CHALMERS
UNIVERSITY OF TECHNOLOGY

Dietary Environmental Impact in the SWITCH Study

Assessing Sustainability Indicators, Dietary Target Adherence, and Blood Lipid Biomarkers

Master's thesis in Biotechnology, Food and Nutrition Science

EMMA KLARSON

DEPARTMENT OF LIFE SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

MASTER'S THESIS 2026

Dietary Environmental Impact in the SWITCH Study

Assessing Sustainability Indicators, Dietary Target Adherence, and Blood Lipid Biomarkers

EMMA KLARSON

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Life Sciences
Division of Food and Nutrition Science
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026

Dietary Environmental Impact in the SWITCH Study
Assessing Sustainability Indicators, Dietary Target Adherence, and Blood Lipid Biomarkers
EMMA KLARSON

Supervisor: Mari Wollmar, Life Sciences
Examiner: Rikard Landberg, Life Sciences

Master's Thesis 2026
Department of Life Sciences
Division of Food and Nutrition Science
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Illustration of sustainable dietary choices and environmental impact, generated using ChatGPT.

Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Dietary Environmental Impact in the SWITCH Study
Assessing Sustainability Indicators, Dietary Target Adherence, and Blood Lipid Biomarkers
EMMA KLARSON
Department of Life Sciences
Chalmers University of Technology

Abstract

Food systems contribute substantially to environmental degradation and are closely linked to non-communicable diseases. Sustainable dietary transitions are therefore important for improving both planetary and human health. This thesis evaluated the environmental impact of dietary intake within the SWITCH study and examined associations with dietary target adherence, blood lipid biomarkers, and environmental database choice.

Dietary intake data from 113 participants were analysed across three study visits during a 13-week intervention period. Environmental impact was estimated by linking 24-hour dietary recalls to the RISE Climate Database and the SAFAD framework. Linear mixed-effects models were used to assess changes over time and associations between dietary carbon footprint, adherence to SWITCH dietary targets, and blood lipid biomarkers. Agreement between RISE and SAFAD carbon footprint estimates was evaluated using correlation and Bland–Altman analysis.

Dietary carbon footprint decreased from baseline to week 7 and increased again by week 13 in both RISE and SAFAD, although these changes were not statistically significant after correction for multiple testing. In contrast, several SAFAD indicators changed significantly during the intervention. Cropland use, nitrogen input, ammonia emissions, and biogenic methane decreased by week 7, whereas carbon dioxide emissions and animal welfare loss increased. By week 13, biodiversity loss, pesticide use, and carbon dioxide emissions had increased from baseline, while ammonia emissions remained reduced. Higher overall adherence to the SWITCH dietary targets was not associated with lower dietary carbon footprint, and no significant associations were observed between dietary carbon footprint and blood lipid biomarkers. This may partly reflect that the adherence score focused on increased intake of selected foods and did not account for reductions in carbon-intensive foods such as red and ruminant meat. SAFAD consistently produced higher carbon footprint estimates than RISE, although the two databases were strongly correlated.

Overall, dietary environmental impact changed across study visits, with significant changes in several SAFAD indicators but not in dietary carbon footprint. Higher adherence to the SWITCH dietary targets was not associated with lower dietary carbon footprint, nor was carbon footprint associated with blood lipid biomarkers. SAFAD produced consistently higher carbon footprint estimates than RISE, although the databases were strongly correlated.

Keywords: carbon footprint, blood lipid biomarkers, dietary adherence, dietary sustainability, environmental impact assessment, RISE Climate Database, SAFAD framework, sustainable diets, SWITCH study.

Acknowledgements

I would like to thank my supervisor Mari Wollmar for her guidance, support and valuable feedback throughout this thesis project. I would also like to thank Claudia for the collaboration during the laboratory work within the SWITCH study.

I am grateful to my partner, family and friends for their encouragement, patience and support throughout my studies. Your support has meant a lot to me.

Emma Klarson, Gothenburg, May 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis.

BMI	Body mass index
BMR	Basal metabolic rate
CO_2	Carbon dioxide
CO_2e	Carbon dioxide equivalent
CH_4	Methane
EER	Estimated energy requirement
EI	Energy intake
FDR	False discovery rate
GDPR	General Data Protection Regulation
HDL	High-density lipoprotein
LCA	Life cycle assessment
LDL	Low-density lipoprotein
N_2O	Nitrous oxide
NH_3	Ammonia
PAL	Physical activity level
REDCap	Research Electronic Data Capture
RISE	Research Institutes of Sweden
SAFAD	Sustainability Assessment of Foods and Diets
SDG	Sustainable Development Goal
SLU	Swedish University of Agricultural Sciences
SLV	Swedish Food Agency
SWITCH	Switching European food systems for a just transition towards healthy and sustainable dietary behaviour through knowledge and innovation

Nomenclature

Below is the nomenclature of indices, parameters, and variables used throughout this thesis.

Indices

i	Index for participants or food items, depending on context
j	Index for study visits or food groups, depending on context
k	Index for environmental indicators

Parameters

e_i	Environmental impact factor per unit of food item i
CF_i	Carbon footprint factor per unit of food item i
$Target_j$	Predefined SWITCH intake target for food group j
CV_{wEI}	Within-subject coefficient of variation in energy intake (%)
CV_{wB}	Coefficient of variation in basal metabolic rate (%)
CV_{tP}	Between-subject coefficient of variation in physical activity level (%)
d	Number of dietary assessment days
z	Z-value used for the Goldberg confidence limits

Variables

$Adherence_{ij}$	Adherence score for participant i and food group j
$Intake_{ij}$	Reported intake for participant i and food group j
BMR	Basal metabolic rate (kcal/day)
BMI	Body mass index (kg/m^2)
CF	Carbon footprint (kg CO_2e/day)
CF_{ij}	Dietary carbon footprint for participant i at visit j
CO_2e	Carbon dioxide equivalents
EER	Estimated energy requirement (kcal/day)
EI	Reported energy intake (kcal/day)
EI/BMR	Ratio between reported energy intake and basal metabolic rate
PAL	Physical activity level
W	Body weight (kg)
Y_{ij}	Environmental impact outcome for participant i at visit j
u_i	Participant-specific random intercept
ε_{ij}	Residual error term for participant i at visit j
β	Regression coefficient from linear mixed-effects models

Contents

List of Figures

List of Tables

1
Introduction

1.1 Background

The global food system contributes substantially to environmental pressure and accounts for approximately 30% of global greenhouse gas emissions, about 70% of freshwater use, and occupies more than 30% of all potentially cultivable land [1]. Of nine planetary boundaries, the food system exceeds five, including land system change, biosphere integrity, freshwater change, biogeochemical flows and greenhouse gas emissions [2].

In parallel with its environmental impacts, the food system has a strong influence on human health [3]. There is substantial evidence linking dietary patterns to the development of non-communicable diseases, which account for the majority of global mortality. These include cardiovascular disease, type 2 diabetes, obesity, and certain cancers, and are often long-lasting conditions influenced by genetic, physiological, environmental, and behavioural factors. Diets characterised by low intake of fruits, vegetables, nuts, seeds, and dietary fibre, together with high consumption of red and processed meat, are associated with an increased risk of major non-communicable diseases. Among these, cardiovascular diseases account for the largest share of mortality and are the leading cause of death in Europe. Importantly, up to 50% of cardiovascular disease cases are estimated to be preventable through dietary changes and other modifiable lifestyle factors [4].

To address these challenges in a European context, the EU-funded project SWITCH (Switching European food systems for a just transition towards healthy and sustainable dietary behaviour through knowledge and innovation) aims to promote healthier and more sustainable dietary habits. Within this project, the SWITCH intervention study evaluates the effects of dietary and behavioural interventions designed to facilitate a transition to a more sustainable diet, the SWITCH diet.

To quantify the environmental impact of dietary intake, environmental assessment tools and databases are required. Life Cycle Assessment (LCA) is commonly used to estimate the environmental impact of food products across the production chain, often expressed as carbon dioxide equivalents (CO_2e). The RISE Climate Database provides estimates of climate impact expressed as CO_2e. In contrast, the Sustainability Assessment of Foods and Diets (SAFAD) framework, developed by the Swedish University of Agricultural Sciences (SLU), includes a broader set of environmental indicators, such as cropland occupation, water use, biodiversity loss, and pesticide use, as well as social indicators such as antibiotic use and animal welfare loss. Hereafter, the RISE Food Climate Database and the SAFAD framework are referred to as the RISE database and the SAFAD database, respectively.

1.2 Aim and specific objectives

The overall aim of this thesis was to evaluate dietary environmental impact within the SWITCH study and to explore how environmental impact estimates relate to dietary adherence, blood lipid biomarkers, and database choice. To achieve this, dietary intake data were linked with sustainability assessment data from RISE and SAFAD.

The specific objectives of the thesis are to:

1. assess changes in the environmental impact of dietary intake from baseline to week 7 and week 13,
2. examine the association between adherence to the SWITCH dietary targets and dietary carbon footprint,
3. examine the association between dietary carbon footprint and blood lipid biomarkers,
4. compare dietary carbon footprint estimates from the RISE and SAFAD databases, including agreement between databases and food-level differences.

1.3 Limitations

This thesis is limited by the structure and availability of data within the anonymised SWITCH dataset. Information on intervention group allocation was not available, meaning that comparisons between intervention groups, control groups, or different levels of intervention support could not be performed. Therefore, the analyses were conducted at the combined participant level and should not be interpreted as intervention-specific effects.

The biomarker analyses were limited to selected blood lipid biomarkers available within the scope of this thesis, including total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides. Other cardiometabolic biomarkers collected within the SWITCH study were not analysed due to time constraints and the defined scope of the present thesis.

Dietary intake was assessed using self-reported 24-hour dietary recalls, and the dietary recall data did not include detailed information on food origin, production method, brand, or product-specific characteristics. Consequently, generic environmental impact values were used, and for some food items, average values or manually mapped representative items were required.

2

Theory

This chapter presents the theoretical background for the thesis. It first describes the relationship between food systems, planetary boundaries, and sustainable diets, followed by an overview of diet, cardiometabolic health, and the SWITCH intervention. The chapter then introduces methods for dietary environmental assessment, with particular focus on the RISE Database and the SAFAD framework. Finally, key considerations related to dietary assessment, intervention adherence, and food group specific environmental trade-offs are discussed.

2.1 Food Systems and Planetary Boundaries

The planetary boundaries framework defines the biophysical limits within which humanity can operate safely without destabilising the Earth system [5]. It identifies several processes that regulate planetary stability, including climate change, land-system change, biosphere integrity, freshwater change, biogeochemical flows, ocean acidification, stratospheric ozone depletion, atmospheric aerosol loading, and novel entities. Several of these boundaries have already been transgressed, indicating that current production and consumption patterns exceed the planet's safe operating space.

Food systems play a central role in this overshoot [2]. Food production contribute to several environmental pressures linked to planetary boundaries, including greenhouse gas emissions, land-system change, freshwater use, disruption of nitrogen and phosphorus cycles, biodiversity loss, pesticide use, and antimicrobial use in livestock production. Consequently, dietary change has been identified as an important strategy for reducing environmental pressure.

Food production generates greenhouse gas emissions including carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O), which are commonly expressed as carbon dioxide equivalents (CO_2e). Current global food system emissions substantially exceed proposed climate-compatible limits for the food sector [2].

2.2 Sustainable Healthy Diets and Cardiometabolic Health

Cardiovascular diseases are among the largest contributors to mortality from non-communicable diseases globally and in Europe [3], [4]. Cardiovascular diseases include disorders affecting the heart and vascular system, such as coronary heart disease, cerebrovascular disease, peripheral arterial disease, and thromboembolic disease.

Diet is an important modifiable risk factor for cardiometabolic disease. Dietary patterns associated with higher cardiometabolic risk are typically characterised by low intake of fruits, vegetables, whole grains, legumes, nuts, and dietary fibre, combined with high intake of red and processed meat, refined grains, saturated fat, and energy-dense foods [4]. In contrast, dietary patterns rich in plant-based foods, whole grains, legumes, nuts, fish, and unsaturated fats are generally associated with lower cardiometabolic risk.

2.3 The SWITCH Diet

The SWITCH project is an EU-funded research initiative focused on supporting the transition towards healthier and more sustainable dietary behaviours in Europe. The project aims to facilitate behavioural change through dietary guidance, innovation, knowledge, and increased accessibility to sustainable food choices. Overall, SWITCH focuses on the transition towards diets that support both planetary and human health.

The present thesis was conducted as a sub-study within the larger SWITCH intervention study. The overall study design, participant recruitment, intervention structure, dietary targets, and selection of SWITCH foods were established within the main SWITCH project before the start of this thesis.

Within the SWITCH project, the intervention study evaluates a dietary and lifestyle intervention designed to promote both human health and planetary sustainability. Participants were randomised into intervention groups receiving different levels of dietary advice, behavioural support, and provision of key SWITCH foods, as well as a control group receiving dietary advice only. However, due to anonymisation procedures, group allocation was not available in the analytical dataset used in this thesis. Therefore, the analyses were conducted at the combined participant level and should not be interpreted as intervention-specific effects. The intervention was conducted in three sequential blocks.

The predefined SWITCH dietary guidance focuses on five key food groups considered relevant for health and sustainability in a Nordic context:

- Vegetables: $\geq$ 300 g/day
- Fruits and berries: $\geq$ 200 g/day
- Whole grains: $\geq$ 90 g/day
- Legumes: $\geq$ 150 g/week, corresponding to $\geq$ 21.43 g/day
- Fish and seafood: $\geq$ 450 g/week, corresponding to $\geq$ 64.29 g/day

This thesis uses dietary data from participants in the SWITCH intervention study to evaluate dietary carbon footprint and broader environmental impact indicators, and to examine associations with adherence to the predefined SWITCH dietary targets and selected blood lipid biomarkers.

2.4 Environmental Assessment of Diets

Assessing the environmental impact of diets requires methods that translate reported food intake into carbon footprint estimates and other environmental impact indicators. Life cycle assessment (LCA) is commonly used for this purpose, as it estimates environmental pressures.

Depending on the database and method used, LCA-based estimates may include primary production, agricultural inputs, processing, packaging, transport, retail, preparation, consumption, and waste.

The choice of system boundary is central in LCA because it determines which life-cycle stages are included in the estimate. A cradle-to-gate approach includes impacts up to the farm, factory, or industry gate, but excludes later stages such as retail, household preparation, consumption, and waste. In contrast, a cradle-to-plate approach includes additional post-production stages, such as processing, packaging, transport, retail, and preparation, depending on the specific assessment method used. Consequently, two databases can produce different absolute carbon footprint estimates for the same dietary intake data if they apply different system boundaries.

Climate impact is commonly expressed as carbon footprint in carbon dioxide equivalents (CO_2e). This means that emissions of different greenhouse gases are converted into the amount of CO_2 that would have the same warming effect. Methane and nitrous oxide have a stronger warming effect than CO_2, and are therefore multiplied by global warming potential values. According to the IPCC AR6 report, 1 kg of biogenic methane corresponds to 27.0 kg CO_2e, 1 kg of fossil methane to 29.8 kg CO_2e, and 1 kg of nitrous oxide to 273 kg CO_2e [6].

Although carbon footprint is commonly used in studies of sustainable diets, it captures only one part of dietary environmental impact [7], [8]. Food production also affects land use, water use, biodiversity, nutrient pollution, pesticide use, antibiotic use, and animal welfare [9]. Therefore, using only carbon footprint may hide trade-offs between different sustainability dimensions, and multi-indicator approaches are needed for a more complete assessment.

2.4.1 The RISE Climate Database

The RISE Food Climate Database was developed by RISE to provide generic carbon footprint values for food products relevant to Swedish consumption [10]. The database was developed to increase awareness of the climate impact of foods and to support more informed food choices, meal planning, product development, food purchasing, and climate reporting. The database is not open-source, and full access requires leasing.

The RISE database contains carbon footprint values for approximately 800 food products and is designed for use on the Swedish market [10]. Food items in the database are linked to identifications codes in the Swedish Food Agency (SLV) database.

The values are based on LCA methodology and are expressed as kg CO_2e per kg food [10]. They represent generic carbon footprints rather than producer-specific values. The RISE database follows a cradle-to-gate approach, with values reported at the industrial or factory gate. This means that later life-cycle stages, such as packaging, retail, household or food service preparation, consumption, and waste are not included. For some foods, several carbon footprint values are available for the same SLV identification code depending on region of origin and whether the production system is conventional or organic. This allows a more specific emission factor to be selected when such information is available in the dietary data.

2.4.2 The SAFAD Framework

The SAFAD framework was developed by SLU and Stockholm Resilience Centre to assess environmental and social footprints of foods and diets [9]. SAFAD was developed in response to the need for dietary assessment tools that can evaluate more than carbon footprint. The framework is open-source.

SAFAD contains footprints for approximately 1800 food items and is designed to be compatible with dietary survey data and food consumption datasets [9]. The framework includes raw commodities, processed foods, composite foods, ready-to-eat meals, and drinks. Food items in the database are linked to food items in the SLV database and the European Food Safety Authority.

SAFAD is based on an attributional LCA approach [9]. Compared with the cradle-to-gate approach used in RISE, the SAFAD carbon footprint applies broader system boundaries that are closer to a cradle-to-plate perspective, including processing, preparation, packaging, and transport to the warehouse. For composite foods and meals, recipes can be used to decompose foods into ingredients or raw primary commodities before environmental impacts are calculated. Food origins can be represented using weighted averages based on trade data, and country-specific production data are used where available.

SAFAD includes several environmental and social indicators [9]. The indicators included in this thesis were:

- Carbon footprint, expressed as kg CO_2e
- Carbon dioxide emissions
- Fossil methane emissions
- Biogenic methane emissions
- Nitrous oxide emissions
- Cropland occupation
- Water use
- New nitrogen input
- New phosphorus input
- Ammonia emissions
- Biodiversity loss
- Pesticide use
- Antibiotic use
- Animal welfare loss

The system boundaries differ between the carbon footprint indicator and the other SAFAD indicators [9]. The SAFAD carbon footprint includes emissions from primary production, including farming or fishing and the production of farm inputs, as well as land-use change, processing, preparation, packaging, and transport to the warehouse. Carbon footprint results can also be disaggregated into individual greenhouse gases and life cycle stages. For the other indicators, such as cropland occupation, water use, nutrient inputs, ammonia emissions, biodiversity loss, pesticide use, antibiotic use, and animal welfare loss, the focus is mainly on impacts occurring during primary production.

2.5 Cardiometabolic Risk Assessment of Diets

Cardiometabolic risk can be assessed using circulating biomarkers such as total cholesterol, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, glucose, insulin, and inflammatory markers [11], [12]. In this thesis, the analyses were limited to blood lipid biomarkers, specifically total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides. These biomarkers were selected because they are clinically relevant indicators of lipid metabolism and cardiovascular disease risk, and because they may respond to dietary modification within the time frame of the SWITCH intervention.

Cholesterol is transported in the blood by lipoproteins, mainly including LDL and HDL particles [12]. Total cholesterol reflects the overall concentration of cholesterol in the blood, while LDL cholesterol contributes to cholesterol accumulation in the arterial wall. In contrast, HDL cholesterol is involved in reverse cholesterol transport, carrying excess cholesterol back to the liver for excretion. Triglycerides represent circulating fats used for energy metabolism and storage, and elevated levels may indicate an unfavourable cardiometabolic profile, particularly together with low HDL cholesterol. Together, elevated total cholesterol, elevated LDL cholesterol, elevated triglycerides, and low HDL cholesterol are generally considered to reflect a more unfavourable cardiometabolic risk profile.

Blood lipid biomarkers are relevant in this context because they are closely related to both diet quality and cardiometabolic health. Lipid concentrations can respond to dietary modification within weeks, although the magnitude of response depends on the type and intensity of dietary change, baseline metabolic status, medication use, body weight, physical activity, individual variation, and adherence to the intervention [13], [14]. Clinical guidelines also recommend reassessing lipid measurements after 4 to 12 weeks to evaluate response to lipid-lowering treatment and lifestyle changes [15]. These characteristics make lipid biomarkers suitable outcomes for a short-term dietary intervention study such as SWITCH.

In contrast, markers of glucose metabolism, such as glucose and insulin, may be less sensitive to short-term dietary change in individuals without diabetes or impaired glucose regulation, as glucose homeostasis is tightly regulated. Inflammatory markers may also be influenced by many non-dietary factors and were outside the defined scope of this thesis.

2.6 Dietary Drivers of Environmental Impact and Blood Lipid Biomarkers

The environmental and health-related outcomes examined in this thesis are influenced by different dietary drivers [9], [12]. Dietary carbon footprint and the SAFAD indicators are determined by the amount of each food consumed and the environmental or social impact factor assigned to that food. Blood lipid biomarkers, on the other hand, are influenced by dietary composition, metabolism, and lifestyle factors.

2.6.1 Drivers of Dietary Carbon Footprint

Dietary carbon footprint is expressed as CO_2e and reflects the combined climate impact of different greenhouse gases, including CO_2, CH_4, and N_2O [2], [4], [9]. In food systems, CO_2 emis-

sions are often related to energy use, processing, transport, packaging, and land-use change, while CH_4 is mainly associated with ruminant digestion, manure, and rice production, and N_2O with fertiliser use, soil processes, and manure.

Dietary carbon footprint is determined by both the amount of each food consumed and the emission factor assigned to that food. Animal-source foods, especially ruminant meat such as beef and lamb, are often major contributors because livestock production is associated with methane emissions from enteric fermentation, manure management, feed production, and land use [1], [2], [4]. Dietary patterns with lower climate impact are therefore often characterised by lower intake of animal-source foods, particularly beef and lamb, and higher intake of plant-based foods.

However, plant-based foods are not environmentally uniform. Their impacts can vary depending on crop type, yield, production region, irrigation needs, fertiliser use, pesticide use, processing, storage, and transport [1], [9]. For example, some nuts, coffee, and imported or highly processed plant-based products may have higher impacts than other plant-based foods depending on production conditions and supply chains. Composite foods and ready-to-eat meals can also have variable carbon footprints because their impact depends on ingredient composition, processing, packaging, and preparation.

2.6.2 Drivers of SAFAD Environmental and Social Indicators

SAFAD includes several indicators beyond total carbon footprint, capturing different environmental and social pressure points in food systems [9]. These indicators are reported per kg of food product and are influenced by different production processes, food groups, production systems, and geographic conditions.

Carbon dioxide emissions represent total CO_2 emissions, while fossil methane, biogenic methane, and nitrous oxide represent the separate greenhouse gas components behind the carbon footprint [9]. CO_2 emissions are mainly linked to fossil energy use, land-use change, processing, packaging, and transport. Biogenic methane is mainly driven by ruminant products, especially beef, lamb, and dairy, due to enteric fermentation and manure management. Nitrous oxide is mainly linked to fertiliser use, soil processes, and manure management. Plant-based foods and non-ruminant animal products generally contribute less to biogenic methane, although crop production can still contribute to nitrous oxide through fertiliser use.

Cropland occupation represents the amount of cropland required to produce food, including crops for direct human consumption and feed crops for animal production [9]. Animal-source foods can therefore contribute substantially through feed production, especially ruminant meat. Fruits, vegetables, and root crops often have lower cropland occupation because of high yields. In contrast, some nuts and legumes can have higher cropland occupation when yields are low, for example cashew nuts and lentils.

New nitrogen input represents added reactive nitrogen from mineral fertiliser and biological nitrogen fixation, while new phosphorus input represents added phosphorus from mineral fertiliser [9]. These indicators are driven by fertiliser use in crop production and by feed production for animal-source foods. Meat, milk, and eggs can therefore contribute indirectly through feed demand, while fertilised crops contribute directly. High-yielding crops with lower fertiliser requirements generally contribute less per kg food.

Blue water use represents consumptive use of freshwater in food production [9]. This indicator is strongly influenced by irrigation requirements and production location. Nuts can be important drivers because nut production often requires irrigation, and some fruits and vegetables may also contribute when grown in water-demanding regions. Rain-fed crops and foods produced in regions with lower irrigation requirements generally contribute less.

Ammonia emissions represent NH_3 emissions, mainly from manure management and fertiliser use [9]. Livestock products, particularly cattle and other ruminant products, can therefore be important drivers through manure production and handling. Crop production can also contribute through fertiliser application, while plant-based foods with lower fertiliser requirements generally contribute less.

Biodiversity loss represents biodiversity impact from land use and is expressed as extinctions per million species years [9]. This indicator depends on both the amount of land required and the ecological sensitivity of the production location. Ruminant meat and low-yielding crops, including some nuts, seeds, and legumes, can contribute more strongly to biodiversity loss. Tropical products such as coffee and bananas can also have high biodiversity impacts depending on origin, while high-yielding crops such as tomatoes, cucumbers, carrots, and cabbages from Northern Europe generally contribute less.

Pesticide use represents the amount of active pesticide ingredients used in crop cultivation, including crops used for animal feed [9]. This indicator is mainly driven by crop type, pest pressure, and agricultural management. Nuts are highlighted as one food group that can have relatively high pesticide use, although the SAFAD article notes uncertainty in this indicator. Fruits, vegetables, and other crop-based foods can also vary substantially depending on production practices.

Antibiotic use is a social indicator representing an index of antibiotic use in animal-based food production and the associated risk of antibiotic resistance development [9]. It varies by animal species, country, and production system. Pork can contribute substantially depending on production practices, while plant-based foods do not contribute directly to this indicator.

Animal welfare loss is a social indicator representing an index of the negative impact on animal welfare per kg of product [9]. The index is based on the number of affected animals, the animals' ability to perceive negative effects, and suffering related to the production system, including mortality, disease frequency, space allowance, and slaughter process. Poultry is especially relevant because chicken can have a lower carbon footprint than ruminant meat but higher animal welfare loss, since many animals are required to produce a given amount of food. Plant-based foods do not contribute directly to animal welfare loss.

Overall, the SAFAD indicators show that different foods drive different types of impact [9]. Ruminant meat is important for methane, cropland occupation, nutrient inputs, ammonia emissions, and biodiversity loss. Nuts and some crop-based foods can be important for blue water use, pesticide use, and biodiversity loss. Pork and poultry are particularly relevant for antibiotic use and animal welfare loss. Therefore, foods with lower climate impact do not necessarily have lower impacts across all SAFAD indicators.

2.6.3 Dietary Drivers of Blood Lipid Biomarkers

Blood lipid biomarkers are influenced by overall dietary pattern, dietary fat quality, fibre intake, carbohydrate quality, and energy balance [2], [4]. A high intake of saturated-fat-rich foods, including butter, high-fat dairy products, cheese, fatty red meat, processed meat, palm oil, and coconut oil, can contribute to higher LDL cholesterol. In contrast, replacing saturated-fat-rich foods with sources of unsaturated fats, such as vegetable oils, nuts, seeds, and fish, is associated with a more favourable lipid profile.

Fibre-rich plant foods are relevant for lipid metabolism partly because they provide dietary fibre, plant protein, and unsaturated fats [13]. Specific cholesterol-lowering components, such as plant sterols, soy protein, and viscous fibre, have also been shown to reduce serum lipids in adults with hypercholesterolaemia.

Triglyceride concentrations are influenced by energy balance and carbohydrate quality, since high intake of refined carbohydrates, added sugars, and alcohol may increase hepatic triglyceride production [4], [12]. Elevated triglyceride concentrations are associated with an unfavourable cardiometabolic profile and increased cardiovascular risk, whereas fish and other sources of long-chain omega-3 fatty acids may contribute to lower triglyceride concentrations.

3

Methods

This chapter describes the study population, study design, data collection, data processing procedures, environmental impact assessment, quality control, statistical analyses, and ethical considerations.

3.1 Study Participants

A total of 174 adults aged 18 to 70 years with a body mass index (BMI) between 25 and 35 kg/m^2 were enrolled in the SWITCH intervention study. Participants were weight stable, had not followed a weight-reducing diet during the preceding four months, and were willing to follow the intervention diet.

Exclusion criteria included clinically relevant disease, recent cardiovascular events, diabetes, or use of medications affecting glucose metabolism. Additional exclusion criteria included gastrointestinal disease or surgery, renal or hepatic dysfunction, anaemia, uncontrolled hypertension, hypercholesterolaemia, or elevated fasting glucose. Individuals with incompatible dietary habits, food allergies related to the intervention, high levels of physical activity, pregnancy or lactation, substance abuse, or participation in other clinical studies were also excluded.

3.2 Study Design and Data Collection

The SWITCH study included 300 participants divided into five blocks. At the time of analysis, data collection had been completed for three of these blocks. The participants in the SWITCH intervention study attended three clinical visits: baseline at week 0, midpoint at week 7, and end of intervention at week 13. At each visit, anthropometric measurements, blood samples, questionnaire data, and dietary intake data were collected. Anthropometric data were used for participant characterisation, while blood samples provided information on blood lipid biomarkers. An overview of the study design is shown in Figure 3.1.

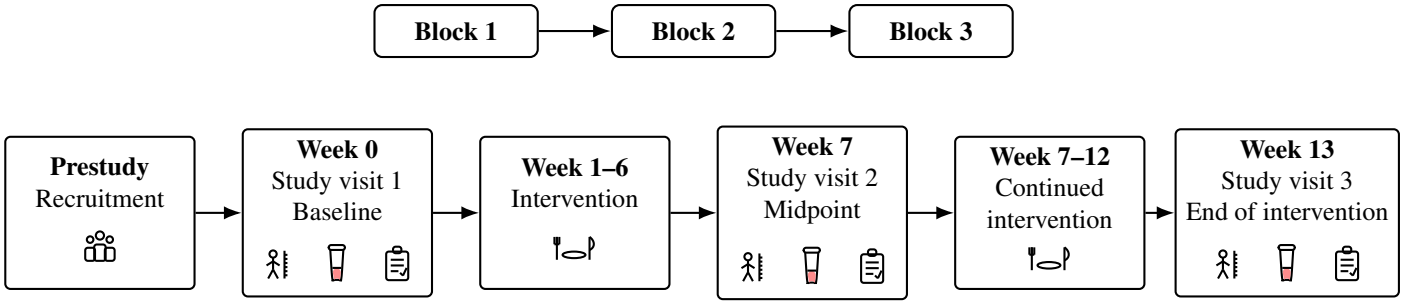


Figure 3.1: Overview of the clinical visits, intervention period, and sequential study blocks in the SWITCH study. The same study design was implemented across three sequential blocks. Participants attended study visits at week 0, week 7, and week 13. At each visit, anthropometric measurements, blood samples, questionnaire data, and dietary recalls were collected.

Dietary intake was assessed using three 24-hour dietary recalls per visit collected with RiksmatenFlex, a web-based dietary assessment tool developed for Swedish dietary surveys. Questionnaire and participant data were collected using REDCap, a secure web-based platform for data collection and study management.

Anthropometric data were used for participant characterisation, while blood samples provided information on blood lipid biomarkers, including total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides. Dietary intake data formed the basis for both environmental impact assessment and analyses of adherence to the SWITCH dietary targets.

Due to anonymisation procedures, group allocation and level of intervention support were not available in the analytical dataset. Consequently, all analyses were conducted on the combined study population, and changes over time were interpreted at the overall participant level rather than as group-specific intervention effects.

3.3 Data Processing and Environmental Impact Assessment

All data processing was performed in R (version 4.4.3) using reproducible scripts. The workflow was initially tested and validated on a subset of the data before being applied to the full dataset.

3.3.1 Data Processing Workflow

The data processing workflow is shown in Figure 3.2.

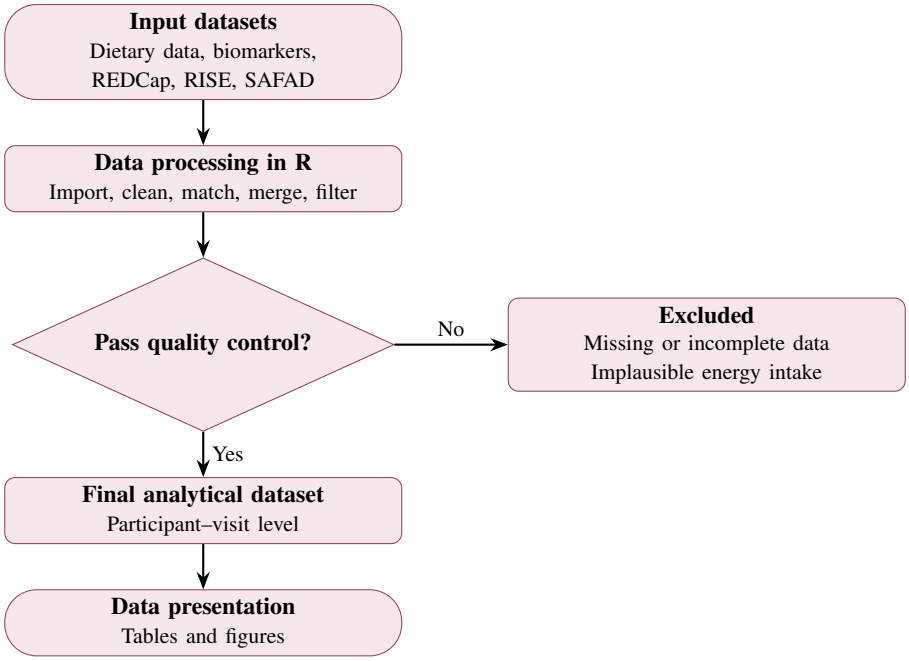


Figure 3.2: Overview of the data processing workflow in R, including data cleaning, quality control, and generation of the final analytical dataset.

