STATISTICAL ANALYSIS PLAN

Study Title:

Cardiovascular Risk Prevention with a Mediterranean Dietary Pattern Reduced in Saturated Fat (CADIMED): Randomized and Controlled Intervention Study with Metabolomic and Gut Microbiota Analyses

Principal Investigator Dr. Carmen Piernas-Sanchez

Department of Biochemistry and Molecular Biology II, Faculty of Pharmacy, Institute of Nutrition and Food Technology, Biomedical Research Center, University of Granada

Trial registration number NCT05778656

Version History

Version:	Version Date:	Changes:
1.0	16 July 2025	
2.0	18 December 2025	Final sample size is stated Added extra sensitivity analyses

Table of Contents

INTRODUCTION

PREFACE

The CADIMED trial team have contributed to and approved this statistical analysis plan (SAP). The SAP supports the study protocol version 1.1 and dated 03/02/2023. Analysis will be carried out using up-to-date versions of Stata and/or R.

PURPOSE AND SCOPE OF THE PLAN

The purpose of this SAP is to complete the main analyses as stated previously in the published protocol (DOI: 10.1016/j.nutres.2025.03.001).

TRIAL OVERVIEW

The Cardiovascular Risk Prevention with a Mediterranean Dietary Pattern Reduced in Saturated Fat study is a two-arm, 8- week parallel randomized controlled intervention trial involving a final sample of 157 adults (≥18 years) with dyslipidemia (not undergoing pharmacological treatment) recruited from healthcare and community settings in Granada (Spain). The primary outcome will assess changes in circulating low-density lipoprotein cholesterol and the fatty acid profile, whilst secondary outcomes will measure changes in CVD-related metabolites/biomarkers, gut microbiome, diet/lifestyle, and intervention feasibility/acceptability

OBJECTIVES

Primary objective

To develop and determine the efficacy of a dietary intervention that promotes the elimination of SFA consumption from red and processed meats and to evaluate its impact on lipid and cardiovascular risk markers in patients with dyslipidemia.

Specific objectives

a. To develop the necessary elements for the implementation of a dietary intervention that promotes the reduction of the consumption of SFA from red and processed meats in patients with dyslipidemia.
b. To test the efficacy of a dietary approach aimed at avoiding red and processed meat, in the context of a Mediterranean dietary pattern, on LDL-C and the fatty acid profile (main outcomes) after eight weeks of intervention, compared to the usual approach consisting of recommendations for a healthier lifestyle.
c. To determine the metabolic alterations parallel to changes in SFA intake from red and processed meat, through a complete lipid and metabolomic profiles, as well as the determination of other intermediate markers of CVD.
d. To evaluate the effect of the intervention on dietary intake and other lifestyle factors.
e. To assess the effect of the intervention on the composition of the gut microbiota and its derived plasma metabolites that could be influenced, given the expected metabolic benefits.
f. To establish a working panel of patients and members of the public (PPI) as well as health care professionals to improve research management and dissemination of project results.

TRIAL DESIGN

The CADIMED study is a randomized, controlled intervention trial, with a two-arm parallel group design, which will evaluate the effect of an eight-week behavioral intervention consisting of a Mediterranean dietary pattern with elimination of red and processed meat, compared to the general advice for decreasing SFA intake in patients with high LDL-C.

For this purpose, 157 patients, aged ≥ 18 years with dyslipidemia were recruited.

A computer-generated randomization with a stratified block design will be used, with sex, age, obesity, and diet (fiber intake) as stratification/minimization variables using block sizes of 2 and 4. The computer program REDCap will reveal the allocation group to the researcher, with no option to visualize the groups that will be assigned to future patients.

It will not be possible to blind patients to the intervention group due to the nature of the intervention. However, the effectiveness measure comparing trial arms (primary outcomes) are objectively measured outcomes (blood lipids and metabolites) which means they are not subject to interpretation by researchers, limiting the effect that researcher knowledge of the intervention group could have on outcomes. Where possible, the study team taking all the outcome measurements will remain blind to group allocation throughout the trial.

OUTCOME MEASURES

PRIMARY OUTCOME

Changes in LDL-C (mg/dL) and the fatty acid profile in erythrocytes (%) in the intervention group vs control after 8 weeks of intervention.

