THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Diet and gut microbiota in cardiometabolic health:

Studies from the Danish Cohort Diet, Cancer and Health – Next Generations

AGNETHA LINN ROSTGAARD-HANSEN

This thesis has been submitted to the graduate school in Bioscience, Chalmers University of Technology and to the Faculty of Science, University of Copenhagen for a double degree, November 2023

Food and Nutrition Science
Department of Life Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023

Diet and gut microbiota in cardiometabolic health:
Studies from the Danish Cohort Diet, Cancer and Health – Next Generations

AGNETHA LINN ROSTGAARD-HANSEN
ISBN 978-91-7905-952-1

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr. 5418
ISSN 0346-718X

Department of Life Sciences
Food and Nutrition Science
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover: Image used under license from Shutterstock.com

Printed by Chalmers digitaltryck
Gothenburg, Sweden 2023

Dedication

To Hektor, Atlas and Emil ♥

Diet and gut microbiota in cardiometabolic health:
Studies from the Danish Cohort Diet, Cancer and Health – Next Generations

AGNETHA LINN ROSTGAARD-HANSEN

Department of Life Sciences
Chalmers University of Technology

ABSTRACT

Cardiometabolic diseases (CMD) are the leading cause of death globally. Diet is a key preventive factor of CMD and a determinant of gut microbiota. Gut microbiota, diet and their interactions have been associated with CMD. In observational studies, diet is measured by self-reported instruments, that need to be validated before use. Gut bacteria vary over time within an individual, making it challenging to study their relationship with health outcomes. Moreover, different dietary patterns may be associated differently with gut microbiota, but few studies exist.

The overall aim of this thesis work was to investigate the role of diet, gut microbiota and their interplay in cardiometabolic health. The MAX sub-cohort from the Diet, Cancer and Health – Next generations (DCH-NG) cohort was established to: validate the DCH-NG food frequency questionnaire (FFQ), validate a dietary quality score (DQS) and associate it with CMD risk factors, investigate gut microbiota temporal variability and associate these with dietary patterns and investigate the direct and indirect effects of a healthy Nordic and Mediterranean diet on CMD risk factors, mediated by gut microbiota.

The FFQ provided satisfactory ranking of individuals according to energy and nutrient intakes. The DQS was useful to rank individuals into groups of having unhealthy, average and healthy dietary habits. Healthy dietary habits were associated with lower levels of several CMD risk factors. Among bacterial genera, 39% had moderate to good reproducibility (ICC>0.5). Gut microbial subgroups (*Bacteroides*, *Prevotella* 9 and *Ruminococcaceae*) were identified and adherence to plant-based dietary patterns differed between subgroups. Healthy Nordic and Mediterranean diets were associated with lower levels of adiposity, but no indirect effect mediated by gut microbiota (*Prevotella*-to-*Bacteroides* ratio) was found. However, healthy Nordic and Mediterranean diets were associated with lower levels of lipidemia and hs-CRP, mediated by adiposity.

In conclusion, the DCH-NG FFQ can be used to rank individuals according to dietary intake in epidemiological studies and the DQS is a good indicator of overall diet quality. Different dietary patterns associated differently with gut microbial subgroups and specific genera. There was an effect of diet on CMD risk factors, though this effect was not mediated by the gut microbiota.

Keywords: food frequency questionnaire, 24-hour dietary recall, validity, reproducibility, dietary patterns, nutrients, gut microbiota, temporal variation, risk factors, cardiometabolic diseases, cohort study, epidemiology

RESUMÈ

Kardiometaboliske sygdomme (CMD) er den mest dominerende dødsårsag på globalt plan. Kosten spiller en afgørende rolle i forbindelse med forebyggelse af CMD og har også en stor indflydelse på tarmmikrobiotaen. Undersøgelser peger på, at der er en sammenhæng mellem tarmmikrobiotaen, kosten samt deres interaktion i forhold til risiko for CMD. I observationelle studier anvendes selvrapporteringsværktøjer, som fx spørgeskemaer, til måling af kostindtag. Disse bør valideres før brug. Tarmbakterier varierer over tid hos individer, hvilket er en udfordring i studier, der undersøger tarmbakterierne i relation til kost og sygdom. Derudover er forskellige kostmønstre muligvis relateret til tarmmikrobiotaen på forskellig vis, dog er der stadig få studier som har undersøgt dette i en dansk population.

Det overordnede formål var at undersøge kosten, tarmmikrobiotaen og deres indbyrdes samspil i forhold til kardiometabolisk sundhed. MAX-sub-kohorten blev etableret som en del af Kost, Kræft og Helbred – Næste Generationer (KKH-NG) kohorten for at validere KKH-NG fødevarefrekvensspørgeskemaet (FFQ); undersøge en kostscore i relation til risikofaktorer for CMD; undersøge den tidsmæssige variation af tarmmikrobiotaen og dens sammenhæng med plantebaserede kostmønstre samt at undersøge de direkte og indirekte effekter af en sund kost (Nordisk- og Middelhavskost) på risikofaktorer for CMD, medieret af tarmmikrobiotaen.

FFQen viste en acceptabel rangering af individer baseret på energi- og næringsstofindtag. Kostscoren var i stand til at kategorisere individer i grupper med henholdsvis usunde, mellem og sunde kostvaner. Sunde kostvaner var associeret med lavere niveauer af flere risikofaktorer for CMD. 39% af de identificerede bakterieslægter viste moderat til god reproducerbarhed (ICC>0,5). Der blev desuden identificeret tarmmikrobiota subgrupper (*Bacteroides*, *Prevotella* 9 og *Ruminococcaceae*). Overholdelse af plantebaserede kostmønstrer var forskellig for disse subgrupper. En sund Nordisk- og Middelhavskost var associeret med lavere niveauer af adipositas, men der var ingen indirekte effekt via tarmmikrobiotaen (*Prevotella*-til-*Bacteroides* ratio). Dog var disse kostmønstre associeret med lavere niveauer af dyslipedæmi og hs-CRP og en del af denne effekt var indirekte via adipositas.