3.3.2 Input Data and Data Cleaning

All datasets were imported from Excel files into R. Variable names were standardised and translated when necessary, and relevant variables were selected for further analysis.

Dietary intake data were obtained from RiksmatenFlex and included reported food items, food identification codes, consumed amounts in grams, and dietary recall days for each participant. Additional dietary data containing energy and nutrient composition were used to estimate total energy intake and support quality control procedures. Dietary status data were used to identify completed and valid dietary registrations.

Participant characteristics, including age, sex, body weight, height, and physical activity level, were obtained from REDCap and used for descriptive statistics and estimation of energy requirements. Biomarker data from the SWITCH study included several cardiometabolic markers. In the present thesis, the analyses were limited to blood lipid biomarkers, including total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides.

Environmental impact data were obtained from the RISE database and the SAFAD database. These datasets provided environmental impact values for food items and were used to estimate the environmental impact of reported dietary intake.

3.3.3 Environmental Impact Calculation

Dietary intake data from RiksmatenFlex were linked to environmental impact values from RISE and SAFAD using SLV food identification codes. For RISE food items with multiple conventional emission factors for the same food identification code, the mean value across countries of origin was used, as the dietary recall data did not include information on food origin. Organic

values were excluded from this average. Food items without direct matches in the environmental datasets were manually mapped to the closest available representative item. The extent of manual mapping and the largest manually mapped food items are described in Appendix B.

Environmental impact was calculated at the food item level by multiplying the reported intake amount of each food item, converted from grams to kilograms, by its corresponding environmental impact value. These food-level estimates were then summed to obtain total daily environmental impact for each participant and registration day. Daily totals were subsequently averaged across valid registration days within each participant visit.

Finally, the environmental impact estimates were merged with the dietary and participant-level datasets to create a single analytical dataset at the participant–visit level.

3.4 Quality Control and Exclusion Criteria

A quality control procedure was applied to define the final analytical sample. Participants were excluded if they were test entries, had incomplete registrations, reported fewer than two valid dietary recall days for a visit, or had missing dietary intake data.

Implausible energy intake was identified using the Goldberg cut-off method [16], which compares reported energy intake with estimated energy requirements based on basal metabolic rate (BMR) and physical activity level (PAL). BMR was estimated using the Schofield equations, and PAL was assigned from self-reported physical activity data. Both lower and upper cut-offs were applied, allowing identification of both underreporters and overreporters. Participants classified as underreporters or overreporters were excluded, and only participants with plausible reported energy intake were retained in the final analytical dataset.

3.4.1 Basal Metabolic Rate

Basal metabolic rate (BMR) was estimated using the Schofield equations [17]. These equations are commonly used in the Goldberg cut-off method and estimate BMR based on sex, age, and body weight.

Men:

$$\text{30–60 years:} \quad BMR = 11.472 \cdot W + 873.1 \tag{3.1}$$

$$\text{>60 years:} \quad BMR = 11.711 \cdot W + 587.7 \tag{3.2}$$

Women:

$$\text{18–30 years:} \quad BMR = 14.818 \cdot W + 486.6 \tag{3.3}$$

$$\text{30–60 years:} \quad BMR = 8.126 \cdot W + 845.6 \tag{3.4}$$

$$\text{>60 years:} \quad BMR = 9.082 \cdot W + 658.5 \tag{3.5}$$

where W represents body weight in kilograms and BMR is expressed in kcal/day.

3.4.2 Physical Activity Level

Physical activity level (PAL) was used to estimate total energy requirements. Estimated energy requirement was calculated as:

$$EER = BMR \times PAL \tag{3.6}$$

where EER represents estimated energy requirement in kcal/day.

PAL values were assigned based on self-reported physical activity levels obtained from REDCap. Participants were categorised into four activity levels corresponding to PAL values of 1.4, 1.6, 1.8, and 2.0. For each participant, the average PAL across available visits was used in the Goldberg cut-off calculation.

3.4.3 Goldberg Cut-off Method

The Goldberg cut-off method was applied to assess the plausibility of reported energy intake. The method is based on the ratio between reported energy intake (EI) and basal metabolic rate (BMR):

$$\frac{EI}{BMR} \tag{3.7}$$

where EI represents the mean reported energy intake across valid dietary assessment days.

To account for variation in energy intake, BMR, and physical activity level, lower and upper confidence limits were calculated. The lower confidence limit was used to identify underreporters, while the upper confidence limit was used to identify overreporters.

The lower confidence limit for acceptable reporting was calculated as:

$$\frac{EI}{BMR} > PAL \times \exp\left(-2 \times S\right) \tag{3.8}$$

The upper confidence limit for acceptable reporting was calculated as:

$$\frac{EI}{BMR} < PAL \times \exp\left(2 \times S\right) \tag{3.9}$$

where PAL represents physical activity level and S is the combined coefficient of variation.

The coefficient S was calculated as:

$$S = \sqrt{\frac{CV_{wEI}^2}{d} + CV_{wB}^2 + CV_{tP}^2} \tag{3.10}$$

where CV_{wEI} represents within-subject variation in energy intake, CV_{wB} represents variation in estimated BMR, CV_{tP} represents between-subject variation in physical activity level, and d is the number of valid dietary assessment days.

Standard values for the coefficients of variation were applied according to [16], where $CV_{wEI} = 23\%$, $CV_{wB} = 8.5\%$, and $CV_{tP} = 15\%$.

Participants with an EI/BMR ratio below the lower confidence limit were classified as underreporters, while participants with an EI/BMR ratio above the upper confidence limit were classified as overreporters. Both underreporters and overreporters were excluded from the final analytical dataset.

3.5 SWITCH Adherence Score

Adherence to the SWITCH dietary targets was assessed at each visit using five predefined food groups: vegetables, fruits and berries, whole grains, legumes, and fish and seafood.

In the main analysis, adherence was assessed using a graded score to capture the degree of target achievement. For each food group, adherence was calculated as the ratio between reported intake and the predefined intake target:

$$Adherence_{ij} = \frac{Intake_{ij}}{Target_j} \tag{3.11}$$

where $Intake_{ij}$ represents the reported intake for participant i in food group j, and $Target_j$ represents the corresponding SWITCH dietary target. Adherence values above 1 were truncated to 1, meaning that participants who met or exceeded a target received the maximum score for that food group.

The total SWITCH adherence score was calculated by summing the five food group specific adherence scores:

$$Total\ adherence_i = \sum_{j=1}^{5} Adherence_{ij} \tag{3.12}$$

The total score ranged from 0 to 5, where higher values indicated greater adherence to the SWITCH dietary targets.

In addition, a binary adherence variable was calculated for each food group to describe the proportion of participants meeting each predefined target. Participants were classified as meeting the target if their intake was equal to or above the target value, and as not meeting the target if their intake was below the target.

The dietary targets were defined as follows:

- Vegetables: $\geq$ 300 g/day
- Fruits and berries: $\geq$ 200 g/day

- Whole grains: $\geq$ 90 g/day
- Legumes: $\geq$ 21.43 g/day, corresponding to $\geq$ 150 g/week
- Fish and seafood: $\geq$ 64.29 g/day, corresponding to $\geq$ 450 g/week

Adherence was summarised as mean intake, mean adherence score, and percentage of participants meeting each target at baseline, visit 2, and visit 3.

It should be noted that the SWITCH adherence score did not include meat consumption as a separate component. Therefore, participants could achieve higher adherence to the included SWITCH targets while still maintaining intake of meat or other carbon-intensive foods.

3.6 Blood Lipid Biomarkers

Blood lipid biomarkers were used to explore associations between dietary carbon footprint and health-related outcomes. Biomarker data included total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides. Biomarker data were linked to dietary intake and environmental impact data at the participant–visit level. At the time of analysis, complete biomarker data were only available for participants in block 1 and block 2 of the SWITCH study, corresponding to 67 participants and 173 participant-visits.

3.7 Statistical Analysis

All statistical analyses and data visualisations were performed in R using the packages `readxl`, `dplyr`, `tidyr`, `readr`, `here`, `tibble`, `lme4`, `lmerTest`, `purrr`, `stringr`, `openxlsx`, and `ggplot2`.

Descriptive statistics, including means and standard deviations, were calculated to summarise participant characteristics, dietary intake, SWITCH adherence, environmental impact measures, and blood lipid biomarkers.

Normality was assessed using the Shapiro–Wilk test together with visual inspection of Q–Q plots and histograms. Environmental impact variables showed deviations from normality before transformation. After log-transformation, the distributions more closely approximated normality and therefore, log-transformed values were used in the statistical models. Q–Q plots used to assess normality before and after log-transformation are provided in Appendix A.2.

Linear mixed-effects models were used to analyse changes in environmental impact across visits and to examine associations between dietary carbon footprint, SWITCH adherence, and blood lipid biomarkers. Participant was included as a random intercept in all models to account for repeated measurements.

Changes in environmental impact across visits were analysed using the following model:

$$\log(Y_{ij}) = \beta_0 + \beta_1 Visit2_{ij} + \beta_2 Visit3_{ij} + u_i + \varepsilon_{ij} \tag{3.13}$$

where Y_{ij} represents the environmental impact outcome for participant i at visit j, and baseline was used as the reference category.

Associations between SWITCH adherence and carbon footprint were analysed using:

$$\log(CF_{ij}) = \beta_0 + \beta_1 Adherence_{ij} + \beta_2 Visit_{ij} + \beta_3 Age_i + \beta_4 Sex_i + \beta_5 BMI_i + \beta_6 Energy_{ij} + u_i + \varepsilon_{ij} \tag{3.14}$$

where CF_{ij} represents carbon footprint, and $Adherence_{ij}$ represents either the total SWITCH adherence score or one of the five SWITCH adherence components (vegetables, fruits and berries, whole grains, legumes, or fish and seafood). Each adherence component was analysed in a separate model (single-component analysis).

Associations between dietary carbon footprint and blood lipid biomarkers were analysed using:

$$Biomarker_{ij} = \beta_0 + \beta_1 Z[\log(CF_{ij})] + \beta_2 Visit_{ij} + \beta_3 Age_i + \beta_4 Sex_i + \beta_5 BMI_i + \beta_6 Energy_{ij} + u_i + \varepsilon_{ij} \tag{3.15}$$

where $Biomarker_{ij}$ represents the blood lipid biomarker for participant i at visit j, and $Z[\log(CF_{ij})]$ represents standardised log-transformed dietary carbon footprint. The coefficient β_1 represents the absolute difference in the biomarker per one standard deviation higher log-transformed dietary carbon footprint. In all models, u_i represents the participant-specific random intercept and ε_{ij} represents the residual error. Models were fitted separately for each biomarker and separately for RISE and SAFAD carbon footprint estimates.

Agreement between RISE and SAFAD carbon footprint estimates was assessed using correlation analysis and Bland–Altman analysis. Mean differences between RISE and SAFAD were calculated across visits to assess systematic differences in daily carbon footprint estimates. In addition, food-level comparisons were performed to identify food items with the largest differences in emission factors and total carbon footprint contributions between databases.

Results are presented as means with standard deviations, model coefficients, confidence intervals, and p-values where appropriate. For log-transformed outcomes, model coefficients were converted to percentage change using the following formula:

$$\%\Delta = (\exp(\beta) - 1) \times 100 \tag{3.16}$$

To account for multiple testing, p-values were adjusted using the Benjamini–Hochberg false discovery rate method within each analysis family. Statistical significance was defined as an FDR-adjusted p-value below 0.05.

3.8 Ethics

This thesis was conducted within the framework of the SWITCH intervention study. The overall intervention aimed to improve public health outcomes while reducing environmental impacts associated with food consumption.

The SWITCH intervention study received ethical approval from the Swedish Ethical Review Authority, and no additional data collection was conducted within this thesis. Participants were informed about the study, and participation was voluntary. Participant data were entered into an electronic case report form (eCRF) in REDCap and coded using unique study identification numbers to limit the risk of individual identification.

Biological samples were collected as part of the main SWITCH study protocol and stored in the hospital biobank, which is part of Biobank West. Sample handling and storage were conducted in accordance with the Swedish Biobanks in Medical Care Act. Blood samples were coded to ensure traceability within the biobank, while other biological samples and study data were coded using unique study identification numbers. All data were handled in accordance with ethical guidelines and the EU General Data Protection Regulation (GDPR).

In addition to data protection and biobank-related considerations, the dietary intervention itself raises ethical considerations. Changing dietary habits may affect participants' daily routines, food preferences, cultural practices, social eating situations, and perceived quality of life. Although the SWITCH diet was designed to support healthier and more sustainable eating patterns, dietary changes may still have been experienced as burdensome by some participants, particularly if they required changes in shopping habits, cooking practices, food availability, costs, or family meals.

The dietary advice was intended to support gradual dietary change rather than impose restrictive eating patterns. Potential risks related to the dietary intervention were considered limited, but participants may still have experienced practical challenges related to satiety, adapting meals to personal preferences, or maintaining dietary changes over time. These aspects are relevant when interpreting adherence to the intervention, as lower adherence may reflect practical or personal barriers rather than unwillingness to follow the dietary advice.

The research context of this thesis is closely aligned with the 2030 Agenda for Sustainable Development, developed by the United Nations, which comprises 17 Sustainable Development Goals (SDGs) aimed at guiding global development toward sustainability [18]. Food systems are a central component of this framework, as they are directly linked to several SDGs, including SDG 2 (Zero Hunger), SDG 3 (Good Health and Well-being), SDG 12 (Responsible Consumption and Production), SDG 13 (Climate Action), SDG 14 (Life Below Water), and SDG 15 (Life on Land).

4

Results

This chapter presents the main results of the thesis. It begins with a description of participant characteristics after quality control, followed by analyses of changes in environmental impact across study visits, adherence to the SWITCH dietary targets, and associations with blood lipid biomarkers for both RISE and SAFAD. Finally, the RISE and SAFAD databases are compared in terms of their carbon footprint estimates.

4.1 Participant characteristics after quality control

After quality control, 113 of the 174 initially available participants remained in the analytical dataset. Participants were excluded due to incomplete or insufficient dietary registrations, missing dietary intake data, test registrations, or implausible energy intake according to the Goldberg cut-off method. The analytical sample consisted of 95 women and 18 men, with a mean age of 55.2 years and a mean BMI of 28.8 kg/m^2. The number of participants with dietary data decreased across visits, with 110 participants at baseline, 97 at visit 2, and 87 at visit 3. Biomarker data were available for participants from Blocks 1 and 2, corresponding to 67 participants, 59.3% of the analytical sample, and 173 participant-visits. Participant characteristics are summarised in Table 4.1, while a more detailed overview before and after quality control is provided in Appendix A.1.

Table 4.1: Participant characteristics after quality control.

Characteristic	After QC
Participants	113
Female / male	95 / 18
Age (years)	55.2 ± 9.0
BMI (kg/m^2)	28.8 ± 2.7
Energy intake (kcal/day)	2105 ± 455
Food intake (g/day)	3139 ± 768
Participants per visit, V1 / V2 / V3	110 / 97 / 87
Participants with 1 / 2 / 3 visits	12 / 21 / 80
Visits with 1 / 2 / 3 recall days	0 / 48 / 246
Block 1 / 2 / 3	37 / 30 / 46
Dietary data coverage, %	100
Biomarker coverage, %	58.8

4.2 Adherence to SWITCH targets

Adherence to the predefined SWITCH dietary targets across study visits is presented in Table 4.2. Overall, intake increased from baseline to visit 2 for all food groups. At visit 3, intake declined slightly for fruit and berries, legumes, and fish and seafood, but generally remained higher than at baseline.

Table 4.2: Adherence to SWITCH dietary targets across visits. Intake values are presented as mean ± standard deviation and percentage of participants meeting the predefined target. The number of participants included was 110 at baseline, 97 at visit 2, and 87 at visit 3.

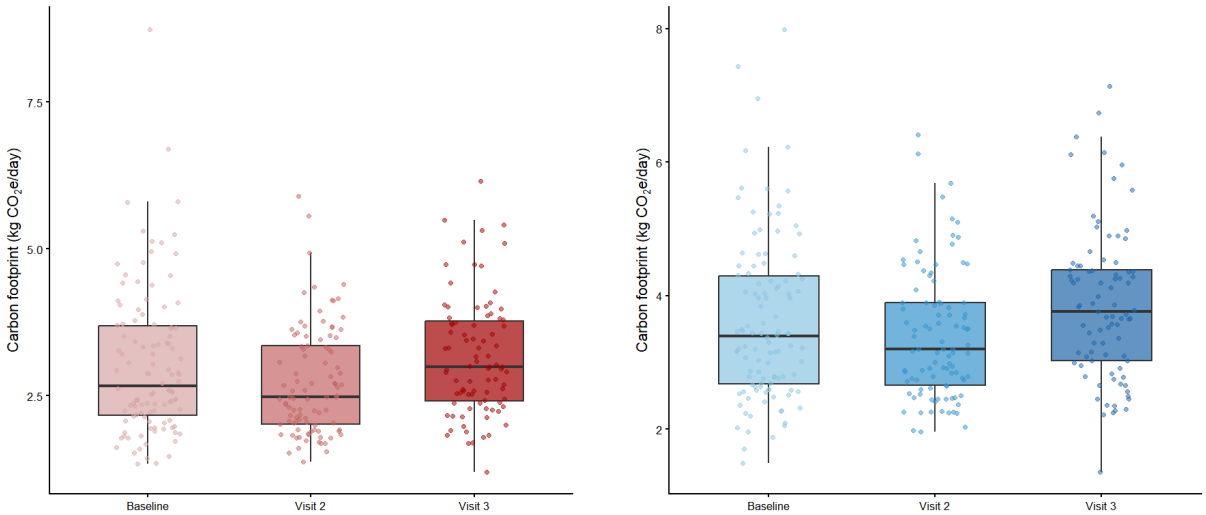
		Baseline		**Visit 2**		**Visit 3**	
Food group	**SWITCH target**	**Mean ± SD**	**Met (%)**	**Mean ± SD**	**Met (%)**	**Mean ± SD**	**Met (%)**
Vegetables	≥300 g/day	201.0 ± 104.7	14.5	287.0 ± 113.2	45.4	287.1 ± 122.7	47.1
Fruit & berries	≥200 g/day	177.4 ± 118.4	35.5	254.1 ± 151.6	64.9	248.6 ± 158.9	59.8
Whole grains	≥90 g/day	48.6 ± 23.1	5.5	79.7 ± 33.1	30.9	82.4 ± 52.9	32.2
Legumes	≥21.43 g/day	24.3 ± 33.3	40.0	57.8 ± 49.1	70.1	48.3 ± 56.1	62.1
Fish & seafood	≥64.29 g/day	34.8 ± 46.6	19.1	73.7 ± 49.3	54.6	67.4 ± 57.7	46.0

Additional descriptive subgroup contribution analyses are presented in Appendix C. These analyses show mean daily intake and carbon footprint contributions for selected food subgroups across visits and were used to support interpretation of the observed changes in environmental impact.

4.3 Changes in environmental impact across visits

Changes in environmental impact across study visits are shown in Figure 4.1, Table 4.3, and Figure 4.2. Figure 4.1 presents CO_2e estimates from both the RISE Climate Database and the SAFAD framework, whereas Table 4.3 and Figure 4.2 include CO_2e together with the full set of environmental impact indicators from SAFAD.

Figure 4.1 shows a decrease in CO_2e from baseline to visit 2, followed by an increase at visit 3, for both RISE and SAFAD. The boxplots show some outlying values, indicating substantial variation in dietary carbon footprint between participants. These values were retained because they passed the predefined quality control procedures and may reflect plausible variation in intake of carbon-intensive foods. However, remaining misreporting cannot be fully excluded despite Goldberg filtering.



(a) RISE Climate Database **(b)** SAFAD framework

Figure 4.1: Carbon footprint (kg CO_2e/day) across baseline, visit 2, and visit 3, estimated using the RISE Climate Database (a) and the SAFAD framework (b).

Table 4.3 shows that several environmental indicators changed significantly after adjustment for multiple testing using the Benjamini–Hochberg false discovery rate method.

At visit 2, significant reductions were observed for cropland use ($\beta = -0.09$), nitrogen input ($\beta = -0.14$), ammonia emissions ($\beta = -0.35$), and biogenic methane ($\beta = -0.34$). In contrast, carbon dioxide emissions ($\beta = 0.10$) and animal welfare loss ($\beta = 0.45$) increased significantly. The increase in animal welfare loss was notable in the context of an intervention aimed at promoting a more sustainable dietary pattern.

At visit 3, significant increases were observed for biodiversity loss ($\beta = 0.12$), pesticide use ($\beta = 0.19$), and carbon dioxide emissions ($\beta = 0.17$), while ammonia emissions remained significantly reduced compared with baseline ($\beta = -0.24$).

Table 4.3: Environmental impact indicators across study visits. Values are presented as mean ± standard deviation per day. The number of participants included was 110 at baseline, 97 at visit 2, and 87 at visit 3. Asterisks indicate statistical significance after Benjamini–Hochberg false discovery rate adjustment: * $p_{\text{FDR}} < 0.05$, ** $p_{\text{FDR}} < 0.01$, *** $p_{\text{FDR}} < 0.001$.

	Baseline	**Visit 2**		**Visit 3**	
Indicator	Mean ± SD	Mean ± SD	β	Mean ± SD	β
Carbon footprint, RISE (kg CO_2e/day)	3.04 ± 1.25	2.74 ± 0.92	-0.09	3.16 ± 1.00	0.06
Carbon footprint, SAFAD (kg CO_2e/day)	3.62 ± 1.23	3.40 ± 0.96	-0.05	3.85 ± 1.11	0.08
Cropland use (m^2 yr/day)	5.25 ± 1.51	4.83 ± 1.45	-0.09*	5.42 ± 1.77	0.03
Water use (m^3/day)	0.21 ± 0.14	0.20 ± 0.09	0.00	0.23 ± 0.13	0.07
New nitrogen input (g N/day)	63.7 ± 24.4	54.5 ± 18.1	-0.14**	63.0 ± 23.7	-0.01
New phosphorus input (g P/day)	6.5 ± 1.7	6.4 ± 1.7	-0.02	7.0 ± 2.1	0.07
Ammonia emissions (g NH_3/day)	11.0 ± 6.7	7.6 ± 4.6	-0.35***	8.9 ± 5.9	-0.24*
Biodiversity loss ($\times 10^{-12}$ E/MSY/day)	1.15 ± 0.35	1.19 ± 0.36	0.03	1.30 ± 0.44	0.12**
Pesticide use (g a.i./day)	0.93 ± 0.37	0.99 ± 0.43	0.05	1.17 ± 0.55	0.19***
Use of antibiotics (index/day)	5.53 ± 3.42	6.70 ± 4.19	0.18	7.03 ± 4.81	0.13
Animal welfare loss (index/day)	6.15 ± 7.33	9.02 ± 10.85	0.45*	9.13 ± 12.90	0.28
Carbon dioxide (kg CO_2/day)	1.70 ± 0.47	1.88 ± 0.50	0.10**	2.04 ± 0.61	0.17***
Methane, fossil (kg CH_4/day)	0.004 ± 0.001	0.003 ± 0.001	-0.07	0.004 ± 0.001	0.05
Methane, biogenic (kg CH_4/day)	0.040 ± 0.026	0.027 ± 0.016	-0.34***	0.034 ± 0.021	-0.15
Nitrous oxide (kg N_2O/day)	0.003 ± 0.001	0.002 ± 0.001	-0.05	0.003 ± 0.001	0.07

Figure 4.2 shows the estimated percentage changes in environmental impact indicators at visit 2 and visit 3 compared with baseline. Values above zero indicate an increase, while values below zero indicate a decrease relative to baseline. These results are consistent with those presented in Table 4.3.

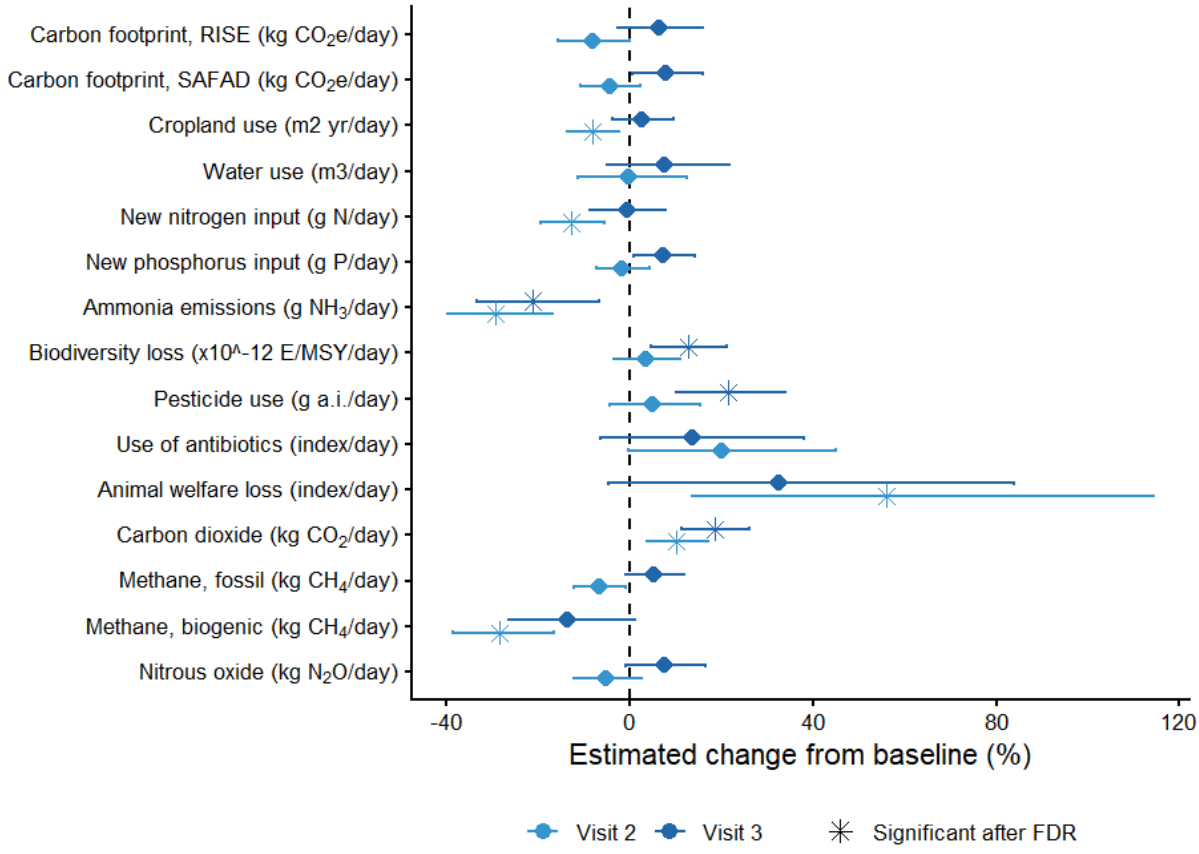


Figure 4.2: Estimated percentage changes in environmental impact indicators at visit 2 and visit 3 compared with baseline. Points represent model estimates and horizontal lines represent 95% confidence intervals. Asterisks indicate statistical significance after Benjamini–Hochberg false discovery rate adjustment.

4.4 Association between SWITCH adherence and carbon footprint

Associations between adherence to the SWITCH diet and dietary carbon footprint were analysed using both a graded adherence score and binary target adherence variables. The graded score captured the degree of target achievement, whereas the binary variables classified participants as either meeting or not meeting each predefined dietary target.

4.4.1 Graded adherence score

Associations between graded adherence and dietary carbon footprint are shown in Table 4.4 and Figure 4.3. Table 4.4 presents estimated associations between the graded SWITCH adherence variables and both RISE and SAFAD carbon footprint, while Figure 4.3 illustrates the corresponding percentage differences.

The total graded SWITCH adherence score was not significantly associated with carbon footprint after FDR adjustment. Among the individual adherence components, higher vegetable adherence was associated with a higher SAFAD carbon footprint after FDR adjustment ($\beta = 0.18$),

corresponding to a 19.8% higher carbon footprint per one-unit higher adherence score. No other statistically significant associations were observed.

Table 4.4: Associations between graded SWITCH adherence and dietary carbon footprint estimated using RISE and SAFAD. Analyses included 294 participant-visits from 113 participants. Asterisks indicate statistical significance after Benjamini–Hochberg false discovery rate adjustment: * $p_{\text{FDR}} < 0.05$, ** $p_{\text{FDR}} < 0.01$, *** $p_{\text{FDR}} < 0.001$.

	RISE carbon footprint		SAFAD carbon footprint	
SWITCH target	β	**% change (95% CI)**	β	**% change (95% CI)**
SWITCH adherence score	-0.02	-2.3 (-5.9, 1.4)	0.00	0.5 (-2.6, 3.6)
Vegetables	0.08	8.8 (-6.1, 26.0)	0.18*	19.8 (6.4, 34.9)
Fruit and berries	-0.09	-8.8 (-19.6, 3.5)	-0.02	-2.4 (-12.1, 8.3)
Whole grains	-0.13	-12.0 (-23.7, 1.4)	-0.09	-8.4 (-18.5, 2.9)
Legumes	-0.06	-6.1 (-12.8, 1.1)	-0.05	-5.3 (-10.7, 0.6)
Fish and seafood	0.02	2.4 (-6.3, 11.9)	0.08	8.8 (1.3, 16.9)

Figure 4.3 shows the estimated percentage changes in carbon footprint across graded adherence variables. These patterns are consistent with the results presented in Table 4.4.

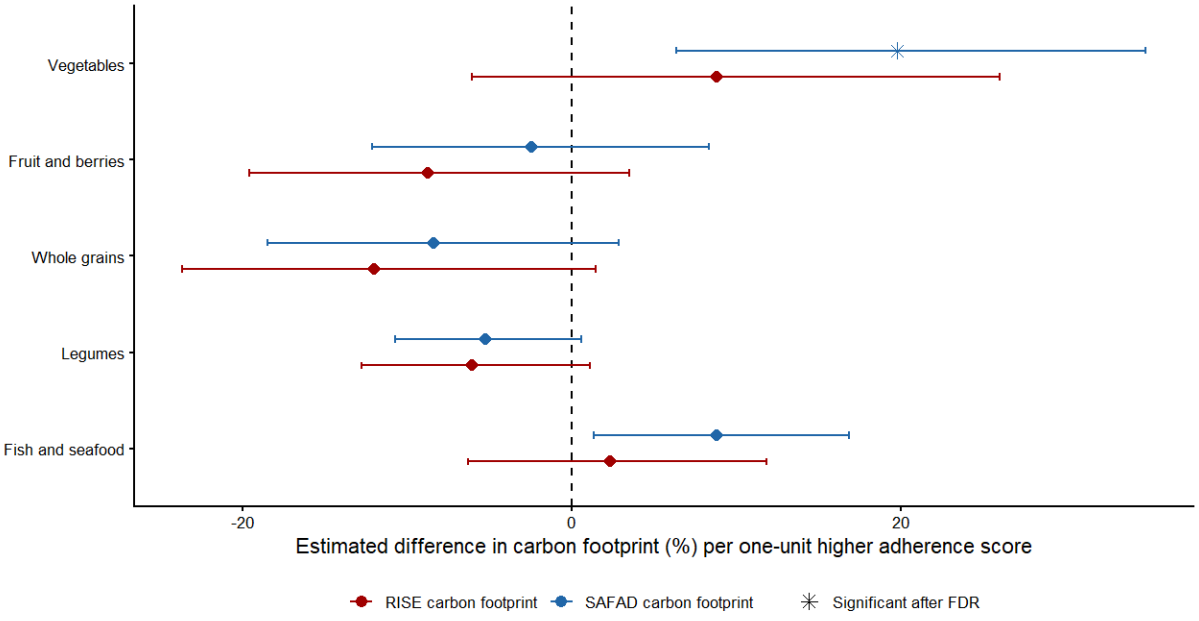


Figure 4.3: Associations between graded SWITCH adherence and dietary carbon footprint estimated using RISE and SAFAD. Points represent percentage change estimates and horizontal lines represent 95% confidence intervals. Asterisks indicate statistical significance after Benjamini–Hochberg false discovery rate adjustment.

4.4.2 Binary target adherence

Associations between binary target adherence and dietary carbon footprint are shown in Table 4.5 and Figure 4.4. Table 4.5 presents estimated associations between the binary SWITCH

adherence variables and both RISE and SAFAD carbon footprint, while Figure 4.4 illustrates the corresponding percentage differences.

