
3.3.2. Sensitivity of results due to potential imbalance of confounders at baseline 24
3.3.2. Sensitivity of results due to delay of treatment Session start: DIFFERENCES Between randomization and session 0 24
3.3.4. Sensitivity of results due to UNBLINDING of researchers at follow-up 25
3.3.5. Sensitivity of results due to insufficent wearable recordings at baseline 25
3.3.6. Sensitivity of results due to initials errors of calculating QRISK2 scores at baseline and violation of study protocol 25
3.3.7. Sensitivity of results due to other potentail violations of the protocol 26
3.3.8: Sensitivity of results to missing data 26
Analysis 8a: Sensitivity of results due to missing data by including covariates predictive of missingness in analyses model 26

Analysis 8b: Sensitivity of results due to missing data by using multiple imputations 26

Analysis 8c: Sensitivity of results due to missing data by assuming missing not at random 27

3.4. Secondary outcome analyses 28

3.5. Complier average causal effects (CACE) 29

3.6. Adverse EVENTS (AE) 30

3.7. Compliance 31

Reference List 32

1. BRIEF DESCRIPTION OF THE TRIAL

1.1 RESEARCH OBJECTIVES

Main objectives

The overall aim is to compare in a randomized controlled trial the effectiveness and cost effectiveness of two formats of a motivational interviewing intervention, group and individual, with usual care, in increasing healthy dietary habits and increasing physical activity in people at high risk of cardiovascular disease over a 24 months follow up.

Primary objective:

To examine whether group motivational interviewing (G-MI+) delivered by healthy lifestyle facilitators is more effective than usual care in increasing healthy dietary habits and **increasing physical activity (using an accelerometer)** 24 months later.

Secondary objectives:

i: To examine whether individual motivational interviewing (I-MI+) delivered by healthy lifestyle facilitators is more effective than usual care in increasing healthy dietary habits or increased physical activity 24 months later (data will be analysed together with the Primary objective)

b) Further secondary hypotheses

ii: to examine whether group motivational interviewing (G-MI+) is more cost-effective than individual motivational interviewing, and usual care, in terms of **quality-adjusted life years** gained over the 24 month follow-up period

iii; to examine whether individual and group motivational interviewing delivered by healthy lifestyle facilitators are more effective than usual care in reducing the **cardiovascular disease risk score**.

1.2 WHAT IS THE TRIAL DESIGN?

This is a 3 parallel arm multi-centre randomised controlled trial for people screened as at high risk for CVD.

The three arms consist of:

1. **Usual treatment:** Standard care as recommended by NICE and adapted for the local population.

2. **G-MI+:** Group delivered motivational interviewing intervention,

3. **I-MI+:** Individual motivational interviewing intervention

Data of main outcomes will be collected at baseline before randomization and at 12 months and at 24 months after baseline measurements (and not after randomization because randomization may be delayed and certain measurements may change on time of year, or holiday season). As participants of the group motivational interviewing arm, but not in the other two arms, are clustered within groups we have a partially clustered (or nested) design. To account for the possible correlation within the G-MI+-arm sample size will be increased in this arm by 25% and the number of patients will differ between arms.

1.3 RANDOMIZATION

Originally we proposed to stratify randomisation by general practice and by ethnicity (white and African/Caribbean/ south Asian/Chinese) within practice. However, this method was not feasible because we the number of patients in many surgeries is not large enough to stratify by both factors. Because of the large sample size we decided after consultation and approval of the TSC to perform simple randomization and to assess the effect of possible

imbalance between the arms in a sensitivity analyses by including practice and ethnicity as factors in the analyses model.

Randomization will be performed by block randomization with blocks of sizes of 10. In each block 10 subjects will be randomized to G-MI+, I-MI+ and TAU in a 4:3:3 ratios. The unequal allocation ratio of 40%:30%:30% ensures that G-MI+ arm will have approximately 25% more patients. (The exact ratio based on the power analyses is 0.38:0.31:0.31).

1.4. SETTING

At the time of submission for funding the South London Health Innovation and Education Cluster (HIEC) was used as the setting. We have chosen the HIEC because it includes 12 CCGs across South London (Bexley, Bromley, Croydon, Greenwich, Kingston, Lambeth, Lewisham, Richmond & Twickenham, Southwark, Sutton, Merton, Wandsworth) that are linked to each other by an educational and training infrastructure and inherent in this infrastructure is an efficient method for recruitment.

1.5 TARGET POPULATION

The case definition includes adults age 40-74 years who screen positive for high CVD risk (Qrisk2 score ≥20% over the next ten years) on the NHS health check or identified by other cardiovascular screening methods means and not known to have cardiovascular disease or to be on the diabetes, kidney, atrial fibrillation or stroke register.

1.6 ELIGIBILITY SCREENING

INCLUSION CRITERIA

Participants are eligible if aged between 40-74 years old, fluent in conversational English, permanent residents and planning to stay in UK for at least ¾ of year and at high cardiovascular risk according to GP records (QRisk2 score ≥ 20%).

EXCLUSION CRITERIA

The revised exclusion criteria are: Established cardiovascular disease including congenital heart disease, angina, myocardial infarction, coronary revascularization procedures, peripheral artery disease, CABG or angioplasty; has a pacemaker; on a register for diabetes, kidney disease, arterial fibrillation or stroke (either ischemic or haemorrhagic,

including transient ischemic attacks); disabling COPD; disabling neurological disorder (MS/ Parkinson's); severe mental illness such as psychosis, learning disability, dementia and cognitive impairment; registered blind; housebound or resident in nursing home; unable to move about independently or not ambulatory; > 3 falls in past year; pregnancy, advanced cancer and morbid obesity BMI >50 kg/m2. When in doubt we will seek the GP opinion and approval.

SELECTION OF PARTICIPANTS

The sampling frame will be GP practices with list sizes greater than 5,000 patients. This represents around a quarter to third of all practices in the HIEC. To recruit patients from every practice in the HIEC is not cost beneficial, as smaller practices are less likely to be operational with the Health Checks and to have fewer patients to recruit from. Unless there is a Health Checks register, patients with CVD score >20% risk will be identified by screening the databases using a search strategy based on a range of terms including READ codes which will be validated and tested for test-retest reliability before inception. The GP or the health care provider will invite those who have a CVD risk score >20% to participate in the study and after the patient has given permission to be contacted a researcher will invite the patient into the study.