Konklusion: KKH-NG FFQen kan anvendes til at rangere individer baseret på deres kostindtag i epidemiologiske studier, og kostscoren fungerer som en god indikator for den samlede kvalitet af kosten. Kostmønstrene var forbundet på forskellig vis med subgrupperne af tarmmikrobiota og specifikke bakterieslægter. Der var en effekt af kosten på risikofaktorer for CMD, dog var denne ikke medieret gennem tarmmikrobiotaen.

LIST OF PUBLICATIONS

This doctoral thesis is based on the work contained in the following papers:

I. **Rostgaard-Hansen AL,** Rosthøj S, Brunius C, Olsen SF, Bjerregaard AA, Cade JE, Tjønneland A, Landberg R, Halkjær J. Relative Validity and Reproducibility of a Web-Based Semi-Quantitative Food Frequency Questionnaire in the Danish Diet, Cancer, and Health—Next Generations MAX Study. *Nutrients*, 2023. DOI: https://doi.org/10.3390/nu15102389

II. **Rostgaard-Hansen AL**, Lau CJ, Halkjær J, Olsen A, Toft U. An updated validation of the Dietary Quality Score: associations with risk factors for cardiometabolic diseases in a Danish population. *European Journal of Nutrition*, 2023. DOI: https://doi.org/10.1007/s00394-023-03100-4

III. **Rostgaard-Hansen AL***, Esberg A*, Dicksved J, Hansen T, Pelve E, Brunius C, Halkjær J, Tjønneland A, Johansson I, Landberg R. Gut microbiota composition in the Danish Diet, Cancer, and Health – Next Generations MAX sub-cohort: Temporal variability and association with dietary patterns. Submitted.

IV. **Rostgaard-Hansen AL**, Grand MK, Esberg A, Rosthøj S, Dicksved J, Hansen T, Pelve E, Brunius C, Halkjær J, Tjønneland A, , Johansson I, Landberg R. Direct and indirect effects of healthy Nordic and Mediterranean diet patterns via gut microbiota on intermediate risk factors for cardiometabolic health. Manuscript.

*Joint first authorship.

Published papers not included in the thesis:

Palmnäs M, Brunius C, Shi L, **Rostgaard-Hansen A**, Torres NE, González-Domínguez R, Zamora-Ros R, Ye YL, Halkjær H, Tjønneland A, Riccardi G, Giacco R, Costabile G, Vetrani C, Nielsen J, Andres-Lacueva C, Landberg R (2020). Perspective: Metabotyping-A Potential Personalized Nutrition Strategy for Precision Prevention of Cardiometabolic Disease. *Adv Nutr*, 11(3): p. 524-532.

Andersen JLM, Halkjær J, **Rostgaard-Hansen AL**, Martinussen N, Lund AQ, Kyrø C, Tjønneland A, Olsen A (2021). Intake of whole grain and associations with lifestyle and demographics: a cross-sectional study based on the Danish Diet, Cancer and Health-Next Generations cohort. *Eur J Nutr*, 60(2): p. 883-895.

Lanuza F, Bondonno NP, Zamora-Ros R, **Rostgaard-Hansen AL**, Tjønneland A, Landberg R, Halkjær J, Andres-Lacueva C (2022). Comparison of Flavonoid Intake Assessment Methods Using USDA and Phenol Explorer Databases: Subcohort Diet, Cancer and Health Next Generations-MAX Study. *Front Nutr*, 9: p. 873774.

Lanuza F, Zamora-Ros R, Bondonno N.P, Meroño T, **Rostgaard-Hansen AL**, Riccardi G, Tjønneland A, Landberg R, Halkjær J, Andres-Lacueva C (2023). Dietary polyphenols, metabolic syndrome and cardiometabolic risk factors: An observational study based on the DCH-NG subcohort. *Nutr Metab Cardiovasc Dis*, 33(6): p. 1167-1178.

Lanuza F, Zamora-Ros R, **Rostgaard-Hansen AL**, Tjønneland A, Landberg R, Halkjær J, Andres-Lacueva C (2023). Descriptive analysis of dietary (poly)phenol intake in the subcohort MAX from DCH-NG: "Diet, Cancer and Health-Next Generations cohort". *Eur J Nutr*, 62(1): p. 337-350.

Mostafa H, Meroño T, Miñarro A, Sánchez-Pla A, Lanuza F, Zamora-Ros R, **Rostgaard Hansen AL**, Estanyol-Torres N, Cubedo-Culleré M, Tjønneland A, Landberg R, Halkjær J, Andres-Lacueva C (2023). Dietary Sources of Anthocyanins and Their Association with Metabolome Biomarkers and Cardiometabolic Risk Factors in an Observational Study. *Nutrients*, 15(5).

Lanuza F, Meroño T, Zamora-Ros R, Bondonno NP, **Rostgaard-Hansen AL**, Sánchez-Pla A, Miro B, Carmona-Pontaque F, Riccardi G, Tjønneland A, Landberg R, Halkjær J, Andres-Lacueva C (2023). Plasma metabolomic profiles of plant-based dietary indices reveal potential pathways for metabolic syndrome associations. *Atherosclerosis*, 9 (382).

CONTRIBUTION REPORT

Paper I: Agnetha Linn Rostgaard-Hansen (ALRH) designed the study together with co-authors, coordinated the data collection, participated in data processing, conducted the data analyses with supervision, interpreted the data and was responsible for writing the manuscript.

Paper II: ALRH designed the study together with co-authors, coordinated the data collection and data preparation, interpreted the data and was responsible for writing the manuscript.