The total binary SWITCH adherence score was not significantly associated with carbon footprint after FDR adjustment. Among the individual dietary targets, meeting the fish and seafood target was associated with a higher SAFAD carbon footprint after FDR adjustment ($\beta = 0.10$), corresponding to a 10.4% higher carbon footprint. No other statistically significant associations were observed.

Table 4.5: Associations between binary SWITCH target adherence and dietary carbon footprint estimated using RISE and SAFAD. Analyses included 294 participant-visits from 113 participants. Asterisks indicate statistical significance after Benjamini–Hochberg false discovery rate adjustment: * $p_{\text{FDR}} < 0.05$, ** $p_{\text{FDR}} < 0.01$, *** $p_{\text{FDR}} < 0.001$.

	RISE carbon footprint		**SAFAD carbon footprint**	
SWITCH target	β	**% change (95% CI)**	β	**% change (95% CI)**
SWITCH binary adherence score	-0.01	-1.3 (-4.0, 1.5)	0.00	0.5 (-1.8, 2.8)
Vegetables	0.02	2.0 (-5.4, 9.9)	0.03	3.2 (-2.8, 9.7)
Fruit and berries	-0.02	-2.3 (-9.6, 5.5)	0.00	0.1 (-6.1, 6.7)
Whole grains	-0.07	-7.0 (-14.7, 1.3)	-0.06	-5.6 (-12.0, 1.2)
Legumes	-0.05	-4.8 (-11.2, 2.0)	-0.05	-5.2 (-10.3, 0.2)
Fish and seafood	0.02	2.3 (-4.8, 10.0)	0.10*	10.4 (4.2, 17.0)

Figure 4.4 shows the estimated percentage changes in carbon footprint across binary adherence variables. These patterns are consistent with the results presented in Table 4.5.

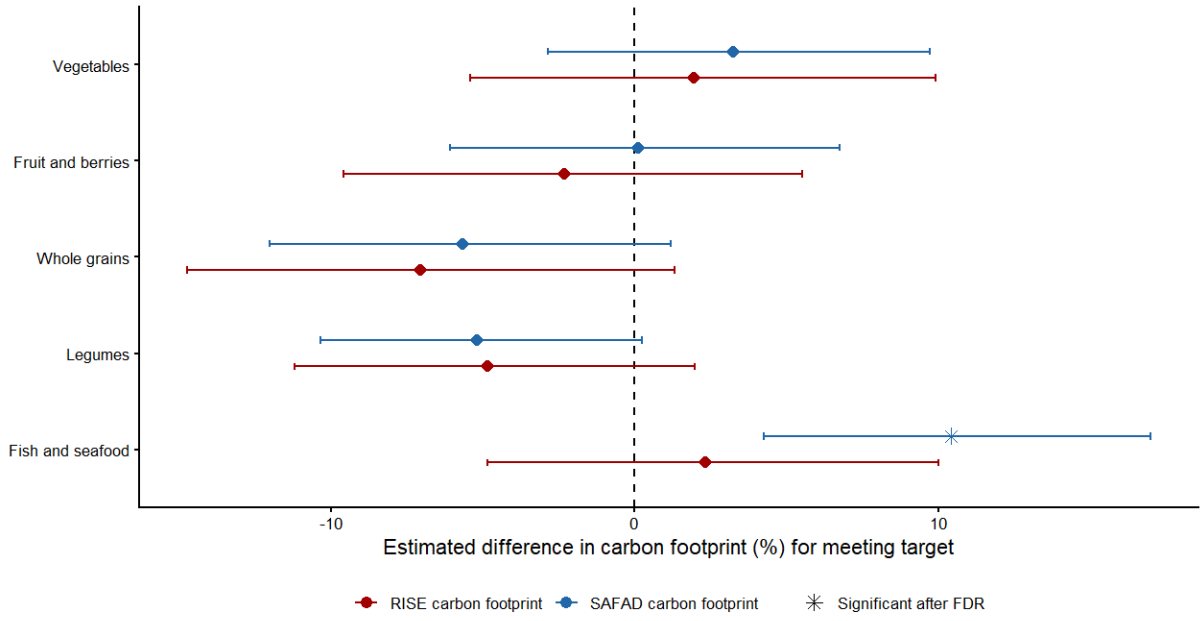


Figure 4.4: Associations between binary SWITCH target adherence and dietary carbon footprint estimated using RISE and SAFAD. Points represent percentage change estimates and horizontal lines represent 95% confidence intervals. Asterisks indicate statistical significance after Benjamini–Hochberg false discovery rate adjustment.

The graded and binary adherence analyses showed partly different association patterns. The graded analysis identified vegetable adherence as positively associated with SAFAD carbon footprint, whereas the binary analysis identified meeting the fish and seafood target as positively associated with SAFAD carbon footprint.

It should be noted that meat consumption was not included as a separate component of the SWITCH adherence score. Therefore, higher adherence to the included dietary targets did not necessarily imply lower intake of meat or other carbon-intensive foods.

4.5 Associations with biomarkers

Associations between dietary carbon footprint and blood lipid biomarkers are shown in Table 4.6 and Figure 4.5. Table 4.6 presents estimated associations for both RISE and SAFAD carbon footprint, while Figure 4.5 illustrates the corresponding estimated differences.

Table 4.6 shows that, after adjustment for multiple testing using the Benjamini–Hochberg false discovery rate method, no statistically significant associations were observed between dietary carbon footprint and total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, or the LDL/HDL ratio, for either the RISE or SAFAD carbon footprint estimates.

Table 4.6: Associations between dietary carbon footprint and blood lipid biomarkers. Analyses included 173 participant-visits with available biomarker, dietary, and covariate data. Estimates represent differences in biomarker values per one standard deviation higher log-transformed dietary carbon footprint. Asterisks indicate statistical significance after Benjamini–Hochberg false discovery rate adjustment: * $p_{\text{FDR}} < 0.05$, ** $p_{\text{FDR}} < 0.01$, *** $p_{\text{FDR}} < 0.001$.

	RISE carbon footprint		SAFAD carbon footprint	
Biomarker	β	**Estimate (95% CI)**	β	**Estimate (95% CI)**
Total cholesterol	0.02	0.02 (-0.10, 0.13)	-0.00	-0.00 (-0.13, 0.12)
LDL cholesterol	-0.02	-0.02 (-0.10, 0.06)	-0.00	-0.00 (-0.10, 0.09)
HDL cholesterol	0.00	0.00 (-0.03, 0.03)	0.00	0.00 (-0.03, 0.04)
Triglycerides	0.01	0.01 (-0.04, 0.06)	-0.02	-0.02 (-0.07, 0.04)
LDL/HDL ratio	-0.02	-0.02 (-0.09, 0.05)	0.01	0.01 (-0.07, 0.08)

Figure 4.5 shows the estimated differences in blood lipid biomarkers across carbon footprint measures. Values are close to zero and confidence intervals include zero for all biomarkers, indicating no significant associations. These patterns are consistent with the results presented in Table 4.6.

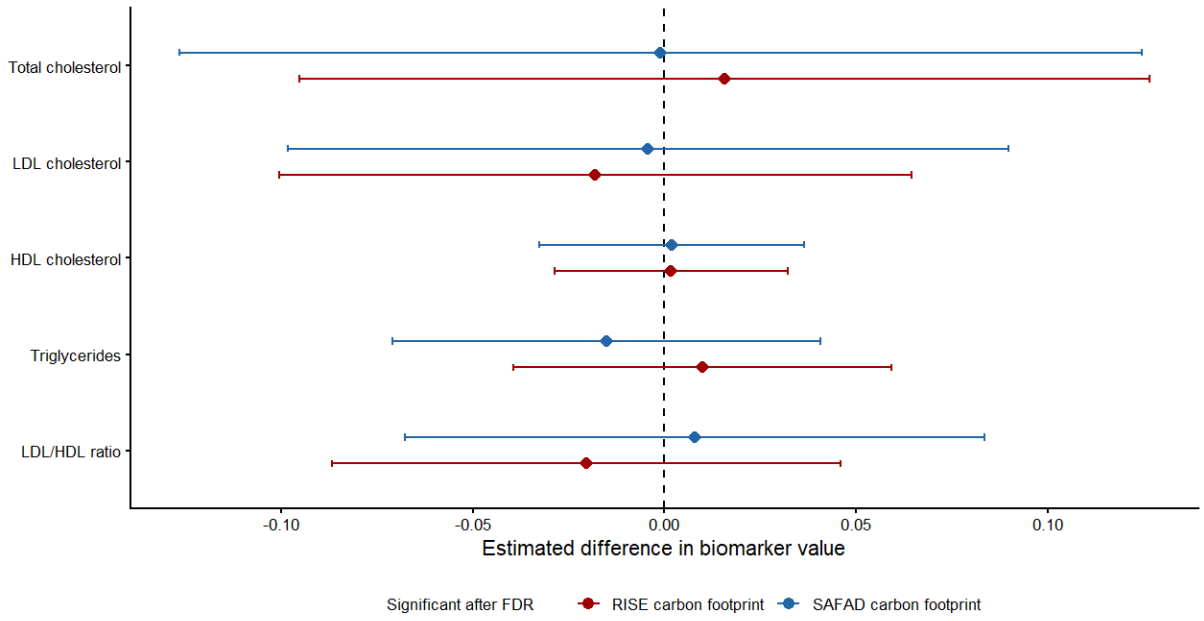


Figure 4.5: Associations between dietary carbon footprint and blood lipid biomarkers. Points represent estimated differences in blood lipid biomarker values per one standard deviation higher log-transformed carbon footprint, and horizontal lines show 95% confidence intervals. Asterisks indicate statistical significance after Benjamini–Hochberg false discovery rate adjustment.

4.6 Comparison between RISE and SAFAD

The comparison between RISE and SAFAD is presented at both the participant and food-item levels. Figure 4.6 and Table 4.7 show mean daily carbon footprint estimates across visits, while Figure 4.7 shows agreement between the databases at the participant level. Food-level agreement is shown in Figure 4.8, while detailed food-level differences in emission factors and total carbon footprint contributions are presented in Appendix D.

Figure 4.6 shows mean daily carbon footprint estimates across study visits. SAFAD consistently yielded higher estimates than RISE at all visits.

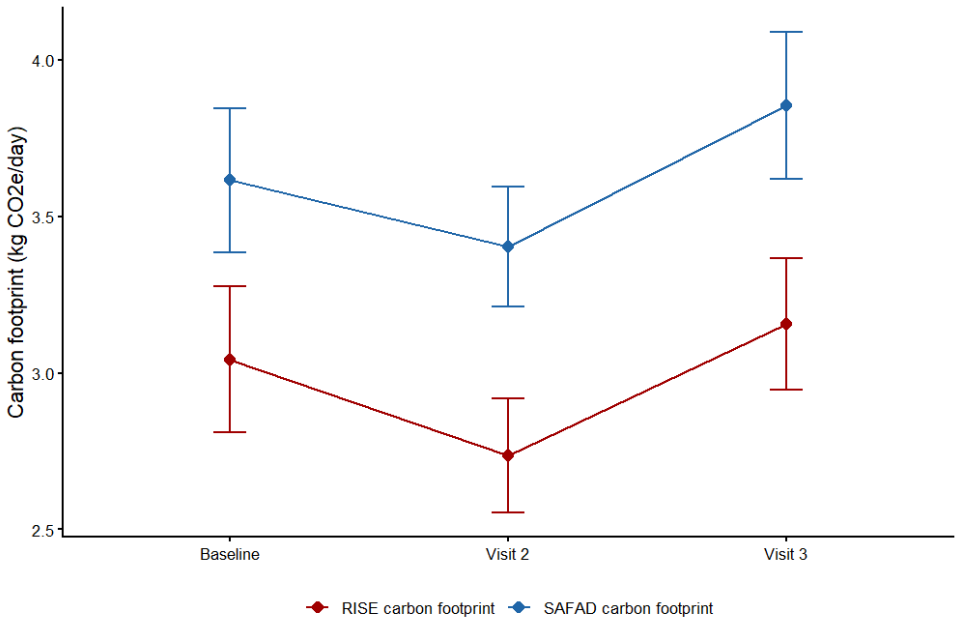


Figure 4.6: Mean daily carbon footprint estimated using RISE and SAFAD across baseline, visit 2, and visit 3. Error bars represent 95% confidence intervals.

Table 4.7 summarises daily carbon footprint estimates across visits. SAFAD yielded higher mean values than RISE at all time points, with consistent differences ranging from 0.57 to 0.70 kg CO_2e/day.

Table 4.7: Comparison of daily CO_2e estimates between RISE and SAFAD across study visits. Values are presented as mean ± standard deviation. The number of participants included was 110 at baseline, 97 at visit 2, and 87 at visit 3.

	Carbon footprint (kg CO_2e/day)		**Mean difference**
Visit	**RISE**	**SAFAD**	**(kg CO_2e/day)**
Baseline	3.04 ± 1.25	3.62 ± 1.23	0.57
Visit 2	2.74 ± 0.92	3.40 ± 0.96	0.67
Visit 3	3.16 ± 1.00	3.85 ± 1.11	0.70

Figure 4.7 shows the Bland–Altman analysis comparing RISE and SAFAD estimates of daily dietary carbon footprint at the participant level. SAFAD produced higher estimates than RISE on average, with a mean difference of 0.64 kg CO_2e/day. The 95% limits of agreement ranged from -0.69 to 1.97 kg CO_2e/day.

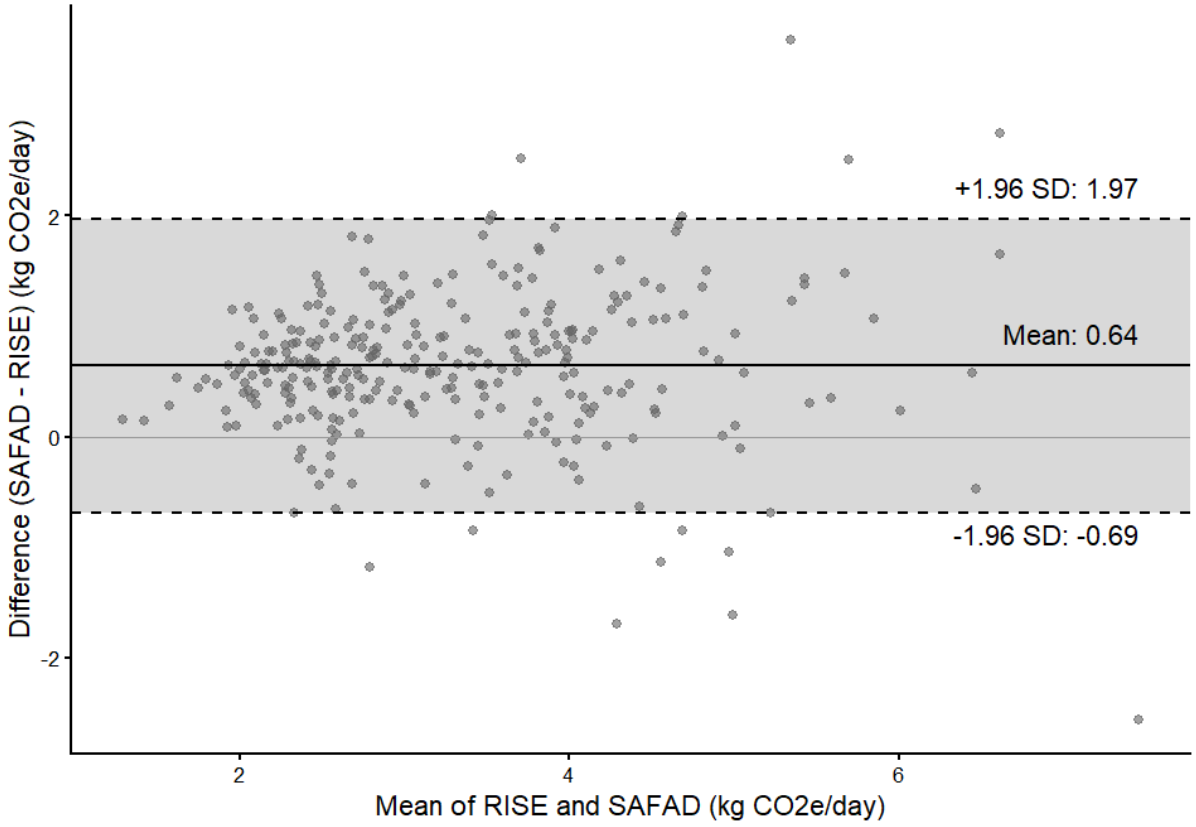


Figure 4.7: Bland–Altman plot showing agreement between RISE and SAFAD estimates of daily carbon footprint (kg CO_2e/day). The solid line represents the mean difference, and dashed lines indicate the 95% limits of agreement.

Figure 4.8 shows the corresponding Bland–Altman analysis at the food-item level, comparing RISE and SAFAD emission factors. The mean difference was -0.38 kg CO_2e/kg food when calculated as RISE minus SAFAD, indicating that SAFAD produced slightly higher emission factors on average. The 95% limits of agreement ranged from -7.62 to 6.86 kg CO_2e/kg food.

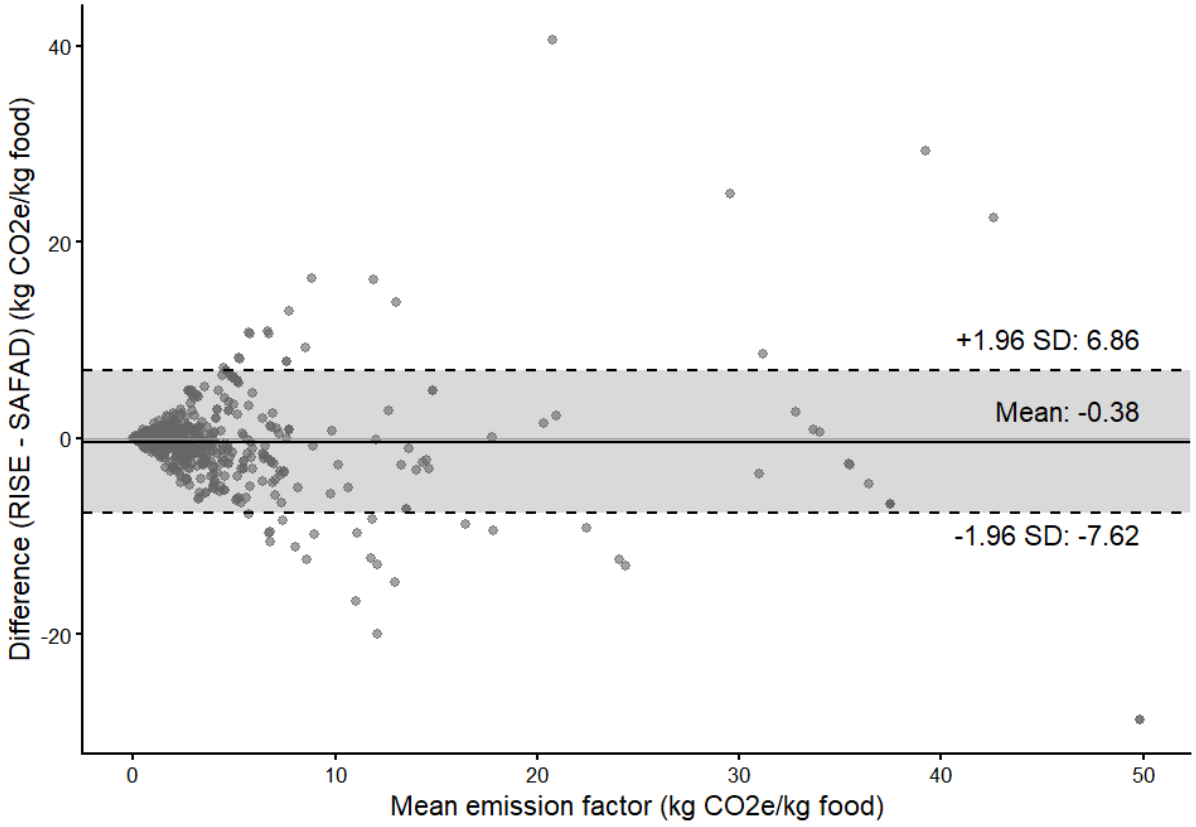


Figure 4.8: Bland–Altman plot showing agreement between RISE and SAFAD food-level emission factors (kg CO_2e/kg food). The solid line represents the mean difference, and dashed lines indicate the 95% limits of agreement.

Detailed food-level differences in emission factors and total carbon footprint contributions are presented in Appendix D.

5

Discussion

This chapter interprets the findings of the thesis in relation to the study aim, research questions, and theoretical background.

5.1 Methodological Considerations and Data Characteristics

A key methodological consideration in this study is that the data were fully anonymised, making it impossible to distinguish between intervention and control groups or between different levels of support. As a result, the observed trends reflect average effects across the entire study population rather than the specific effects of the intervention. This may mask effects within subgroups and limits the ability to draw conclusions about the intervention-specific effects.

Quality control procedures were applied using the Goldberg cut-off to exclude implausible energy reporters. While this improves the reliability of dietary intake data, it also reduced the sample size from 174 to 113 participants. This reduction may introduce selection bias, as underreporting is not randomly distributed. Therefore, the final analytical sample may overrepresent participants with more plausible or consistent dietary reporting behaviour.

Participant characteristics before and after quality control are summarised in Appendix A.1. The final analytical sample consisted of 113 participants after quality control, the majority of whom were women, with a mean age of approximately 55 years and a mean BMI of 28.8 kg/m^2. Compared with the general Swedish population, the sample was older and had a mean BMI within the overweight range [19], [20].

The analytical dataset was unbalanced after quality control. Among the 113 participants included in the final analyses, 110 contributed data at baseline, 97 at visit 2, and 87 at visit 3. In total, 12 participants contributed data from one visit, 21 from two visits, and 80 from all three visits. This should be considered when interpreting changes across visits, as the estimates are based on partly different numbers of observations at each time point. However, the use of linear mixed-effects models allowed participants with incomplete follow-up data to be retained in the analyses. These differences in both participant characteristics and dietary context should be considered when interpreting the generalisability of the findings.

Changes across the intervention period provide further context for interpreting the results. As shown in Appendix A.1, mean energy intake decreased from baseline to visit 2 and increased again at visit 3, while total food intake increased over time. A similar pattern is observed for adherence to the SWITCH dietary targets in Table 4.2, where adherence improved at visit 2

and declined slightly at visit 3. These patterns indicate an initial improvement in adherence to the recommended dietary targets, followed by a partial decline at visit 3. This may reflect difficulties in maintaining dietary changes over time, although the mechanisms underlying this pattern were not assessed in the present study.

Such trends are consistent with findings from dietary intervention studies, which suggest that adherence may decrease over time and that maintaining dietary changes can be challenging [21], [22]. Declining motivation and a return towards previous dietary habits have been proposed as possible explanations in previous research, but these mechanisms cannot be established from the present data. These behavioural patterns are nevertheless important for interpreting the results across all research questions.

In addition to the main mixed-effects models, descriptive subgroup contribution analyses were included in Appendix C to provide further context for interpreting changes in environmental impact across visits. These analyses were exploratory and were not used for statistical inference.

5.2 Changes in Environmental Impact Across Visits

Changes in environmental impact across visits are shown in Figure 4.1, Table 4.3, and Figure 4.2. Overall, both RISE and SAFAD indicated a tendency towards reduced climate impact from baseline to visit 2, followed by an increase at visit 3. However, after adjustment for multiple testing using the Benjamini–Hochberg false discovery rate method, no statistically significant changes were observed for the aggregated carbon footprint measures. This suggests that changes in total climate impact were less robust than changes in several individual SAFAD indicators.

The descriptive subgroup contribution analyses in Appendix C provide additional context for these patterns. From baseline to visit 2, intake increased for several SWITCH-related food groups, including fish, fruit, legumes, shellfish, and vegetables. At the same time, carbon footprint contributions decreased for several animal-source or higher-impact food groups, including cheese, liquid dairy, poultry, and processed meat. This suggests that the decrease in total carbon footprint at visit 2 may partly reflect changes in the relative contribution of different food groups, where increased intake of recommended foods coincided with lower contributions from several higher-impact food groups.

At visit 3, carbon footprint increased again. The subgroup contribution analyses show that several food groups contributed more to carbon footprint at visit 3 than at visit 2, including cheese, liquid dairy, meat, shellfish, vegetables, and the heterogeneous “Other” category. This suggests that the initial reduction observed at visit 2 may have been partly offset at visit 3 by changes in overall dietary composition and increased contributions from some food groups with higher or more variable environmental impacts.

Importantly, changes in environmental impact cannot be attributed solely to changes in the five targeted SWITCH food groups. Non-targeted food groups, particularly meat, processed meat, and dairy, also changed across visits and contributed substantially to dietary carbon footprint. As shown in Appendix C, changes in these food groups occurred alongside changes in the targeted food groups. Thus, the observed changes in environmental impact likely reflect changes in the overall dietary composition rather than adherence to the five SWITCH targets alone.

5.2.1 Environmental changes from baseline to visit 2

From baseline to visit 2, several SAFAD indicators showed significant reductions after FDR adjustment, including cropland use, nitrogen input, ammonia emissions, and biogenic methane. These changes suggest that early dietary changes may have reduced several pressures related to agricultural production, particularly livestock production, manure management, ruminant digestion, and feed production.

This interpretation is consistent with Appendix C, where carbon footprint contributions from several animal-source or higher-impact food groups, including cheese, liquid dairy, poultry, processed meat, and the heterogeneous "Other" category, decreased from baseline to visit 2. Meat also contributed less to carbon footprint at visit 2 than at baseline.

However, visit 2 did not show uniform improvement. Carbon dioxide emissions and animal welfare loss increased significantly, suggesting that reductions in some livestock-related pressures may have been accompanied by increases in other pressures. Appendix C shows that intake of fish, shellfish, fruit, legumes, and vegetables increased from baseline to visit 2, and that fish and shellfish made a larger carbon footprint contribution at visit 2 than at baseline. This suggests that the environmental changes at visit 2 reflected trade-offs rather than a uniform shift towards lower-impact foods.

5.2.2 Environmental changes from baseline to visit 3

From baseline to visit 3, the pattern shifted. Significant increases after FDR adjustment were observed for biodiversity loss, pesticide use, and carbon dioxide emissions. In contrast, ammonia emissions remained significantly reduced, indicating that some reduction in livestock-related pressures may have persisted.

The increase in carbon footprint at visit 3 is consistent with Appendix C, which shows higher contributions at visit 3 than at visit 2 from several food groups, including cheese, liquid dairy, meat, shellfish, vegetables, and the heterogeneous "Other" category. This suggests that the increase at visit 3 may partly reflect a combination of higher total intake and increased contribution from food groups with higher or more variable environmental impacts.

The increases in pesticide use, biodiversity loss, and carbon dioxide emissions also suggest that later dietary changes may have introduced other environmental pressures. This interpretation is consistent with the theory chapter, which emphasises that both plant-based and animal-source foods vary substantially in environmental impact depending on production system, origin, processing, storage, transport, and supply chain. Therefore, the visit 3 pattern should not be interpreted simply as a shift towards either more or less sustainable eating, but rather as a change in the balance between different environmental pressures.

5.2.3 Trade-offs across SAFAD indicators

Taken together, the SAFAD indicators suggest that participants showed short-term improvements in some environmental pressures related to livestock production and feed systems, particularly at visit 2. However, these improvements were not consistent across all indicators or maintained across all visits.

Appendix C helps illustrate why these trade-offs may have occurred. From baseline to visit

2, carbon footprint contributions from several animal-source or higher-impact food groups decreased, including cheese, liquid dairy, processed meat, poultry, meat, and the heterogeneous "Other" category. At the same time, intake increased for fish, shellfish, fruit, legumes, and vegetables. This pattern is consistent with partial substitution, where some recommended foods may have replaced higher-impact foods rather than simply being added. However, since the subgroup analyses are descriptive and aggregated, specific food substitutions cannot be confirmed at the individual level.

This interpretation is in line with previous substitution studies. Rose et al. found that replacing high-impact beef items with calorically equivalent poultry or pork alternatives reduced the dietary carbon footprint by 48.4% while it also improved the diet quality by 3.6% [23]. Harwatt et al. showed that replacing beef with beans could achieve 46–74% of the greenhouse gas reductions needed to meet the US 2020 climate target and could free 42% of US cropland [24]. Similarly, Aleksandrowicz et al. found in a systematic review that sustainable dietary shifts reduced greenhouse gas emissions and land use by median values of 22% and 28%, respectively, with larger reductions when animal-source foods, particularly ruminant meat and dairy, were restricted more strongly [25].

These studies support the interpretation that the lower carbon footprint at visit 2 may partly reflect reduced contributions from animal-source foods. At the same time, they also show that the replacement food matters. Replacing red meat with poultry may reduce climate impact, but can introduce other concerns, such as animal welfare. Replacing red meat with legumes appears more favourable for greenhouse gas emissions and cropland use. In the present study, poultry did not increase; instead, several animal-source food groups decreased while legumes, vegetables, fruit, fish, and shellfish increased. Therefore, the results should be interpreted as being consistent with partial substitution and addition effects, rather than as evidence of one specific substitution pathway.

5.2.4 Comparison with the Average Swedish Diet

Comparison with the Average Swedish Diet, described in Appendix A.3, provides a broader national context. The average Swedish diet was estimated to generate approximately 1.7 tonnes CO_2e per year, corresponding to 4.66 kg CO_2e per day. In comparison, participants had lower baseline values, with 3.04 kg CO_2e/day using RISE and 3.62 kg CO_2e/day using SAFAD.

A broader comparison across SAFAD indicators also showed that several indicators were lower at baseline than the national reference values. This suggests that participants may already have had a lower overall environmental impact than the average Swedish diet before the intervention. One possible explanation is a healthy volunteer effect, where participants in dietary intervention studies may be more health or sustainability conscious than the general population.

However, this comparison should be interpreted with caution. The Average Swedish Diet is based on supply data, whereas the study results are based on self-reported intake. Supply data reflect food availability and may include losses and waste, while reported intake reflects what participants stated that they consumed. These methodological differences may partly explain why the national reference values are higher.

5.3 Association Between SWITCH Adherence and Carbon Footprint

The association between adherence to the SWITCH dietary targets and carbon footprint is presented in Table 4.4, Figure 4.3, Table 4.5, and Figure 4.4. Overall, higher adherence to the five assessed SWITCH dietary targets was not significantly associated with lower carbon footprint after FDR adjustment, regardless of whether adherence was assessed using the graded or binary scale. This finding should be interpreted specifically as an absence of an association between adherence to the assessed SWITCH targets and dietary carbon footprint, rather than as evidence that adherence to a sustainable dietary pattern in general is unrelated to environmental impact.

The lack of association may be explained by the composite nature of the score. The SWITCH adherence score combines several dietary targets with potentially different environmental consequences. Some components, such as vegetables, fruit and berries, whole grains, and legumes, may be associated with lower climate impact, while others, such as fish and seafood, may increase climate impact depending on species, production method, and supply chain. These opposing effects may therefore mask associations when combined into one score.

The absence of meat intake from the adherence score is also important, as meat, particularly red and processed meat, can be a major contributor to dietary greenhouse gas emissions. Participants could therefore meet several SWITCH targets while still maintaining a relatively high intake of meat or other carbon-intensive foods, weakening the association between adherence to the assessed targets and carbon footprint.

This interpretation is consistent with the descriptive subgroup contribution analyses in Appendix C, which suggest that the environmental effect of increasing recommended foods depends on whether these foods replace higher-impact foods or are added to the existing diet. Thus, adherence to the five assessed SWITCH targets does not necessarily represent adherence to an overall sustainable dietary pattern, as the latter also depends on the foods that are reduced or replaced.

5.3.1 Vegetable adherence in the graded analysis

In the graded adherence analysis, higher vegetable adherence was associated with higher SAFAD carbon footprint after FDR adjustment, corresponding to a 19.8% higher carbon footprint per one-unit higher adherence score. No significant association was observed for RISE. This result should not be interpreted as evidence that vegetables are generally high-impact foods. Rather, it may reflect variation in the types of vegetables consumed, as well as differences in production system, origin, seasonality, storage, processing, packaging, and transport.

The graded scale may have made this association more visible because it captured variation in the degree of target achievement rather than only whether participants met the target. Vegetable intake varied substantially among participants, and a continuous score therefore retained more information about differences in intake than a binary classification. In contrast, the binary variable only distinguished between participants below and above the target, which may have reduced sensitivity to intake variation below the threshold. The positive association may also reflect that higher vegetable intake was added to the existing diet rather than replacing meat or other high-impact foods.