PATIENTS FROM SAME HOUSEHOLD

To avoid a contamination of treatment of persons who are living in the same household only one person per household will be randomized to a treatment. If 2 or more persons from one household are recruited at the same time only one person will be randomly selected. The patient ID numbers will be sent to an independent researcher who will use a random number generator to select one person for randomization and inform the researcher. If a patient of a household was randomized to a treatment no other household members will be considered for recruitment for the duration of the study.

1.7. BASELINE MEASURES

This will be collected prior to randomisation.

SOCIODEMOGRAPHIC

Data on age, gender, self-report ethnicity, occupational status, educational attainment, marital status and literacy will be collected.

BIOMEDICAL

Weight, height, body mass index, waist circumference, lipids, HbA1c and fasting glucose will be measured. Weight will be measured in light clothing, without shoes on the Class 3 Tanita SC240 weighing digital scale to 0.01 kg for weight and body fat composition. Height will be measured to 0.1 cm using SECA stadiometers with the supported stretch stature method. These will be used to calculate body mass index (weight/height^2, kg/m^2). Waist circumference will be measured horizontally halfway between the lowest rib and the upper prominence of the pelvis using a non-extensible steel tape against the bare abdomen. Blood pressure and resting heart rate will be measured with the Omron M7 digital BP monitors using standardised procedures of the average of 3 readings 1 minute apart while seated. The QRISK2 score will be the research measure of CVD risk (Bayley et al., 2015).

LIFESTYLE

i) Smoking status: If current how many cigarettes/day, ex-smoker (for how many years) and never smoked.

ii) Alcohol intake will be measured using the Alcohol Use Disorders Identification Test.

iii) Physical activity will be measured objectively using the Actigraph GT3X accelerometer, a tri-axial sensor movement sensor that also records step counts. The Actigraph instrument has been validated in comparison with doubly labelled water. The participant and the outcome assessor are blind to the readings. Participants will be given verbal and written instructions: how to attach the accelerometer, to remove when sleeping or when it might get wet or during contact sports, to wear the accelerometer for seven consecutive days, and keep to their normal routine. Participants will be included in all analyses if they have completed at least t**hree? four weekdays** of accelerometer monitoring. We will ask participants to keep a log of activities including sedentary ones to assist with the qualitative interpretation of the data.

iv) Dietary intake will be assessed using a standardized multiple-pass 24-hour dietary recall as it can be more objective and more reliable as a measure of change in intervention studies; data will be entered and nutrient intakes obtained using a standard validated computerized interview software programme. Researchers will be trained to follow a standardized protocol, ask neutral probing questions to encourage recall of food items, and taught about different methods of food preparations and brands in different cultures. Portion size will be assessed

food photographs to estimate daily calorie intake. Total and non-HDL cholesterol will be measured as a proxy biomarker of change in dietary fat intake (Bayley et al., 2015).

PSYCHOLOGICAL

i) Health beliefs about diet, exercise, perceptions of risk for developing CVD and related conditions will be measured by the Brief Illness Perception Questionnaire adapted for perception of risk.(Broadbent *et al.*, 2006)(Bayley et al., 2015) Self-efficacy measures for diet and physical activity will be included as psychological processes we are seeking to change during the intervention.

ii) Depressive symptoms using the 9-item Patient Health Questionnaire as depression is associated with worse outcomes in CVD.

Research workers will be trained in conducting standardised anthropometric measures with repeated training and supervision at quarterly intervals to ensure reliability. They will be trained by a dietician in the coding of different cultural foods and neutral probing techniques for improving recall. All equipment will be calibrated daily.

FOLLOW-UP MEASURES

Updated sociodemographic, biomedical, lifestyle and psychological data will be collected at 12 and 24 months follow up after baseline by an independent researcher who is blind to the treatment condition

1.7. DURATION OF TREATMENT PERIOD

The programme will consist of 11 sessions spread quarterly over 12 months. Each participant allocated to a treatment arm will have a session 0 as an introduction to the intervention and to become familiar with the healthy lifestyle facilitator but also to receive their workbooks, MEND packs and pedometer. Then the intensive phase will consist of 6 weekly sessions at the beginning of the first quarter. The first 3 sessions will focus on physical activity and the second 3 sessions on diet. The maintenance phase will consist of 4 sessions delivered at 3, 6, 9 and 12 months. Each group will have a maximum of 10-11 participants and sessions will last 120 minutes. Sessions for those participants allocated to I-MI+ will last 40 minutes. We have kept the number of sessions the same and reduced the duration of each session to approximately match for attention in the two groups.

1.8. SAMPLE SIZE AND POWER CALCULATIONS

Table 2 illustrates the distribution of sample sizes (without dropout) for two conservative effect sizes and significance levels. To interpret these we need to multiply the effect sizes by the standard deviation (SD) of the change in our outcome variables to obtain differences in change between two arms.

The power calculation of our main outcome variables-change in physical activity, weight and total cholesterol- are based on a) a meta-analysis of RCTs observed that the use of pedometers increased physical activity by 2,500 steps/day (SD 2,700)(Bravata *et al.*, 2007); b) study by Morgan et al (2009)[(Morgan *et al.*, 2009)]and Whincup (personal communication based on British Regional Heart Study data) which reported a standard deviation for weight change of 5kg and c) Whincup (personal communication based on British Regional Heart Study) which reported a SD of change in total cholesterol change of 1.0 umol/l.

We have selected a very conservative effect size of 0.25 (expressed as the difference in units of pooled SDs, d) which translates to an ability to detect a difference between two groups of 675 steps/day (physical activity), 1.25 kg weight and 0.25 umol/l total cholesterol. We calculated the sample size to detect these differences in our primary hypotheses, weight and physical activity at 90% and two tailed alpha 0.025 to take account of multiple comparisons. We took into account clustering effect within the group intervention (intraclass correlation coefficient=0.05) by using the user-written STATA function Clsampsi (version 1.9) which calculates the optimal sample size in presence of differential clustering effects [(Roberts and Batistatou, 2010)] . We estimated 1,420 participants and at a drop out of 20%, 1,704 participants in total. For those taking the average number of steps of 10,000/day and for those only achieving 5000 steps/day this represents approximately 7% and 14% increase respectively.