Paper III: ALRH designed the study together with co-authors, coordinated the data collection and data preparation, worked in close collaboration with the joint first author who conducted the data analysis, interpreted the data and was responsible for writing the manuscript.

Paper IV: ALRH designed the study together with co-authors, coordinated the data collection and data preparation, worked in close collaboration with the second author who conducted the data analysis, interpreted the data and was responsible for writing the manuscript.

ABBREVIATIONS

AFM	Absolute fat mass
ASV	Amplicon sequence variants
BMI	Body mass index
BP	Blood pressure
CHD	Coronary heart disease
CHO	Total cholesterol
CKD	Chronic kidney disease
CMD	Cardiometabolic diseases
CVD	Cardiovascular diseases
DBP	Diastolic blood pressure
DCH-NG	Diet, Cancer and Health – Next Generations
DLW	Doubly labelled water
DQS	Dietary Quality Score
FDR	False discovery rate
FFQ	Food frequency questionnaire
HbA1c	Glycated haemoglobin
HC	Hierarchical clustering
HDL-C	High-density lipoprotein cholesterol
HNFI	Healthy Nordic food index
hPDI	Healthy plant-based diet index
hs-CRP	high-sensitivity C-reactive protein
GLP-1	Glucagon like peptide 1
ICC	Intra-class correlation coefficient
IL-6	Interleukin-6
LDL-C	Low-density lipoprotein cholesterol
LOA	Limits of agreement
LPS	Lipopolysaccharides
LSQ	Lifestyle questionnaire
MED	Mediterranean diet
NAFLD	Non-alcoholic fatty liver disease
OPLS-DA	Orthogonal partial least squares discriminant analysis
PAD	Peripheral artery disease

P/B	*Prevotella*-to-*Bacteroides*
PCA	Principal component analysis
PDI	Plant-based diet index
pro-veg	Provegetarian diet index
PYY	Peptide YY
qPCR	quantitative polymerase chain reaction
RFM	Relative fat mass
rMED	relative Mediterranean diet score
SBP	Systolic blood pressure
SCFA	Short-chain fatty acids
SEM	Structural equation model
16S rRNA	16S ribosomal ribonucleic acid
24-HDR	24-hour dietary recall
TG	Triglycerides
TMA	Trimethylamine
TMAO	Trimethylamine N-oxide
TNF-α	Tissue necrosis factor alpha
T2D	Type 2 diabetes
uPDI	unhealthy plant-based diet index
VF	Visceral fat
WC	Waist circumference
WFR	Weighted food record

TABLE OF CONTENTS

1. INTRODUCTION

Non-communicable diseases, particularly cardiometabolic diseases (CMD) are the leading cause of death and a major contributor to poor health globally[1-3]. CMD refers to mainly cardiovascular diseases (CVD) and type 2 diabetes (T2D) [4,5], but chronic kidney disease (CKD)[6] and non-alcoholic fatty liver disease (NAFLD)[7] have also been recognized as part of this disease group. Vascular and metabolic dysfunctions are the main characteristics of CMD[4-7]. CMD share several intermediate risk factors including obesity, hypertension, dyslipidaemia, insulin resistance, hyperglycaemia and low-grade inflammation[8]. Several of these risk factors are modifiable through lifestyle modifications such as diet and therefore to a large degree targetable for prevention or intervention. Dietary patterns rich in healthy plant-based foods have been associated with reduced risk of CVD and T2D as well as lower levels of triglycerides (TG), total cholesterol (CHO), low-density lipoprotein cholesterol (LDL-C), blood glucose, C-reactive protein (CRP), body mass index (BMI) and blood pressure (BP)[9]. Diet also has a major influence on gut microbiota composition and its metabolic output with indications that plant-based foods may have a favourable effect on the gut microbiota. Increasing evidence also points to associations of gut microbial alterations with CMD and their intermediate risk factors[10,11] and that the effect of diet on CMD risk may be modulated via the gut microbiota[12-14].

Accurate measures of dietary exposures, validity of aggregated foods into a dietary index and knowledge about the stability of the gut microbiota are important in epidemiological studies prior to investigating the association between diet, the gut microbiota and health outcomes. Assessment of dietary intake is challenging since dietary assessment methods are prone to measurement errors and therefore dietary intake needs to be validated[15]. Likewise, assessment of temporal gut microbiota variability within an individual is important to assess the appropriate sample size in studies investigating the gut microbiota in relation to diet and disease outcomes. However, investigations of temporal gut microbiota variability are scarce[16-18].

Different nutrients, foods and healthy dietary indexes and dietary patterns have been associated with the relative abundance of specific microbes[19-29] and gut microbial community subgroups, referred to as enterotypes including *Bacteroides*, *Prevotella* and *Ruminococcus*[30-32]. Less is known about the association between different healthy dietary patterns and the gut microbiota in a Danish population. In addition, small intervention studies have reported gut microbial interactions with diet on cardiometabolic health[12-14] but observational studies exploring the potential indirect effect of diet on CMD risk factors mediated by the gut microbiota are lacking.

In summary, validation of dietary data and assessment of temporal gut microbial variability is needed in order to improve the design of future prospective studies of diet, the gut microbiota and health outcomes. In addition, exploration of the relationship between dietary patterns and the gut microbiota as well as their interaction in cardiometabolic health will contribute with further knowledge about this complex relation, which is still in its infancy.

2. OBJECTIVES

The overall aim of the thesis was to investigate the role of diet, gut microbiota and their interplay in cardiometabolic health in Danish men and women. A sub-cohort of the Danish Diet, Cancer and Health – Next generations (DCH-NG) cohort, i.e., the MAX sub-cohort was established. This cohort was used to evaluate the validity and reproducibility of dietary assessment and the temporal variability of the gut microbiota composition, as well as to explore the role of dietary patterns on the gut microbiota and their interplay in relation to CMD risk factors were explored.

