UNIVERSITY OF CAMBRIDGE
Department of Public Health and Primary Care

MRC Epidemiology Unit

Statistical Analysis Plan

A randomised trial to determine if communicating different forms of coronary heart disease risk information and lifestyle advice for risk reduction results in health-related behaviour change (ISRCTN 17721237).

Information And Risk Modification Trial (INFORM).

FINAL – 18 August 2015.

1 Introduction

This is the plan for the trial analyses of the primary and secondary outcome variables from a randomised controlled trial to evaluate the effectiveness of providing three different forms of coronary heart disease (CHD) risk information and lifestyle advice on health behaviours. The analyses described in this document will be performed by Stephen Sharp, Senior Statistician, University of Cambridge MRC Epidemiology Unit, once the data have been entered, cleaned and released for use.

This analysis plan refers to the main quantitative outcomes from INFORM. Analyses of data arising from some other quantitative and qualitative components of INFORM will be performed separately by members of the trial team/PhD students/post-doctoral researchers and defined in future analysis plans.

The analyses described in this document are compatible with the recommendations of the CONSORT 2010 statement (www.consort-statement.org).

2 Study outcomes

2.1 Primary outcomes (recorded at baseline and 12 weeks follow-up)

Objectively measured physical activity (m/s^2).

2.2 Secondary outcomes (recorded at baseline and 12 weeks follow-up)

2.2.1 Objective measures

Serum carotenoids (µmol/l):

- total.
- alpha-carotene.
- beta-carotene.
- lutein.
- lycopene.
- beta-cryptoxanthin.
- zeaxanthin.

Total cholesterol (mmol/l).
HDL cholesterol (mmol/l).
LDL cholesterol (mmol/l).
Triglycerides (mmol/l).
Fructosamine (µmol/l).

2.2.2 Self-reported measures of anthropometry, diet and lifestyle

Weight (kg).

Fruit intake (servings/day).
Vegetable intake (servings/day).
Whole grain intake (servings/day).
Fish intake (servings/week).
Alcohol intake (units/week).
Red and processed meat intake (servings/week).

Physical activity level (inactive, moderately inactive, moderately active, active).

Smoking status (yes/no).

2.2.3 Perceived risk

Comparative measure.
Absolute measure.
Accuracy of risk perception

2.2.4 Psychological outcomes

Overall stress.
Mood.
Coronary heart disease related worry.

3 Analysis populations

All trial analyses will be based on the **Intention To Treat (ITT)** principle, i.e. all participants will be included in the group to which they were randomised, regardless of the intervention actually received.

An analysis of the **primary outcome** will also be performed in a **Per Protocol (PP)** population. Individuals will be included in this population if they accessed risk score and lifestyle information; the precise criteria will be defined based on the distributions of uptake, and applied prior to unblinding of the trial.

4 Descriptive analyses

The following baseline characteristics of the study population will be summarised separately within each randomised group:

Age.
Sex.
Ethnicity.
Marital status.
Socioeconomic status.
Family history of CHD.
History of genetic testing.
History of CHD risk assessment.
Genetic risk related worry/anxiety.

Outcome variables which were measured at baseline will be summarised at baseline and 12 weeks follow-up as described in Section 5.

For continuous variables, means and standard deviations will be presented, unless the variable has a highly skewed distribution, in which case medians, 25th and 75th percentiles will be presented. For categorical variables, the number and percentage of participants within each category will be presented. For each variable (continuous or categorical), the % of missing values will be reported.

No p-values will be calculated for these descriptive analyses.

5 Analyses of study outcomes

5.1 Primary outcome

Change (12 weeks follow-up minus baseline) in the primary outcome, objectively measured physical activity, will be summarised separately in each of the 4 randomised groups using means and standard deviations; the means and standard deviations in each group at baseline and follow-up will also be reported.

Intervention effects will be estimated using analysis of covariance (ANCOVA). The outcome in the ANCOVA model will be change (12 weeks follow-up minus baseline) in physical activity, with the baseline value included as a covariate in the model. An F-test will be performed of the null hypothesis that there is no difference in mean change adjusted for baseline between the 4 randomised groups. The ANCOVA model will also be used to derive estimates of the differences in mean change and 95% confidence intervals for 4 pairwise comparisons defined below:

1. Group 1 (control group) vs. Group 2 (lifestyle advice only) – to estimate the effect of providing lifestyle advice compared with no intervention.
2. Group 3 (phenotypic risk score and lifestyle advice) vs. Group 4 (phenotypic and genetic risk scores and lifestyle advice) – to estimate the effect of providing a genetic risk score in addition to lifestyle advice and a phenotypic risk score.
3. Group 3 + Group 4 vs. Group 1 – to estimate the effect of providing risk score information (phenotypic or genetic or both) and lifestyle advice compared with no intervention.
4. Group 3 + Group 4 vs. Group 2 – to estimate the effect of providing risk score information (phenotypic or genetic or both) in addition to lifestyle advice.

p-values for each of these pairwise comparisons will not be calculated.