Assuming a common SD of the change score of 5 on the Framingham risk engine[(Wister *et al.*, 2007)]we would also have 90% power at p=0.025 to detect a difference in change between two arms of 1.25, which would represent a ~5% reduction in the 10 year CVD risk.

Table 2: Distribution of sample sizes at 90% power by significance level and effect size (d)

	p=0.05				p=0.025				p=0.01			
d	G	I	U	Tota	G	I	U	Tota	G	I	U	Tota

				l				l				l
0.25	460	369	369	1438	540	440	440	1704	650	527	527	2046
0.33	270	211	211	830	320	249	249	982	380	304	304	1186
G=group intervention and includes clustering effect (intraclass correlation coefficient=0.05); I=individual intervention; U=usual care; Total= includes 20% dropout.												

1.9. THE ORIGINAL PROPOSED STATISTICAL ANALYSES OF MAIN OUTCOME VARIABLES

A description of the sample will be presented using means and their standard deviations (SD) or counts (proportions). Baseline characteristics of refusers and drop-outs will be compared with participants who complete the study.

An intention to treat analysis will be conducted using STATA 11. The differences in treatment effect between the three arms at 12 and 24 months of this partially nested design will be analysed using mixed effects models with pre-randomisation values as a covariate.(Bauer *et al.*, 2008) Stratification variables (practice as random effect and ethnicity as fixed effect) and other possible confounder (such as gender, PCT) will be included as further covariates. This approach provides valid inferences under the assumption that the missing data mechanism can be ignored (or missing at random). Further sensitivity analysis will be carried out to assess the effect of relaxing the missing at random assumption to allow informative dropout, that is letting missingness also depend on the unobserved value (Carpenter *et al.*, 2007).

2. DATA ANALYSIS PLAN: DATA DESCRIPTION

The analysis will follow the guidelines of the Consort statement for randomized trials (Campbell et al. 2004, 2007) and recommendations of Roberts and Roberts (2005) for the analysis of partially clustered randomized trials to present and analyse the data. The trial statistician will remain blind whenever possible until the main analyses are complete.

2.1 DESCRIPTION OF AVAILABLE DATA

The patterns of availability of baseline and follow-up data will be summarised overall and separately for the three intervention groups for each assessment visit and scale; presented by descriptive summary statistics of clinical and demographical assessments. Data will also be assessed for each phase.

QQ plots and histograms will be used to assess data distribution of continuous measures.

Appropriate summary statistics will be applied, mean and standard deviation for all normally distributed measures; median and quartiles for skewed distributions and ordinal data. Categorical outcomes will be described using both the number and proportion (percentage).

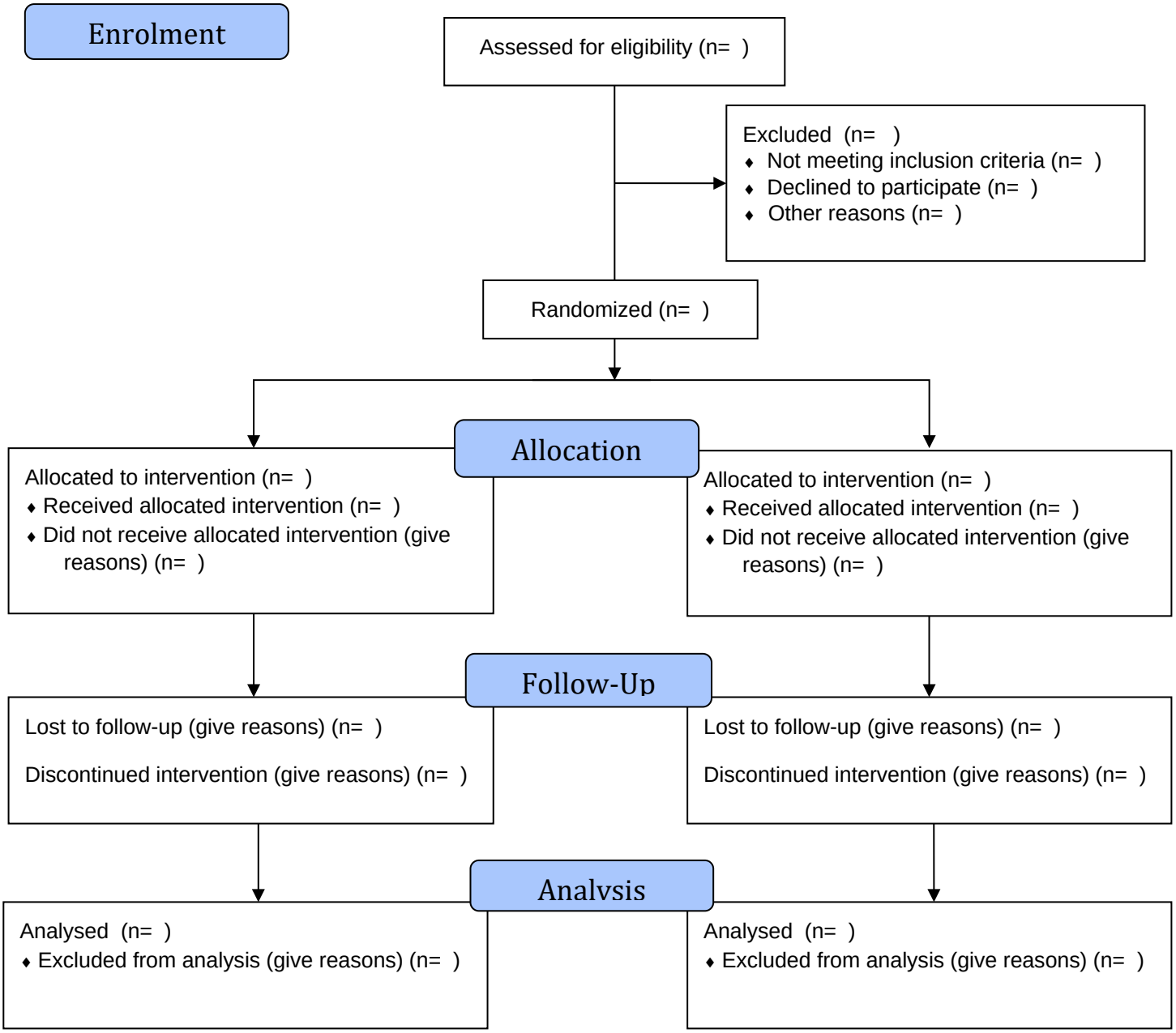
2.2 RECRUITMENT: GP AND PARTICIPANT FLOW

Flow of GP practices *and* individual participants through each stage will be summarized in a CONSORT diagram, including for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analysed for the primary outcome

We will describe protocol deviations from the study as planned together with reasons.