The specific objectives of the thesis were:

- To assess the relative validity of the DCH-NG 376-item food frequency questionnaire (FFQ) with three 24-hour dietary recalls (24-HDRs) for energy and nutrient intakes and to assess the reproducibility of the FFQ over one year for energy, nutrient, and food group intakes (Paper I)

- To assess the validity of the dietary quality score (DQS) based on a 23-item FFQ against the 376-item FFQ, to examine whether the DQS remains a good indicator of overall dietary quality and to investigate whether the DQS is associated with risk factors for cardiometabolic diseases (Paper II)

- To assess the variability of the gut microbiota composition over one year and identify gut microbial community subgroups and to investigate the association between dietary patterns and their food constituents with gut microbial community subgroups and genera (Paper III)

- To investigate whether adherence to the healthy Nordic or Mediterranean diet directly or indirectly affects intermediate risk factors for CMD via mediation by gut microbiota. (Paper IV)

3. BACKGROUND

3.1 Cardiometabolic diseases and their intermediate risk factors

CMD constitute the leading cause of death worldwide, accounting for more than 20 million deaths in 2019[1]. The prevalence of CMDs have nearly doubled in the last two decades, owing to an increase in age, population growth and obesity. Thus, CMD represents a major global health burden today[2,3]. CMDs includes primarily CVD and T2D[4,5], but other disorders have also been regarded as CMD, such as CKD[6] and NAFLD[7]. Metabolic dysfunctions affecting the heart, blood vessels, kidneys and liver are common characteristics of CMDs. CVD comprises a broad cluster of disorders affecting the heart and blood vessels, including notably coronary heart disease (CHD), stroke and peripheral artery disease (PAD)[33]. Most CVDs are caused by artherosclerosis[34]. T2D is characterised by abnormally high blood glucose levels due to inadequate utilization of insulin by the body (insulin resistance) and lack of sufficient insulin production by the pancreas (insulin deficiency)[35]. CKD comprises conditions induced by long-term damage to the renal parenchyma resulting in reduced renal function i.e., impaired ability of the kidneys to filter degradation products from the blood[36]. NAFLD is a condition characterized by accumulation of fat in the liver, for reasons other than alcohol intake[37].

The various conditions within CMDs are strongly correlated. For instance, among patients with T2D, the prevalence of CVD is 30% and of all their deaths 50% are attributable to CVD[38]. Moreover, T2D is the leading cause of CKD[39] and patients with CKD also experience a higher risk of CVD[40]. Furthermore, patients with NAFLD have a two-fold risk of T2D[41] and 40% of the patients die from CVD[42]. Due to overlap in aetiology and pathophysiology, CMDs share a number of intermediate risk factors such as obesity, hypertension, dyslipidaemia, hyperglycaemia, insulin resistance and low-grade inflammation[8]. Age, sex and genetics are non-modifiable risk factors common to CMDs [43-48], whereas poor diet[49], physical inactivity[50], smoking[3] and socioeconomic status[51] constitutes modifiable risk factors of CMDs.

3.1.1 Obesity

Overweight and obesity are characterized by excessive or abnormal fat accumulation, which pose an important health risk. The World Health Organization (WHO) defines overweight, and obesity based on BMI (kg/m^2), an indirect measure of body fat. Accordingly, among adults, a BMI ≥25 is classified as overweight and a BMI ≥30 is classified as obesity[52]. Visceral fat (VF) and waist circumference (WC) are indicators of abdominal obesity, a condition where excessive fat deposits are concentrated around the abdominal organs. In fact, abdominal obesity is thought to be a stronger predictor of cardiometabolic risk than BMI, since BMI does not differentiate between types of body tissue[53]. Development of obesity involves an imbalance between energy intake and energy expenditure i.e., having a larger energy intake than the energy expended resulting in weight gain. This dysregulation of the energy balance is multifactorial and derives from a complex and to some extent unknown interplay between genetic, lifestyle, socioeconomic and psychological factors and the gut microbiome[54].

Besides being one of the main components of metabolic syndrome and a risk factors of CMD, obesity is also a key determinant of several other CMD risk factors[55]. Abdominal obesity has been shown to increase the risk of developing hypertension[56]. This relationship is complex and involves several pathways, including the renin-angiotensin-aldosterone system, sympathetic nervous system, hyperleptinemia, insulin resistance and renal function impairment[57]. Lipid disturbances including elevated levels of TG, CHO and LDL-C, as well as lower levels of high-density lipoprotein cholesterol (HDL-C) are more prevalent in individuals with obesity compared to individuals of normal weight[58,59]. In addition, obesity has also been associated with the development of insulin resistance[60] like insulin resistance increases the risk of developing abdominal obesity[61]. Abdominal obesity increases the release of fatty acids, hormones and proinflammatory cytokines, which may promote low-grade inflammation[62].

Obesity is considered preventable through lifestyle changes. Besides restricting calorie intake, evidence indicates that healthy eating behaviours including diets rich in wholegrains, vegetables and fruits seems to contribute to weight control[63-65], as well as limiting intake of sugary drinks known to contribute to weight gain[66,67]. Physical activity also helps in weight maintenance, although with limited results if not combined with calorie-restricted diet[68].

3.1.2 Hypertension

Hypertension is a common risk factor for CMDs and particularly a key risk factor for CVDs[8,69]. The specific aetiology of hypertension, characterised by a higher blood BP than normal, is multifactorial and most often without a known cause (essential hypertension)[70]. In Europe, the definition of hypertension is a systolic BP ≥140 mmHg or a diastolic BP ≥90 mmHg[71]. However, BP levels below these cut-offs, have also shown to be associated with an increased risk of CVD[72]. In the United States, hypertension guidelines were recently revised with a systolic BP ≥130-139 mmHg and a diastolic BP ≥80-89 mmHg defined as stage I hypertension, and BP readings above the traditional cut off values defined as stage II hypertension[73].