5.3.2 Fish and seafood adherence in the binary analysis

In the binary adherence analysis, meeting the fish and seafood target was associated with a higher SAFAD carbon footprint after FDR adjustment, corresponding to a 10.4% higher carbon footprint. No significant association was observed for RISE. This highlights that health-oriented recommendations do not always align with lower climate impact. Seafood has a highly variable carbon footprint depending on species, production method, fishing method, processing, storage, transport, and supply chain [26]. Recommendations to increase fish intake may therefore need to be more specific regarding species, origin, and production method.

5.3.3 Influence of adherence definition

The graded and binary adherence analyses showed partly different association patterns. The graded analysis identified vegetable adherence as positively associated with SAFAD carbon footprint, whereas the binary analysis identified meeting the fish and seafood target as positively associated with SAFAD carbon footprint. This difference reflects that the two approaches capture different aspects of adherence. The graded score reflects the degree of target achievement and therefore retains information about variation below and above the target. In contrast, the binary adherence variable only distinguishes between participants who met or did not meet each predefined target. The way adherence was defined therefore influenced which dietary components were most strongly associated with carbon footprint.

5.3.4 Influence of database choice

The significant associations for both vegetables and fish and seafood were observed for SAFAD but not for RISE. This may reflect methodological differences between the databases. SAFAD applies broader system boundaries for carbon footprint, including processing, preparation, packaging, and transport to the warehouse, whereas RISE reports values at the industrial or factory gate. Food groups such as vegetables and seafood may be particularly sensitive to these differences because their carbon footprint can depend strongly on production system, origin, storage, processing, transport, and supply chain.

Overall, these findings show that higher adherence to the five assessed SWITCH dietary targets did not necessarily correspond to lower dietary carbon footprint.

5.4 Association Between Biomarkers and Carbon Footprint

The association between dietary carbon footprint and blood lipid biomarkers is presented in Table 4.6 and Figure 4.5. After adjustment for multiple testing using the Benjamini–Hochberg false discovery rate method, no statistically significant associations were observed between dietary carbon footprint and total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, or the LDL/HDL ratio. The estimated effects were close to zero, and confidence intervals included the null value.

One possible explanation is the relatively short intervention period. Although dietary intake changed across visits, these changes may not have been large or sustained enough to produce measurable changes in blood lipid biomarkers within the timeframe. This is particularly relevant

because the dietary changes appeared strongest at visit 2 and were less consistently maintained at visit 3.

Another factor is the availability of biomarker data. Biomarker data were only available for two subsets of participants at the time of analysis, as data collection and processing for the remaining participants were still ongoing. This reduced the sample size and may have limited the precision of the estimates, particularly for detecting small associations between dietary carbon footprint and blood lipid biomarkers.

Finally, dietary carbon footprint and blood lipid biomarkers represent different dimensions of diet. Carbon footprint reflects environmental impact, whereas blood lipid concentrations are influenced by dietary composition and other factors, including dietary fat quality, energy balance, medication use, and individual metabolic characteristics. Therefore, the absence of observed associations in this study does not imply that environmental impact and health-related dietary characteristics are unrelated, but indicates that no associations were observed between dietary carbon footprint and the measured blood lipid biomarkers in this sample.

5.5 Comparison Between RISE and SAFAD

The comparison between RISE and SAFAD is presented in Figure 4.6, Figure 4.7, Figure 4.8, and Table 4.7. Detailed food-level differences between the databases are presented in Table D.1 and Table D.2, showing differences in emission factors and total carbon footprint contributions, respectively.

5.5.1 Overall agreement between databases

SAFAD consistently produced higher estimates of dietary carbon footprint than RISE across all visits, with mean differences ranging from 0.57 to 0.70 kg CO_2e/day. The Bland–Altman analysis showed a mean difference of 0.64 kg CO_2e/day, indicating that SAFAD produced higher values on average.

The systematic difference may be explained by differences in emission factors, system boundaries, food categorisation, and assumptions used for mixed dishes and processed foods. The RISE database reports carbon footprint values at the industrial or factory gate and excludes packaging, later transport stages, and household or restaurant preparation. In contrast, SAFAD includes a broader set of life cycle stages in its carbon footprint calculations, including processing, preparation, packaging, and transport to the warehouse. These methodological differences can lead to different absolute estimates even when the same dietary intake data are used.

5.5.2 Food-level differences

Food-level comparisons showed that some items had very large differences in emission factors between RISE and SAFAD. For example, dried reindeer meat had the largest discrepancy, with an emission factor of 41.00 kg CO_2e/kg food in RISE compared with 0.40 kg CO_2e/kg food in SAFAD. This difference is exceptionally large and should be interpreted with caution. It may reflect differences in LCA system boundaries, food classification, allocation assumptions, or how wild and semi-wild meat production is represented in the databases. It may also indicate

uncertainty in the database values or in the matching of the reported food item to representative database items.

Such discrepancies illustrate that environmental databases may assign substantially different climate impact values to the same reported food item. This is particularly relevant for foods with less standardised production systems or limited underlying LCA data. Therefore, large food-level differences should not be interpreted as precise estimates of true production differences, but rather as indicators of uncertainty and methodological sensitivity.

Although such large per-kg discrepancies are important for understanding uncertainty in food-level comparisons, they do not necessarily have a large influence on total dietary carbon footprint. Dried reindeer meat was reported only once and in a small amount, meaning that its effect on the total dietary estimates was limited. In contrast, frequently consumed foods, such as brewed coffee and hard cheese, contributed more to the overall differences between RISE and SAFAD because of their high consumption frequency or total intake. This shows that database differences should be interpreted at both the emission-factor level and the total-consumption level.

5.5.3 Manual mapping and database compatibility

Manual mapping introduced additional uncertainty, particularly for RISE. In total, 335 food items had to be mapped manually for RISE, compared with only 19 for SAFAD. This difference is important because manual mapping requires assigning reported food items to the closest available representative item, which may not fully capture the original food's ingredients, preparation method, or degree of processing. As a result, manually mapped foods may be systematically under- or overestimated if the selected proxy item has a lower or higher emission factor than the actual reported food.

This issue is especially relevant for composite foods and mixed dishes, where the carbon footprint depends strongly on ingredient composition, preparation method, and portion structure. For example, dishes containing meat, dairy products, or seafood may be underestimated if they are mapped to a less carbon-intensive proxy, while plant-based or vegetable-rich mixed dishes may be overestimated if mapped to a more intensive representative item. Because RISE required substantially more manual mapping than SAFAD, mapping-related uncertainty was likely greater for RISE estimates.

The lower number of manually mapped items in SAFAD suggests that SAFAD was more directly compatible with the dietary recall data used in this study. This likely reflects that SAFAD includes a broader range of foods as consumed, including composite foods, meals, and drinks. In contrast, RISE required more extensive interpretation and matching, particularly for mixed dishes and unspecified foods. Therefore, differences between RISE and SAFAD may reflect not only differences in LCA methodology and system boundaries, but also differences in database coverage and compatibility with dietary intake data.

5.5.4 Implications for interpretation

The two databases also differ in practical use. The RISE database can include several carbon footprint values for the same SLV identification code, depending on region of origin and whether the production system is conventional or organic. This can improve specificity when

detailed food origin and production information are available. However, such information was generally not available in the dietary recall data. SAFAD, in contrast, provides multiple environmental and social indicators and allows for a broader assessment of sustainability beyond climate impact alone.

Overall, the comparison shows that database choice influences absolute carbon footprint estimates. While relative trends appeared broadly robust, RISE and SAFAD should not be treated as interchangeable, especially at the individual food item level. These findings highlight the need for transparency and harmonisation in environmental databases used for dietary assessment, particularly regarding emission factors, system boundaries, food categorisation, and assumptions for composite foods.

5.6 Strengths and Limitations

This study has several strengths. First, the longitudinal design with three visits allowed changes in dietary intake and environmental impact to be assessed over time. Second, quality control procedures were applied to improve the reliability of the dietary data. Third, the use of detailed dietary assessment data enabled linkage between reported food intake and environmental impact estimates. Another important strength is the comparison of two environmental databases, RISE and SAFAD, which made it possible to examine both consistency and uncertainty in carbon footprint estimates. The inclusion of multiple SAFAD indicators also allowed a broader assessment of sustainability beyond climate impact alone.

However, several limitations should be considered. Dietary intake was assessed using self-reported 24-hour recalls. This method provides detailed information on food consumption with relatively low respondent burden, but relies on participants' memory and is therefore subject to recall bias, misreporting, and underreporting [27]. Underreporting may affect both energy intake estimates and calculated environmental impacts. Although quality control reduced some of this uncertainty, it also reduced the sample size and may have introduced selection bias.

Another limitation is the relatively short intervention period, which may not be sufficient to capture long-term dietary behaviour or biomarker responses. Biomarker data were only available for two of the three blocks at the time of analysis, which may have limited the precision of the estimates and the ability to detect small associations. In addition, the anonymisation of the dataset meant that intervention and control groups, as well as different levels of support, could not be separated. This limits the ability to assess the specific effect of the intervention.

The use of Benjamini–Hochberg false discovery rate adjustment reduced the risk of false-positive findings due to multiple testing. However, this also made the analyses more conservative and may have reduced statistical power, particularly in the biomarker analyses where the sample size was limited.

Another limitation concerns the construction of the SWITCH adherence score. Meat and processed meat were not included as separate components of the score, despite being important contributors to dietary carbon footprint. As discussed in Section 5.3 and illustrated by the subgroup contribution analysis in Appendix B, changes in intake of carbon-intensive foods may influence environmental impact independently of adherence to the five included dietary targets. Therefore, the adherence score may not fully capture the dietary changes relevant to environ-

mental sustainability, which limits the interpretation of the association between SWITCH adherence and dietary carbon footprint.

Finally, environmental impact estimates are subject to uncertainty. Differences in emission factors, system boundaries, and assumptions between databases may influence results. Manual mapping of food items, particularly mixed dishes and unspecified foods, introduced additional uncertainty. These limitations should be considered when interpreting the absolute values reported in this study.

5.7 Future Research

Future research should include longer follow-up periods to assess whether dietary changes and environmental improvements can be maintained over time. Studies should also aim to distinguish between intervention and control groups, as well as different levels of support, in order to better evaluate intervention effects.

Future analyses should explicitly examine changes in red and processed meat intake alongside adherence to the predefined SWITCH dietary targets. The present findings indicate that environmental impact may depend not only on increases in recommended food groups, but also on whether these foods replace foods with higher environmental impacts. Meat and processed meat were not included in the SWITCH adherence score, despite contributing substantially to dietary carbon footprint.

Individual dietary substitution analyses would therefore be particularly valuable. Such analyses could assess whether increases in vegetables, legumes, whole grains, or fish and seafood are accompanied by reductions in higher-impact foods, such as red and processed meat and certain dairy products, or whether the recommended foods are instead added to the existing diet. This would provide a better understanding of the dietary changes underlying changes in environmental impact and help distinguish between increased consumption of recommended foods and actual substitution towards a lower-impact dietary pattern. The subgroup contribution analysis presented in Appendix C provides an initial descriptive basis for such analyses.

More complete biomarker data would strengthen analyses of the relationship between environmental impact and health outcomes. Future studies could also investigate a broader range of cardiometabolic biomarkers and assess whether dietary changes associated with lower environmental impact are accompanied by changes in cardiometabolic health.

Future studies should also include more detailed dietary information, such as food origin, production method, brand, and species-level information for fish and seafood. This would improve the precision of environmental assessments and reduce uncertainty related to the use of generic emission factors and manual food mapping.

Finally, future analyses could explore individual-level adherence patterns and dietary trajectories to better understand why some participants achieve sustained dietary change while others do not. Combining adherence analyses with dietary substitution patterns could provide a more comprehensive assessment of how behavioural changes translate into both environmental and health outcomes.

6

Conclusion

This thesis evaluated dietary environmental impact within the SWITCH study and examined how environmental impact estimates were related to dietary target adherence, blood lipid biomarkers, and database choice. Across study visits, carbon footprint tended to decrease from baseline to visit 2 and increase again at visit 3, although these changes were not statistically significant after correction for multiple testing. In contrast, several SAFAD indicators changed significantly, showing that dietary change may reduce some environmental pressures while increasing others. This highlights the importance of assessing sustainable diets using multiple environmental indicators rather than carbon footprint alone.

Higher adherence to the five assessed SWITCH dietary targets was not significantly associated with lower dietary carbon footprint. The graded and binary adherence analyses also identified different food group-specific patterns. Higher vegetable adherence in the graded analysis and meeting the fish and seafood target in the binary analysis were both associated with higher SAFAD carbon footprint. These findings indicate that adherence to the assessed SWITCH dietary targets did not necessarily correspond to lower dietary climate impact. However, because the adherence score did not include reductions in meat or processed meat intake, despite these foods being important contributors to dietary carbon footprint, the results should not be interpreted as evidence that adherence to a broader sustainable dietary pattern is unrelated to environmental impact.

No significant associations were observed between dietary carbon footprint and the measured blood lipid biomarkers in the available biomarker sample. These findings indicate that no associations were detected between dietary carbon footprint and the measured lipid outcomes in this study. Environmental impact and health-related dietary characteristics represent different dimensions of diet and should therefore not be considered equivalent. The comparison between RISE and SAFAD showed that SAFAD consistently produced higher carbon footprint estimates than RISE, although the two databases were strongly correlated. Overall, the findings show that database choice, system boundaries, food mapping, and food substitutions are important considerations when evaluating the environmental impact of diets. Future dietary sustainability assessments should therefore combine transparent environmental databases with multi-indicator approaches and more detailed information on food origin, production method, and dietary substitutions.

Bibliography

[1] L. Aleksandrowicz, R. Green, E. J. Joy, P. Smith, and A. Haines, "The impacts of dietary change on greenhouse gas emissions, land use, water use, and health: A systematic review," *PloS one*, vol. 11, no. 11, e0165797, 2016.

[2] J. Rockström, S. H. Thilsted, W. C. Willett, *et al.*, "The eat–lancet commission on healthy, sustainable, and just food systems," *The Lancet*, vol. 406, no. 10512, pp. 1625–1700, 2025.

[3] World Health Organization, *Noncommunicable diseases*, Accessed: 2026-05-13, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases.

[4] A. Giosuè, F. Recanati, I. Calabrese, *et al.*, "Good for the heart, good for the earth: Proposal of a dietary pattern able to optimize cardiovascular disease prevention and mitigate climate change," *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 32, no. 12, pp. 2772–2781, 2022.

[5] Stockholm Resilience Centre, *Planetary boundaries*, Accessed: 2026-05-13, 2025. [Online]. Available: https://www.stockholmresilience.org/research/planetary-boundaries.html.

[6] P. Forster, T. Storelvmo, K. Armour, *et al.*, "The earth's energy budget, climate feedbacks and climate sensitivity," in *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, 2021, pp. 923–1054, ISBN: 9781009157889. DOI: 10.1017/9781009157896.009. [Online]. Available: https://www.cambridge.org/core/books/climate-change-2021-the-physical-science-basis/earths-energy-budget-climate-feedbacks-and-climate-sensitivity/AE57C97E588FF3060C7C7E47DD4F3C6E.

[7] M. R. Harrison, G. Palma, T. Buendia, M. Bueno-Tarodo, D. Quell, and F. Hachem, "A scoping review of indicators for sustainable healthy diets," *Frontiers in Sustainable Food Systems*, vol. 5, p. 822 263, 2022. DOI: 10.3389/fsufs.2021.822263.

[8] Y. Ran, C. Cederberg, M. Jonell, *et al.*, "Environmental assessment of diets: Overview and guidance on indicator choice," *The Lancet Planetary Health*, vol. 8, no. 3, e172–e187, 2024. DOI: 10.1016/S2542-5196(24)00006-8.

[9] E. Röös, M. Jacobsen, L. Karlsson, *et al.*, "Introducing a comprehensive and configurable tool for calculating environmental and social footprints for use in dietary assessments," *Journal of Cleaner Production*, vol. 519, p. 146 002, 2025.

[10] Research Institutes of Sweden, *RISE food climate database – for sustainable consumption*, Document dated 2025-10-01, Oct. 2025.

[11] S. Thupakula, S. S. R. Nimmala, H. Ravula, S. Chekuri, and R. Padiya, "Emerging biomarkers for the detection of cardiovascular diseases," *The Egyptian Heart Journal*, vol. 74, no. 1, p. 77, 2022.

[12] P. W. Wilson, T. A. Jacobson, S. S. Martin, *et al.*, "Lipid measurements in the management of cardiovascular diseases: Practical recommendations a scientific statement from the national lipid association writing group," *Journal of clinical lipidology*, vol. 15, no. 5, pp. 629–648, 2021.

[13] D. J. A. Jenkins, C. W. C. Kendall, A. Marchie, *et al.*, "Effects of a dietary portfolio of cholesterol-lowering foods vs lovastatin on serum lipids and c-reactive protein," *JAMA*, vol. 290, no. 4, pp. 502–510, 2003. DOI: 10.1001/jama.290.4.502.

[14] J. L. Tang, J. M. Armitage, T. Lancaster, C. A. Silagy, G. H. Fowler, and H. A. W. Neil, "Systematic review of dietary intervention trials to lower blood total cholesterol in free-living subjects," *BMJ*, vol. 316, no. 7139, pp. 1213–1220, 1998. DOI: 10.1136/bmj.316.7139.1213.

[15] S. M. Grundy, N. J. Stone, A. L. Bailey, *et al.*, "2018 aha/acc/aacvpr/aapa/abc/acpm/ada/ags/apha/aspc/nla/pcna guideline on the management of blood cholesterol: Executive summary," *Journal of the American College of Cardiology*, vol. 73, no. 24, pp. 3168–3209, 2019. DOI: 10.1016/j.jacc.2018.11.002.

[16] A. E. Black, "Critical evaluation of energy intake using the goldberg cut-off for energy intake: Basal metabolic rate. a practical guide to its calculation, use and limitations," *International journal of obesity*, vol. 24, no. 9, pp. 1119–1130, 2000.

[17] FAO/WHO/UNU, "Human energy requirements: Report of a joint fao/who/unu expert consultation," Food and Agriculture Organization of the United Nations, Rome, Tech. Rep., 2004, Accessed: 2026-05-13. [Online]. Available: https://www.fao.org/4/y5686e/y5686e00.pdf.

[18] United Nations, *Transforming our world: The 2030 agenda for sustainable development*, United Nations General Assembly Resolution A/RES/70/1, Accessed: 2026-05-08, 2015. [Online]. Available: https://sdgs.un.org/2030agenda.

[19] Statistics Sweden, *Average age and median age in sweden by sex. year 1968–2025*, Statistical database, latest update 24 February 2026, 2026. [Online]. Available: https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0101__BE0101B/BefolkMedianAlder/.

[20] Public Health Agency of Sweden, *Overweight and obesity*, Updated 22 May 2018, 2018. [Online]. Available: https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/living-conditions-and-lifestyle/obesity/.

[21] C. J. Greaves, K. E. Sheppard, C. Abraham, *et al.*, "Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions," *BMC Public Health*, vol. 11, p. 119, 2011. DOI: 10.1186/1471-2458-11-119.

[22] M. L. Dansinger, J. A. Gleason, J. L. Griffith, H. P. Selker, and E. J. Schaefer, "Comparison of the atkins, ornish, weight watchers, and zone diets for weight loss and heart disease risk reduction: A randomized trial," *JAMA*, vol. 293, no. 1, pp. 43–53, 2005. DOI: 10.1001/jama.293.1.43.

[23] D. Rose, A. M. Willits-Smith, and M. C. Heller, "Single-item substitutions can substantially reduce the carbon and water scarcity footprints of us diets," *The American journal of clinical nutrition*, vol. 115, no. 2, pp. 378–387, 2022.

[24] H. Harwatt, J. Sabaté, G. Eshel, S. Soret, and W. Ripple, "Substituting beans for beef as a contribution toward us climate change targets," *Climatic Change*, vol. 143, no. 1, pp. 261–270, 2017.

[25] L. Aleksandrowicz, R. Green, E. J. Joy, P. Smith, and A. Haines, "The impacts of dietary change on greenhouse gas emissions, land use, water use, and health: A systematic review," *PloS one*, vol. 11, no. 11, e0165797, 2016.

[26] E. Hallström, K. Bergman, K. Mifflin, *et al.*, "Combined climate and nutritional performance of seafoods," *Journal of Cleaner Production*, vol. 230, pp. 402–411, 2019. DOI: 10.1016/j.jclepro.2019.04.229.

[27] J.-S. Shim, K. Oh, and H. C. Kim, "Dietary assessment methods in epidemiologic studies," *Epidemiology and health*, vol. 36, e2014009, 2014.

A

Appendix A: Study Sample, Data Transformation, and Reference Values

This appendix provides supplementary information on the analytical sample, quality control, normality assessment, and reference values used for comparison with the Average Swedish Diet.

A.1 Participant Characteristics

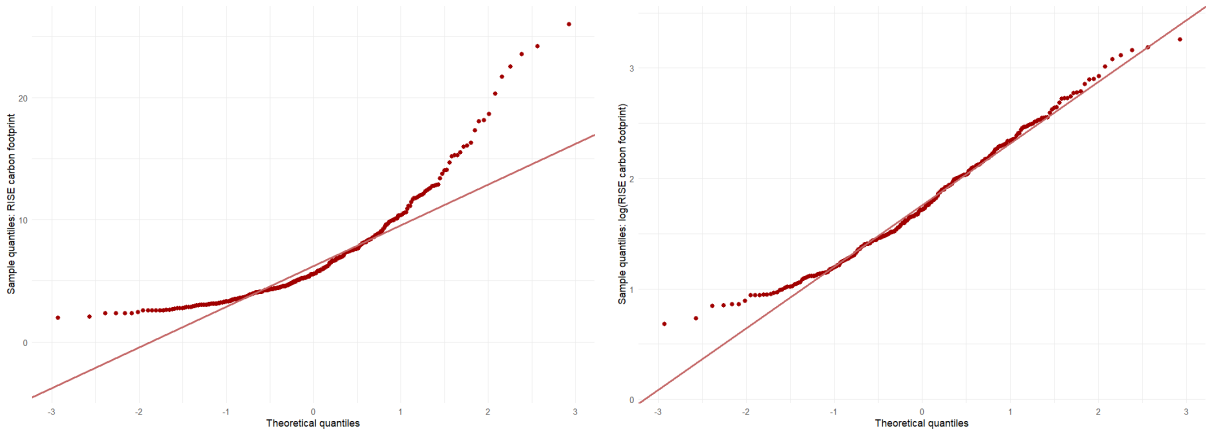
A detailed overview of participant characteristics before and after quality control is presented in Table A.1. The main results section presents a shortened version of this table.

Table A.1: Detailed participant characteristics before and after quality control.

Characteristic	Before QC	After QC
Participants	174	113
Female / male	142 / 32	95 / 18
Age (years)	55.0 ± 9.2	55.2 ± 9.0
BMI (kg/m^2)	28.8 ± 2.6	28.8 ± 2.7
Energy intake (kcal/day)	1928 ± 632	2105 ± 455
Food intake (g/day)	2999 ± 819	3139 ± 768
Energy intake visit 1 (kcal/day)	1935 ± 763	2073 ± 414
Energy intake visit 2 (kcal/day)	1861 ± 425	2048 ± 376
Energy intake visit 3 (kcal/day)	1994 ± 638	2209 ± 561
Food intake visit 1 (g/day)	2911 ± 799	3022 ± 747
Food intake visit 2 (g/day)	3019 ± 783	3157 ± 745
Food intake visit 3 (g/day)	3092 ± 877	3268 ± 803
Participants per visit, V1 / V2 / V3	174 / 151 / 132	110 / 97 / 87
Participants with 1 / 2 / 3 visits	20 / 25 / 129	12 / 21 / 80
Participant-visits	457	294
Visits with 1 / 2 / 3 recall days	0 / 0 / 457	0 / 48 / 246
Block 1 / 2 / 3	50 / 49 / 75	37 / 30 / 46
Dietary data coverage (%)	100	100
Biomarker coverage (%)	65.0	58.8

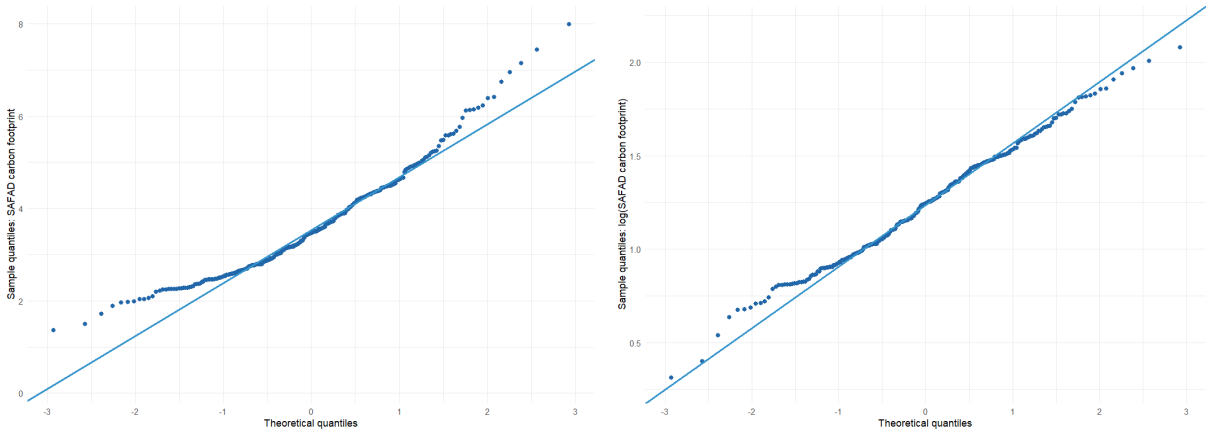
A.2 Assessment of Normality and Data Transformation

Quantile–Quantile plots were used to assess the distribution of RISE and SAFAD CO_2e variables. Both variables deviated from normality on the original scale but more closely approximated normality after log-transformation, as shown in Figure A.1 and Figure A.2.



(a) Original scale **(b)** Log-transformed

Figure A.1: Q–Q plots of RISE CO_2e values before and after log-transformation.



(a) Original scale **(b)** Log-transformed

Figure A.2: Q–Q plots of SAFAD CO_2e values before and after log-transformation.

A.3 Average Swedish Diet

Environmental impact indicators for the Average Swedish Diet were used as a reference point for comparison with the observed dietary intake in the SWITCH study [9]. Annual values were converted to daily estimates by dividing by 365.

This comparison should be interpreted cautiously because the Average Swedish Diet values are based on supply data, whereas the SWITCH estimates are based on self-reported dietary intake. Supply data may include food availability, losses, and waste, while self-reported intake may be affected by recall bias and underreporting.

Table A.2: Environmental impact indicators for the Average Swedish Diet based on SAFAD. Annual values were converted to daily estimates for comparison with observed dietary data.

Indicator	Per year	Per day
Carbon footprint ($t\ CO_2e$/year; $kg\ CO_2e$/day)	1.7	4.66
Cropland use (ha/year; m^2 yr/day)	0.24	6.58
Blue water use (m^3/year; m^3/day)	54	0.148
New nitrogen input (kg N/year; g N/day)	31	85
New phosphorus input (kg P/year; g P/day)	3.9	10.7
Ammonia emissions (kg NH_3/year; g NH_3/day)	5.2	14
Biodiversity loss ($\times 10^{-12}$ E/MSY/year; $\times 10^{-12}$ E/MSY/day)	560	1.53
Pesticide use (g a.i./year; g a.i./day)	489	1.34
Use of antibiotics (index/year; index/day)	1832	5.02
Animal welfare loss (index/year; index/day)	1735	4.75

B

Appendix B: Environmental Database Mapping

This appendix provides supplementary information on manual food mapping in RISE and SAFAD.

B.1 Manual Mapping

Manual mapping was required when reported food items did not have a direct food code match in RISE or SAFAD. Unmatched food items were assigned to the closest available representative item based on food type, main ingredient, preparation method, and degree of processing. The extent of manual mapping differed substantially between RISE and SAFAD, as shown in Table B.1.

Table B.1: Summary of manual food item mapping in RISE and SAFAD. Percentages refer to the share of total reported intake by weight.

Mapping characteristic	**RISE**	**SAFAD**
Total unique food IDs	948	948
Direct matches before mapping	613	929
Food IDs requiring manual mapping	335	19
Still unmatched after mapping	0	0
Amount requiring manual mapping (g)	738936	48028
Amount requiring manual mapping (%)	19.4	1.3
Matched amount after mapping (%)	100	100

RISE required manual mapping for 335 food IDs, corresponding to 19.4% of the total reported intake by weight. In contrast, SAFAD required manual mapping for 19 food IDs, corresponding to 1.3% of the total reported intake by weight. After manual mapping, all reported food IDs and grams were matched in both databases.

B.2 Manual Mapping for RISE

Table B.2 presents the 20 manually mapped RISE food items with the highest total reported intake.

Table B.2: Food items with the highest total reported intake that required manual mapping in the RISE Climate Database.

Old		New		Reported intake	
ID	**Name**	**ID**	**Name**	**n**	**Total amount (g)**
2816	Caffelatte	100178	Coffee with milk	256	69700
6616	Chicken stew with dairy products	2863	Chicken stew with cream/crème fraiche	294	60375
6558	Wine, unspecified	1907	Red wine, 14% alcohol	263	60015
6543	Beer, unspecified	1901	Low-alcohol beer, 2.3% alcohol	121	57575
5614	Semi-skimmed milk	150	Semi-skimmed milk, 1.5% fat, fortified	810	52466.5
6751	Pasta salad, unspecified	5804	Pasta carbonara with pasta, pork and cream	187	44000
5912	Homemade pizza, various flavours	769	Homemade pizza with shrimp and mussels	144	43803
799	Pizza, unspecified	6628	Restaurant pizza with meat	35	32550
2862	Chicken stew with coconut milk	6391	Chicken tikka masala stew	156	32100
5573	Bread roll with toppings	782	Sandwich with smoked turkey, cream cheese, sun-dried tomato and lettuce	245	29447.5
6555	Dark wholegrain bread >30%	207	Wholegrain rye-sifted bread, approx. 7% fibre, kavring type	369	29185.5
3099	Salad with beans and chickpeas	387	Vegetable mix with peas, beans, carrot and cauliflower, frozen summer vegetable type	148	28650
6620	Vegetable soup with dairy products	438	Thickened vegetable soup, vegetarian	108	25935
5778	Hard cheese, 27% fat	79	Hard cheese, 26% fat	885	25850.5
5838	Restaurant hamburger with bread and accompaniments	2512	Restaurant hamburger with bread and cheese	91	25515
6602	Tomato-based vegetable stew with legumes	6408	Provençal bean stew with potato, celeriac, white wine and crème fraiche, vegetarian	150	22926
3859	Fried or wok-fried vegetables and root vegetables	2672	Oven-roasted vegetables and root vegetables	144	22272
5613	Low-fat milk	151	Low-fat milk, 0.5% fat, fortified	206	21976.5
5577	Wrap with toppings	792	Wheat tortilla wrap with feta cheese, olives, lettuce and garlic dressing	60	21400
854	Wholegrain pasta, boiled with salt	846	Pasta, boiled without salt	212	19432

C

Subgroup Contribution Analysis

This appendix presents a descriptive subgroup contribution analysis used to support interpretation of changes in dietary carbon footprint across visits. Food items were aggregated into broader food subgroups and summarised as mean daily intake and mean daily carbon footprint contribution per visit. The analysis is descriptive and was not used for statistical inference.

Table C.1 shows mean daily intake for selected food subgroups across visits. The selected subgroups include foods that were relevant to the SWITCH dietary targets, foods with relatively large carbon footprint contributions, and food groups that changed across visits.

Table C.1: Mean daily intake by selected food subgroup across visits. Values are presented as g/day.

Subgroup	Baseline	Visit 2	Visit 3
Berries	18.4	27.3	30.9
Bread	93.3	97.0	93.2
Butter	2.0	3.1	2.4
Cereals	10.9	16.2	18.8
Cheese	62.0	50.0	60.6
Fish	36.4	67.1	57.4
Fruit	163.0	225.0	219.0
Legumes	22.5	63.1	64.6
Liquid dairy	255.0	230.0	264.0
Meat	18.0	16.7	16.0
Other	1940.0	1924.0	1999.0
Poultry	46.2	33.3	30.2
Processed meat	17.5	9.8	13.1
Shellfish	9.7	18.7	20.9
Vegetables	106.0	167.0	168.0

Table C.2 shows the contribution of selected food subgroups to total dietary carbon footprint estimated using the RISE Climate Database. The table includes both the absolute contribution in kg CO_2e/day and the percentage contribution to total RISE carbon footprint within each visit. The ΔB columns show the absolute change from baseline.

Table C.2: RISE carbon footprint contribution by selected food subgroup across visits. Values are presented as kg CO_2e/day with percentage contribution to total RISE carbon footprint. ΔB indicates absolute change from baseline in kg CO_2e/day.