Participant throughput will be summarised in a CONSORT diagram, including the stages of enrolment, allocation, follow-up and analysis.

CONSORT 2010 Flow Diagram



Schulz et al. 2010: From: http://www.consort-statement.org/consort-statement/flow-diagram0/

2.3. BASELINE COMPARABILITY OF RANDOMISED GROUPS

Participant level baseline variables will be described both overall and between randomised groups. No statistical significance tests or confidence interval will be calculated for the difference between randomised groups on any participant level baseline variables to assess for baseline differences. The randomisation of intervention groups to participants should have ensured that any imbalance over all measured and unmeasured baseline characteristics is due to chance. In a secondary sensitivity analyses we will include potential confounding variables which differ substantially between the three groups. Variables measured at baseline which may affect primary outcome will be assessed for potential imbalance. These following variables will be assessed:

1.) Age
2.) Gender
3.) Ethnicity
4.) Index of deprivation
5.) Education status
6.) Marital status
7.) Smoking status

2.4. ADHERENCE TO ALLOCATED TREATMENT

We describe the distribution of the number of treatment sessions and any deviations from protocol within the two treatment arms.

Results will be stated in terms of proportion who attended each session and average number of sessions attended. In this context a patient is deemed to have received therapy if they attended at least 1 session out of the sessions 1-10 during their allocated intervention period The data are used in the CONSORT flow chart. Reasons given for withdrawing from therapy are summarised.

Drop-outs from intervention without attending any session: The baseline demographic data will be explored to investigate how this sub-group form the rest of the population.

2.5. ADHERENCE TO THE PROTOCOL

The clinical psychologist (NdZ), research manager (KW), trial administrator (CB), healthy lifestyle facilitators and participants are unblinded to treatment allocation. The statistician (DS), principal investigator (KI), trial manager (KT) and the research assistants are blind to treatment allocation.

How is adherence defined?

Intervention compliance will be reported as numbers (% of intervention group) of patients attending each session (Table 2).

	Number of patients (% of intervention group) attending session		
Therapy session	Group Enhanced MI N= (%)	Individual Enhanced MI N= (%)	Usual care N= (%)
0			
1			
2			
3			
4			
5			
6			
7			
8			
8			
9			
10			

Table 2: Intervention compliance: numbers (% of intervention group) of patients attending each session.

2.6 LOSS TO FOLLOW UP AND OTHER MISSING DATA

It is the aim of the trial to minimise withdrawal of participants within a treatment arm from treatment and follow-up.

Withdrawal from the trial follow-up will be reported by intervention group. Summaries will be given of the reasons for withdrawal (grouped as appropriate). The distribution of times between randomisation and withdrawal from follow-up will be summarised in a histogram. The therapy session cohort number, and number of session attended will be summarised.

Time since baseline (months)	Number of patients (% of intervention group) with body weight and physical activity measured		
Time since baseline (months)	Usual care (N=)	G-EMI (N=)	I-EMI (N=)
Pre-baseline	n= (%)	n= (%)	n= (%)
Baseline			
12			
24			

Table 4: % Example of data collected

Where available, the reasons for missing baseline and follow-up data will be summarised overall and by randomised group at the visits and scale levels. Missing data in the database will be presented in compliance with CONSORT diagram (Boutron *et al.*, 2008).

MISSINGNESS PATTERN OF MAIN OUTCOME VARIABLE

The relationship between characteristics (demographic and clinical) and missing data will be summarised at 12 and 24 months follow-up (graphically) and the potential demographic and clinical factors that may be related to be missingness of bodyweight and physical activity at 24 months, such as whether a patient completed therapy or the number of session a patient will be examined.

2.7 ADVERSE EVENT REPORTING

Summarise adverse events (AE), adverse reactions (AR), serious adverse events (SAE) and serious adverse reactions (SAR), by intervention arm and time point (prior to 12 and 24 months) will be presented.

3. STATISTICAL DATA ANALYSES PLAN: COMPARISON OF BODY WEIGHT AND PHYSICAL ACTIVITY BETWEEN INTERVENTION GROUPS (PRIMARY OUTCOME ANALYSIS)

3.1. GENERAL COMMENTS

The main statistical analyses are targeted at estimating the difference in the mean outcomes between participants randomised to i) G-EMI or usual care ii) I-EMI and usual care and iii) G-MI+ and I-MI+ at the 24 months post-treatment observation time. The main outcome variables are cholesterol, **body weight** and **physical activity** which are measured at baseline, 12 months and 24 months follow-up.

The analyses of effectiveness outlined in this strategy will be pragmatic, based on intention-to-treat and will utilise all available follow-up data from all randomised participants. The trial statistician will remain blind whenever possible until the main analyses are complete.

The significance level will be 2.5%% (2 sided) for specified analyses.

The study design is complex: Patients are nested within groups in the G-MI+ treatment arm but not in the two other arms (partially nested design). Furthermore, patient are nested within surgery with patients within surgeries belong to different treatment arms. In addition, we expect differences between boroughs and between ethnic groups and both factors need to be considered in the analysis model. Post-randomization measurements are taken at 12 month and 24 months follow-up. To account for the repeated observations over time there is the need of specifying the structure of the residual errors within the lowest-level groups (patients) to be correlated.

Finally, some patients live in the same household and may be randomized to the same or different treatment arms. To avoid this situation only one person per household will be randomized. However, a small number of couples patient (January 2014: 2 pairs) were randomized. If the number of patients of the same household remains small (<1%) the possible clustering effect will be ignored to avoid a further complication of the statistical analysis model.

3.2. PRIMARY OUTCOME "PHYSICAL ACTIVITY"

Physical activity will be recorded using actigraphy which involves the use of a portable device the size of a large wristwatch to record movement over time in the form of activity counts. Each participant was asked to wear an actigraph for seven days during the day and only to remove when sleeping or when it might get wet or during contact sports.