In a large cohort study of more than one million individuals, those diagnosed with hypertension had a lifetime risk of overall CVD of 63% compared to 46% among those with normal BP at 30 years of age. In addition, those with hypertension developed CVD five years earlier than those with normal BP[69]. Hypertension is also closely related to the risk of developing CKD, since high BP may induce kidney damage[8,74]. Furthermore, decreasing renal function has been reported to increase the occurrence of hypertension[75]. While hypertension has also been associated with higher risk of T2D and NAFLD[76,77], a recent study using Mendelian randomization of more than 300,000 individuals from the UK Biobank study suggested that having T2D increases the risk of hypertension and not vice versa[78].

Several risk factors are known to increase the risk of hypertension including age, family history of hypertension, unhealthy diet (including high sodium intake, low potassium intake and high alcohol intake), smoking, physical inactivity and obesity[79,80]. In this regard, hypertension is to some extent preventable through healthy dietary habits, physical activity, smoking avoidance and restriction of alcohol intake, as well as maintaining a normal weight[79].

3.1.3 Dyslipidaemia

Dyslipidaemia refers to abnormal levels of lipids in the blood, including high levels of TG, total cholesterol and LDL-C, and low levels of HDL-C[81]. In a recent meta-analysis, total cholesterol and LDL-C were associated with increased CVD mortality, whereas high levels of HDL-C was associated with decreased CVD mortality[82]. Coexistence of dyslipidaemia and hypertension seems to accelerate atherosclerosis and thereby synergistically increase the risk of CVD[83]. Particularly, LDL has been shown to play a central role in the initiation of atherosclerosis where LDL particles are involved in plaque formation[84]. Abnormal lipid levels are also common among individuals with T2D and NAFLD. These abnormalities are mainly characterised by high levels of TG and small dense LDL particles, low levels of HDL-C, but with normal or only slightly increased levels of LDL-C[85,86]. Besides the use of lipid-lowering medication, a healthy lifestyle is also being promoted in order to prevent or reduce dyslipidaemia and thus risk of CMD. This includes consuming a healthy diet, including reduction in intake of saturated fats, increasing the intake of fibre-rich foods, and being physically active[87].

3.1.4 Insulin resistance and hyperglycaemia

Insulin is a hormone produced by the pancreas whose role of which is to facilitate uptake of glucose from the blood into fat-, muscle- and liver cells. Insulin resistance is a condition where the insulin response is reduced (primarily in muscle and liver) leading to increased production of insulin by the pancreas (hyperinsulinemia). Insulin resistance over time will exhaust the pancreas and ultimately impair insulin production leading to high blood glucose levels (hyperglycaemia)[88].

The long-term consequences of hyperglycaemia are many and include development of T2D and CVD, renal damage, neuropathy and vascular damage[89]. The process of developing insulin resistance has been related to cytokines, such as leptin, resistin, tumour necrosis factor alpha (TNF-α), interleukin-6 (IL-6), which are regulated by adipose tissue and adiposity consequently has a large effect on glucose metabolism and insulin resistance[90].

3.1.5 Low-grade inflammation

Inflammation can be acute or chronic. Acute inflammation is the immediate and temporary physiological response to an infection or tissue damage, whereas chronic inflammation is characterised by slightly to moderately elevated inflammatory markers persisting for months or years[91]. Increasing evidence points to a role of chronic low-grade inflammation in the development of CMDs[92]. A recent meta-analysis reported an increased incidence of CVD with several inflammatory biomarkers including CRP, IL-6, fibrinogen and galectine-3[92]. Moreover, inflammatory biomarkers (e.g., CRP and IL-6) have been associated with increased risk of T2D[93]. Several in- or extrinsic factors are believed to increase the risk of inflammation including smoking, excess visceral adipose tissue, reactive oxygen species and specific gut microbial patterns[94]. Whether inflammation induces insulin resistance, or whether it is the other

way around is still debated. For instance, insulin resistance in adipose tissue has been demonstrated to induce inflammation in mice[95], whereas the inflammatory marker TNF-α induced by excess adipose tissue has been shown to decrease insulin signalling and thereby promote insulin resistance[96,97]. For CVD, the atherosclerosis cascade has been intensively studied and shown to involve inflammation in several steps from early lesions in the inner wall of the arteries to plaque formation[98].

To sum up, a considerable proportion of CMDs is considered preventable through a healthy lifestyle including eating a healthy diet but also being physical active and avoiding smoking[3]. Reducing cardiometabolic risk factors, through lifestyle changes is an important strategy towards the prevention of CMDs.

3.2 Prevention of cardiometabolic diseases by means of diet

Diet is one of the key factors in the prevention and mitigation of CMDs[99]. Extensive investigations have been carried out to elucidate which foods or nutrients may have beneficial or adverse effects on CMD development[100]. Listed below are examples of major food groups and nutrients and their influence on CMD risk.

Fruit and vegetables

Fruits and vegetables contain dietary fibres, vitamins (vitamin B, C and E), phytochemicals, selenium and potassium[101]. Higher intakes of fruits and vegetables have consistently been associated with lower risk of CVD in observational studies[102]. The biological mechanisms by which nutrients from fruits and vegetables may prevent or mitigate CMDs are many and involve modulation of hormone metabolism, alteration in cholesterol metabolism, decrease in platelet aggregation and BP, stimulation of the immune system, modulation of detoxification enzymes, as well as antiviral and antibacterial activity[103].