Subgroup	Baseline		Visit 2			Visit 3		
	CO_2e	%	CO_2e	%	ΔB	CO_2e	%	ΔB
Berries	0.023	0.7	0.035	1.2	+0.014	0.042	1.3	+0.020
Bread	0.074	2.4	0.073	2.6	-0.003	0.096	2.9	+0.022
Cheese	0.354	11.2	0.245	8.6	-0.109	0.303	9.2	-0.050
Fish	0.125	4.0	0.214	7.5	+0.089	0.185	5.6	+0.059
Fruit	0.106	3.4	0.134	4.7	+0.023	0.134	4.1	+0.022
Legumes	0.047	1.5	0.079	2.8	+0.031	0.071	2.2	+0.024
Liquid dairy	0.437	13.9	0.351	12.3	-0.086	0.388	11.8	-0.049
Meat	0.231	7.3	0.211	7.4	-0.019	0.252	7.7	+0.021
Other	0.995	31.6	0.784	27.5	-0.211	0.979	29.8	-0.016
Poultry	0.140	4.4	0.126	4.4	-0.014	0.103	3.1	-0.042
Processed meat	0.103	3.3	0.067	2.4	-0.036	0.075	2.3	-0.031
Shellfish	0.071	2.3	0.157	5.5	+0.086	0.230	7.0	+0.159
Vegetables	0.166	5.3	0.142	5.0	-0.026	0.184	5.6	+0.017

Table C.3 shows the corresponding subgroup contribution analysis using the SAFAD framework.

Table C.3: SAFAD carbon footprint contribution by selected food subgroup across visits. Values are presented as kg CO_2e/day with percentage contribution to total SAFAD carbon footprint. ΔB indicates absolute change from baseline in kg CO_2e/day.

Subgroup	Baseline		Visit 2			Visit 3		
	CO_2e	%	CO_2e	%	ΔB	CO_2e	%	ΔB
Berries	0.022	0.6	0.028	0.8	+0.006	0.032	0.8	+0.010
Bread	0.119	3.3	0.117	3.4	-0.002	0.136	3.5	+0.017
Cheese	0.487	13.5	0.338	9.9	-0.149	0.422	11.0	-0.064
Fish	0.225	6.2	0.439	12.9	+0.214	0.386	10.0	+0.161
Fruit	0.124	3.4	0.164	4.8	+0.040	0.162	4.2	+0.039
Legumes	0.015	0.4	0.045	1.3	+0.030	0.050	1.3	+0.035
Liquid dairy	0.496	13.7	0.442	13.0	-0.054	0.460	11.9	-0.036
Meat	0.319	8.8	0.268	7.9	-0.051	0.340	8.8	+0.021
Other	0.950	26.3	0.784	23.0	-0.166	0.973	25.3	+0.023
Poultry	0.113	3.1	0.084	2.5	-0.029	0.073	1.9	-0.039
Processed meat	0.153	4.2	0.092	2.7	-0.061	0.112	2.9	-0.041
Shellfish	0.063	1.7	0.143	4.2	+0.080	0.184	4.8	+0.121
Vegetables	0.219	6.1	0.218	6.4	-0.001	0.261	6.8	+0.042

D

Food-level differences between RISE and SAFAD

Table D.1 shows food items with the largest differences in emission factors between RISE and SAFAD among those reported in the study. Large differences were observed for both animal-based foods and mixed dishes. The largest difference was observed for dried reindeer meat, with estimates of 41.00 kg CO_2e/kg food in RISE compared with 0.40 kg CO_2e/kg food in SAFAD. Several of these foods were infrequently consumed, indicating that large differences in emission factors did not necessarily translate into large differences in total dietary carbon footprint.

Table D.1: Food items with the largest differences in emission factors between RISE and SAFAD. Emission factors are expressed as kg CO_2e per kg food, while total carbon footprint values are shown for context.

Food ID	Food name	n	Amount (g)	Emission factors (kg CO_2e/kg food)			Total carbon footprint (kg CO_2e)		
				RISE	SAFAD	Diff.	RISE	SAFAD	Diff.
1002	Dried reindeer meat	1	18	41.00	0.43	40.60	0.7	0.0	0.7
3993	Beef tartare	4	520	53.90	24.60	29.30	28.0	12.8	15.2
2908	Lamb ribs	1	66	35.50	64.20	-28.70	2.3	4.2	-1.9
3593	Fried lamb fillet	1	50	35.50	64.20	-28.70	1.8	3.2	-1.4
1403	Boiled lobster	7	540	42.00	17.10	24.90	22.7	9.2	13.4
4174	Beef carpaccio	1	60	53.90	31.40	22.50	3.2	1.9	1.3
1484	Fried liver	1	90	2.10	22.10	-20.00	0.2	2.0	-1.8
3038	Grilled squid	2	120	2.70	19.40	-16.70	0.3	2.3	-2.0
1080	Reindeer meat, sliced and fried	2	51	17.00	0.71	16.30	0.9	0.0	0.8
776	Fish burger with bun and assorted toppings	2	560	20.00	3.82	16.20	11.2	2.1	9.1

Table D.2 shows the food items contributing most to the total differences in carbon footprint based on reported consumption in the study.

Highly consumed foods contributed substantially to overall differences between databases. Hard cheese showed the largest total difference, while foods such as brewed coffee also contributed notably despite relatively small differences in emission factors, due to high reported consumption.

Table D.2: Food items with the largest differences in total carbon footprint between RISE and SAFAD. Total carbon footprint values are summed across all reported consumption, while emission factors are expressed as kg CO_2e per kg food.

Food ID	Food name	n	Amount (g)	Emission factors (kg CO_2e/kg food)			Total carbon footprint (kg CO_2e)		
				RISE	SAFAD	Diff.	RISE	SAFAD	Diff.
5778	Hard cheese, 27% fat	840	24658	5.80	9.19	-3.39	143.0	238.0	-94.5
2941	Boiled white fish	56	6402	1.90	11.60	-9.65	12.2	74.0	-61.8
5809	Bolognese sauce with beef and pork	37	5955	1.80	9.59	-7.79	10.7	59.3	-48.6
1395	Boiled shrimp	75	8524	17.30	12.40	4.89	147.0	106.0	41.7
1316	Fried salmon with salt	112	13966	5.64	8.26	-2.62	78.8	116.0	-37.7
1190	Cooked chicken	140	12934	6.12	3.37	2.75	79.1	44.5	34.6
6543	Beer, unspecified	107	51575	0.64	1.13	-0.49	32.8	64.8	-31.9
5741	Green beans, unspecified	34	3012	12.00	1.45	10.60	36.3	4.4	31.9
1957	Brewed coffee	1487	411612	0.45	0.28	0.17	185.0	155.0	30.1
5927	Tacos with assorted toppings, unspecified	12	2550	13.10	3.91	9.19	33.4	10.0	23.4

E

R Scripts

E.1 Data Processing

```
source("Script/Data Processing/1.Import packages and create maps.R")
source("Script/Data Processing/2.Import data.R")
source("Script/Data Processing/3.Rename and clean columns.R")
source("Script/Data Processing/4.RISE.R")
source("Script/Data Processing/5.SAFAD.R")
source("Script/Data Processing/6.Merge.R")
source("Script/Data Processing/7.Selection.R")
```

E.1.1 1.Import packages and create maps.R

```
# Import packages
library(readxl)
library(dplyr)
library(tidyr)
library(readr)
library(here)
library(tibble)
library(lme4)
library(lmerTest)
library(purrr)
library(stringr)
library(openxlsx)
library(ggplot2)
library(tidyverse)
library(rstatix)
library(ggpubr)
library(emmeans)
library(knitr)

# Create maps
path_in  <- here("IN_DATA")
path_out <- here("OUT_DATA")
path_results <- here("RESULTS")

if (!dir.exists(path_in)) {
  dir.create(path_in, recursive = TRUE)
}

if (!dir.exists(path_out)) {
  dir.create(path_out, recursive = TRUE)
```

```
}

if (!dir.exists(path_results)) {
  dir.create(path_results, recursive = TRUE)
}

cat("\n1.Import packages and create maps\n")
```

E.1.2 2.Import data.R

```
# Import data
diet_status_orginal <- read_excel(file.path(path_in, "kost_status.xlsx"))
diet_data_orginal <- read_excel(file.path(path_in, "kostdata.xlsx"))
diet_data_kcal_orginal <- read_excel(file.path(path_in, "kostdata_kcal_osv.xlsx"))
safad_orginal <- read_excel(file.path(path_in, "safad_original.xlsx"))
rise_orginal <- read_excel(file.path(path_in, "rise_original.xlsx"))
biomarkers_orginal <- read_excel(file.path(path_in, "biomarkers.xlsx"))
redcap_orginal <- read_excel(file.path(path_in, "redcap.xlsx"))

cat("\n2.Import data\n")
cat("\nQC AFTER IMPORT\n")
cat("Rows in diet_status_orginal:", nrow(diet_status_orginal), "\n")
cat("Rows in diet_data_orginal:", nrow(diet_data_orginal), "\n")
cat("Rows in diet_data_kcal_orginal:", nrow(diet_data_kcal_orginal), "\n")
cat("Rows in safad_orginal:", nrow(safad_orginal), "\n")
cat("Rows in rise_orginal:", nrow(rise_orginal), "\n")
cat("Rows in biomarkers_orginal:", nrow(biomarkers_orginal), "\n")
cat("Rows in redcap_orginal:", nrow(redcap_orginal), "\n")
```

E.1.3 3.Rename and clean columns.R

```
clean_chr <- function(x) {
  trimws(as.character(x))
}

clean_num <- function(x) {
  as.numeric(gsub(",", ".", trimws(as.character(x))))
}

clean_id <- function(x) {
  x <- clean_chr(x)
  x[x %in% c("", "NA", "NaN", "NULL")] <- NA_character_
  stringr::str_replace(x, "^([0-9]+)\\.0+$", "\\1")
}

diet_status <- diet_status_orginal %>%
  rename(
    `User ID` = `Användarnamn`,
    Status = `Kostregistreringsstatus`,
    `Registration Note` = `Registreringsdagens beskrivning`,
    Done = `Klar`,
    `Finish Date` = `Klarmarkerad datum`,
    `Start Date` = `Startdatum`,
    Day = `Dagnummer`
  ) %>%
  mutate(
    across(c(`User ID`, Status, `Registration Note`, Done), clean_chr),
```

```
    Day = clean_num(Day)
  ) %>%
  select(
    `User ID`,
    Status,
    `Registration Note`,
    Done,
    Day
  )

diet_data_english <- diet_data_orginal %>%
  rename(
    `User ID` = `Användarnamn`,
    `Food ID` = `LivsmedelsId`,
    `Food Name` = `Livsmedelsnamn`,
    `Food Amount` = `Livsmedel mängd`,
    `Main Group` = `Huvudgrupp`,
    Subgroup = `Delgrupp`,
    `Subgroup Amount` = `Delgrupp mängd`,
    Day = `Dag`
  )

diet_data_for_SWITCH <- diet_data_english %>%
  mutate(
    across(c(`User ID`, Subgroup), clean_chr),
    across(c(`Subgroup Amount`, Day), clean_num)
  ) %>%
  select(
    `User ID`,
    Subgroup,
    `Subgroup Amount`,
    Day
  )

diet_data_for_SAFAD <- diet_data_english %>%
  mutate(
    across(c(`User ID`, `Food ID`, `Food Name`), clean_chr),
    across(c(`Food Amount`, Day), clean_num)
  ) %>%
  select(
    `User ID`,
    Day,
    `Food ID`,
    `Food Name`,
    `Food Amount`
  )

diet_data_for_RISE <- diet_data_for_SAFAD

diet_data_kcal <- diet_data_kcal_orginal %>%
  rename(
    `User ID` = `Användarnamn`,
    Day = `Dag`,
    `Food ID` = `LivsmedelsId`,
    `Food Name` = `Livsmedelsnamn`,
    `Food Amount` = `Livsmedel mängd`,
    `Energy (kJ)` = `Ener (kJ)`,
    `Energy (kcal)` = `Ener (kcal)`,
```

```
    `Carbs (g)` = `Kolh (g)`,
    `Fat (g)` = `Fett (g)`,
    `Protein (g)` = `Prot (g)`,
    `Whole grains (g)` = `Fullk/tot (g)`
  ) %>%
  mutate(
    across(c(`User ID`, `Food ID`, `Food Name`), clean_chr),
    across(
      c(
        Day,
        `Food Amount`,
        `Energy (kJ)`,
        `Energy (kcal)`,
        `Carbs (g)`,
        `Fat (g)`,
        `Protein (g)`,
        `Whole grains (g)`
      ),
      clean_num
    )
  ) %>%
  select(
    `User ID`,
    Day,
    `Food ID`,
    `Food Name`,
    `Food Amount`,
    `Energy (kJ)`,
    `Energy (kcal)`,
    `Carbs (g)`,
    `Fat (g)`,
    `Protein (g)`,
    `Whole grains (g)`
  )

safad_indicators <- c(
  "Carbon footprint, total (kg CO2e) - 1000 g",
  "Carbon dioxide, total (kg CO2) - 1000 g",
  "Methane, fossil, total (kg CH4) - 1000 g",
  "Methane, biogenic, total (kg CH4) - 1000 g",
  "Nitrous oxide, total (kg N2O) - 1000 g",
  "Cropland (m2*year) - 1000 g",
  "New N input (kg N) - 1000 g",
  "New P input (kg P) - 1000 g",
  "Water (m3) - 1000 g",
  "Pesticides (g a.i) - 1000 g",
  "Biodiversity (E/MSY) - 1000 g",
  "Ammonia (kg NH3) - 1000 g",
  "Animal welfare (index) - 1000 g",
  "Antibiotics (index) - 1000 g"
)

safad <- safad_orginal %>%
  rename(
    `Food ID - SAFAD` = `SFA Code`,
    `Food Name - SAFAD` = `SFA Name`,
    `Carbon footprint, total (kg CO2e) - 1000 g` = `Carbon footprint, total (kg CO2e)`,
    `Carbon dioxide, total (kg CO2) - 1000 g` = `Carbon dioxide, total (kg CO2)`,
```

```
    `Methane, fossil, total (kg CH4) - 1000 g` = `Methane, fossil, total (kg CH4)`,
    `Methane, biogenic, total (kg CH4) - 1000 g` = `Methane, biogenic, total (kg CH4)`,
    `Nitrous oxide, total (kg N2O) - 1000 g` = `Nitrous oxide, total (kg N2O)`,
    `Cropland (m2*year) - 1000 g` = `Cropland (m2*year)`,
    `New N input (kg N) - 1000 g` = `New N input (kg N)`,
    `New P input (kg P) - 1000 g` = `New P input (kg P)`,
    `Water (m3) - 1000 g` = `Water (m3)`,
    `Pesticides (g a.i) - 1000 g` = `Pesticides (g a.i)`,
    `Biodiversity (E/MSY) - 1000 g` = `Biodiversity (E/MSY)`,
    `Ammonia (kg NH3) - 1000 g` = `Ammonia (kg NH3)`,
    `Animal welfare (index) - 1000 g` = `Animal welfare (index)`,
    `Antibiotics (index) - 1000 g` = `Antibiotics (index)`
  ) %>%
  mutate(
    across(c(`Food ID - SAFAD`, `Food Name - SAFAD`, `Ingredient Name`), clean_chr),
    across(c(`Net Amount (g)`, all_of(safad_indicators)), clean_num)
  ) %>%
  select(
    `Food ID - SAFAD`,
    `Food Name - SAFAD`,
    `Ingredient Name`,
    `Net Amount (g)`,
    all_of(safad_indicators)
  )

rise <- rise_orginal %>%
  rename(
    `Food ID - RISE` = `SLV ID`,
    `Food Name - RISE` = `SLV namn`,
    `RISE - Carbon footprint, total (kg CO2e) - 1000 g` =
      `Klimatavtryck kg CO2e/kg`
  ) %>%
  mutate(
    across(
      c(
        `Food ID - RISE`,
        `Food Name - RISE`,
        `Region ID`,
        Region,
        `Region källa`,
        `Konventionell/ekologisk`,
        `Konventionell/ekologisk källa`
      ),
      clean_chr
    ),
    `RISE - Carbon footprint, total (kg CO2e) - 1000 g` =
      clean_num(`RISE - Carbon footprint, total (kg CO2e) - 1000 g`)
  ) %>%
  select(
    `Food ID - RISE`,
    `Food Name - RISE`,
    `RISE - Carbon footprint, total (kg CO2e) - 1000 g`,
    `Region ID`,
    Region,
    `Region källa`,
    `Konventionell/ekologisk`,
    `Konventionell/ekologisk källa`
  )
```

```
biomarker_variables <- c(
  "Total Cholesterol",
  "LDL",
  "HDL",
  "Glucose",
  "Insulin",
  "Triglycerides"
)

biomarkers <- biomarkers_orginal %>%
  rename(
    `User ID - A` = ID,
    `User ID - B` = Visit,
    `Total Cholesterol` = `P-Kolesterol`,
    LDL = `P-LDL`,
    HDL = `P-HDL`,
    Insulin = `P-Insulin`,
    Triglycerides = `P-Triglycerider`,
    Glucose = `P-Glukos`
  ) %>%
  mutate(
    `User ID` = paste0(`User ID - A`, "-", `User ID - B`),
    `User ID` = clean_chr(`User ID`),
    across(all_of(biomarker_variables), clean_num)
  ) %>%
  select(
    `User ID`,
    all_of(biomarker_variables)
  )

redcap <- redcap_orginal %>%
  rename(
    `Participant ID` = `Användarnamn`
  ) %>%
  mutate(
    `Participant ID` = clean_chr(`Participant ID`),
    sex = clean_chr(sex),
    across(c(Block, age, height, weight, BMI, sgpal), clean_num)
  ) %>%
  select(
    Block,
    `Participant ID`,
    age,
    sex,
    height,
    weight,
    BMI,
    sgpal
  )

cat("\n3.Rename and clean columns\n")
cat("\nQC AFTER CLEANING\n")
participants <- n_distinct(sub("-.*$", "", diet_status$`User ID`))
visits <- n_distinct(diet_status$`User ID`)
cat("Number of participants (visits) in diet_status:", participants, "(", visits, ")\n")
```

```
participants <- n_distinct(sub("-.*$", "", diet_data_for_SWITCH$‘User ID‘))
visits <- n_distinct(diet_data_for_SWITCH$‘User ID‘)
cat("Number of participants (visits) in diet_data_for_SWITCH:", participants, "(", visits, ")\n")

participants <- n_distinct(sub("-.*$", "", diet_data_kcal$‘User ID‘))
visits <- n_distinct(diet_data_kcal$‘User ID‘)
cat("Number of participants (visits) in diet_data_kcal:", participants, "(", visits, ")\n")

participants <- n_distinct(sub("-.*$", "", biomarkers$‘User ID‘))
visits <- n_distinct(biomarkers$‘User ID‘)
cat("Number of participants (visits) in biomarkers:", participants, "(", visits, ")\n")

participants <- n_distinct(redcap$‘Participant ID‘)
cat("Number of participants in redcap:", participants, "\n")
```

E.1.4 4.RISE.R

```
manual_mapping_file <- file.path(path_in, "RISE_manual_mapping_full.xlsx")

# 4.1 Prepare and average duplicated RISE Food IDs
rise_prepared <- rise %>%
  mutate(
    ‘Food ID - RISE‘ = clean_id(‘Food ID - RISE‘)
  )

rise_duplicates <- rise_prepared %>%
  filter(!is.na(‘Food ID - RISE‘)) %>%
  count(‘Food ID - RISE‘) %>%
  filter(n > 1)

rise_unique <- rise_prepared %>%
  filter(!is.na(‘Food ID - RISE‘)) %>%
  group_by(‘Food ID - RISE‘) %>%
  summarise(
    ‘Food Name - RISE‘ = dplyr::first(
      stats::na.omit(‘Food Name - RISE‘),
      default = NA_character_
    ),
    ‘RISE - Carbon footprint, total (kg CO2e) - 1000 g‘ = if_else(
      all(is.na(‘RISE - Carbon footprint, total (kg CO2e) - 1000 g‘)),
      NA_real_,
      mean(‘RISE - Carbon footprint, total (kg CO2e) - 1000 g‘, na.rm = TRUE)
    ),
    n_RISE_rows_averaged = n(),
    .groups = "drop"
  ) %>%
  distinct(‘Food ID - RISE‘, .keep_all = TRUE)

rise_duplicate_average_log <- rise_prepared %>%
  semi_join(rise_duplicates, by = "Food ID - RISE")

cat("\n4.RISE\n")
cat("\nQC: RISE DATABASE AFTER AVERAGING\n")
cat("Rows in rise_unique:", nrow(rise_unique), "\n")
cat("Unique RISE Food IDs:", n_distinct(rise_unique$‘Food ID - RISE‘), "\n")

# 4.2 Prepare dietary data for RISE and remove exact duplicate rows
```

```
diet_data_for_RISE_raw_n <- nrow(diet_data_for_RISE)

diet_data_for_RISE_duplicates <- diet_data_for_RISE %>%
  mutate(
    `User ID` = clean_chr(`User ID`),
    `Food ID` = clean_id(`Food ID`),
    `Food Name` = clean_chr(`Food Name`),
    `Food Amount` = clean_num(`Food Amount`),
    Day = clean_num(Day)
  ) %>%
  count(`User ID`, Day, `Food ID`, `Food Name`, `Food Amount`) %>%
  filter(n > 1) %>%
  arrange(desc(n))

diet_data_for_RISE <- diet_data_for_RISE %>%
  mutate(
    `User ID` = clean_chr(`User ID`),
    `Food ID` = clean_id(`Food ID`),
    `Food Name` = clean_chr(`Food Name`),
    `Food Amount` = clean_num(`Food Amount`),
    Day = clean_num(Day)
  ) %>%
  distinct(`User ID`, Day, `Food ID`, `Food Name`, `Food Amount`, .keep_all = TRUE)

cat("\nQC: RISE DIET DATA DEDUPLICATION\n")
cat("Rows before deduplication:", diet_data_for_RISE_raw_n, "\n")
cat("Rows after deduplication:", nrow(diet_data_for_RISE), "\n")
cat("Exact duplicate rows removed:", diet_data_for_RISE_raw_n - nrow(diet_data_for_RISE), "\n")
cat("Duplicate food records remaining:",
    nrow(
      diet_data_for_RISE %>%
        count(`User ID`, Day, `Food ID`, `Food Name`, `Food Amount`) %>%
        filter(n > 1)
    ),
    "\n"
)

# 4.3 Match diet data to RISE before manual mapping
diet_data_before_manual <- diet_data_for_RISE %>%
  left_join(rise_unique, by = c("Food ID" = "Food ID - RISE")) %>%
  mutate(
    has_rise_match_before_manual =
      !is.na(`RISE - Carbon footprint, total (kg CO2e) - 1000 g`)
  )

cat("\nQC: BEFORE MANUAL RISE MAPPING\n")
cat("Matched rows:", sum(diet_data_before_manual$has_rise_match_before_manual), "\n")
cat("Unmatched rows:", sum(!diet_data_before_manual$has_rise_match_before_manual), "\n")
cat(
  "Unmatched Food IDs:",
  n_distinct(diet_data_before_manual$`Food ID`[
    !diet_data_before_manual$has_rise_match_before_manual
  ], na.rm = TRUE),
  "\n"
)

# 4.4 Create manual mapping file
rise_unmatched_mapping <- diet_data_before_manual %>%
```

```
  filter(!has_rise_match_before_manual) %>%
  group_by(`Food ID`, `Food Name`) %>%
  summarise(
    n_rows = n(),
    total_g = sum(`Food Amount`, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  arrange(desc(total_g)) %>%
  mutate(
    `Manual Food ID - RISE` = NA_character_,
    `Manual Food Name - RISE` = NA_character_,
    `Mapping comment` = NA_character_
  )

cat("\nQC: MANUAL RISE MAPPING LIST\n")
cat("Food IDs in manual mapping list:", nrow(rise_unmatched_mapping), "\n")
cat(
  "Total unmatched grams:",
  round(sum(rise_unmatched_mapping$total_g, na.rm = TRUE), 1),
  "\n"
)

if (!file.exists(manual_mapping_file)) {
  openxlsx::write.xlsx(
    rise_unmatched_mapping,
    manual_mapping_file,
    overwrite = FALSE
  )

  cat("Manual mapping file was created:", manual_mapping_file, "\n")
} else {
  cat("Existing manual mapping file was not overwritten\n")
}

# 4.5 Import manual mapping
manual_mapping_raw <- readxl::read_excel(manual_mapping_file)

mapping_rise <- manual_mapping_raw %>%
  transmute(
    `Food ID` = clean_id(`Food ID`),
    `Manual Food ID - RISE` = clean_id(`Manual Food ID - RISE`),
    `Manual Food Name - RISE` = clean_chr(`Manual Food Name - RISE`)
  ) %>%
  filter(
    !is.na(`Food ID`),
    !is.na(`Manual Food ID - RISE`)
  ) %>%
  distinct(`Food ID`, .keep_all = TRUE)

mapping_rise_unmatched <- mapping_rise %>%
  anti_join(
    rise_unique,
    by = c("Manual Food ID - RISE" = "Food ID - RISE")
  )

cat("\nQC: IMPORTED MANUAL RISE MAPPING\n")
cat("Manual mappings imported:", nrow(mapping_rise), "\n")
cat("Manual mappings without RISE match:", nrow(mapping_rise_unmatched), "\n")
```

```
# 4.6 Match diet data to RISE after manual mapping
diet_data_with_RISE <- diet_data_for_RISE %>%
  left_join(mapping_rise, by = "Food ID") %>%
  mutate(
    `Food ID Final - RISE` = coalesce(
      `Manual Food ID - RISE`,
      `Food ID`
    ),
    manually_mapped_to_RISE = !is.na(`Manual Food ID - RISE`)
  ) %>%
  left_join(rise_unique, by = c("Food ID Final - RISE" = "Food ID - RISE")) %>%
  mutate(
    has_rise_match =
      !is.na(`RISE - Carbon footprint, total (kg CO2e) - 1000 g`),
    `RISE Carbon e (kg)` =
      (`Food Amount` / 1000) *
      `RISE - Carbon footprint, total (kg CO2e) - 1000 g`
  )

# 4.7 QC after manual mapping
cat("\nQC: AFTER MANUAL RISE MAPPING\n")
cat("Matched rows:", sum(diet_data_with_RISE$has_rise_match), "\n")
cat("Unmatched rows:", sum(!diet_data_with_RISE$has_rise_match), "\n")
cat(
  "Unmatched Food IDs:",
  n_distinct(diet_data_with_RISE$`Food ID`[
    !diet_data_with_RISE$has_rise_match
  ], na.rm = TRUE),
  "\n"
)

cat(
  "Manually mapped Food IDs:",
  n_distinct(diet_data_with_RISE$`Food ID`[
    diet_data_with_RISE$manually_mapped_to_RISE
  ], na.rm = TRUE),
  "\n"
)

cat(
  "Total RISE carbon e (kg), row level:",
  round(sum(diet_data_with_RISE$`RISE Carbon e (kg)`, na.rm = TRUE), 3),
  "\n"
)

# 4.8 Create day-level RISE output
day_rise <- diet_data_with_RISE %>%
  group_by(`User ID`, Day) %>%
  summarise(
    `RISE Carbon e (kg)` = sum(`RISE Carbon e (kg)`, na.rm = TRUE),
    RISE_matched_g = sum(`Food Amount`[has_rise_match], na.rm = TRUE),
    RISE_unmatched_g = sum(`Food Amount`[!has_rise_match], na.rm = TRUE),
    .groups = "drop"
  )

# 4.9 QC day-level RISE output
cat("\nQC: DAY-LEVEL RISE OUTPUT\n")
```

```
cat("Rows in day_rise:", nrow(day_rise), "\n")
cat("Unique User IDs:", n_distinct(day_rise$`User ID`), "\n")
cat(
  "Total RISE carbon e (kg), day level:",
  round(sum(day_rise$`RISE Carbon e (kg)`, na.rm = TRUE), 3),
  "\n"
)
cat(
  "Total RISE unmatched grams:",
  round(sum(day_rise$RISE_unmatched_g, na.rm = TRUE), 1),
  "\n"
)
```