Upon completion of the 7 day monitoring period, data will be downloaded using the manufacturer's software producing a file containing minute-by-minute movement counts for each participant. Data will be cleaned and summarized using standard procedures provided by ActiGraph software following recommendations by Ward et al 2005.

Patients need to have at least four days of at least 540 min recorded activity to include the respective measurement in the analysis. The definition of "Non wear definition" is 60 minutes of consecutive zeroes for as recommended by Winkler et al 2009, Evenson 2009).

The raw accelerometer counts are summarized as steps, accelerometer wear time, counts per minute (CPM), time spent in moderate to vigorous physical activity (MVPA (>1952 Counts per Minute, >3 METs) using standard Freedson (1998) cut points and time spent in >= 10 min bouts of MVPA. A tolerance of 1 minutes of interruption will be allowed. Counts

Counts per minute per day are defined as the average of 'sum of activity counts per day' divided by the' number of minutes of wear time in that day.

The main outcome will be the average number of steps per day measured over seven days. The secondary outcome variable will be time spent in >=10 min bouts at least moderate activity (MVPA) per day.

Missing activity days:

Because activity may depend on day of the week the mean of less than seven activity days may be biased and we will impute intermittent missing values using the Expectation Maximization (EM) algorithm as recommended by Catellier et al. (2005) to impute missing daily values. The mean of the observed and imputed data will then be used for the primary analyses. The single imputation method treats imputed values as they were actual observed

and hence ignore the uncertainty of the correct value is not taken into account. However, Catellier demonstrated that EM imputation methods with intermittent missing activity data resulted produced similar results as multiple imputation as simulation studies showed that there were no difference in estimates of means and standard deviations. Both methods produced unbiased under MCAR conditions (with EM imputation sometimes performing slightly better) and hence their use is regarded as standard for intermittent missing in days of activity data (Lee and Gill, 2016, Staudenmayer *et al.*, 2012)

Day of the week, month, year included as indicator variables in the imputation model. Activity data will be imputed within each treatment arm separately. Standard clinical and demographic variables will be considered as potential predictors.

Days will only be imputed for cases with at least **three** days of at least 540 min recorded to avoid bias by deliberate removing the wearable during the day which can bias subsequent analyses (Evenson *et al.*, 2015). Data with less activity measured will be regarded as missing for this observation period and procedures to handle complete missing observations of activity will be described in the next chapter. Potential selection bias attributable to insufficient valid accelerometer data will be examined by comparing baseline characteristics of participants with **three** or more valid days of data against the characteristics of those who did not met the criterion (see 2.6.).

3.3. OVERVIEW OF ANALYSES OF PRIMARY OUTCOME:

1. Primary Intention to treat analysis
2. Sensitivity of results due to potential imbalance of confounders at baseline
3. Sensitivity of results due to missing data by
    - 3.1. including covariates predictive of missingness in analyses model
    - 3.2. using multiple imputation
    - 3.3. assuming missing not at random
4. Estimation of the treatment effect impact only for compliers (Complier average causal effects, CACE)

3.3.1. PRIMARY INTENTION TO TREAT ANALYSIS

The main statistical analyses are targeted at estimating the difference in the mean weight and physical activity outcome i) between participants randomised to G-EMI and Treatment as usual ii) between participants randomised to I-MI+ and Treatment as usual and iii) between participants randomised to I-MI+ and Treatment as usual at 24 months follow-up observation time point.

The main analyses will be done for weight and physical activity separately; the estimated differences between the treatment arms at 24 months follow-up will be done within one model.

An analysis of covariance (ANCOVA) approach is utilised for these analyses as the model accounts for the possible imbalance due to random sampling in baseline measurement of the outcome variable to control for pre-treatment differences. An analysis of covariance approach is preferred because of a usually increased statistical power to detect any treatment effects, since baseline and post-treatment measurements are assumed to be correlated. Furthermore, the ANCOVA approach is known to deal better with possible regression to the mean effects.

The main outcome values are assumed to arise from normal distributions (this will be checked - see Model assumptions checks - and if necessary appropriate transformations will be used). A linear mixed model using STATA's XTMIXED command will be used for estimation.

The outcome variables are body weight/physical activity at 12 and 24 months after randomization. This allows us to include subjects with measurements with at least one observation assuming missingness at random.

In the linear mixed model body weight/physical activity levels constitutes therefore the dependent variable. "Treatment randomisation group", "time (with two levels 12 and 24 months post-randomization)", the interaction between "treatment group and time", borough, ethnicity and the "baseline values of body weight/physical activity" are the fixed part of the model. Time" will be entered as a categorical variable to avoid making a parametric assumption of the body weight/physical activity over time. The random parts of the models "surgery" (patients are nested in surgeries) and "therapy group" (patients within the G-MI+ are clustered within the same group).

To account for the partially nested design of “therapy group” we will use an approach which matches the non-parallel data structure as recommended by Roberts and Roberts 2005, Lee and Thompspn 2005, Bauer et al 2008, which has been shown to be superior to other analyses strategies, such as a fully clustered design with clusters of sample size 1 for individual treatment arms because of its control of type 1 errors (Baldwin et al 2010):

In the partially nested design a cluster variable for therapy group needs to be defined where subjects of the two individual treatment arms are treated as clusters of size 1. A further dummy coded variable for group treatment indicates whether a subject is assigned to group treatment (1) versus individual treatment (0, here I-MI+ or TAU). Therapist group will then be used as a random effect with “group treatment” as a random slope which introduces random intercepts for the clustered condition only. Table 5 shows an example of coding to account for the partially nested design: Using STATAs xtmixed the partially nested design would be specified in the following way:

xtmixed DV baseline i.treatment || cluster ID: GroupTreatment

Surgery ID	Cluster ID *"Therapy group"*	Subject ID	*Group Treatment* (Group =1 versus non-group=0)	I-EMI *treatment* (I-EMI=1 Other =0)	*Baseline* (Individual covariate)	Ethnicity (Individual covariate)	Outcome at 24 months *(DV)*
1	1	1	1	0	89	1	84
1	1	2	1	0	90	2	87
1	1	3	1	0	78	1	75
1	2	4	1	0	99	1	95
1	2	5	1	0	75	2	77
2	3	6	0	1	87	3	85
2	4	7	0	0	69	1	67
2	5	8	0	0	92	2	89
3	6	9	0	0	77	3	79

Table 5: Example of coding to account for the partially nested design: In the partially nested design a cluster variable for "therapy group" needs to be defined where subjects of the two individual treatment arms are treated as clusters of size 1 (Column 2). A further dummy coded variable for group treatment indicates whether a subject is assigned to group treatment (1) versus individual treatment (0, here I-EMI or TAU), see column 3. Therapy group will then be used as a random effect with "group treatment" as a random slope which introduces random intercepts for the clustered condition only. A second random effect for surgery (column 1) will used to account for the possible dependency of patients within a surgery.