An analysis will be performed to check whether adjusting for age (≤60 vs >60) and sex (the randomisation stratifiers) in the ANCOVA model has any impact on the conclusions; if it has no impact, then they will not be included in the model.

5.2 Secondary outcomes

5.2.1 Objective measures

For each outcome, the same analysis will be performed as described in section 5.1. Triglycerides will be log transformed prior to analysis, due to the skewed distribution of this variable.

5.2.2 Self-reported measures of anthropometry, diet and lifestyle

For each continuous outcome, the same analysis will be performed as described in section 5.1. Intakes of fruit, vegetables, whole grain, fish, red and processed meat will all be analysed as continuous variables, derived as described in section 6.1. Physical activity level will also be analysed as a continuous variable with values 0,1,2 and 3.

Smoking status (yes/no) at 12 weeks follow-up will be analysed using a binomial regression model, with adjustment for smoking status at baseline. A likelihood ratio test will be performed of the null hypothesis that there is no difference in risk of smoking adjusted for baseline smoking status between the 4 randomised groups. Estimates of the difference in risk of smoking and 95%

confidence intervals for the 4 pairwise comparisons defined in section 5.1 will be derived from this model.

5.2.3 Perceived risk

The comparative measure of risk will be analysed as a continuous variable, coded as follows: much less likely = -2, less likely = -1, about the same = 0, more likely = +1, much more likely = +2. This variable, and the continuous absolute cardiovascular risk (range 0 to 100) will be analysed using the method described in section 5.1.

Accuracy of risk perception will be defined as perceived risk being either ≤ or > the risk estimate, and analysed using a binomial regression model as described for smoking status in section 5.2.2.

5.2.4 Psychological outcomes

Overall stress will be analysed as a continuous variable, coded as follows: very low=1, low=2, moderate=3, high=4, very high=5. This variable, and the continuous mood variable (range 0 to 2), will be analysed using the method described in section 5.1.

Coronary heart disease (CHD) related worry is a continuous variable (range 6-24) only measured at 12 weeks follow-up. Intervention effects will be estimated using linear regression, in which the outcome will be CHD related worry at 12 weeks follow-up. An F-test will be performed of the null hypothesis that there is no difference in mean level of CHD related worry at 12 weeks follow-up between the 4 randomised groups. The model will be used to derive estimates of the differences in mean values of CHD related worry and 95% confidence intervals for the 4 pairwise comparisons defined in section 5.1.

6 Considerations for analysis

6.1 Derivation of diet outcomes

Fruit intake and vegetable intake will each be analysed as number of servings/day, where “None”=0, “One a day”=1, “2-3 per day”=2.5, “4 per day”=4, “5 or more per day”=6.

Whole grain intake will be analysed as number of servings/day, where “None”=0, “One a day”=1, “2 per day”=2, “3-4 per day”=3.5, “5 or more per day”=6.

Fish intake will be analysed as number of servings/week, where “None”=0, “One a week”=1, “2-4 per week”=3, “5-6 per week”=5.5, “7 or more per week”=8.

Red and processed meat intake will be analysed as number of servings/week, where “None”=0, “One a week”=1, “2-3 per week”=2.5, “4-6 per week”=5, “7 or more per week”=8.

6.2 Missing data

Missing values of primary outcome

If an individual has a missing value for objectively measured physical activity at follow-up, they will be excluded from the analysis.

The percentage of individuals with missing values of objectively measured physical activity at follow-up will be calculated. Levels of missing data are expected to be low, but if this is not the case, the

potential impact of missing data will be explored in sensitivity analyses using a pattern mixture model (White 2012).

Missing baseline values of outcomes

For continuous outcomes, those participants with a missing baseline value of the variable will be included in the analysis using the missing indicator method (White 2005), which is a valid method for pre-randomisation measures in trials, ensuring that no further participants are excluded while maintaining the advantage of improved precision.

6.3 Subgroup analyses

For the primary outcome, the following subgroup analyses will be performed:

(1) age (below/above median value).

(2) baseline phenotypic coronary heart disease risk score (below/above median value).

(3) sex (men/women).

(4) BMI (below/above median value).

(5) self-perceived risk below/above estimated risk.

For each subgroup analysis, a 3 degrees of freedom F-test of the null hypothesis that there is no interaction between any of the randomised groups and the subgroup variable will be performed. If the p-value for a particular interaction is <0.05, then estimates and 95% confidence intervals for the 4 pairwise comparisons defined in section 5.1 will be derived within each subgroup defined by that particular variable.

6.4 Multiplicity

Given the number of outcomes, randomised groups and therefore comparisons, the results for the primary outcome will be regarded as convincing if the p-value from the 3 d.f. test is <0.01, while the results for each secondary outcome will be regarded as convincing if the relevant p-value from the 3 d.f. test is <0.001.

7 References

White IR, Thompson SG (2005) Adjusting for partially missing baseline measurements in randomized trials. Stat.Med 24: 993-1007.

White IR, Carpenter J, Horton NJ (2012) Including all individuals is not enough: lessons for intention-to-treat analysis. Clinical Trials 9: 396-407.