Whole grains

Whole grains are seeds from cereal plants containing the bran, endosperm and germ. They are plentiful in dietary fibres, vitamins, minerals, phytochemicals as well as polyunsaturated fatty acids[104]. Evidence from prospective cohort studies demonstrates a lower risk of CVD with higher intakes of whole grains[105-107]. The potential mechanisms for the protective role of whole grains in the development of CMDs are thought to be due, in part, to dietary fibres such as soluble fibres and resistant starch. These nutrients can be fermented in the gut to short-chain fatty acids (SCFA)[108]. SCFA have been associated with improved gut health[109] and have been suggested to play a role in glucose and lipid metabolism as well as inflammation[110].

Dairy

Dairy products are in general high in protein, calcium and some also in fat. Overall, total dairy consumption (i.e., milk, yoghurt and cheese) has been associated with lower risk of T2D, but not with CVD. In addition, intakes of high-fat compared to low-fat dairy products have likewise not been associated with CVD risk. Previously the high fat content in dairy, particularly saturated fat, has been coupled with adverse effects on blood lipids, but recent evidence does

not support this[111]. However, evidence indicates that fermented dairy products such as yogurt, sour milk products and cheese, are associated with lower risk of CVD and yoghurt for T2D[111,112]. The consumption of fermented dairy products is suggested to relate to a beneficial effect on the gut microbiota[113].

Fish

Fatty fish contains omega-3 fatty acids, vitamin D and minerals[114]. Eating fatty fish has been shown to associate with lower risk of CHD and overall CVD in observational studies[115,116]. However, the association between fish intake and T2D risk is debated and results from meta-analyses are inconclusive[117,118]. The beneficial properties of fish in relation to heart disease seems to be mediated by polyunsaturated omega-3 fatty acids such as docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA)[119]. For instance, supplements with DHA and EPA have been shown to reduce heart rate[120], BP[121], blood triglycerides[122] and platelet aggregation[123] and increase HDL-C and interestingly also LDL-C[122]. Conflictingly, fish also contains pollutants and heavy metals (in particular predatory fish), which are harmful to our heathy. Still, the benefits of eating fish a couple of times a week is expected to outweigh the harmful effects of contaminants[124].

Meat

Red meat (pork, beef, veal and lamb) and processed meat have been associated with higher risk of CMDs in observational studies[125]. In a recent umbrella review of meta-analyses higher intakes of red meat and processed meat were associated with higher risk of CVD (CHD, stroke and heart failure) and T2D, especially for processed meat[126]. Red and processed meat contain B vitamins, minerals, saturated fatty acids and cholesterol[127]. Processed meat also contains high amounts of sodium and some also nitrite and nitrate[128]. The adverse effects of a higher red meat intake (including processed forms) has been suggested to be attributable to several components such as saturated fatty acids, salt, nitrate as well as dietary Trimethylamine N-oxide (TMAO) precursors (choline, L-carnitine, and betaine). Higher intakes of saturated fatty acids has been associated with higher levels of LDL-C[129], a higher salt intake was shown to increase BP[130], nitrate or byproducts of nitrate may contribute to insulin resistance, endothelial dysfunction and atherosclerosis[131] whereas blood levels of TMAO, a gut-derived metabolite produced by bacterial fermentation of choline and L-carnitine, have been linked with CVD risk[132].

Dietary fats

Intakes of particular types of dietary fats (i.e., trans-fatty acids, saturated, monounsaturated or polyunsaturated fatty acids) have been associated with varying CVD risk[133]. Trans-fatty acids have consistently been associated with increased risk of CVD and consequently industrially produced trans-fatty acids are today under regulation[134]. Animal-based foods as well as coconut and palm oils have a high content of saturated fatty acids[135]. Vegetable oils (e.g., olive oil and rapeseed oil), nuts and avocados are high in monounsaturated fatty acids[136], while fish, vegetable oils (e.g. safflower and sunflower oils), nuts and seeds are rich in polyunsaturated fatty acids[114,135]. Studies have shown that replacement of saturated fatty acids with mono- or polyunsaturated fatty acids is associated with a reduction in CVD risk as well as LDL-C levels, whereas replacement with carbohydrates shows no reduction in CVD risk[133].

Sugar-Sweetened beverages

These types of beverages have little nutritional value and are high in added sugar and calories. Due to the high energy intake and perhaps lower satiety feeling, sugar-sweetened beverages may lead to higher total energy intake and subsequently weight gain[67]. Sugar-sweetened beverages have been associated with a higher risk of obesity, T2D, CHD and stroke[137].

Alcohol

Alcohol itself is a hazardous chemical substance and therefore no amount of alcohol is considered healthy. A high consumption of alcoholic beverages has been associated with higher risk of CVD[138]. Still, results from systematic reviews and meta-analyses have conversely shown a reduction in CVD and T2D risk among women and men with low to moderate intakes of alcoholic beverages[139-141]. However, there is no evidence pointing in the direction that the beneficial properties could be ascribed to alcohol alone. In fact, a recent large prospective study has found an increased risk of CVD also at low to moderate intakes of alcoholic beverages[142].

To sum up, investigating single nutrients and foods in relations to risk of CMDs have provided great insights into which dietary components can be ascribed as being healthy and unhealthy. However, since humans do not eat single nutrients or foods, investigating dietary patterns in relation to risk of disease may be in favour in terms of mimicking eating habits of free-living individuals. Also, foods contain nutrients in various amounts and combinations, where some nutrients are intercorrelated and interacts with each other. Therefore, to incorporate the complexity of the human diet, studying dietary patterns may be more closely associated with overall disease risk than that of single foods or nutrients[143].