To model the dependency of the repeated observations of the same subjects at 12 and 24 months we model the covariance between the residuals within the lowest level group" patients" to be correlated by using an unstructured covariance pattern model (Bayley et al., 2015).

For the final model the group difference estimates and associated confidence intervals will be reported for 12 (for secondary analyses) and 24 months after randomization (Bayley et al., 2015).

Further explanation to the independent variables "Surgery", "Borough" and "ethnicity":

Surgery (or GP): is used as a measure of GP organization and proxy for socio-demographic differences within boroughs which are regarded as a possible prognostic factor and to account for the possible clustering effect of surgery. Surgery will therefore be included as a random factor in the analyses model.

Borough

Clinical commissioning group (CCG) boroughs is regarded as a possible prognostic and/or predictive factor and therefore analyses of intervention effect will be adjusted by including "borough" ((Bexley, Bromley, Croydon, Greenwich, Kingston, Lambeth, Lewisham, Richmond & Twickenham, Southwark, Sutton, Merton, Wandsworth) as a fixed, categorical covariate. The borough that randomises the largest number of participants will be the reference category for the effect.

Ethnicity

Ethnicity is regarded as an important factor that influences prognosis or treatment responsiveness. Because stratified randomization was not feasible it is possible to obtain some imbalance between treatment groups for ethnicity and therefore analyses of intervention effect will be adjusted by including "ethnicity" in the analysis model.

MODEL ASSUMPTION CHECKS

For the mixed effects model the following assumptions will be checked:

Normality

A histogram of level one standardised residuals against their normal scores will be produced to verify that the residuals have a normal distribution.

No outlier

Any obvious outliers or influential observations (based on their leverage) will be identified by the residuals and the analysis will be repeated excluding them to check the robustness of the conclusions of the primary analysis. Any differences when outliers are excluded will be reported.

Homogeneity of variances

A plot will be produced of the standardized residuals against the fitted values to examine the assumption of constant variability of the residuals across the range of fitted values. If necessary we further relax the heteroskedascity assumption (error variance is the same between the two treatment groups) treatment specific error variances may be used.

3.3.2. SENSITIVITY OF RESULTS DUE TO POTENTIAL IMBALANCE OF CONFOUNDERS AT BASELINE

The large sample size should ensure that all possible confounding variables are equally distributed between treatment arms. However, in a sensitivity analysis we extend the analyses model of the primary analysis described in the previous section by including by including baseline variables with substantial imbalance at baseline thought to be important in determining outcome into the model (Table 1). The analysis of this model will give intervention effect estimates adjusted for imbalances. We will report the changes in predicted outcome differences after controlling for baseline covariates in addition to the main analysis model. Variables measured at baseline which we believe may affect outcome are:

Age
Gender
Ethnicity (already in main analysis)
Index of deprivation
Education status
Marital status
Smoking status

3.3.2. SENSITIVITY OF RESULTS DUE TO DELAY OF TREATMENT SESSION START: DIFFERENCES BETWEEN RANDOMIZATION AND SESSION 0

The time between randomization of patient and start of intervention with session 0 may differ between participants due to study flow, scheduling conflicts and issues with HLF contracts. In a further sensitivity analysis we extend the analyses model of the primary analysis described in section 3.3.1 by including time between randomisation and session 0

into the model. The analysis of this model will give intervention effect estimates adjusted for imbalances in time between randomization and start of intervention. We will report the changes in predicted outcome differences in addition to the main analysis model.

3.3.4. SENSITIVITY OF RESULTS DUE TO UNBLINDING OF RESEARCHERS AT FOLLOW-UP

Researchers collecting data should be blinded to treatment but unblinding may take place. In a further sensitivity analysis we extend the analyses model of the primary analysis described in section 3.3.1 by including unblinding at follow-up time (yes/no) into the model. The analysis of this model will give intervention effect estimates adjusted for potential effects of unblinding. We will report the changes in predicted outcome differences in addition to the main analysis

3.3.5. SENSITIVITY OF RESULTS DUE TO INSUFFICENT WEARABLE RECORDINGS AT BASELINE

To be randomized, participants were required to wear the accelerometer for at least 5 days for at least 540 min. A number of patients did not fulfil the criteria due to software problems which overestimated initial estimates of wear time. In a further sensitivity analysis we extend the analyses model of the primary analysis described in section 3.3.1 by including number of days correctly worn for at least 540 min at baseline into the model. The analysis of this model will give intervention effect estimates adjusted for imbalances in days worn at baseline. We will report the changes in predicted outcome differences in addition to the main analysis model.

3.3.6. SENSITIVITY OF RESULTS DUE TO INITIALS ERRORS OF CALCULATING QRISK2 SCORES AT BASELINE AND VIOLATION OF STUDY PROTOCOL

To be randomized, participants needed a Qrisk2 score of at least 20%. Due to errors at the calculations of Qrisk scores due to updates of the online algorithm and human errors some patients were randomized with a Qrisk2 score < 20%.

In a further sensitivity analysis we extend the analyses model of the primary analysis described in section 3.3.1 by including Qrisk2 score >=20 % (yes/no) at baseline into the model. The analysis of this model will give intervention effect estimates adjusted for Qrisk2 score violations. We will report the changes in predicted outcome differences in addition to the main analysis model.

3.3.7. SENSITIVITY OF RESULTS DUE TO OTHER POTENTAIL VIOLATIONS OF THE PROTOCOL

Similar sensitivity analyses as described above for other potential violations of the study protocol will be performed if needed.