E.1.5 5.SAFAD.R

```
manual_mapping_file <- file.path(path_in, "SAFAD_manual_mapping_full.xlsx")

# 5.1 Prepare SAFAD total-only table
safad_total <- safad %>%
  filter(`Ingredient Name` == "(total)") %>%
  mutate(
    `Food ID - SAFAD` = clean_id(`Food ID - SAFAD`)
  ) %>%
  distinct(`Food ID - SAFAD`, .keep_all = TRUE)

# 5.2 Prepare dietary data for SAFAD
diet_data_for_SAFAD <- diet_data_kcal %>%
  mutate(
    `Food ID` = clean_id(`Food ID`),
    `Food Amount` = clean_num(`Food Amount`)
  )

# 5.3 Match diet data to SAFAD before manual mapping
diet_data_before_manual <- diet_data_for_SAFAD %>%
  left_join(
    safad_total,
    by = c("Food ID" = "Food ID - SAFAD")
  ) %>%
  mutate(
    has_safad_match_before_manual =
      !is.na(`Carbon footprint, total (kg CO2e) - 1000 g`)
  )

cat("\n5.SAFAD\n")
cat("\nQC: BEFORE MANUAL SAFAD MAPPING\n")
cat("Matched rows:", sum(diet_data_before_manual$has_safad_match_before_manual), "\n")
cat("Unmatched rows:", sum(!diet_data_before_manual$has_safad_match_before_manual), "\n")
cat(
  "Unmatched Food IDs:",
  n_distinct(diet_data_before_manual$`Food ID`[
    !diet_data_before_manual$has_safad_match_before_manual
  ], na.rm = TRUE),
  "\n"
)

# 5.4 Create manual mapping file
safad_unmatched_mapping <- diet_data_before_manual %>%
```

```
  filter(!has_safad_match_before_manual) %>%
  group_by(`Food ID`, `Food Name`) %>%
  summarise(
    n_rows = n(),
    total_g = sum(`Food Amount`, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  arrange(desc(total_g)) %>%
  mutate(
    `Manual Food ID - SAFAD` = NA_character_,
    `Manual Food Name - SAFAD` = NA_character_,
    `Mapping comment` = NA_character_
  )

cat("\nQC: MANUAL SAFAD MAPPING LIST\n")
cat("Food IDs in manual mapping list:", nrow(safad_unmatched_mapping), "\n")
cat(
  "Total unmatched grams:",
  round(sum(safad_unmatched_mapping$total_g, na.rm = TRUE), 1),
  "\n"
)

if (!file.exists(manual_mapping_file)) {
  openxlsx::write.xlsx(
    safad_unmatched_mapping,
    manual_mapping_file,
    overwrite = FALSE
  )

  cat("Manual mapping file was created:", manual_mapping_file, "\n")
} else {
  cat("Existing manual mapping file was not overwritten\n")
}

# 5.5 Import manual mapping
manual_mapping_raw <- readxl::read_excel(manual_mapping_file)

mapping_safad <- manual_mapping_raw %>%
  transmute(
    `Food ID` = clean_id(`Food ID`),
    `Manual Food ID - SAFAD` = clean_id(`Manual Food ID - SAFAD`),
    `Manual Food Name - SAFAD` = clean_chr(`Manual Food Name - SAFAD`)
  ) %>%
  filter(
    !is.na(`Food ID`),
    !is.na(`Manual Food ID - SAFAD`)
  ) %>%
  distinct(`Food ID`, .keep_all = TRUE)

cat("\nQC: IMPORTED MANUAL SAFAD MAPPING\n")
cat("Manual mappings imported:", nrow(mapping_safad), "\n")

# 5.6 Match diet data to SAFAD after manual mapping
diet_data_with_SAFAD <- diet_data_for_SAFAD %>%
  left_join(mapping_safad, by = "Food ID") %>%
  mutate(
    `Food ID Final - SAFAD` = coalesce(
      `Manual Food ID - SAFAD`,
```

```
      `Food ID`
    ),
    manually_mapped_to_SAFAD = !is.na(`Manual Food ID - SAFAD`)
  ) %>%
  left_join(
    safad_total,
    by = c("Food ID Final - SAFAD" = "Food ID - SAFAD")
  ) %>%
  mutate(
    has_safad_match =
      !is.na(`Carbon footprint, total (kg CO2e) - 1000 g`)
  )

# 5.7 QC after manual mapping
cat("\nQC: AFTER MANUAL SAFAD MAPPING\n")
cat("Matched rows:", sum(diet_data_with_SAFAD$has_safad_match), "\n")
cat("Unmatched rows:", sum(!diet_data_with_SAFAD$has_safad_match), "\n")
cat(
  "Unmatched Food IDs:",
  n_distinct(diet_data_with_SAFAD$`Food ID`[
    !diet_data_with_SAFAD$has_safad_match
  ], na.rm = TRUE),
  "\n"
)
cat(
  "Manually mapped Food IDs:",
  n_distinct(diet_data_with_SAFAD$`Food ID`[
    diet_data_with_SAFAD$manually_mapped_to_SAFAD
  ], na.rm = TRUE),
  "\n"
)

# 5.8 Calculate SAFAD values for consumed amoun
diet_data_with_SAFAD <- diet_data_with_SAFAD %>%
  mutate(
    scale_factor = if_else(
      !is.na(`Net Amount (g)`) & `Net Amount (g)` > 0,
      `Food Amount` / `Net Amount (g)`,
      NA_real_
    ),
    carbon_kg =
      scale_factor * `Carbon footprint, total (kg CO2e) - 1000 g`,
    carbon_dioxide_kg =
      scale_factor * `Carbon dioxide, total (kg CO2) - 1000 g`,
    methane_fossil_kg =
      scale_factor * `Methane, fossil, total (kg CH4) - 1000 g`,
    methane_biogenic_kg =
      scale_factor * `Methane, biogenic, total (kg CH4) - 1000 g`,
    nitrous_oxide_kg =
      scale_factor * `Nitrous oxide, total (kg N2O) - 1000 g`,
    cropland_m2yr =
      scale_factor * `Cropland (m2*year) - 1000 g`,
    new_n_kg =
      scale_factor * `New N input (kg N) - 1000 g`,
    new_p_kg =
      scale_factor * `New P input (kg P) - 1000 g`,
    water_m3 =
      scale_factor * `Water (m3) - 1000 g`,
```

```
    pesticides_g =
      scale_factor * `Pesticides (g a.i) - 1000 g`,
    biodiversity =
      scale_factor * `Biodiversity (E/MSY) - 1000 g`,
    ammonia_kg =
      scale_factor * `Ammonia (kg NH3) - 1000 g`,
    animal_welfare =
      scale_factor * `Animal welfare (index) - 1000 g`,
    antibiotics =
      scale_factor * `Antibiotics (index) - 1000 g`
  )

# 5.9 Keep relevant row-level columns
diet_data_with_SAFAD <- diet_data_with_SAFAD %>%
  select(
    `User ID`,
    Day,
    `Food ID`,
    `Food ID Final - SAFAD`,
    `Food Name`,
    `Food Amount`,
    has_safad_match,
    manually_mapped_to_SAFAD,
    carbon_kg,
    carbon_dioxide_kg,
    methane_fossil_kg,
    methane_biogenic_kg,
    nitrous_oxide_kg,
    cropland_m2yr,
    new_n_kg,
    new_p_kg,
    water_m3,
    pesticides_g,
    biodiversity,
    ammonia_kg,
    animal_welfare,
    antibiotics
  ) %>%
  rename(
    `Carbon e (kg)` = carbon_kg,
    `Carbon dioxide (kg)` = carbon_dioxide_kg,
    `Methane fossil (kg)` = methane_fossil_kg,
    `Methane biogenic (kg)` = methane_biogenic_kg,
    `Nitrous oxide (kg)` = nitrous_oxide_kg,
    `Cropland (m2yr)` = cropland_m2yr,
    `New n (kg)` = new_n_kg,
    `New p (kg)` = new_p_kg,
    `Water (m3)` = water_m3,
    `Pesticides (g)` = pesticides_g,
    Biodiversity = biodiversity,
    `Ammonia (kg)` = ammonia_kg,
    `Animal welfare` = animal_welfare,
    Antibiotics = antibiotics
  )

# 5.10 Create day-level SAFAD output
day_safad <- diet_data_with_SAFAD %>%
  group_by(`User ID`, Day) %>%
```

```
  summarise(
    `Carbon e (kg)` = sum(`Carbon e (kg)`, na.rm = TRUE),
    `Carbon dioxide (kg)` = sum(`Carbon dioxide (kg)`, na.rm = TRUE),
    `Methane fossil (kg)` = sum(`Methane fossil (kg)`, na.rm = TRUE),
    `Methane biogenic (kg)` = sum(`Methane biogenic (kg)`, na.rm = TRUE),
    `Nitrous oxide (kg)` = sum(`Nitrous oxide (kg)`, na.rm = TRUE),
    `Cropland (m2yr)` = sum(`Cropland (m2yr)`, na.rm = TRUE),
    `New n (kg)` = sum(`New n (kg)`, na.rm = TRUE),
    `New p (kg)` = sum(`New p (kg)`, na.rm = TRUE),
    `Water (m3)` = sum(`Water (m3)`, na.rm = TRUE),
    `Pesticides (g)` = sum(`Pesticides (g)`, na.rm = TRUE),
    Biodiversity = sum(Biodiversity, na.rm = TRUE),
    `Ammonia (kg)` = sum(`Ammonia (kg)`, na.rm = TRUE),
    `Animal welfare` = sum(`Animal welfare`, na.rm = TRUE),
    Antibiotics = sum(Antibiotics, na.rm = TRUE),
    SAFAD_matched_g = sum(`Food Amount`[has_safad_match], na.rm = TRUE),
    SAFAD_unmatched_g = sum(`Food Amount`[!has_safad_match], na.rm = TRUE),
    .groups = "drop"
  )

# 5.11 QC day-level output
cat("\nQC: DAY-LEVEL SAFAD OUTPUT\n")
cat("Rows in day_safad:", nrow(day_safad), "\n")
cat("Unique User IDs:", n_distinct(day_safad$`User ID`), "\n")
cat(
  "Total SAFAD carbon e (kg):",
  round(sum(day_safad$`Carbon e (kg)`, na.rm = TRUE), 3),
  "\n"
)
```

E.1.6 6.Merge.R

```
# 6.1 Group definitions for SWITCH
switch_groups <- list(
  vegetables = c("Grönsaker", "Rotfrukter", "Svamp"),
  fruit_berries = c("Frukt", "Bär"),
  legumes = c("Baljväxter"),
  fish_shellfish = c("Fisk", "Skaldjur"),
  meat = c("Kött", "Processat kött, charkvaror"),
  poultry = c("Fågel")
)

sum_switch_group <- function(subgroup, amount, target_group) {
  sum(amount[subgroup %in% target_group], na.rm = TRUE)
}

# 6.2 Summarise SWITCH food groups to day level
switch_day <- diet_data_for_SWITCH %>%
  mutate(
    `User ID` = clean_chr(`User ID`),
    Subgroup = clean_chr(Subgroup),
    `Subgroup Amount` = clean_num(`Subgroup Amount`),
    Day = clean_num(Day)
  ) %>%
  group_by(`User ID`, Day) %>%
  summarise(
    switch_vegetables_g =
      sum_switch_group(Subgroup, `Subgroup Amount`, switch_groups$vegetables),
```

```
    switch_fruit_berries_g =
      sum_switch_group(Subgroup, `Subgroup Amount`, switch_groups$fruit_berries),
    switch_legumes_g =
      sum_switch_group(Subgroup, `Subgroup Amount`, switch_groups$legumes),
    switch_fish_shellfish_g =
      sum_switch_group(Subgroup, `Subgroup Amount`, switch_groups$fish_shellfish),
    switch_meat_g =
      sum_switch_group(Subgroup, `Subgroup Amount`, switch_groups$meat),
    switch_poultry_g =
      sum_switch_group(Subgroup, `Subgroup Amount`, switch_groups$poultry),
    .groups = "drop"
  )

# 6.3 Summarise energy and nutrients to day level
day_kcal <- diet_data_kcal %>%
  mutate(
    `User ID` = clean_chr(`User ID`),
    Day = clean_num(Day),
    `Food Amount` = clean_num(`Food Amount`),
    `Energy (kJ)` = clean_num(`Energy (kJ)`),
    `Energy (kcal)` = clean_num(`Energy (kcal)`),
    `Carbs (g)` = clean_num(`Carbs (g)`),
    `Fat (g)` = clean_num(`Fat (g)`),
    `Protein (g)` = clean_num(`Protein (g)`),
    `Whole grains (g)` = clean_num(`Whole grains (g)`)
  ) %>%
  group_by(`User ID`, Day) %>%
  summarise(
    `Food Amount Total (g)` = sum(`Food Amount`, na.rm = TRUE),
    `Energy Total (kJ)` = sum(`Energy (kJ)`, na.rm = TRUE),
    `Energy Total (kcal)` = sum(`Energy (kcal)`, na.rm = TRUE),
    `Carbs Total (g)` = sum(`Carbs (g)`, na.rm = TRUE),
    `Fat Total (g)` = sum(`Fat (g)`, na.rm = TRUE),
    `Protein Total (g)` = sum(`Protein (g)`, na.rm = TRUE),
    `Whole Grains Total (g)` = sum(`Whole grains (g)`, na.rm = TRUE),
    .groups = "drop"
  )

# 6.4 Prepare diet status
diet_status_std <- diet_status %>%
  mutate(
    `User ID` = clean_chr(`User ID`),
    Status = clean_chr(Status),
    `Registration Note` = clean_chr(`Registration Note`),
    Done = clean_chr(Done),
    Day = clean_num(Day)
  )

# 6.5 Merge all day-level datasets
day_final <- diet_status_std %>%
  left_join(switch_day, by = c("User ID", "Day")) %>%
  left_join(day_kcal, by = c("User ID", "Day")) %>%
  left_join(day_safad, by = c("User ID", "Day")) %>%
  left_join(day_rise, by = c("User ID", "Day")) %>%
  left_join(biomarkers, by = "User ID") %>%
  tidyr::separate(
    `User ID`,
    into = c("Participant ID", "Visit"),
```

```
    sep = "-",
    remove = FALSE
  ) %>%
  mutate(
    `Participant ID` = clean_chr(`Participant ID`),
    Visit = clean_num(Visit)
  )

# 6.6 QC after merge
cat("\n6.Merge\n")
cat("\nQC: MERGED DAY-LEVEL DATA\n")
cat("Rows in day_final:", nrow(day_final), "\n")
cat("Unique User IDs:", n_distinct(day_final$`User ID`), "\n")
cat("Unique Participant IDs:", n_distinct(day_final$`Participant ID`), "\n")
```

E.1.7 7.Selection.R

```
# 7.1 Select completed and valid diet records
excluded_participants <- c("test_8", "test_10", "1224")

day_selected <- day_final %>%
  mutate(
    `User ID` = clean_chr(`User ID`),
    `Participant ID` = clean_chr(`Participant ID`),
    Done = clean_chr(Done),
    Day = clean_num(Day),
    Visit = clean_num(Visit),
    `Energy Total (kcal)` = clean_num(`Energy Total (kcal)`)
  ) %>%
  filter(
    !`Participant ID` %in% excluded_participants,
    !is.na(`User ID`),
    !is.na(`Participant ID`),
    !is.na(Day),
    !is.na(Visit),
    !is.na(`Energy Total (kcal)`),
    Done == "Ja"
  )

cat("\n7.Selection\n")
cat("\nQC: VALID COMPLETED DIET DAYS\n")
cat("Rows after Done = Ja filter:", nrow(day_selected), "\n")
cat("Unique User IDs:", n_distinct(day_selected$`User ID`), "\n")
cat("Unique Participant IDs:", n_distinct(day_selected$`Participant ID`), "\n")


# 7.2 Count valid days per participant-visit
participant_day_count <- day_selected %>%
  group_by(`User ID`, Visit) %>%
  summarise(
    valid_days = n_distinct(Day),
    .groups = "drop"
  )

# 7.3 Keep participant-visits with at least 2 valid days
day_final_selected <- day_selected %>%
  left_join(participant_day_count, by = c("User ID", "Visit")) %>%
  filter(valid_days >= 2)
```

```
cat("\nQC: PARTICIPANT-VISITS WITH AT LEAST 2 VALID DAYS\n")
cat("Rows retained:", nrow(day_final_selected), "\n")
cat("Participant-visits retained:", n_distinct(day_final_selected$`User ID`), "\n")
cat("Participant IDs retained:", n_distinct(day_final_selected$`Participant ID`), "\n")

# 7.4 Create average per participant-visit
average_columns <- c(
  "switch_vegetables_g" = "Average Switch Vegetables (g)",
  "switch_fruit_berries_g" = "Average Switch Fruit & Berries (g)",
  "Whole Grains Total (g)" = "Average Switch Whole Grains (g)",
  "switch_legumes_g" = "Average Switch Legumes (g)",
  "switch_fish_shellfish_g" = "Average Switch Fish & Shellfish (g)",
  "switch_meat_g" = "Average Meat (g)",
  "switch_poultry_g" = "Average Poultry (g)",

  "Food Amount Total (g)" = "Average Food Amount (g)",
  "Energy Total (kJ)" = "Average Energy (kJ)",
  "Energy Total (kcal)" = "Average Energy (kcal)",
  "Carbs Total (g)" = "Average Carbs (g)",
  "Fat Total (g)" = "Average Fat (g)",
  "Protein Total (g)" = "Average Protein (g)",

  "RISE Carbon e (kg)" = "Average Carbon e RISE (kg)",
  "Carbon e (kg)" = "Average Carbon e (kg)",
  "Carbon dioxide (kg)" = "Average Carbon dioxide (kg)",
  "Methane fossil (kg)" = "Average Methane fossil (kg)",
  "Methane biogenic (kg)" = "Average Methane biogenic (kg)",
  "Nitrous oxide (kg)" = "Average Nitrous oxide (kg)",
  "Cropland (m2yr)" = "Average Cropland (m2yr)",
  "New n (kg)" = "Average New n (kg)",
  "New p (kg)" = "Average New p (kg)",
  "Water (m3)" = "Average Water (m3)",
  "Pesticides (g)" = "Average Pesticides (g)",
  "Biodiversity" = "Average Biodiversity",
  "Ammonia (kg)" = "Average Ammonia (kg)",
  "Animal welfare" = "Average Animal welfare",
  "Antibiotics" = "Average Antibiotics",

  "Glucose" = "Glucose",
  "Total Cholesterol" = "Total Cholesterol",
  "LDL" = "LDL",
  "HDL" = "HDL",
  "Insulin" = "Insulin",
  "Triglycerides" = "Triglycerides"
)

day_average_selected_v1 <- day_final_selected %>%
  group_by(`User ID`, `Participant ID`, Visit) %>%
  summarise(
    valid_days = first(valid_days),
    across(
      all_of(names(average_columns)),
      ~ mean(.x, na.rm = TRUE)
    ),
    .groups = "drop"
  ) %>%
  rename_with(
```

```
    ~ unname(average_columns[.x]),
    all_of(names(average_columns))
  )

# 7.5 Create average energy value for Goldberg
average_energy_goldberg <- day_final_selected %>%
  group_by(`Participant ID`) %>%
  summarise(
    `Average Energy Goldberg (kcal)` =
      mean(`Energy Total (kcal)`, na.rm = TRUE),
    valid_days_goldberg = n_distinct(paste(Visit, Day)),
    .groups = "drop"
  )

# 7.6 Add REDCap variables
redcap_std <- redcap %>%
  mutate(
    `Participant ID` = clean_chr(`Participant ID`)
  )

day_average_selected <- day_average_selected_v1 %>%
  mutate(
    `Participant ID` = clean_chr(`Participant ID`)
  ) %>%
  left_join(redcap_std, by = "Participant ID") %>%
  left_join(average_energy_goldberg, by = "Participant ID")

# 7.7 Goldberg classification using Schofield
SWITCH_post_data_processing_all <- day_average_selected %>%
  mutate(
    `Participant ID` = clean_chr(`Participant ID`),
    `User ID` = clean_chr(`User ID`),
    sex = clean_chr(sex),
    age = clean_num(age),
    weight = clean_num(weight),
    BMI = clean_num(BMI),
    sgpal = clean_num(sgpal),
    `Average Energy (kcal)` = clean_num(`Average Energy (kcal)`),
    `Average Energy Goldberg (kcal)` =
      clean_num(`Average Energy Goldberg (kcal)`)
  ) %>%
  mutate(
    PAL = case_when(
      sgpal == 1 ~ 1.4,
      sgpal == 1.5 ~ 1.5,
      sgpal == 2 ~ 1.6,
      sgpal == 2.5 ~ 1.7,
      sgpal == 3 ~ 1.8,
      sgpal == 3.5 ~ 1.9,
      sgpal == 4 ~ 2.0,
      TRUE ~ NA_real_
    )
  ) %>%
  mutate(
    BMR_Schofield = case_when(
      sex == "Male" & age >= 30 & age < 60 ~ 11.472 * weight + 873.1,
      sex == "Male" & age >= 60 ~ 11.711 * weight + 587.7,
      sex == "Female" & age >= 18 & age < 30 ~ 14.818 * weight + 486.6,
```

```
    sex == "Female" & age >= 30 & age < 60 ~ 8.126 * weight + 845.6,
    sex == "Female" & age >= 60 ~ 9.082 * weight + 658.5,
    TRUE ~ NA_real_
  ),
  BMR_method = "Schofield",
  BMR_selected = BMR_Schofield
) %>%
mutate(
  energy_requirement_kcal = BMR_selected * PAL,
  EI_BMR = `Average Energy Goldberg (kcal)` / BMR_selected
) %>%
mutate(
  CVwEI = 23,
  CVwB = 8.5,
  CVtP = 15,
  d_goldberg = valid_days_goldberg,
  z = 2,
  S = sqrt((CVwEI^2 / d_goldberg) + CVwB^2 + CVtP^2),
  lower_cut = PAL * exp(-z * (S / 100)),
  upper_cut = PAL * exp(z * (S / 100))
) %>%
mutate(
  goldberg = case_when(
    is.na(EI_BMR) |
      is.na(PAL) |
      is.na(BMR_selected) |
      is.na(d_goldberg) ~ NA_character_,
    EI_BMR < lower_cut ~ "Under-reporter",
    EI_BMR > upper_cut ~ "Over-reporter",
    TRUE ~ "Plausible"
  )
) %>%
mutate(
  `Carbon e RISE per 1000 req kcal` =
    (`Average Carbon e RISE (kg)` / energy_requirement_kcal) * 1000,
  `Carbon e per 1000 req kcal` =
    (`Average Carbon e (kg)` / energy_requirement_kcal) * 1000,
  `Carbon dioxide per 1000 req kcal` =
    (`Average Carbon dioxide (kg)` / energy_requirement_kcal) * 1000,
  `Methane fossil per 1000 req kcal` =
    (`Average Methane fossil (kg)` / energy_requirement_kcal) * 1000,
  `Methane biogenic per 1000 req kcal` =
    (`Average Methane biogenic (kg)` / energy_requirement_kcal) * 1000,
  `Nitrous oxide per 1000 req kcal` =
    (`Average Nitrous oxide (kg)` / energy_requirement_kcal) * 1000,
  `Cropland per 1000 req kcal` =
    (`Average Cropland (m2yr)` / energy_requirement_kcal) * 1000,
  `New n per 1000 req kcal` =
    (`Average New n (kg)` / energy_requirement_kcal) * 1000,
  `New p per 1000 req kcal` =
    (`Average New p (kg)` / energy_requirement_kcal) * 1000,
  `Water per 1000 req kcal` =
    (`Average Water (m3)` / energy_requirement_kcal) * 1000,
  `Pesticides per 1000 req kcal` =
    (`Average Pesticides (g)` / energy_requirement_kcal) * 1000,
  `Biodiversity per 1000 req kcal` =
    (`Average Biodiversity` / energy_requirement_kcal) * 1000,
  `Ammonia per 1000 req kcal` =
```

```
      (‘Average Ammonia (kg)‘ / energy_requirement_kcal) * 1000,
    ‘Animal welfare per 1000 req kcal‘ =
      (‘Average Animal welfare‘ / energy_requirement_kcal) * 1000,
    ‘Antibiotics per 1000 req kcal‘ =
      (‘Average Antibiotics‘ / energy_requirement_kcal) * 1000
  )

# 7.8 Keep plausible reporters
SWITCH_post_data_processing <- SWITCH_post_data_processing_all %>%
  filter(goldberg == "Plausible")

# 7.9 QC after Goldberg selection
cat("\nQC: GOLDBERG CLASSIFICATION\n")
cat("All participant-visits:", nrow(SWITCH_post_data_processing_all), "\n")
cat("Plausible:", sum(SWITCH_post_data_processing_all$goldberg == "Plausible", na.rm = TRUE), "\n")
cat("Under-reporters:", sum(SWITCH_post_data_processing_all$goldberg == "Under-reporter", na.rm = TRU
cat("Over-reporters:", sum(SWITCH_post_data_processing_all$goldberg == "Over-reporter", na.rm = TRUE)
cat("Missing Goldberg classification:", sum(is.na(SWITCH_post_data_processing_all$goldberg)), "\n")

cat("\nQC: FINAL SELECTED DATASET\n")
cat("Rows retained:", nrow(SWITCH_post_data_processing), "\n")
cat("Unique Participant IDs retained:", n_distinct(SWITCH_post_data_processing$‘Participant ID‘), "\n
```

E.2 Data Presentation

```
# RESULTS
source("Script/Data Presentation/Adherence_to_SWITCH_targets.R")
source("Script/Data Presentation/RQ1.R")
source("Script/Data Presentation/RQ2.R")
source("Script/Data Presentation/RQ3.R")
source("Script/Data Presentation/RQ4.R")

# APPENDIX
source("Script/Data Presentation/Participant_Characteristics.R")
source("Script/Data Presentation/QQ_plot_RISE.R")
source("Script/Data Presentation/QQ_plot_SAFAD.R")
source("Script/Data Presentation/Subgroup_Contribution.R")
```

E.2.1 Q–Q plot for SAFAD

```
cat("\nQQ Plot for SAFAD\n")
QQ_SAFAD_original <- ggplot(SWITCH_post_data_processing, aes(sample = ‘Average Carbon e (kg)‘)) +
  stat_qq(color = "#2066a8", size = 2) +
  stat_qq_line(color = "#3594cc", linewidth = 1) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold")) +
  labs(
    x = "Theoretical quantiles",
    y = "Sample quantiles: SAFAD carbon footprint"
  )

SWITCH_post_data_processing_log_SAFAD <- SWITCH_post_data_processing %>%
  mutate(
    log_SAFAD = log(‘Average Carbon e (kg)‘),
  )
```

```
QQ_SAFAD_log <- ggplot(SWITCH_post_data_processing_log_SAFAD, aes(sample = log_SAFAD)) +
  stat_qq(color = "#2066a8", size = 2) +
  stat_qq_line(color = "#3594cc", linewidth = 1) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold")) +
  labs(
    x = "Theoretical quantiles",
    y = "Sample quantiles: log(SAFAD carbon footprint)"
  )

print(QQ_SAFAD_original)
print(QQ_SAFAD_log)

print(shapiro.test(SWITCH_post_data_processing$`Average Carbon e (kg)`))
print(shapiro.test(SWITCH_post_data_processing_log_SAFAD$`log_SAFAD`))
```

E.2.2 Q–Q plot for RISE

```
cat("\nQQ Plot for RISE\n")
QQ_RISE_original <- ggplot(SWITCH_post_data_processing, aes(sample = `Average Carbon e RISE (kg)`)) +
  stat_qq(color = "#a00000", size = 2) +
  stat_qq_line(color = "#c46666", linewidth = 1) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold")) +
  labs(
    x = "Theoretical quantiles",
    y = "Sample quantiles: RISE carbon footprint"
  )

SWITCH_post_data_processing_log_RISE <- SWITCH_post_data_processing %>%
  mutate(
    log_RISE = log(`Average Carbon e RISE (kg)`)
  )

QQ_RISE_log <- ggplot(SWITCH_post_data_processing_log_RISE, aes(sample = log_RISE)) +
  stat_qq(color = "#a00000", size = 2) +
  stat_qq_line(color = "#c46666", linewidth = 1) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold")) +
  labs(
    x = "Theoretical quantiles",
    y = "Sample quantiles: log(RISE carbon footprint)"
  )

print(QQ_RISE_original)
print(QQ_RISE_log)

print(shapiro.test(SWITCH_post_data_processing$`Average Carbon e RISE (kg)`))
print(shapiro.test(SWITCH_post_data_processing_log_RISE$`log_RISE`))
```

E.2.3 Participant Characteristics

```
# Participant characteristics and QC table
first_non_missing <- function(x) {
```

```
  x <- x[!is.na(x)]
  if (length(x) == 0) return(NA)
  x[1]
}

mean_sd <- function(x, digits = 0) {
  paste0(
    round(mean(x, na.rm = TRUE), digits), " ± ",
    round(sd(x, na.rm = TRUE), digits)
  )
}

count_distribution <- function(x) {
  paste0(
    sum(x == 1, na.rm = TRUE), " / ",
    sum(x == 2, na.rm = TRUE), " / ",
    sum(x == 3, na.rm = TRUE)
  )
}

before <- day_final %>%
  mutate(
    `Participant ID` = trimws(as.character(`Participant ID`)),
    Day = as.numeric(Day),
    Visit = as.numeric(Visit),
    `Energy Total (kcal)` = as.numeric(`Energy Total (kcal)`),
    `Food Amount Total (g)` = as.numeric(`Food Amount Total (g)`)
  ) %>%
  left_join(redcap, by = "Participant ID") %>%
  group_by(`User ID`, `Participant ID`, Visit) %>%
  summarise(
    valid_days = n_distinct(Day),

    `Average Energy (kcal)` = mean(`Energy Total (kcal)`, na.rm = TRUE),
    `Average Food Amount (g)` = mean(`Food Amount Total (g)`, na.rm = TRUE),

    age = first_non_missing(age),
    sex = first_non_missing(sex),
    BMI = first_non_missing(BMI),
    Block = first_non_missing(Block),

    LDL = first_non_missing(LDL),
    HDL = first_non_missing(HDL),
    `Total Cholesterol` = first_non_missing(`Total Cholesterol`),
    Triglycerides = first_non_missing(Triglycerides),

    .groups = "drop"
  )

after <- SWITCH_post_data_processing %>%
  mutate(
    `Participant ID` = trimws(as.character(`Participant ID`)),
    Visit = as.numeric(Visit),
    `Average Energy (kcal)` = as.numeric(`Average Energy (kcal)`),
    `Average Food Amount (g)` = as.numeric(`Average Food Amount (g)`)
  )

summary_qc <- function(df) {
```

```
df_participant <- df %>%
  distinct(`Participant ID`, .keep_all = TRUE) %>%
  mutate(
    sex_clean = trimws(as.character(sex)),
    block_clean = trimws(as.character(Block))
  )

visits_per_participant_dist <- df %>%
  distinct(`Participant ID`, Visit) %>%
  count(`Participant ID`) %>%
  pull(n)

days_per_visit_dist <- df$valid_days

participants_per_visit <- df %>%
  distinct(`Participant ID`, Visit) %>%
  count(Visit) %>%
  tidyr::complete(Visit = 1:3, fill = list(n = 0)) %>%
  arrange(Visit) %>%
  pull(n)

tibble::tibble(
  age = mean_sd(df_participant$age, digits = 1),

  BMI = mean_sd(df_participant$BMI, digits = 1),

  female_male = paste0(
    sum(df_participant$sex_clean == "Female", na.rm = TRUE), " / ",
    sum(df_participant$sex_clean == "Male", na.rm = TRUE)
  ),

  participants = n_distinct(df$`Participant ID`),

  participants_per_visit = paste0(
    participants_per_visit[1], " / ",
    participants_per_visit[2], " / ",
    participants_per_visit[3]
  ),

  visits_dist = count_distribution(visits_per_participant_dist),

  days_dist = count_distribution(days_per_visit_dist),

  energy_v1 = mean_sd(df$`Average Energy (kcal)`[df$Visit == 1], digits = 0),
  energy_v2 = mean_sd(df$`Average Energy (kcal)`[df$Visit == 2], digits = 0),
  energy_v3 = mean_sd(df$`Average Energy (kcal)`[df$Visit == 3], digits = 0),

  food_v1 = mean_sd(df$`Average Food Amount (g)`[df$Visit == 1], digits = 0),
  food_v2 = mean_sd(df$`Average Food Amount (g)`[df$Visit == 2], digits = 0),
  food_v3 = mean_sd(df$`Average Food Amount (g)`[df$Visit == 3], digits = 0),

  dietary_data_coverage = paste0(
    round(
      100 * mean(
        !is.na(df$`Average Energy (kcal)`) &
          !is.na(df$`Average Food Amount (g)`) &
          !is.na(df$valid_days)
```

```
        ),
        1
      ),
      "%"
    ),

    block = paste0(
      sum(df_participant$block_clean == "Block 1", na.rm = TRUE), " / ",
      sum(df_participant$block_clean == "Block 2", na.rm = TRUE), " / ",
      sum(df_participant$block_clean == "Block 3", na.rm = TRUE)
    ),

    biomarker_data = paste0(
      round(
        100 * mean(
          !is.na(df_participant$LDL) &
            !is.na(df_participant$HDL) &
            !is.na(df_participant$`Total Cholesterol`) &
            !is.na(df_participant$Triglycerides)
        ),
        1
      ),
      "%"
    )
  )
}

before_stats <- summary_qc(before)
after_stats  <- summary_qc(after)

participant_characteristics <- tibble::tibble(
  Variable = c(
    "Female / Male",
    "Age",
    "BMI",
    "Participants",
    "Participants per visit (V1 / V2 / V3)",
    "Visits per participant (1 / 2 / 3)",
    "Days per visit (1 / 2 / 3)",
    "Block 1 / Block 2 / Block 3",
    "Energy intake V1 (kcal/day)",
    "Energy intake V2 (kcal/day)",
    "Energy intake V3 (kcal/day)",
    "Food intake V1 (g/day)",
    "Food intake V2 (g/day)",
    "Food intake V3 (g/day)",
    "Dietary data coverage",
    "Biomarker data"
  ),

  `Before QC` = c(
    before_stats$female_male,
    before_stats$age,
    before_stats$BMI,
    before_stats$participants,
    before_stats$participants_per_visit,
    before_stats$visits_dist,
    before_stats$days_dist,
```

```
    before_stats$block,
    before_stats$energy_v1,
    before_stats$energy_v2,
    before_stats$energy_v3,
    before_stats$food_v1,
    before_stats$food_v2,
    before_stats$food_v3,
    before_stats$dietary_data_coverage,
    before_stats$biomarker_data
  ),

  `After QC` = c(
    after_stats$female_male,
    after_stats$age,
    after_stats$BMI,
    after_stats$participants,
    after_stats$participants_per_visit,
    after_stats$visits_dist,
    after_stats$days_dist,
    after_stats$block,
    after_stats$energy_v1,
    after_stats$energy_v2,
    after_stats$energy_v3,
    after_stats$food_v1,
    after_stats$food_v2,
    after_stats$food_v3,
    after_stats$dietary_data_coverage,
    after_stats$biomarker_data
  )
)

View(participant_characteristics)
print(participant_characteristics)
```