3.3 Dietary patterns

Dietary patterns represent the overall variety, frequency and amounts of beverages and foods consumed by individuals over time[144]. To investigate dietary patterns, dietary indexes can be constructed based on either data-driven or hypothesis-driven analysis also referred to as *a posteriori* or *a priori* approaches[145]. With the *a posterior* approach, diet indexes are typically constructed based on dimensionality reduction techniques, such as factor or cluster analysis, using dietary data collected from FFQs, 24-HDRs or weighed food records (WFR) from the study population. With these techniques dietary variables are thus aggregated into factors which represents different dietary patterns. Those factors can then be used in further diet-disease association analyses[146]. However, diet indexes derived from data-driven techniques may be specific to the population at study and may therefore be irreproducible in other populations[145]. Conversely, with the *a priori* approach the diet index is constructed based on current evidence regarding the association of foods or nutrients with disease risk. Some indexes are also established based on specific national dietary guidelines[146]. A limitation of the *a priori* approach is that it is restricted to the level of current nutritional evidence[145].

Both the data-driven and hypothesis-driven methods have been widely applied and numerous diet indexes exist. Of data-driven indexes known are the prudent and western diet indexes based on foods of plant-based origin, fish and fermented dairy products (prudent) and conversely foods that are energy-dense and with high amounts of sugar, fat and salt (western)[147]. Many hypothesis-driven indexes represent overall dietary patterns for instance the healthy eating index[148], the Mediterranean Diet score[149] and the Plant-based Diet Indexes[150,151], where others only constitute healthy food elements of a diet, also referred to as dietary quality indexes such as the Scandinavian indexes: the Dietary Quality Score[152,153] and the Healthy Nordic food index[154]. Dietary patterns with focus on plant-based foods have been associated with lower CMD risk[149,155,156], which will be described in more detail below.

3.3.1 The Dietary Quality score

In contrast to several other dietary indexes, the DQS was developed as a simple tool to investigate the overall quality of dietary habits in large populations. Therefore, the DQS was calculated based on a short FFQ (also referred to as a screener) which is easy to administer and complete within a few minutes and without comprehensive backend nutrient analysis. The food groups in the DQS are based on nutritional and health aspects related to CVD and construction of the score was based on the Danish dietary guidelines at the time for fruit, vegetables, fish and fats. Adherence to the DQS has been associated with lower levels of several CMD risk factors[152,153]. The DQS is currently also used in the Danish National report regarding the disease burden of risk factors at the societal level[157]. Since the first validation of the DQS, the screener has been updated and therefore a new validation is needed to ensure that the DQS is still a valid tool for assessing overall dietary quality.

3.3.2 The Mediterranean diet score

Lower incidence of heart disease and all-cause mortality was found in populations of the Mediterranean area which have been ascribed to the cardioprotective effects of their diet[158]. This inspired the development of the Mediterranean diet score to further investigate the beneficial health effects of such diet[159]. The Mediterranean diet is found in various versions, but common features include high intakes of olive oil, vegetables, legumes, fruits, whole grain cereals, nuts and seeds, moderate to high intake of fish, moderate intake of dairy products and alcohol and low intakes of meat (specifically red meat)[160,161]. The Mediterranean diet has been extensively studied in relation to all-cause mortality and CVD. In a meta-analysis of observational studies, a lower risk of CVD was reported for subjects with the highest adherence compared with subjects with the lowest adherence[162]. Furthermore, a large randomized controlled intervention study investigating the long-term effects of the Mediterranean diet on CVD risk, a decrease in CVD incidence after approximately 5 years was found for those who consumed a Mediterranean diet compared with those consuming a control diet[163]. A similar preventive effect was reported for incidence of T2D at 4 years of follow-up[164] as well as a reduction in the prevalence of metabolic syndrome after 1 year[165].

3.3.3 The healthy Nordic dietary patterns and food indexes

The Mediterranean diet has been promoted worldwide due to its beneficial health effects. However, adherence to the diet outside the Mediterranean region may be challenged by differences in food culture and availability of foods among other things. As an alternative, healthy dietary patterns of other regions where the type and variety of crops are different from that of the Mediterranean region have led to the development of the Nordic diet[154]. Several different healthy Nordic dietary patterns and indexes exist: the Healthy Nordic Food Index[154], New Nordic Diet[166], NORDIET[167], the Baltic Sea Diet[168] and SYSDIET[169]. The indexes vary in their composition, but all of them include the main healthy elements of a Nordic diet such as whole grains (rye and oats), fruits (mainly apples, pear and berries), vegetables (mainly root vegetables and cabbage) as well as fish. In a meta-analysis regarding the association between Nordic diet and CMDs outcomes and risk markers it was shown that a Nordic diet was associated with reductions (though small) in risk of CVD and T2D as well as lower levels of CMD risk markers such as LDL-C, apolipoprotein B, non-HDL-cholesterol, body weight, BMI, insulin and systolic blood pressure (SBP)[170].

3.3.4 The provegetarian and plant-based diet indexes

CVD mortality has been reported to be lower among vegetarians compared with non-vegetarians[171]. In order to mimic a vegetarian diet, the provegetarian index (pro-veg) was developed. Different from previous dietary indexes, the pro-veg weights plant-based foods positively and weights animal-based foods negatively. Hence a high intake of vegetables, fruit, legumes, cereals, potatoes, nuts and olive oil will be given a high score, where a high intake of meats and meat products, animal fats, eggs, fish and seafood and dairy products will be given a low score i.e., the food score is reversed for animal-based[150]. Since then, different plant-based indexes have been developed based on the pro-veg, with the addition that some plant-based foods may be less healthy such as sugar-sweetened beverages, refined grains and sweets and desserts. The plant-based diet index (PDI) therefore also contains more food groups than the pro-veg[151]. Three variations of the PDI exist:

The *overall PDI*, where , vegetables, fruits, nuts, legumes, whole grains, vegetable oils, tea and coffee, fruit and vegetable juices, refined grains, potatoes, sugar sweetened beverages, sweet and desserts are scored positively and animal fat, dairy, eggs, fish and seafood, meat and miscellaneous animal-based foods are scored reversely. Animal-based foods are reversely scored in all the PDIs[151]. The *healthy PDI* (hPDI) includes the same food groups as in the PDI but for vegetable juices, refined grains, potatoes, sugar sweetened beverages, sweet and desserts a reversed score is applied instead, scoring only healthy plant-based foods positively[151]. The *unhealthy PDI* (uPDI) scores unhealthy plant-based foods positively and healthy plant-based foods reversely i.e., high intakes of vegetable juices, refined grains, potatoes, sugar sweetened beverages, sweets and desserts are given a high score and high intakes of fruits, vegetables, whole grains, nuts, legumes, vegetable oils, tea and coffee are scored reversely[151].