3.3.8: SENSITIVITY OF RESULTS TO MISSING DATA

There will be a certain proportion of patients for whom it is not possible to measure the main outcome variables at 12 or 24 months and the data are missing. We only expect very little or no missing data for at baseline. In the case of missing data at baseline we will use mean imputation of the baseline value as recommended by (White and Thompson, 2005). Joint modelling of baseline and outcome will be considered if the number of baseline values is large as this is the most efficient method (White and Thompson, 2005).

The described mixed model will be fitted using maximum likelihood methods that are valid under the missing at random (MAR) assumption. However, this assumption relates to the variables that are included in the model, to allow for a variable predicting missingness this variable needs to be included as either one of the explanatory or dependent variables of the mixed model (Gaughran et al., 2013). This assumption will be investigated (see 2.6). Three types of sensitivity analysis for violations of the assumptions of MAR will be performed to assess the sensitivity of the results to missing outcome data:

ANALYSIS 3.3.8A: SENSITIVITY OF RESULTS DUE TO MISSING DATA BY INCLUDING COVARIATES PREDICTIVE OF MISSINGNESS IN ANALYSES MODEL

Should the investigation described in 2.4 indicate any demographic or clinical baseline variables that are predictors of outcome missingness, then such variables will be included as further covariates in the previous model and post-treatment group difference estimates and associated confidence intervals will be reported

ANALYSIS 3.3.8B: SENSITIVITY OF RESULTS DUE TO MISSING DATA BY USING MULTIPLE IMPUTATIONS

A missing data analysis using the method of multiple imputation (Shafer, 1997). Data handled using multiple imputation will be imputed m 100 or more times applying a set seed using STATA version 13.1 (StataCorp 2013). In this procedure missing data is filled in using other information which has been observed on patients. The imputation process will be run

separately in the three treatment arms, to allow for the relationships between variables in the imputation model and Outcomes at 24 months to differ according to arm. We account for clustering within G-EMI by including cluster indicators in our imputation model (van Buuren 2010). By including clusters as indicator variables the regression function of the imputed variable is allowed to vary by cluster. Each of these completed datasets can then be analysed using the proposed statistical modelling and the estimates from the linear mixed model will be drawn from the average of analysis of each of the completed datasets using Rubin's Rule.

Multiple imputations allow us to include all randomised patients, not just those who completed at least one follow-up visit. Furthermore, multiple imputations give unbiased estimates under the assumption that data are missing at random (MAR). This assumption states that the probability that main outcome variables, weight and physical activity, are missing at 12 and 24 months occurs randomly after taking into account observed data which is part of the imputation model.

For our analysis, the imputation model will include all variables which we believe may contain information about the missingness mechanism at 24 months and must include all variables that will be used in the analysis model (Shafer 1997), including measurements at 12 months. Thus the imputation model will include variables such as weight and physical health measurements at baseline and 12 months, randomised intervention group, borough, demographic and clinical characteristics, treatment adherence, etc.

It is known that imputation method with a cluster indicator variable can have computational problems for small clusters and Markov chain Monte Carlo methods using Gibbs sample may need to be used (implemented for example in the user-written R package mice by van Buuren and Groothuis-Oudshoorn 2011).

ANALYSIS 3.3.8C: SENSITIVITY OF RESULTS DUE TO MISSING DATA BY ASSUMING MISSING NOT AT RANDOM

Analysis of data where the outcome is incomplete always requires untestable assumptions about the missing data commonly that they are missing at random. We will perform sensitivity analyses to explore the effect of departures (varied over a plausible range)

from the assumption of missing at random made in the main analysis as recommended by White et al (2011) (Gaughran et al., 2013). We will use multiple imputation to impute the missing body weight and physical activity outcomes at 24 months using the same imputation model as described before except that the intermediate measurements (including 12 months) of main outcome variables will not be used because drop-out is likely to vary according to the availability of intermediate measurements. We then add a fixed amount of weight (or physical activity) to the imputed 24 months follow-up levels to assess possible non-ignorable drop-out mechanism and reanalyse the data sets and combining the results of the ANCOVA using Rubin's rule. For example, to assess for differential drop-out in patients between treatment groups a sensitivity analysis would assume that the cases lost to follow-up have systematically worse outcome than completers and outcome will be worse in treatment group and we will assess the effect of different values on the treatment differences at 24 months (Gaughran et al., 2013).

3.4. SECONDARY OUTCOME ANALYSES

Treatment effects on secondary outcomes at 12 and 24 months such as cardiovascular risk score and other outcomes (see Table 6) will be assessed in a similar way, using generalisations of the linear mixed model to allow for non-normal distributed data where necessary.

Table 6 of secondary outcome variables:

- LDL cholesterol
- CVD risk score (the QRisk2 measurement of CVD risk will be sensitive to changes in weight, cholesterol, blood pressure, diabetes status and smoking status).
- The number of fatal and non-fatal CVD events and hospital admissions
- A secondary physical activity outcome will be the amount of at least moderate physical activity (MVPA) in > 10 minute bouts
- Changes in dietary habits will be measured via analysis of dietary recall data-

Health beliefs (illness perception and self-efficacy) and depression at 12 and 24 months will be assessed as measures of mediating processes.

CONSIDERING INCOMPLETE MEASURES OF SCALE (QUESTIONNAIRES) DATA:

This paragraph is not relevant for the primary outcome variable but only for some of the secondary outcome and baseline variables which are based on questionnaires or similar item based measurement scales.

The planned strategy for handling missing data at the item and scales will depend on the amount of missing data observed and the planned analyses for the outcomes.

Missing covariate item data will be imputed using prorating, that is by replacing the missing item score by the mean of the observed items if less than 20% of the items scores are missing. Items within each scale are indicators of a specific concept and as a result assumed to be closely and positively correlated and therefore regarded as a particularly applicable technique if less than 20 % are missing [8, 9]. Simulation studies have shown that pro-rating (or case mean substitution) is a robust method when data are missing on less than 20% of items in both random and systematic patterns [9].

Pro rating is implemented across items within a scale, or subscale, for each assessment and participant. In the unlikely event that more than 20% missing items on at least some of the data collected multiple imputations will be implemented at the scale level. If this is carried forward it is essential that the item level data is not imputed by prorating at any point. Multiple imputation will also considered if an item is missing in more than 5% of the patients who were administered the questionnaire.