E.2.4 Subgroup Contribution

```
cat("\nSubgroup Contribution\n")

num_clean <- function(x) {
  as.numeric(gsub(",", ".", trimws(as.character(x))))
}

chr_clean <- function(x) {
  trimws(as.character(x))
}

# 1. Subgroup order for final tables
visit_levels <- c("Baseline", "Visit 2", "Visit 3")

subgroup_order <- c(
  "Berries",
  "Bread",
  "Butter",
  "Cereals",
  "Cheese",
  "Fat and oil",
  "Fish",
```

```
  "Fruit",
  "High-fat dishes",
  "Legumes",
  "Liquid dairy",
  "Meat",
  "Missing subgroup",
  "Mushrooms",
  "Nuts and seeds",
  "Offal",
  "Oil",
  "Other",
  "Other fats",
  "Pasta",
  "Pastries",
  "Plant protein/tofu",
  "Plant-based dairy",
  "Potatoes",
  "Poultry",
  "Processed meat",
  "Root vegetables",
  "Shellfish",
  "Spreads",
  "Vegetables"
)

# 2. Translate subgroup names to English
recode_subgroup_to_english <- function(x) {
  x <- str_squish(as.character(x))

  case_when(
    is.na(x) | x == "" | x %in% c("Övrigt / Tomma", "(Tomma)", "Tomma") ~ "Missing subgroup",

    x %in% c("Bär", "Berries") ~ "Berries",
    x %in% c("Bröd", "Bread") ~ "Bread",
    x %in% c("Smör", "Butter") ~ "Butter",
    x %in% c("Cerealier", "Spannmål", "Cereals") ~ "Cereals",
    x %in% c("Ost", "Cheese") ~ "Cheese",
    x %in% c("Fett och olja", "Fat and oil") ~ "Fat and oil",
    x %in% c("Fisk", "Fish") ~ "Fish",
    x %in% c("Frukt", "Fruit") ~ "Fruit",
    x %in% c("Feta rätter", "Högfetträtter", "High-fat dishes") ~ "High-fat dishes",
    x %in% c("Baljväxter", "Legumes") ~ "Legumes",
    x %in% c("Flytande mejeri", "Liquid dairy") ~ "Liquid dairy",
    x %in% c("Kött", "Meat") ~ "Meat",
    x %in% c("Svamp", "Mushrooms") ~ "Mushrooms",
    x %in% c("Nötter och frön", "Nuts and seeds") ~ "Nuts and seeds",
    x %in% c("Inälvsmat", "Offal") ~ "Offal",
    x %in% c("Olja", "Oil") ~ "Oil",
    x %in% c("Övrigt", "Other") ~ "Other",
    x %in% c("Övriga fetter", "Övrigt fett", "Other fats") ~ "Other fats",
    x %in% c("Pasta") ~ "Pasta",
    x %in% c("Bakverk", "Pastries") ~ "Pastries",
    x %in% c("Växtprotein/tofu", "Plant protein/tofu") ~ "Plant protein/tofu",
    x %in% c("Växtbaserad mejeri", "Plant-based dairy") ~ "Plant-based dairy",
    x %in% c("Potatis", "Potatoes") ~ "Potatoes",
    x %in% c("Fågel", "Poultry") ~ "Poultry",
    x %in% c("Processat kött, charkvaror", "Processed meat") ~ "Processed meat",
    x %in% c("Rotfrukter", "Root vegetables") ~ "Root vegetables",
```

```
    x %in% c("Skaldjur", "Shellfish") ~ "Shellfish",
    x %in% c("Pålägg", "Spreads") ~ "Spreads",
    x %in% c("Grönsaker", "Vegetables") ~ "Vegetables",

    TRUE ~ x
  )
}

# 3. Food ID to subgroup map
food_subgroup_map <- diet_data_english %>%
  mutate(
    `Food ID` = chr_clean(`Food ID`),
    Subgroup = chr_clean(Subgroup)
  ) %>%
  filter(
    !is.na(`Food ID`),
    !is.na(Subgroup),
    `Food ID` != "",
    Subgroup != ""
  ) %>%
  distinct(`Food ID`, Subgroup) %>%
  group_by(`Food ID`) %>%
  slice(1) %>%
  ungroup()

all_subgroups <- tibble(Subgroup = subgroup_order)

# 4. Valid visits and included days after final selection
final_visit_ids <- SWITCH_post_data_processing %>%
  transmute(
    `User ID` = chr_clean(`User ID`),
    `Participant ID` = chr_clean(`Participant ID`),
    Visit = as.numeric(Visit)
  ) %>%
  distinct()

valid_day_counts <- day_final_selected %>%
  mutate(
    `User ID` = chr_clean(`User ID`),
    `Participant ID` = chr_clean(`Participant ID`),
    Visit = as.numeric(Visit),
    Day = as.numeric(Day)
  ) %>%
  semi_join(
    final_visit_ids,
    by = c("User ID", "Participant ID", "Visit")
  ) %>%
  distinct(`User ID`, `Participant ID`, Visit, Day) %>%
  count(`User ID`, `Participant ID`, Visit, name = "valid_days")

valid_visits <- final_visit_ids %>%
  left_join(
    valid_day_counts,
    by = c("User ID", "Participant ID", "Visit")
  ) %>%
  filter(
    !is.na(valid_days),
    valid_days >= 2
```

```
  )

included_days <- day_final_selected %>%
  mutate(
    `User ID` = chr_clean(`User ID`),
    Day = as.numeric(Day)
  ) %>%
  semi_join(valid_visits, by = "User ID") %>%
  distinct(`User ID`, Day)

# 5. Mean daily intake per participant-visit-subgroup
amount_long <- diet_data_with_RISE %>%
  mutate(
    `User ID` = chr_clean(`User ID`),
    `Food ID` = chr_clean(`Food ID`),
    Day = as.numeric(Day),
    `Food Amount` = num_clean(`Food Amount`)
  ) %>%
  left_join(food_subgroup_map, by = "Food ID") %>%
  mutate(
    Subgroup = if_else(
      is.na(Subgroup) | Subgroup == "",
      "Missing subgroup",
      recode_subgroup_to_english(Subgroup)
    ),
    Subgroup = if_else(Subgroup %in% subgroup_order, Subgroup, "Other")
  ) %>%
  semi_join(included_days, by = c("User ID", "Day")) %>%
  inner_join(valid_visits, by = "User ID") %>%
  group_by(`User ID`, `Participant ID`, Visit, valid_days, Subgroup) %>%
  summarise(
    Total_Amount_g = sum(`Food Amount`, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(
    Avg_Amount_g = Total_Amount_g / valid_days
  ) %>%
  select(
    `User ID`,
    `Participant ID`,
    Visit,
    valid_days,
    Subgroup,
    Avg_Amount_g
  )

# 6. RISE CO2e per participant-visit-subgroup
rise_long <- diet_data_with_RISE %>%
  mutate(
    `User ID` = chr_clean(`User ID`),
    `Food ID` = chr_clean(`Food ID`),
    Day = as.numeric(Day),
    `RISE Carbon e (kg)` = num_clean(`RISE Carbon e (kg)`)
  ) %>%
  left_join(food_subgroup_map, by = "Food ID") %>%
  mutate(
    Subgroup = if_else(
      is.na(Subgroup) | Subgroup == "",
```

```
      "Missing subgroup",
      recode_subgroup_to_english(Subgroup)
    ),
    Subgroup = if_else(Subgroup %in% subgroup_order, Subgroup, "Other")
  ) %>%
  semi_join(included_days, by = c("User ID", "Day")) %>%
  inner_join(valid_visits, by = "User ID") %>%
  group_by(`User ID`, `Participant ID`, Visit, valid_days, Subgroup) %>%
  summarise(
    Total_RISE_CO2e_kg = sum(`RISE Carbon e (kg)`, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(
    Avg_RISE_CO2e_kg = Total_RISE_CO2e_kg / valid_days
  ) %>%
  select(
    `User ID`,
    `Participant ID`,
    Visit,
    valid_days,
    Subgroup,
    Avg_RISE_CO2e_kg
  )

# 7. SAFAD CO2e per participant-visit-subgroup
safad_long <- diet_data_with_SAFAD %>%
  mutate(
    `User ID` = chr_clean(`User ID`),
    `Food ID` = chr_clean(`Food ID`),
    Day = as.numeric(Day),
    `Carbon e (kg)` = num_clean(`Carbon e (kg)`)
  ) %>%
  left_join(food_subgroup_map, by = "Food ID") %>%
  mutate(
    Subgroup = if_else(
      is.na(Subgroup) | Subgroup == "",
      "Missing subgroup",
      recode_subgroup_to_english(Subgroup)
    ),
    Subgroup = if_else(Subgroup %in% subgroup_order, Subgroup, "Other")
  ) %>%
  semi_join(included_days, by = c("User ID", "Day")) %>%
  inner_join(valid_visits, by = "User ID") %>%
  group_by(`User ID`, `Participant ID`, Visit, valid_days, Subgroup) %>%
  summarise(
    Total_SAFAD_CO2e_kg = sum(`Carbon e (kg)`, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(
    Avg_SAFAD_CO2e_kg = Total_SAFAD_CO2e_kg / valid_days
  ) %>%
  select(
    `User ID`,
    `Participant ID`,
    Visit,
    valid_days,
    Subgroup,
    Avg_SAFAD_CO2e_kg
```

```
  )

# 8. Combine and include zero rows for non-consumers
subgroup_consumers_only <- amount_long %>%
  full_join(
    rise_long,
    by = c("User ID", "Participant ID", "Visit", "valid_days", "Subgroup")
  ) %>%
  full_join(
    safad_long,
    by = c("User ID", "Participant ID", "Visit", "valid_days", "Subgroup")
  ) %>%
  mutate(
    Avg_Amount_g = replace_na(Avg_Amount_g, 0),
    Avg_RISE_CO2e_kg = replace_na(Avg_RISE_CO2e_kg, 0),
    Avg_SAFAD_CO2e_kg = replace_na(Avg_SAFAD_CO2e_kg, 0)
  )

subgroup_long_complete <- tidyr::expand_grid(
  valid_visits,
  all_subgroups
) %>%
  left_join(
    subgroup_consumers_only,
    by = c("User ID", "Participant ID", "Visit", "valid_days", "Subgroup")
  ) %>%
  mutate(
    Avg_Amount_g = replace_na(Avg_Amount_g, 0),
    Avg_RISE_CO2e_kg = replace_na(Avg_RISE_CO2e_kg, 0),
    Avg_SAFAD_CO2e_kg = replace_na(Avg_SAFAD_CO2e_kg, 0)
  )

# 9. Group means across participant-visits
subgroup_summary <- subgroup_long_complete %>%
  mutate(
    Visit_label = case_when(
      Visit == 1 ~ "Baseline",
      Visit == 2 ~ "Visit 2",
      Visit == 3 ~ "Visit 3",
      TRUE ~ as.character(Visit)
    ),
    Visit_label = factor(Visit_label, levels = visit_levels),
    Subgroup = factor(Subgroup, levels = subgroup_order)
  ) %>%
  group_by(Visit_label, Subgroup) %>%
  summarise(
    Mean_Amount_g = mean(Avg_Amount_g, na.rm = TRUE),
    Mean_RISE_CO2e_kg = mean(Avg_RISE_CO2e_kg, na.rm = TRUE),
    Mean_SAFAD_CO2e_kg = mean(Avg_SAFAD_CO2e_kg, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  complete(
    Visit_label = factor(visit_levels, levels = visit_levels),
    Subgroup = factor(subgroup_order, levels = subgroup_order),
    fill = list(
      Mean_Amount_g = 0,
      Mean_RISE_CO2e_kg = 0,
      Mean_SAFAD_CO2e_kg = 0
```

```
    )
  )

# 10. Formatting functions
fmt_num <- function(x, digits) {
  formatC(x, format = "f", digits = digits)
}

fmt_contribution <- function(value, percent) {
  paste0(
    fmt_num(value, 3),
    " (",
    fmt_num(percent, 1),
    "%)"
  )
}

# 11. Table C.1: Mean daily intake
table_C1 <- subgroup_summary %>%
  mutate(value = fmt_num(Mean_Amount_g, 1)) %>%
  select(Subgroup, Visit_label, value) %>%
  pivot_wider(
    names_from = Visit_label,
    values_from = value
  ) %>%
  arrange(Subgroup) %>%
  mutate(Subgroup = as.character(Subgroup))

# 12. Table C.2: RISE contribution
table_C2 <- subgroup_summary %>%
  group_by(Visit_label) %>%
  mutate(
    Total_RISE = sum(Mean_RISE_CO2e_kg, na.rm = TRUE),
    Percent_RISE = if_else(
      Total_RISE > 0,
      100 * Mean_RISE_CO2e_kg / Total_RISE,
      0
    ),
    value = fmt_contribution(Mean_RISE_CO2e_kg, Percent_RISE)
  ) %>%
  ungroup() %>%
  select(Subgroup, Visit_label, value) %>%
  pivot_wider(
    names_from = Visit_label,
    values_from = value
  ) %>%
  arrange(Subgroup) %>%
  mutate(Subgroup = as.character(Subgroup))

# 13. Table C.3: SAFAD contribution
table_C3 <- subgroup_summary %>%
  group_by(Visit_label) %>%
  mutate(
    Total_SAFAD = sum(Mean_SAFAD_CO2e_kg, na.rm = TRUE),
    Percent_SAFAD = if_else(
      Total_SAFAD > 0,
      100 * Mean_SAFAD_CO2e_kg / Total_SAFAD,
      0
```

```
    ),
    value = fmt_contribution(Mean_SAFAD_CO2e_kg, Percent_SAFAD)
  ) %>%
  ungroup() %>%
  select(Subgroup, Visit_label, value) %>%
  pivot_wider(
    names_from = Visit_label,
    values_from = value
  ) %>%
  arrange(Subgroup) %>%
  mutate(Subgroup = as.character(Subgroup))

# 14. Print tables
cat("\nTable C.1: Mean daily intake by food subgroup across visits\n")
print(table_C1, n = Inf)

cat("\nTable C.2: RISE carbon footprint contribution by food subgroup across visits\n")
print(table_C2, n = Inf)

cat("\nTable C.3: SAFAD carbon footprint contribution by food subgroup across visits\n")
print(table_C3, n = Inf)
```

E.2.5 Adherence to SWITCH targets

```
cat("\nTABLE. ADHERENCE TO SWITCH TARGETS ACROSS VISITS\n")

# 1. Define SWITCH targets
switch_targets <- tibble::tibble(
  food_group = c(
    "Vegetables",
    "Fruit & Berries",
    "Whole Grains",
    "Legumes",
    "Fish & Shellfish"
  ),
  target_min = c(
    300,
    200,
    90,
    21.43,
    64.29
  ),
  target_max = c(
    400,
    400,
    NA,
    NA,
    NA
  )
)

# 2. Select variables
switch_data <- SWITCH_post_data_processing %>%
  select(
    `Participant ID`,
    Visit,
    `Average Switch Vegetables (g)`,
```

```
    `Average Switch Fruit & Berries (g)`,
    `Average Switch Whole Grains (g)`,
    `Average Switch Legumes (g)`,
    `Average Switch Fish & Shellfish (g)`) %>%
  mutate(
    `Participant ID` = trimws(as.character(`Participant ID`)),
    Visit = as.numeric(Visit)
  )

# 3. Reshape and join targets
switch_long <- switch_data %>%
  pivot_longer(
    cols = -c(`Participant ID`, Visit),
    names_to = "food_group",
    values_to = "intake"
  ) %>%
  mutate(
    food_group = recode(
      food_group,
      `Average Switch Vegetables (g)` = "Vegetables",
      `Average Switch Fruit & Berries (g)` = "Fruit & Berries",
      `Average Switch Whole Grains (g)` = "Whole Grains",
      `Average Switch Legumes (g)` = "Legumes",
      `Average Switch Fish & Shellfish (g)` = "Fish & Shellfish"
    )
  ) %>%
  left_join(switch_targets, by = "food_group") %>%
  mutate(
    has_target = !is.na(target_min),

    meets_target = case_when(
      !has_target ~ NA,
      !is.na(target_max) ~ intake >= target_min & intake <= target_max,
      has_target & is.na(target_max) ~ intake >= target_min,
      TRUE ~ NA
    ),

    target_text = case_when(
      !has_target ~ "No SWITCH target",
      !is.na(target_max) ~ paste0(target_min, "-", target_max, " g/day"),
      has_target & is.na(target_max) ~ paste0("\geq", round(target_min, 2), " g/day"),
      TRUE ~ NA_character_
    )
  )

# 4. Summarise mean, SD and % meeting target per visit
switch_summary <- switch_long %>%
  group_by(food_group, target_text, Visit) %>%
  summarise(
    n = sum(!is.na(intake)),
    mean_intake = mean(intake, na.rm = TRUE),
    sd_intake = sd(intake, na.rm = TRUE),

    percent_meeting_num = ifelse(
      all(is.na(meets_target)),
      NA_real_,
      mean(meets_target, na.rm = TRUE) * 100
    ),
```

```
    .groups = "drop"
  ) %>%
  mutate(
    mean_sd = ifelse(
      is.na(sd_intake),
      sprintf("%.1f", mean_intake),
      sprintf("%.1f ± %.1f", mean_intake, sd_intake)
    ),

    percent_meeting = ifelse(
      is.na(percent_meeting_num),
      "-",
      sprintf("%.0f%%", percent_meeting_num)
    )
  )

# 5. Final presentation table
switch_table <- switch_summary %>%
  select(food_group, target_text, Visit, mean_sd, percent_meeting) %>%
  pivot_wider(
    names_from = Visit,
    values_from = c(mean_sd, percent_meeting),
    names_glue = "Visit {Visit}_{.value}"
  ) %>%
  rename(
    `Food group` = food_group,
    `SWITCH target` = target_text,
    `Baseline mean ± SD` = `Visit 1_mean_sd`,
    `Visit 2 mean ± SD` = `Visit 2_mean_sd`,
    `Visit 3 mean ± SD` = `Visit 3_mean_sd`,
    `Baseline (%) meeting target` = `Visit 1_percent_meeting`,
    `Visit 2 (%) meeting target` = `Visit 2_percent_meeting`,
    `Visit 3 (%) meeting target` = `Visit 3_percent_meeting`
  )

switch_table <- switch_table %>%
  mutate(
    `Food group` = factor(
      `Food group`,
      levels = c(
        "Vegetables",
        "Fruit & Berries",
        "Whole Grains",
        "Legumes",
        "Fish & Shellfish"
      )
    )
  ) %>%
  arrange(`Food group`)

cat("\nAdherence to SWITCH dietary targets across visits:\n")
print(switch_table, n = Inf, width = Inf)
```

E.2.6 RQ1

```
cat("\nRQ1\n")
```

```
# 1. Prepare data
data_rq1 <- SWITCH_post_data_processing %>%
  mutate(
    Visit = case_when(
      Visit %in% c(1, "1", "Baseline") ~ "Baseline",
      Visit %in% c(2, "2", "Visit 2") ~ "Visit 2",
      Visit %in% c(3, "3", "Visit 3") ~ "Visit 3",
      TRUE ~ as.character(Visit)
    ),
    Visit = factor(Visit, levels = c("Baseline", "Visit 2", "Visit 3"))
  )

# 2. Variables and labels
environmental_vars <- c(
  "Average Carbon e RISE (kg)",
  "Average Carbon e (kg)",
  "Average Cropland (m2yr)",
  "Average Water (m3)",
  "Average New n (kg)",
  "Average New p (kg)",
  "Average Ammonia (kg)",
  "Average Biodiversity",
  "Average Pesticides (g)",
  "Average Antibiotics",
  "Average Animal welfare",
  "Average Carbon dioxide (kg)",
  "Average Methane fossil (kg)",
  "Average Methane biogenic (kg)",
  "Average Nitrous oxide (kg)"
)

indicator_labels <- c(
  "Average Carbon e RISE (kg)" = "Carbon footprint, RISE (kg CO2e/day)",
  "Average Carbon e (kg)" = "Carbon footprint, SAFAD (kg CO2e/day)",
  "Average Cropland (m2yr)" = "Cropland use (m2 yr/day)",
  "Average Water (m3)" = "Water use (m3/day)",
  "Average New n (kg)" = "New nitrogen input (g N/day)",
  "Average New p (kg)" = "New phosphorus input (g P/day)",
  "Average Ammonia (kg)" = "Ammonia emissions (g NH3/day)",
  "Average Biodiversity" = "Biodiversity loss (x10^-12 E/MSY/day)",
  "Average Pesticides (g)" = "Pesticide use (g a.i./day)",
  "Average Antibiotics" = "Use of antibiotics (index/day)",
  "Average Animal welfare" = "Animal welfare loss (index/day)",
  "Average Carbon dioxide (kg)" = "Carbon dioxide (kg CO2/day)",
  "Average Methane fossil (kg)" = "Methane, fossil (kg CH4/day)",
  "Average Methane biogenic (kg)" = "Methane, biogenic (kg CH4/day)",
  "Average Nitrous oxide (kg)" = "Nitrous oxide (kg N2O/day)"
)

display_scales <- c(
  "Average Carbon e RISE (kg)" = 1,
  "Average Carbon e (kg)" = 1,
  "Average Cropland (m2yr)" = 1,
  "Average Water (m3)" = 1,
  "Average New n (kg)" = 1000,
  "Average New p (kg)" = 1000,
  "Average Ammonia (kg)" = 1000,
```

```
  "Average Biodiversity" = 1e12,
  "Average Pesticides (g)" = 1,
  "Average Antibiotics" = 1,
  "Average Animal welfare" = 1,
  "Average Carbon dioxide (kg)" = 1,
  "Average Methane fossil (kg)" = 1,
  "Average Methane biogenic (kg)" = 1,
  "Average Nitrous oxide (kg)" = 1
)

indicator_order <- unname(indicator_labels[environmental_vars])

# 3. Descriptive statistics
rq1_long <- data_rq1 %>%
  select(`Participant ID`, Visit, all_of(environmental_vars)) %>%
  pivot_longer(
    cols = all_of(environmental_vars),
    names_to = "variable",
    values_to = "value_original"
  ) %>%
  mutate(
    indicator = recode(variable, !!!indicator_labels),
    indicator = factor(indicator, levels = indicator_order),
    display_scale = display_scales[variable],
    value = value_original * display_scale
  )

rq1_descriptive <- rq1_long %>%
  group_by(indicator, variable, Visit) %>%
  summarise(
    n = sum(!is.na(value)),
    mean = mean(value, na.rm = TRUE),
    sd = sd(value, na.rm = TRUE),
    median = median(value, na.rm = TRUE),
    q1 = quantile(value, 0.25, na.rm = TRUE),
    q3 = quantile(value, 0.75, na.rm = TRUE),

    mean_sd = case_when(
      first(variable) %in% c(
        "Average New n (kg)",
        "Average New p (kg)",
        "Average Ammonia (kg)"
      ) ~ paste0(round(mean, 1), " ± ", round(sd, 1)),

      first(variable) == "Average Biodiversity" ~
        paste0(round(mean, 2), " ± ", round(sd, 2)),

      first(variable) %in% c(
        "Average Methane fossil (kg)",
        "Average Methane biogenic (kg)",
        "Average Nitrous oxide (kg)"
      ) ~ paste0(
        formatC(mean, format = "f", digits = 3),
        " ± ",
        formatC(sd, format = "f", digits = 3)
      ),

      TRUE ~ paste0(round(mean, 2), " ± ", round(sd, 2))
```

```
    ),
    .groups = "drop"
  ) %>%
  arrange(indicator, Visit)

rq1_descriptive_wide <- rq1_descriptive %>%
  select(indicator, variable, Visit, mean_sd) %>%
  pivot_wider(
    names_from = Visit,
    values_from = mean_sd
  )

# 4. Log-transformed LMM
# Model: log(value) ~ Visit + (1 | Participant ID)
run_log_lmm <- function(var_name) {

  tmp <- data_rq1 %>%
    select(`Participant ID`, Visit, all_of(var_name)) %>%
    rename(value = all_of(var_name)) %>%
    filter(
      !is.na(value),
      value > 0,
      !is.na(Visit),
      !is.na(`Participant ID`)
    ) %>%
    mutate(log_value = log(value))

  model <- lmer(
    log_value ~ Visit + (1 | `Participant ID`),
    data = tmp
  )

  emm <- emmeans(model, ~ Visit)

  contrast_table <- as.data.frame(
    contrast(
      emm,
      method = list(
        "Visit 2 - Baseline" = c(-1, 1, 0),
        "Visit 3 - Baseline" = c(-1, 0, 1),
        "Visit 3 - Visit 2"  = c(0, -1, 1)
      ),
      adjust = "none"
    )
  ) %>%
    mutate(
      variable = var_name,
      indicator = recode(var_name, !!!indicator_labels),
      ratio = exp(estimate),
      percent_change = (exp(estimate) - 1) * 100,
      lower_ratio = exp(estimate - 1.96 * SE),
      upper_ratio = exp(estimate + 1.96 * SE),
      lower_percent = (lower_ratio - 1) * 100,
      upper_percent = (upper_ratio - 1) * 100,
      n_model = nrow(tmp)
    )

  emm_table <- as.data.frame(emm) %>%
```

```
    mutate(
      variable = var_name,
      indicator = recode(var_name, !!!indicator_labels),
      geometric_mean = exp(emmean),
      lower_CL_original = exp(lower.CL),
      upper_CL_original = exp(upper.CL),
      n_model = nrow(tmp)
    )

  list(
    model = model,
    emmeans = emm_table,
    contrasts = contrast_table
  )
}

rq1_results_list <- lapply(environmental_vars, run_log_lmm)
names(rq1_results_list) <- environmental_vars

rq1_emmeans <- bind_rows(lapply(rq1_results_list, `[[`, "emmeans"))
rq1_contrasts <- bind_rows(lapply(rq1_results_list, `[[`, "contrasts"))

# 5. FDR correction for baseline comparisons
rq1_contrasts_baseline <- rq1_contrasts %>%
  filter(contrast %in% c("Visit 2 - Baseline", "Visit 3 - Baseline")) %>%
  mutate(
    p_fdr = p.adjust(p.value, method = "BH"),

    p_value_formatted = if_else(
      p.value < 0.001,
      "<0.001",
      as.character(round(p.value, 3))
    ),

    p_fdr_formatted = if_else(
      p_fdr < 0.001,
      "<0.001",
      as.character(round(p_fdr, 3))
    ),

    beta = round(estimate, 2),

    sig = case_when(
      p.value < 0.001 ~ "***",
      p.value < 0.01  ~ "**",
      p.value < 0.05  ~ "*",
      TRUE ~ ""
    ),

    sig_fdr = case_when(
      p_fdr < 0.001 ~ "***",
      p_fdr < 0.01  ~ "**",
      p_fdr < 0.05  ~ "*",
      TRUE ~ ""
    ),

    beta_sig = paste0(
      formatC(beta, format = "f", digits = 2),
```

```
      sig
    ),

    beta_sig_fdr = paste0(
      formatC(beta, format = "f", digits = 2),
      sig_fdr
    )
  )

# 6. Final table for thesis
rq1_table_final <- rq1_descriptive_wide %>%
  left_join(
    rq1_contrasts_baseline %>%
      filter(contrast == "Visit 2 - Baseline") %>%
      select(
        indicator,
        variable,
        `Visit 2 beta` = beta_sig,
        `Visit 2 p-value` = p_value_formatted,
        `Visit 2 beta FDR` = beta_sig_fdr,
        `Visit 2 FDR p-value` = p_fdr_formatted
      ),
    by = c("indicator", "variable")
  ) %>%
  left_join(
    rq1_contrasts_baseline %>%
      filter(contrast == "Visit 3 - Baseline") %>%
      select(
        indicator,
        variable,
        `Visit 3 beta` = beta_sig,
        `Visit 3 p-value` = p_value_formatted,
        `Visit 3 beta FDR` = beta_sig_fdr,
        `Visit 3 FDR p-value` = p_fdr_formatted
      ),
    by = c("indicator", "variable")
  ) %>%
  mutate(
    indicator = factor(indicator, levels = indicator_order)
  ) %>%
  arrange(indicator) %>%
  select(
    Indicator = indicator,
    Baseline,
    `Visit 2`,
    `Visit 2 beta`,
    `Visit 2 p-value`,
    `Visit 2 beta FDR`,
    `Visit 2 FDR p-value`,
    `Visit 3`,
    `Visit 3 beta`,
    `Visit 3 p-value`,
    `Visit 3 beta FDR`,
    `Visit 3 FDR p-value`
  ) %>%
  mutate(
    Indicator = as.character(Indicator)
  )
```

```
print(rq1_table_final)

# 7. Figure data: RISE and SAFAD CO2e
rq1_co2e_plot_data <- data_rq1 %>%
  select(
    `Participant ID`,
    Visit,
    `Average Carbon e RISE (kg)`,
    `Average Carbon e (kg)`
  ) %>%
  pivot_longer(
    cols = c(`Average Carbon e RISE (kg)`, `Average Carbon e (kg)`),
    names_to = "Method",
    values_to = "CO2e"
  ) %>%
  mutate(
    Method = recode(
      Method,
      `Average Carbon e RISE (kg)` = "RISE",
      `Average Carbon e (kg)` = "SAFAD"
    )
  ) %>%
  group_by(Visit, Method) %>%
  summarise(
    n = sum(!is.na(CO2e)),
    mean = mean(CO2e, na.rm = TRUE),
    sd = sd(CO2e, na.rm = TRUE),
    se = sd / sqrt(n),
    lower_ci = mean - 1.96 * se,
    upper_ci = mean + 1.96 * se,
    lower_sd = mean - sd,
    upper_sd = mean + sd,
    .groups = "drop"
  ) %>%
  mutate(
    MethodVisit = paste(Method, Visit)
  )

custom_colors <- c(
  "RISE Baseline" = "#d8a6a6",
  "RISE Visit 2"  = "#c46666",
  "RISE Visit 3"  = "#a00000",
  "SAFAD Baseline" = "#8cc5e3",
  "SAFAD Visit 2"  = "#3594cc",
  "SAFAD Visit 3"  = "#2066a8"
)

# 8. Boxplot data
rq1_boxplot_data <- data_rq1 %>%
  select(
    `Participant ID`,
    Visit,
    `Average Carbon e RISE (kg)`,
    `Average Carbon e (kg)`
  ) %>%
  pivot_longer(
    cols = c(`Average Carbon e RISE (kg)`, `Average Carbon e (kg)`),
```

```
    names_to = "Method",
    values_to = "CO2e"
  ) %>%
  mutate(
    Method = recode(
      Method,
      `Average Carbon e RISE (kg)` = "RISE",
      `Average Carbon e (kg)` = "SAFAD"
    ),
    MethodVisit = paste(Method, Visit)
  ) %>%
  filter(!is.na(CO2e), CO2e > 0)

rq1_plot_boxplot_rise <- rq1_boxplot_data %>%
  filter(Method == "RISE") %>%
  ggplot(aes(x = Visit, y = CO2e, fill = MethodVisit)) +
  geom_boxplot(outlier.shape = NA, alpha = 0.7, width = 0.6) +
  geom_jitter(aes(color = MethodVisit), width = 0.2, alpha = 0.5, size = 1.5) +
  scale_fill_manual(values = custom_colors) +
  scale_color_manual(values = custom_colors) +
  labs(
    x = NULL,
    y = expression("Carbon footprint (kg CO"[2]*"e/day)")
  ) +
  theme_classic(base_size = 13) +
  theme(
    legend.position = "none"
  )

print(rq1_plot_boxplot_rise)

rq1_plot_boxplot_safad <- rq1_boxplot_data %>%
  filter(Method == "SAFAD") %>%
  ggplot(aes(x = Visit, y = CO2e, fill = MethodVisit)) +
  geom_boxplot(outlier.shape = NA, alpha = 0.7, width = 0.6) +
  geom_jitter(aes(color = MethodVisit), width = 0.2, alpha = 0.5, size = 1.5) +
  scale_fill_manual(values = custom_colors) +
  scale_color_manual(values = custom_colors) +
  labs(
    x = NULL,
    y = expression("Carbon footprint (kg CO"[2]*"e/day)")
  ) +
  theme_classic(base_size = 13) +
  theme(
    legend.position = "none"
  )

print(rq1_plot_boxplot_safad)

# 9. Coefficient plot with FDR significance
rq1_coef_data <- rq1_contrasts_baseline %>%
  mutate(
    contrast = recode(
      contrast,
      "Visit 2 - Baseline" = "Visit 2",
      "Visit 3 - Baseline" = "Visit 3"
    ),
    indicator = factor(indicator, levels = rev(indicator_order)),
```

```
    significant = p_fdr < 0.05
  )

pd <- position_dodge(width = 0.55)

indicator_axis_labels <- c(
  "Carbon footprint, RISE (kg CO2e/day)" =
    'paste("Carbon footprint, RISE (kg ", CO[2], "e/day)")',
  "Carbon footprint, SAFAD (kg CO2e/day)" =
    'paste("Carbon footprint, SAFAD (kg ", CO[2], "e/day)")',
  "Cropland use (m2 yr/day)" =
    '"Cropland use (m2 yr/day)"',
  "Water use (m3/day)" =
    '"Water use (m3/day)"',
  "New nitrogen input (g N/day)" =
    '"New nitrogen input (g N/day)"',
  "New phosphorus input (g P/day)" =
    '"New phosphorus input (g P/day)"',
  "Ammonia emissions (g NH3/day)" =
    'paste("Ammonia emissions (g ", NH[3], "/day)")',
  "Biodiversity loss (x10^-12 E/MSY/day)" =
    '"Biodiversity loss (x10^-12 E/MSY/day)"',
  "Pesticide use (g a.i./day)" =
    '"Pesticide use (g a.i./day)"',
  "Use of antibiotics (index/day)" =
    '"Use of antibiotics (index/day)"',
  "Animal welfare loss (index/day)" =
    '"Animal welfare loss (index/day)"',
  "Carbon dioxide (kg CO2/day)" =
    'paste("Carbon dioxide (kg ", CO[2], "/day)")',
  "Methane, fossil (kg CH4/day)" =
    'paste("Methane, fossil (kg ", CH[4], "/day)")',
  "Methane, biogenic (kg CH4/day)" =
    'paste("Methane, biogenic (kg ", CH[4], "/day)")',
  "Nitrous oxide (kg N2O/day)" =
    'paste("Nitrous oxide (kg ", N[2]*O, "/day)")'
)

rq1_coef_plot <- ggplot(
  rq1_coef_data,
  aes(
    x = percent_change,
    y = indicator,
    color = contrast,
    shape = significant
  )
) +
  geom_vline(
    xintercept = 0,
    linetype = "dashed",
    linewidth = 0.7
  ) +
  geom_errorbar(
    aes(xmin = lower_percent, xmax = upper_percent),
    orientation = "y",
    width = 0.15,
    linewidth = 0.75,
    position = pd
```

```
  ) +
  geom_point(
    size = 3,
    position = pd
  ) +
  scale_color_manual(
    values = c(
      "Visit 2" = "#3594cc",
      "Visit 3" = "#2066a8"
    )
  ) +
  scale_shape_manual(
    values = c(
      `FALSE` = 16,
      `TRUE` = 8
    ),
    breaks = c(TRUE),
    labels = c(
      `TRUE` = "Significant after FDR"
    )
  ) +
  scale_y_discrete(
    labels = function(x) parse(text = unname(indicator_axis_labels[x]))
  ) +
  labs(
    x = "Estimated change from baseline (%)",
    y = NULL,
    color = NULL,
    shape = NULL
  ) +
  theme_classic(base_size = 13) +
  theme(
    legend.position = "bottom",
    legend.box = "horizontal",
    legend.title = element_blank()
  ) +
  guides(
    color = guide_legend(order = 1),
    shape = guide_legend(order = 2)
  )

print(rq1_coef_plot)
```