SECONDARY OUTCOME

- Changes in other metabolites derived from lipid and metabolomic profile analysis (mg/dL, %), including HDL-C, total cholesterol, triglycerides, and endogenous and exogenous metabolites related to CVD risk and meat consumption (e.g., TMAO, etc.).
- Changes in other CVD risk related markers, including biomarkers of inflammation and endothelial damage, insulin resistance, and diabetes: (glucose (mg/dL), insulin (mU/mL); hemoglobin A1c (%); C-reactive protein, IL-6, IL8/CXCL8, IL10, IL23, TNF-alpha, IFN gamma, ICAM1, VCAM1, ELAM1, MPO, endothelin I and angiotensin II (ng/mL)); blood pressure (mmHg); weight, BMI (kg/m2) and waist circumference (cm).

- Changes in gut microbiota and derived metabolites, including lipopolysaccharides, lipoprotein binding lipopolysaccharides, and other relevant metabolites (microgram/mL).
- Changes in dietary intakes, measured by 24-hour dietary recall, FFQ and 14-point Mediterranean Diet Adherence Screener (MEDAS), including total energy intake (kJ and kcal/d); SFA intake, total fat, sugar, carbohydrates and protein (g/d; kcal/d; %); fiber, salt, general eating patterns, and red and processed meat consumption (g/d; kcal/d; over- all and from different sources).
- Changes in other lifestyle factors including physical activity (minutes/d); smoking (cigarettes/d); and alcohol intake (units/d) using self-reported questionnaire data.

Process measures and exploratory outcomes:

A mix of quantitative and qualitative measures will be employed to undertake a process evaluation, including:

- Recruitment and follow-up rates: the total recruited; the number invited, eligible, and consented; the number who completed all baseline assessments; the number randomized and time to recruit; the number of patients completing the study.
- Acceptability and adherence to the intervention measured through questionnaires at the end of follow-up to under- stand the effectiveness of the intervention. This includes measures of acceptance of the intervention, website use, knowledge about diet and CVD and evidence of contamination of the control group. Descriptive analyses will be performed using the answers to these questions (scales 1- 5, 1 indicating low acceptance and 5 indicating high acceptance).
- The feasibility and validity of collecting self-reported dietary data through a food diary (24-hour dietary recall) will also be assessed.
- Exploratory subgroup analyses by socioeconomic status: an exploratory analysis of subgroups based on sociodemographic characteristics such as age group, sex, and education/income group will be conducted.

TARGET POPULATION

The study aimed to include 156 participants. A final sample of 157 participants was achieved.

Inclusion Criteria

1) Signed informed consent form.
2) Age ≥18 years to ≤75 years.
3) With indication to start lifestyle intervention based on recent values (e.g., last 3-6 months) of LDL-C ≥ 116 mg/dl and <190 mg/dl; as long as pharmacological treatment has not been advised and started (ESC/EAS Guide, Rev Esp Cardiol. 2020).
4) Device access (e.g., smartphone, tablet, computer) with internet and digital skills to use the website.
5) Motivation and willingness to be assigned to any group to improve their diet and commitment to perform the designated protocol.

Exclusion Criteria

1) Patients with familial hypercholesterolemia and/or on treatment with lipid-lowering drugs (including statins, ezetimibe, PCSK9 inhibitors, fibrates, bile acid sequestrants, omega-3 fatty acids, nicotinic acid/Vit B3), dietary supplements or functional foods for the treatment of dyslipidemia (plant sterols, monacolin, red yeast rice, fiber supplements 3-10 g (Plantago), policosanol, berberine, and soy protein/lecithin) in the last 3 months.

2) Cardiovascular risk in need of pharmacological treatment, for example, patients with recent and/or established CVD, type 1 and 2 diabetes.

3) Uncompensated thyroid function disorders; relevant comorbidities (including liver failure and cholestasis, chronic kidney disease, Cushing's syndrome, nephrotic syndrome, class III obesity – BMI > 40kg/m2); cancer; psychiatric disorders and/or that in the opinion of the investigator hinder the fulfilment or follow-up of the study.

4) Excessive alcohol consumption: > 40 g/d (4 UBEs/day) in men and > 20-25 g/d (2-2.5 UBEs/day) in women (Ministry of Health, July 2020).

4) Pregnancy, lactation, perimenopause (amenorrhea < 1 year) and women who change, start, or finish their treatment with hormonal contraceptives during the study.

5) Other research enrolment or eating patterns (e.g., vegetarians and vegans) that could interfere with study requirements.

6) Failure to grant informed consent form or not complete the initial assessment.