To ensure the same strategy is followed across all scales reported in the principle paper(s), any guidance given by authors of validated questionnaires will supersede the methods outlined herein.

3.5. COMPLIER AVERAGE CAUSAL EFFECTS (CACE)

In addition to the standard intention-to-treat analysis we will estimate a measure of the treatment impact only for compliers. Specifically, we will employ instrumental variable (IV) methods to evaluate the causal effect of CBT in the subpopulations that are considered compliers to treatment. This complier average causal effect (CACE) is of scientific and policy interest, because it assesses the intervention effectiveness of the treatment when it is in fact taken (treatment *efficacy)*. Noncompliance is a common problem in randomised clinical trials (RCT) which could cause biased results in determining the (non-ITT) effect of the

treatment. Participants who comply with a particular treatment may be a biased sample of participants randomised to that treatment. Standard treatment effect analyses such as per-protocol (PP) analysis, which compares the average treatment effects for participants who comply with the assigned treatment, are therefore subject to bias. We therefore use an instrumental variable approach as suggested by Dunn et al (2005) to estimate the CACE, where randomisation indicator is used as an instrumental variable. This approach provides a direct estimate of treatment efficacy which is protected from selection bias by the randomization. Analyses will be performed using STATA's user written package ivreg2 (Baum 2007). CACE will be estimated separately for G-EMI and I-EMI (versus TAU). Cluster robust standard errors will be used which allows residuals which are not independent within clusters and adjusts for possible heteroscedasticity. However, currently it is not possible to correct standard errors for more than one cluster. We will therefore control standard errors estimations by specifying the cluster with the larger ICC or create a combined cluster variable of "therapy group and surgery"

3.6. ADVERSE EVENTS (AE)

These analyses are based in the safety data and describe the safety outcomes. SAEs will be defined in accordance with the EU clinical trial Directives. Any adverse event deemed to be an SAE will initially be reported to the CI and then to the KCL (sponsor) within 48 hours. SAEs will be recorded by time and intervention group.

The following potential adverse events are monitored:

- Cardiovascular even
- Results in death;
- Is life-threatening;
- Required hospitalisation or prolongation of existing hospitalisation;
- Results in persistent or significant disability or incapacity;

AEs will be tabulated by event type and clinical classification by time and randomised group. Each table will detail the number of participants that were still in the trial at the time points by randomisation group. If any adverse events are selected as being of particular interest they will be further summarised; outlining the severity (mild, moderate, severe) and if

classified serious the expectedness (expected, unexpected) of the event. Any death will be reported as a serious adverse event.

Withdrawal from treatment will be regarded as potentially safety related outcomes and will be reported in terms of the reason, days from randomisation and person responsible for making the decision to withdraw.

3.7. COMPLIANCE

The distribution of the number of weeks completed treatment, i.e. complied with the protocol. We will state results in terms of compliance/ non-compliance and summarise reasons for not completing assessments. All unplanned unblindings of the researchers carrying out the quantitative assessments will be reported.

Chi squared (or Fishers exact) tests will be used to describe any differences in withdrawals and SAEs between intervention groups.

Reference List

Baldwin, S., Bauer, D., Stice, E., & Rohde, P. (2011). Evaluating models for partially clustered designs. Psychological Methods, 16, 149–165.

Bauer, D., Sterba, S. & Hallfors, D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioural Research* **43**, 210-236.
Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., Ravaud, P. & Grp, C. (2008). Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. *Annals of Internal Medicine* **148**, 295-309.
Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., Stave, C. D., Olkin, I. & Sirard, J. R. (2007). Using Pedometers to Increase Physical Activity and Improve Health: A Systematic Review. *JAMA* **298**, 2296-2304.
Broadbent, E., Petrie, K., Main, J. & Weinman, J. (2006). The Brief Illness Perception Questionnaire. *Journal of Psychosomatic Research* **60**, 631-637.
Carpenter, J., Kenward, M. & White, I. (2007). Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Stat Methods Med Res* **16**, 259-75.
Evenson, K. R., Sotres-Alvarez, D., Deng, Y., Marshall, S. J., Isasi, C. R., Esliger, D. W. & Davis, S. (2015). Accelerometer Adherence and Performance in a Cohort Study of US Hispanic Adults. *Medicine and Science in Sports and Exercise* **47**, 725-734.
Lee, J. A. & Gill, J. (2016). Missing value imputation for physical data measured by accelerometer. *SMMR*, 1-20.
Morgan, P., Lubans, D., Collins, C., Warren, J. & Callister, R. (2009). The SHED-IT Randomized Controlled Trial: Evaluation of an Internet-based Weight Loss Program for Men. *Obesity* **17**, 2025 - 2032.
Roberts, C. & Batistatou, E. (2010). Clsampsi.Downloaded from http://personalpages.manchester.ac.uk/staff/Chris.Roberts/.
Shafer, J. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall New York.
Staudenmayer, J., Zhu, W. & Catellier, D. J. (2012). Statistical considerations in the analysis of accelerometry-based activity monitor data. *Med Sci Sports Exerc* **44**, S61-7.
White, I. R. & Thompson, S. G. (2005). Adjusting for partially missing baseline measurements in randomized trials. *Stat Med* **24**, 993-1007.
Wister, A., Loewen, N., Kennedy-Symonds, H., McGowan, B., Mccoy, B. & Singer, J. (2007). One-year follow-up of a therapeutic lifestyle intervention targeting cardiovascular disease risk. *Canadian Medical Association Journal* **177**, 859-865.

Bayley, A., de Zoysa, N., Cook, D., Whincup, P., Stahl, D., Twist, K., … Ismail, K. (2015). Comparing the effectiveness of an enhanced MOtiVational intErviewing InTervention (MOVE IT) with usual care for reducing cardiovascular risk in high risk subjects: study protocol for a randomised controlled trial. *Trials*, *16*(1), 112.

Gaughran, F., Stahl, D., Ismail, K., Atakan, Z., Lally, J., Gardner-Sood, P., … Smith, S. (2013). Improving physical health and reducing substance use in psychosis – randomised control trial (IMPACT RCT): study protocol for a cluster randomised controlled trial. *BMC Psychiatry*, *13*(1). https://doi.org/10.1186/1471-244x-13-263