E.2.7 RQ2

```
cat("\nRQ2\n")

data_rq2 <- SWITCH_post_data_processing %>%
  mutate(
    Visit = factor(
      case_when(
        Visit %in% c(1, "1", "Baseline") ~ "Baseline",
        Visit %in% c(2, "2", "Visit 2") ~ "Visit 2",
        Visit %in% c(3, "3", "Visit 3") ~ "Visit 3",
        TRUE ~ as.character(Visit)
      ),
      levels = c("Baseline", "Visit 2", "Visit 3")
```

```
    ),
    sex = factor(sex),

    veg_score = pmin(`Average Switch Vegetables (g)` / 300, 1),
    fruit_score = pmin(`Average Switch Fruit & Berries (g)` / 200, 1),
    wholegrain_score = pmin(`Average Switch Whole Grains (g)` / 90, 1),
    legume_score = pmin(`Average Switch Legumes (g)` / 21.43, 1),
    fish_score = pmin(`Average Switch Fish & Shellfish (g)` / 64.29, 1),

    total_adherence_score =
      veg_score +
      fruit_score +
      wholegrain_score +
      legume_score +
      fish_score,

    veg_met = if_else(`Average Switch Vegetables (g)` >= 300, 1, 0),
    fruit_met = if_else(`Average Switch Fruit & Berries (g)` >= 200, 1, 0),
    wholegrain_met = if_else(`Average Switch Whole Grains (g)` >= 90, 1, 0),
    legume_met = if_else(`Average Switch Legumes (g)` >= 21.43, 1, 0),
    fish_met = if_else(`Average Switch Fish & Shellfish (g)` >= 64.29, 1, 0),

    total_targets_met =
      veg_met +
      fruit_met +
      wholegrain_met +
      legume_met +
      fish_met
  )

co2_vars <- c(
  "Average Carbon e RISE (kg)" = "RISE carbon footprint",
  "Average Carbon e (kg)" = "SAFAD carbon footprint"
)

graded_adherence_vars <- c(
  total_adherence_score = "SWITCH adherence score",
  veg_score = "Vegetables",
  fruit_score = "Fruit and berries",
  wholegrain_score = "Whole grains",
  legume_score = "Legumes",
  fish_score = "Fish and seafood"
)

binary_adherence_vars <- c(
  total_targets_met = "SWITCH binary adherence score",
  veg_met = "Vegetables",
  fruit_met = "Fruit and berries",
  wholegrain_met = "Whole grains",
  legume_met = "Legumes",
  fish_met = "Fish and seafood"
)

run_rq2_model <- function(co2_var, adherence_var, adherence_label, adherence_type) {

  tmp <- data_rq2 %>%
    select(
      `Participant ID`,
```

```
      Visit,
      age,
      sex,
      BMI,
      `Average Energy (kcal)`,
      CO2e = all_of(co2_var),
      adherence_value = all_of(adherence_var)
    ) %>%
    filter(
      !is.na(CO2e), CO2e > 0,
      !is.na(adherence_value),
      !is.na(Visit),
      !is.na(age),
      !is.na(sex),
      !is.na(BMI),
      !is.na(`Average Energy (kcal)`),
      !is.na(`Participant ID`)
    )

  model <- lmer(
    log(CO2e) ~ adherence_value + Visit + age + sex + BMI +
      `Average Energy (kcal)` + (1 | `Participant ID`),
    data = tmp
  )

  as.data.frame(summary(model)$coefficients) %>%
    tibble::rownames_to_column("term") %>%
    filter(term == "adherence_value") %>%
    transmute(
      adherence_type = adherence_type,
      exposure = adherence_label,
      outcome = co2_vars[[co2_var]],
      beta = Estimate,
      SE = `Std. Error`,
      p_value = `Pr(>|t|)`,
      lower = beta - 1.96 * SE,
      upper = beta + 1.96 * SE,
      percent_change = (exp(beta) - 1) * 100,
      lower_percent = (exp(lower) - 1) * 100,
      upper_percent = (exp(upper) - 1) * 100,
      n_model = nrow(tmp)
    )
}

rq2_graded_results <- purrr::map_dfr(names(co2_vars), function(co2) {
  purrr::map_dfr(names(graded_adherence_vars), function(adherence) {
    run_rq2_model(
      co2_var = co2,
      adherence_var = adherence,
      adherence_label = graded_adherence_vars[[adherence]],
      adherence_type = "Graded"
    )
  })
})

rq2_binary_results <- purrr::map_dfr(names(co2_vars), function(co2) {
  purrr::map_dfr(names(binary_adherence_vars), function(adherence) {
    run_rq2_model(
```

```
      co2_var = co2,
      adherence_var = adherence,
      adherence_label = binary_adherence_vars[[adherence]],
      adherence_type = "Binary"
    )
  })
})

rq2_results_fdr <- bind_rows(
  rq2_graded_results,
  rq2_binary_results
) %>%
  group_by(adherence_type) %>%
  mutate(
    p_fdr = p.adjust(p_value, method = "BH"),
    significant_fdr = p_fdr < 0.05,
    stars_fdr = case_when(
      p_fdr < 0.001 ~ "***",
      p_fdr < 0.01 ~ "**",
      p_fdr < 0.05 ~ "*",
      TRUE ~ ""
    ),
    Beta = paste0(formatC(beta, format = "f", digits = 2), stars_fdr),
    `% change (95% CI)` = paste0(
      formatC(percent_change, format = "f", digits = 1), " (",
      formatC(lower_percent, format = "f", digits = 1), ", ",
      formatC(upper_percent, format = "f", digits = 1), ")"
    ),
    `FDR p-value` = if_else(
      p_fdr < 0.001,
      "<0.001",
      as.character(round(p_fdr, 3))
    )
  ) %>%
  ungroup()

make_rq2_table <- function(type) {
  rq2_results_fdr %>%
    filter(adherence_type == type) %>%
    select(
      `SWITCH target` = exposure,
      Outcome = outcome,
      Beta,
      `% change (95% CI)`
    ) %>%
    pivot_wider(
      names_from = Outcome,
      values_from = c(Beta, `% change (95% CI)`)
    ) %>%
    transmute(
      `SWITCH target`,
      `RISE beta` = `Beta_RISE carbon footprint`,
      `RISE % change (95% CI)` =
        `% change (95% CI)_RISE carbon footprint`,
      `SAFAD beta` = `Beta_SAFAD carbon footprint`,
      `SAFAD % change (95% CI)` =
        `% change (95% CI)_SAFAD carbon footprint`
    )
```

```
}

rq2_graded_table <- make_rq2_table("Graded")
rq2_binary_table <- make_rq2_table("Binary")

cat("\nTable 4.4: Associations between graded SWITCH adherence and dietary carbon footprint\n")
print(rq2_graded_table, n = Inf, width = Inf)

cat("\nTable 4.5: Associations between binary SWITCH target adherence and dietary carbon footprint\n"
print(rq2_binary_table, n = Inf, width = Inf)

custom_colors_rq2 <- c(
  "RISE carbon footprint" = "#a00000",
  "SAFAD carbon footprint" = "#2066a8"
)

make_rq2_plot <- function(type) {

  plot_data <- rq2_results_fdr %>%
    filter(
      adherence_type == type,
      exposure != "SWITCH adherence score",
      exposure != "SWITCH binary adherence score"
    ) %>%
    mutate(
      exposure = factor(
        exposure,
        levels = rev(c(
          "Vegetables",
          "Fruit and berries",
          "Whole grains",
          "Legumes",
          "Fish and seafood"
        ))
      ),
      outcome = factor(
        outcome,
        levels = c("RISE carbon footprint", "SAFAD carbon footprint")
      ),
      significant = factor(significant_fdr, levels = c(FALSE, TRUE))
    )

  x_lab <- if (type == "Graded") {
    "Estimated difference in carbon footprint (%) per one-unit higher adherence score"
  } else {
    "Estimated difference in carbon footprint (%) for meeting target"
  }

  ggplot(
    plot_data,
    aes(
      x = percent_change,
      y = exposure,
      color = outcome,
      shape = significant
    )
  ) +
    geom_vline(
```

```
      xintercept = 0,
      linetype = "dashed",
      linewidth = 0.7
    ) +
    geom_errorbar(
      aes(xmin = lower_percent, xmax = upper_percent),
      orientation = "y",
      width = 0.15,
      linewidth = 0.75,
      position = position_dodge(width = 0.55)
    ) +
    geom_point(
      size = 3,
      position = position_dodge(width = 0.55)
    ) +
    scale_color_manual(values = custom_colors_rq2) +
    scale_shape_manual(
      values = c(`FALSE` = 16, `TRUE` = 8),
      breaks = c(TRUE),
      labels = c(`TRUE` = "Significant after FDR"),
      drop = FALSE
    ) +
    labs(
      x = x_lab,
      y = NULL,
      color = NULL,
      shape = NULL
    ) +
    theme_classic(base_size = 13) +
    theme(
      legend.position = "bottom",
      legend.box = "horizontal",
      legend.title = element_blank()
    )
}

rq2_plot_graded <- make_rq2_plot("Graded")
rq2_plot_binary <- make_rq2_plot("Binary")

print(rq2_plot_graded)
print(rq2_plot_binary)
```

E.2.8 RQ3

```
cat("\nRQ3\n")

data_rq3 <- SWITCH_post_data_processing %>%
  mutate(
    Visit = factor(
      case_when(
        Visit %in% c(1, "1", "Baseline") ~ "Baseline",
        Visit %in% c(2, "2", "Visit 2") ~ "Visit 2",
        Visit %in% c(3, "3", "Visit 3") ~ "Visit 3",
        TRUE ~ as.character(Visit)
      ),
      levels = c("Baseline", "Visit 2", "Visit 3")
    ),
```

```
    sex = factor(sex),
    `LDL/HDL ratio` = if_else(!is.na(LDL) & !is.na(HDL) & HDL > 0, LDL / HDL, NA_real_)
  )

co2_vars <- c(
  "Average Carbon e RISE (kg)" = "RISE carbon footprint",
  "Average Carbon e (kg)" = "SAFAD carbon footprint"
)

lipid_vars <- c(
  "Total Cholesterol" = "Total cholesterol",
  LDL = "LDL cholesterol",
  HDL = "HDL cholesterol",
  Triglycerides = "Triglycerides",
  "LDL/HDL ratio" = "LDL/HDL ratio"
)

# Descriptive biomarker table
rq3_lipid_descriptive_wide <- data_rq3 %>%
  select(Visit, all_of(names(lipid_vars))) %>%
  pivot_longer(-Visit, names_to = "biomarker", values_to = "value") %>%
  mutate(Biomarker = recode(biomarker, !!!lipid_vars)) %>%
  group_by(Biomarker, Visit) %>%
  summarise(
    mean_sd = paste0(round(mean(value, na.rm = TRUE), 2), " ± ",
                     round(sd(value, na.rm = TRUE), 2)),
    .groups = "drop"
  ) %>%
  pivot_wider(names_from = Visit, values_from = mean_sd)

print(rq3_lipid_descriptive_wide)

# Model function
run_rq3_model <- function(co2_var, lipid_var) {
  tmp <- data_rq3 %>%
    select(
      `Participant ID`, Visit, age, sex, BMI, `Average Energy (kcal)`,
      CO2e = all_of(co2_var),
      biomarker_value = all_of(lipid_var)
    ) %>%
    filter(
      !is.na(CO2e), CO2e > 0,
      !is.na(biomarker_value),
      !is.na(Visit), !is.na(age), !is.na(sex),
      !is.na(BMI), !is.na(`Average Energy (kcal)`),
      !is.na(`Participant ID`)
    ) %>%
    mutate(log_CO2e_z = as.numeric(scale(log(CO2e))))

  model <- lmer(
    biomarker_value ~ log_CO2e_z + Visit + age + sex + BMI +
      `Average Energy (kcal)` + (1 | `Participant ID`),
    data = tmp
  )

  as.data.frame(summary(model)$coefficients) %>%
    tibble::rownames_to_column("term") %>%
    filter(term == "log_CO2e_z") %>%
```

```
    transmute(
      exposure = co2_vars[[co2_var]],
      biomarker = lipid_vars[[lipid_var]],
      beta = Estimate,
      SE = `Std. Error`,
      p_value = `Pr(>|t|)`,
      lower = beta - 1.96 * SE,
      upper = beta + 1.96 * SE,
      n_model = nrow(tmp)
    )
}

# Run models + FDR
rq3_results_fdr <- purrr::map_dfr(names(co2_vars), function(co2) {
  purrr::map_dfr(names(lipid_vars), function(lipid) {
    run_rq3_model(co2, lipid)
  })
}) %>%
  mutate(
    p_fdr = p.adjust(p_value, method = "BH"),
    stars_fdr = case_when(
      p_fdr < 0.001 ~ "***",
      p_fdr < 0.01 ~ "**",
      p_fdr < 0.05 ~ "*",
      TRUE ~ ""
    ),
    Beta = paste0(formatC(beta, format = "f", digits = 2), stars_fdr),
    `Absolute difference (95% CI)` = paste0(
      formatC(beta, format = "f", digits = 2), " (",
      formatC(lower, format = "f", digits = 2), ", ",
      formatC(upper, format = "f", digits = 2), ")"
    ),
    `FDR p-value` = if_else(p_fdr < 0.001, "<0.001", as.character(round(p_fdr, 3)))
  )

rq3_table_final_fdr <- rq3_results_fdr %>%
  select(
    Exposure = exposure,
    Biomarker = biomarker,
    Beta,
    `Absolute difference (95% CI)`,
    `FDR p-value`,
    n = n_model
  ) %>%
  arrange(Exposure, Biomarker)

print(rq3_table_final_fdr)

# Coefficient plot
rq3_coef_plot <- rq3_results_fdr %>%
  mutate(
    exposure = factor(exposure, levels = c("RISE carbon footprint", "SAFAD carbon footprint")),
    biomarker = factor(
      biomarker,
      levels = rev(c("Total cholesterol", "LDL cholesterol", "HDL cholesterol",
                     "Triglycerides", "LDL/HDL ratio"))
    ),
    significant = factor(p_fdr < 0.05, levels = c(FALSE, TRUE))
```

```
  ) %>%
  ggplot(aes(x = beta, y = biomarker, color = exposure, shape = significant)) +
  geom_vline(xintercept = 0, linetype = "dashed", linewidth = 0.7) +
  geom_errorbar(
    aes(xmin = lower, xmax = upper),
    orientation = "y",
    width = 0.15,
    linewidth = 0.75,
    position = position_dodge(width = 0.55)
  ) +
  geom_point(size = 3, position = position_dodge(width = 0.55)) +
  scale_color_manual(values = c(
    "RISE carbon footprint" = "#a00000",
    "SAFAD carbon footprint" = "#2066a8"
  )) +
  scale_shape_manual(
    values = c(`FALSE` = 16, `TRUE` = 8),
    breaks = c(TRUE),
    labels = c(`TRUE` = "Significant after FDR"),
    drop = FALSE
  ) +
  labs(
    x = "Estimated difference in biomarker value",
    y = NULL,
    color = NULL,
    shape = NULL
  ) +
  theme_classic(base_size = 13) +
  theme(
    legend.position = "bottom",
    legend.box = "horizontal",
    legend.title = element_blank()
  )

print(rq3_coef_plot)
```

E.2.9 RQ4

```
cat("\nRQ4\n")

# 1. Participant-level comparison
rise_safad_participant_comparison <- SWITCH_post_data_processing %>%
  select(
    `Participant ID`,
    Visit,
    `Average Carbon e RISE (kg)`,
    `Average Carbon e (kg)`
  ) %>%
  drop_na() %>%
  mutate(
    Visit = as.numeric(Visit),
    Visit_label = factor(
      case_when(
        Visit == 1 ~ "Baseline",
        Visit == 2 ~ "Visit 2",
        Visit == 3 ~ "Visit 3",
        TRUE ~ paste0("Visit ", Visit)
```

```
      ),
      levels = c("Baseline", "Visit 2", "Visit 3")
    ),
    RISE = `Average Carbon e RISE (kg)`,
    SAFAD = `Average Carbon e (kg)`,
    difference_RISE_minus_SAFAD = RISE - SAFAD,
    difference_SAFAD_minus_RISE = SAFAD - RISE,
    mean_RISE_SAFAD = (RISE + SAFAD) / 2
  )

cat("\nParticipant-visits included:", nrow(rise_safad_participant_comparison), "\n")
cat("Participants included:", n_distinct(rise_safad_participant_comparison$`Participant ID`), "\n")


# 2. Visit-level comparison table
daily_carbon_footprint_by_database_and_visit <- rise_safad_participant_comparison %>%
  group_by(Visit_label) %>%
  summarise(
    n = n(),
    RISE_mean = mean(RISE, na.rm = TRUE),
    RISE_sd = sd(RISE, na.rm = TRUE),
    SAFAD_mean = mean(SAFAD, na.rm = TRUE),
    SAFAD_sd = sd(SAFAD, na.rm = TRUE),
    mean_difference_SAFAD_minus_RISE =
      mean(difference_SAFAD_minus_RISE, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(
    `RISE mean ± SD` = sprintf("%.2f ± %.2f", RISE_mean, RISE_sd),
    `SAFAD mean ± SD` = sprintf("%.2f ± %.2f", SAFAD_mean, SAFAD_sd),
    `Mean difference SAFAD - RISE` =
      sprintf("%.2f", mean_difference_SAFAD_minus_RISE)
  ) %>%
  select(
    Visit = Visit_label,
    n,
    `RISE mean ± SD`,
    `SAFAD mean ± SD`,
    `Mean difference SAFAD - RISE`
  )

cat("\nDaily CO2e estimates from RISE and SAFAD across visits:\n")
print(daily_carbon_footprint_by_database_and_visit, n = Inf, width = Inf)


# 3. Descriptive statistics
rise_safad_descriptive_statistics <- rise_safad_participant_comparison %>%
  summarise(
    n = n(),
    mean_RISE = mean(RISE, na.rm = TRUE),
    sd_RISE = sd(RISE, na.rm = TRUE),
    mean_SAFAD = mean(SAFAD, na.rm = TRUE),
    sd_SAFAD = sd(SAFAD, na.rm = TRUE),
    mean_difference_RISE_minus_SAFAD =
      mean(difference_RISE_minus_SAFAD, na.rm = TRUE),
    mean_difference_SAFAD_minus_RISE =
      mean(difference_SAFAD_minus_RISE, na.rm = TRUE),
    sd_difference = sd(difference_SAFAD_minus_RISE, na.rm = TRUE),
```

```
    median_difference = median(difference_SAFAD_minus_RISE, na.rm = TRUE),
    min_difference = min(difference_SAFAD_minus_RISE, na.rm = TRUE),
    max_difference = max(difference_SAFAD_minus_RISE, na.rm = TRUE)
  )

cat("\nDescriptive statistics for RISE vs SAFAD:\n")
print(rise_safad_descriptive_statistics, width = Inf)


# 4. Paired t-test and Pearson correlation
rise_safad_paired_t_test <- t.test(
  rise_safad_participant_comparison$RISE,
  rise_safad_participant_comparison$SAFAD,
  paired = TRUE
)

rise_safad_pearson_correlation <- cor.test(
  rise_safad_participant_comparison$RISE,
  rise_safad_participant_comparison$SAFAD,
  method = "pearson"
)

cat("\nPaired t-test for RISE vs SAFAD:\n")
print(rise_safad_paired_t_test)

cat("\nPearson correlation for RISE vs SAFAD:\n")
print(rise_safad_pearson_correlation)


# 5. Bland-Altman statistics, participant level, SAFAD - RISE
bland_altman_mean_difference <- mean(
  rise_safad_participant_comparison$difference_SAFAD_minus_RISE,
  na.rm = TRUE
)

bland_altman_sd_difference <- sd(
  rise_safad_participant_comparison$difference_SAFAD_minus_RISE,
  na.rm = TRUE
)

bland_altman_upper_limit <- bland_altman_mean_difference +
  1.96 * bland_altman_sd_difference

bland_altman_lower_limit <- bland_altman_mean_difference -
  1.96 * bland_altman_sd_difference

cat("\nParticipant-level Bland-Altman statistics:\n")
cat("Mean difference SAFAD - RISE:", round(bland_altman_mean_difference, 3), "\n")
cat("Upper limit:", round(bland_altman_upper_limit, 3), "\n")
cat("Lower limit:", round(bland_altman_lower_limit, 3), "\n")


# 6. Plot: mean daily carbon footprint by database and visit
rq4_mean_daily_carbon_footprint_plot_data <- rise_safad_participant_comparison %>%
  select(Visit_label, RISE, SAFAD) %>%
  pivot_longer(
    cols = c(RISE, SAFAD),
    names_to = "Database",
```

```
    values_to = "Carbon_footprint"
  ) %>%
  mutate(
    Database = recode(
      Database,
      RISE = "RISE carbon footprint",
      SAFAD = "SAFAD carbon footprint"
    ),
    Database = factor(
      Database,
      levels = c("RISE carbon footprint", "SAFAD carbon footprint")
    )
  ) %>%
  group_by(Visit_label, Database) %>%
  summarise(
    n = n(),
    mean = mean(Carbon_footprint, na.rm = TRUE),
    sd = sd(Carbon_footprint, na.rm = TRUE),
    se = sd / sqrt(n),
    ci_lower = mean - 1.96 * se,
    ci_upper = mean + 1.96 * se,
    .groups = "drop"
  )

rq4_mean_daily_carbon_footprint_plot <- ggplot(
  rq4_mean_daily_carbon_footprint_plot_data,
  aes(
    x = Visit_label,
    y = mean,
    group = Database,
    color = Database,
    shape = Database
  )
) +
  geom_line(linewidth = 0.75) +
  geom_point(size = 3) +
  geom_errorbar(
    aes(ymin = ci_lower, ymax = ci_upper),
    width = 0.12,
    linewidth = 0.75
  ) +
  scale_color_manual(
    values = c(
      "RISE carbon footprint" = "#a00000",
      "SAFAD carbon footprint" = "#2066a8"
    )
  ) +
  scale_shape_manual(
    values = c(
      "RISE carbon footprint" = 16,
      "SAFAD carbon footprint" = 16
    )
  ) +
  labs(
    x = NULL,
    y = "Carbon footprint (kg CO2e/day)",
    color = NULL,
    shape = NULL
```

```
  ) +
  theme_classic(base_size = 13) +
  theme(
    legend.position = "bottom",
    legend.box = "horizontal",
    legend.title = element_blank()
  )

print(rq4_mean_daily_carbon_footprint_plot)


# 7. Plot: Bland-Altman, participant level
rq4_bland_altman_plot <- ggplot(
  rise_safad_participant_comparison,
  aes(
    x = mean_RISE_SAFAD,
    y = difference_SAFAD_minus_RISE
  )
) +
  annotate(
    "rect",
    xmin = -Inf,
    xmax = Inf,
    ymin = bland_altman_lower_limit,
    ymax = bland_altman_upper_limit,
    fill = "grey70",
    alpha = 0.5
  ) +
  geom_point(color = "grey40", alpha = 0.6, size = 1.8) +
  geom_hline(
    yintercept = 0,
    color = "grey60",
    linewidth = 0.5
  ) +
  geom_hline(
    yintercept = bland_altman_mean_difference,
    color = "black",
    linewidth = 0.7
  ) +
  geom_hline(
    yintercept = bland_altman_upper_limit,
    linetype = "dashed",
    color = "black"
  ) +
  geom_hline(
    yintercept = bland_altman_lower_limit,
    linetype = "dashed",
    color = "black"
  ) +
  annotate(
    "text",
    x = max(rise_safad_participant_comparison$mean_RISE_SAFAD, na.rm = TRUE),
    y = bland_altman_mean_difference,
    label = paste0("Mean: ", round(bland_altman_mean_difference, 2)),
    hjust = 1,
    vjust = -1
  ) +
  annotate(
```

```
    "text",
    x = max(rise_safad_participant_comparison$mean_RISE_SAFAD, na.rm = TRUE),
    y = bland_altman_upper_limit,
    label = paste0("+1.96 SD: ", round(bland_altman_upper_limit, 2)),
    hjust = 1,
    vjust = -1
  ) +
  annotate(
    "text",
    x = max(rise_safad_participant_comparison$mean_RISE_SAFAD, na.rm = TRUE),
    y = bland_altman_lower_limit,
    label = paste0("-1.96 SD: ", round(bland_altman_lower_limit, 2)),
    hjust = 1,
    vjust = 1.5
  ) +
  labs(
    x = "Mean of RISE and SAFAD (kg CO2e/day)",
    y = "Difference (SAFAD - RISE) (kg CO2e/day)"
  ) +
  theme_classic(base_size = 13)

print(rq4_bland_altman_plot)


# 8. Food-level comparison
rise_food_level_summary <- diet_data_with_RISE %>%
  filter(
    !is.na(`Food ID`),
    !is.na(`Food Name`),
    !is.na(`Food Amount`),
    `Food Amount` > 0,
    !is.na(`RISE Carbon e (kg)`)
  ) %>%
  group_by(`Food ID`, `Food Name`) %>%
  summarise(
    n = n(),
    amount_g = sum(`Food Amount`, na.rm = TRUE),
    total_RISE = sum(`RISE Carbon e (kg)`, na.rm = TRUE),
    RISE_per_kg = total_RISE / (amount_g / 1000),
    .groups = "drop"
  )

safad_food_level_summary <- diet_data_with_SAFAD %>%
  filter(
    !is.na(`Food ID`),
    !is.na(`Food Name`),
    !is.na(`Food Amount`),
    `Food Amount` > 0,
    !is.na(`Carbon e (kg)`)
  ) %>%
  group_by(`Food ID`, `Food Name`) %>%
  summarise(
    amount_g_SAFAD = sum(`Food Amount`, na.rm = TRUE),
    total_SAFAD = sum(`Carbon e (kg)`, na.rm = TRUE),
    SAFAD_per_kg = total_SAFAD / (amount_g_SAFAD / 1000),
    .groups = "drop"
  )
```

```
rise_safad_food_level_comparison <- rise_food_level_summary %>%
  inner_join(
    safad_food_level_summary,
    by = c("Food ID", "Food Name")
  ) %>%
  mutate(
    difference_per_kg_RISE_minus_SAFAD = RISE_per_kg - SAFAD_per_kg,
    absolute_difference_per_kg = abs(difference_per_kg_RISE_minus_SAFAD),
    difference_total_RISE_minus_SAFAD = total_RISE - total_SAFAD,
    absolute_difference_total = abs(difference_total_RISE_minus_SAFAD)
  )

cat(
  "\nFood items included in food-level comparison:",
  nrow(rise_safad_food_level_comparison),
  "\n"
)


# 9. Plot: Bland-Altman, food level
rq4_food_level_bland_altman_data <- rise_safad_food_level_comparison %>%
  filter(
    !is.na(RISE_per_kg),
    !is.na(SAFAD_per_kg)
  ) %>%
  mutate(
    mean_emission_factor = (RISE_per_kg + SAFAD_per_kg) / 2,
    difference_RISE_minus_SAFAD = RISE_per_kg - SAFAD_per_kg
  )

food_bland_altman_mean_difference <- mean(
  rq4_food_level_bland_altman_data$difference_RISE_minus_SAFAD,
  na.rm = TRUE
)

food_bland_altman_sd_difference <- sd(
  rq4_food_level_bland_altman_data$difference_RISE_minus_SAFAD,
  na.rm = TRUE
)

food_bland_altman_upper_limit <- food_bland_altman_mean_difference +
  1.96 * food_bland_altman_sd_difference

food_bland_altman_lower_limit <- food_bland_altman_mean_difference -
  1.96 * food_bland_altman_sd_difference

cat("\nFood-level Bland-Altman statistics:\n")
cat("Number of foods:", nrow(rq4_food_level_bland_altman_data), "\n")
cat(
  "Mean difference RISE - SAFAD:",
  round(food_bland_altman_mean_difference, 3),
  "kg CO2e/kg food\n"
)
cat("Upper limit:", round(food_bland_altman_upper_limit, 3), "\n")
cat("Lower limit:", round(food_bland_altman_lower_limit, 3), "\n")

rq4_food_level_bland_altman_plot <- ggplot(
  rq4_food_level_bland_altman_data,
```

```
  aes(
    x = mean_emission_factor,
    y = difference_RISE_minus_SAFAD
  )
) +
  annotate(
    "rect",
    xmin = -Inf,
    xmax = Inf,
    ymin = food_bland_altman_lower_limit,
    ymax = food_bland_altman_upper_limit,
    fill = "grey70",
    alpha = 0.5
  ) +
  geom_point(
    color = "grey40",
    alpha = 0.6,
    size = 1.8
  ) +
  geom_hline(
    yintercept = 0,
    color = "grey60",
    linewidth = 0.5
  ) +
  geom_hline(
    yintercept = food_bland_altman_mean_difference,
    color = "black",
    linewidth = 0.7
  ) +
  geom_hline(
    yintercept = food_bland_altman_upper_limit,
    linetype = "dashed",
    color = "black"
  ) +
  geom_hline(
    yintercept = food_bland_altman_lower_limit,
    linetype = "dashed",
    color = "black"
  ) +
  annotate(
    "text",
    x = max(rq4_food_level_bland_altman_data$mean_emission_factor, na.rm = TRUE),
    y = food_bland_altman_mean_difference,
    label = paste0("Mean: ", round(food_bland_altman_mean_difference, 2)),
    hjust = 1,
    vjust = -1
  ) +
  annotate(
    "text",
    x = max(rq4_food_level_bland_altman_data$mean_emission_factor, na.rm = TRUE),
    y = food_bland_altman_upper_limit,
    label = paste0("+1.96 SD: ", round(food_bland_altman_upper_limit, 2)),
    hjust = 1,
    vjust = -1
  ) +
  annotate(
    "text",
    x = max(rq4_food_level_bland_altman_data$mean_emission_factor, na.rm = TRUE),
```

```
    y = food_bland_altman_lower_limit,
    label = paste0("-1.96 SD: ", round(food_bland_altman_lower_limit, 2)),
    hjust = 1,
    vjust = 1.5
  ) +
  labs(
    x = "Mean emission factor (kg CO2e/kg food)",
    y = "Difference (RISE - SAFAD) (kg CO2e/kg food)"
  ) +
  theme_classic(base_size = 13)

print(rq4_food_level_bland_altman_plot)


# 10. Food items with largest emission factor differences
food_items_largest_emission_factor_differences <- rise_safad_food_level_comparison %>%
  arrange(desc(absolute_difference_per_kg)) %>%
  slice_head(n = 10) %>%
  transmute(
    `Food ID`,
    `Food name` = `Food Name`,
    n,
    `Amount (g)` = round(amount_g, 0),
    `RISE kg CO2e/kg` = round(RISE_per_kg, 2),
    `SAFAD kg CO2e/kg` = round(SAFAD_per_kg, 2),
    `Difference kg CO2e/kg` =
      round(difference_per_kg_RISE_minus_SAFAD, 2),
    `RISE total kg CO2e` = round(total_RISE, 1),
    `SAFAD total kg CO2e` = round(total_SAFAD, 1),
    `Difference total kg CO2e` =
      round(difference_total_RISE_minus_SAFAD, 1)
  )

cat("\nFood items with largest differences in emission factors:\n")
print(food_items_largest_emission_factor_differences, n = Inf, width = Inf)


# 11. Food items with largest total carbon footprint differences
food_items_largest_total_carbon_differences <- rise_safad_food_level_comparison %>%
  arrange(desc(absolute_difference_total)) %>%
  slice_head(n = 10) %>%
  transmute(
    `Food ID`,
    `Food name` = `Food Name`,
    n,
    `Amount (g)` = round(amount_g, 0),
    `RISE kg CO2e/kg` = round(RISE_per_kg, 2),
    `SAFAD kg CO2e/kg` = round(SAFAD_per_kg, 2),
    `Difference kg CO2e/kg` =
      round(difference_per_kg_RISE_minus_SAFAD, 2),
    `RISE total kg CO2e` = round(total_RISE, 1),
    `SAFAD total kg CO2e` = round(total_SAFAD, 1),
    `Difference total kg CO2e` =
      round(difference_total_RISE_minus_SAFAD, 1)
  )

cat("\nFood items with largest differences in total carbon footprint:\n")
print(food_items_largest_total_carbon_differences, n = Inf, width = Inf)
```