SAMPLE SIZE

A previous randomized controlled trial in a similar population, in which similar dietary approaches for reducing SFA intake were tested to reduce SFA intake, achieved long-term change in LDL-C of approximately 15 mg/dl (SD26). If 78 participants are recruited in each group (n = 156), it will be possible to detect a minimum difference of 15 mg/dl between the intervention and control group, with 90% power, two-sided $\alpha = 0.05$ and 20% attrition rate using intention-to-treat analysis. So, the final sample size includes an additional 20% of participants to mitigate dropouts.

RANDOMIZATION AND BLINDING IN THE ANALYSIS STAGE

A computer-generated randomization with a stratified block design will be employed, using sex, age (< 50 vs ≥50 years), obesity (BMI < 30 vs ≥30), and fruit/vegetable intake as marker of fibre (< 5 vs ≥5 portions/d) as stratification variables using randomly permuted block sizes of 2 and 4. The computer program will reveal the allocation group to the researcher, with no option to visualize the groups that will be assigned to future patients.

It will not be possible to blind patients to the intervention group due to the nature of the intervention. However, the effectiveness measure comparing trial arms (primary outcomes) are objectively measured outcomes (blood lipids and metabolites) which means they are not subject to interpretation by researchers, limiting the effect that researcher knowledge of the intervention group could have on outcomes. Where possible, the study team taking all the outcome measurements will remain blind to group allocation throughout the trial.

ANALYSIS – GENERAL CONSIDERATIONS

DATA CLEANING

Prior to the final data lock, data cleaning will be performed, including checking if outcome variables are in the correct ranges.

DESCRIPTIVE STATISTICS AND PARTICIPANT CHARACTERISTICS

A table will present the baseline characteristics by trial arm and overall (see example templates in Appendix 1). The table will include age, gender, ethnic group, education, employment, income, smoking and alcohol habits. Continuous variables will be summarized using means and standard deviations. Medians with interquartile ranges will be presented where appropriate. Categorical variables will be summarized using counts and percentages. Data will be analyzed using R and/or Stata.

Baseline characteristics will be coded as: age (years); gender (men, women); ethnic group (White Caucasian, Arab, Asian, Black, Latin-American, Mixed, Other); education group (no formal qualifications, secondary education, higher education); monthly income (tertiles); smoking (never, former, current); alcohol intake (never, less than 4 times a month, between 2-3 times a week, more than 4 times a week), and relevant health history, including family history.

DEFINITION OF POPULATION FOR ANALYSIS

All statistical analyses of efficacy outcomes will be carried out on all participants who complete the study (complete case analysis). We will endeavor to obtain full follow-up data on every participant to allow full analysis, but we will inevitably experience the problem of missing data due to withdrawal, loss to follow up, or non-response to questionnaire items. The sample size calculation has accounted for a 20% non-response rate.

DATA MONITORING COMMITTEE AND INTERIM ANALYSES

Due to the low risk of harm and short length of the intervention, a data monitoring committee will not be needed, and an interim analysis will not be conducted.

PRIMARY ANALYSIS

ANALYSIS OF PRIMARY AND SECONDARY OUTCOMES

Statistical analysis will be performed with STATA or R/R studio. Baseline characteristics will be reported using summary descriptive methods (raw count and means with SD or 95% confidence interval), by intervention group. No formal significance testing will be done. Continuous variables will be presented as mean and SD unless the use of the median and interquartile range is needed. Categorical variables will be expressed as frequencies and percentages.

The primary analysis of effectiveness outcomes will be analyzed on an intention-to-treat basis, that is, all participants will be analyzed according to the group to which they were randomly allocated, regardless of their compliance. Analysis of effectiveness measures will use data from all participants for whom a baseline and follow-up measure exist.

Linear regression models will be used to calculate the difference of means with 95% confidence intervals, with adjustment for baseline levels of the dependent variable and recruitment site. The assumption of missing completely at random will be assessed for participants who did not provide follow up data, by comparing their baseline characteristics against participants without missing outcome data. Additionally, the sensitivity of the results to confounding due to differences in baseline characteristics will be examined; as well as the sensitivity of imputing missing outcome data using baseline-observation carried forwards (using the baseline value) and multiple imputation (using group allocation and patient's characteristics to impute the missing data). Another sensitivity analysis will use a constrained baseline (meaning baseline adjusted) linear mixed model, which accounts for baseline differences among the study groups. The model will include fixed effects for time (two levels) and treatment (two levels) as well as the unique participant identifier as a random effect.

An exploratory subgroup analysis of outcomes by sociodemographic characteristics will also be conducted. Exploratory outcomes will be analyzed on the sample of completers, using standard statistical methods for a difference of means or linear regression.
Process measures will use all data available, regardless of whether participants completed the trial or withdrew.

The statistical analysis of metabolomics data will employ various methodologies to identify outliers, clusters, or patterns of metabolites, including principal component analysis for the initial clustering. To discern different patterns of metabolites between the intervention and control groups, both before and after the intervention, a novel multivariate approach, Analysis of Variance Multiblock Orthogonal Partial Least Squares, will be applied. Analysis of

Variance Multiblock Orthogonal Partial Least Squares involves employing kernel-based multiblock orthogonal projections to latent structures approach on the pure and residual-augmented matrices resulting from the ANOVA decomposition of the design response matrix. It provides a comprehensive view of important effects in the model, facilitating their interpretation. The analysis will assess the main effects of time (follow-up vs baseline); intervention group vs control; interaction time x intervention group. Univariate analyses, such as Student t-tests, will be used to compare baseline means between the intervention and control groups. A false discovery rate (*q*) will be calculated to mitigate false positives, with a cut-off of <0.1.

Data analysis for the gut microbiome will be performed using the QUIIME2 v.2019.4 program. The data will be filtered using the q2-dada2 plugin in the R software (DADA2 method), resulting in Amplicon Sequence Variants. Taxonomic assignment will be determined based on the GreenGenes database (v.13.8). Alpha diversity (Shannon index) and beta diversity (Unifrac index) will be determined. The analysis of composition microbiomes method in QUIIME2 will be employed to detect differences in taxonomic composition. Predictions of the metabolic functions of the metagenomes will be made using the PICRUSt2 tool. For these analyses, statistical software such as SIMCA (Umetrics, Umea, Sweden), Stata, and R will be employed. Artificial intelligence methods, including association rule mining and genetic algorithms, will be applied to integrate and analyze phenotypic and metabolomic data. This approach aims to identify predictive patterns associated with better outcomes in this population.

HANDLING MISSING DATA

The percentage and absolute withdrawal and participants lost-to-follow up will be reported for each study arm in the CONSORT flow-chart and reasons for missing data will be documented.

HANDLING OUTLIERS

Outliers will be investigated and if appropriate, a sensitivity analysis will be conducted by excluding these outliers.

MODEL ASSUMPTIONS

Before the analysis, the assumptions of linear regression, specifically homoscedasticity, and the normality of the distribution of residuals will be explored. Where these assumptions are not met, a transformation and/or suitable non-parametric analyses will be employed.

APPENDICES

Appendix 1. Example template tables for presentation of results (final tables and figures may vary)

Baseline characteristics of participants

	Intervention		Control	
	N or mean	% or SD	N or mean	% or SD
Age, years, mean (SD)				
Sex, female				
Ethnic groups				
TBD				
Education groups				
TBD				
Smoking group				
TBD				
Alcohol group				
TDB				
Individual income groups				
TBD				
Relevant Health conditions				
TBD				

Primary outcomes

Outcome	Intervention		Control		Between-group difference (95% CI)	p-value
	Baseline value	Mean change ± SD	Baseline value	Mean change ± SD		
LDL-C (mg/dL)						
Palmitic acid (16:0, %)						
Oleic acid (18:1, %)						
EPA (20:5n-3, %)						
DHA (22:6n-3, %)						
….						

Secondary outcomes

Outcome	Intervention		Control		Between-group difference (95% CI)	p-value
	Baseline value	Mean change ± SD	Baseline value	Mean change ± SD		
HDL-C (mg/dL)						
Total cholesterol (mg/dL)						
Triglycerides (mg/dL)						
….						
Glucose (mg/dL)						
Insulin (mU/mL)						
Hemoglobin A1c (%)						
C-reactive protein (mg/L)						
IL-6 (pg/mL)						
IL-8/CXCL8 (pg/mL)						
....						
Systolic blood pressure (mmHg)						
Diastolic blood pressure (mmHg)						
Weight (kg)						
BMI (kg/m²)						
Waist circumference (cm)						
….						

Recruitment rates

	Intervention		Control	
	n	%	n	%
Potential participants signed up				
Potential participants eligible				
Potential participants consented				
Potential participants completed baseline assessments				
Potential participants randomized				
Participants failed to complete follow up (n, % participants)				