Studies investigating the association between adherence to several of these plant-based dietary patterns have in general shown lower mortality[150] and lower risk of T2D[151] and metabolic syndrome[172]. Conversely, the uPDI was associated with a higher risk of T2D[151] and metabolic syndrome[172]. Thus, the quality of plant-based foods seems to play an important role in risk of disease.

3.4 Dietary assessment methods

The human diet is complex and one of the most difficult exposures to measure in free-living individuals[143]. Foods are made up of dietary nutrients such as fat, protein, carbohydrate, alcohol and micronutrients including vitamins and minerals but also contain phytochemicals, additives, contaminants, and chemicals formed during food processing or cooking, as well as numerous unknown substances, which together make up thousands of different compounds of which many are intercorrelated[143,173]. In epidemiological studies, diet has been and, most often, still is assessed by self-reported methods, such as diet history, FFQ, 24-HDR or WFR[174]. Besides the WFR, these dietary assessment methods all have a retrospective design, meaning that the participants must report their food intake from memory back in time over the last 24-hours or up to a whole year. While the WFR is not retrospective in design, is has a large participant burden since they require the participant to weigh all foods consumed which, in turn, may lead to the participant changing their eating habits. Self-reported dietary assessment methods are prone to substantial random and systematic measurement errors[175] due to day-to-day variation in food intake[176], misreporting of food intake[177-179] and outdated food composition databases among others. Consequently, it is important to consider the errors of the dietary assessment method of choice in the population under study and in relation to the study objectives.

As a complement to self-reported dietary assessment, biomarkers may be used. Biomarkers are measured in body fluids or tissues (such as plasma or urine) and thus represent an objective measurements which does not relying on self-reported intakes[180]. There are several types of dietary exposure biomarkers: recovery-, concentration-, replacement- and prediction biomarkers[181]. Recovery biomarkers represent estimates of absolute nutrient intakes such as doubly labelled water technique (DLW) for total energy expenditure measurements, urinary nitrogen to reflect protein intake as well as potassium and sodium in urine as biomarkers of their intake. However, only these validated recovery biomarkers exist, which limits their application in assessment of usual dietary intake but they can be used to calibrate energy or the above mentioned nutrients from self-reported instruments[182]. Concentration biomarkers are correlated with dietary intake. For instance, plasma vitamin C and serum carotenoids are common biomarkers of fruit and vegetable intake[183]. Concentration biomarkers are typically also affected by metabolism and phenotypic characteristics such as age, sex and obesity, which makes this type of biomarker unsuitable for estimating absolute dietary intake. Replacement biomarkers are similar to concentration biomarkers and used as surrogate markers when information of the desired compound is poor or not available from food composition databases such as phytoestrogens or polyphenols[181]. Prediction biomarkers have a dose-response relationship with intake and therefore have been used to predict intake of e.g., 24-hour urinary

fructose and sucrose. However, the recovery of these nutrients is low which also restricts their use in estimating absolute dietary intake[181]. Recent advancements suggest that concentration biomarkers may also be used to calibrate self-reported dietary intake in a similar way to recovery biomarkers[184].

Biomarkers are also subject to measurement errors, although different from those in self-report assessment methods. Measurement errors in biomarker assessment results from errors related to how the samples are collected and stored, analytical instrument errors and variation within and between subjects[185]. Extensive developments of metabolomics techniques that allow the analysis of thousands of small molecules in biological samples have opened for development of new dietary exposure biomarkers that hold promise for reflection of dietary intake in epidemiological studies[186]. An overview and description of the advantages and disadvantages of the different dietary assessment methods are shown in **Table 1**.

Table 1. Dietary assessment methods in nutrition epidemiology.

	Diet history[174]	**Food frequency questionnaire**[174,187]	**24-hour dietary recall**[174,188]	**Weighted food record**[174,188]	**Recovery biomarkers**[181]
Measurement	Usual intake	Usual intake of overall diet or targeted specific food or nutrient intakes	Actual intake, short-term or usual intake (if several recalls are collected)	Actual intake, short-term or usual intake (if several records are collected)	Absolute intake
Design	Includes three elements: 1) Interview about meal pattern (recall), 2) List of usual frequencies and amounts of foods and beverages (recall) 3) Three-day food record	Recall of usual frequencies of a predefined list of foods and beverages with closed-ended response categories. Reporting portion size may also be incorporated	Recall of foods and beverages including amounts the past 24-hours	Weighing of foods and beverages for one or several days	Body fluids such as urine, blood, saliva, stool or tissue samples
Administration method	Interviewer and paper-based or web-based	Self-administered paper-based or web-based	Interviewer or self-administered paper-based or web-based	Self-administered paper-based or web-based	-
Advantages	Assessment of meal pattern and detailed information about food intake such as cooking methods	Low-cost Suitable for large groups of people Low participant burden	Short recall period If interviewer-based: no literacy requirements If self-administered web-based: low cost and suitable for large groups of people Low participant burden	Most precise method due to weighing of all foods and beverages and detailed information about brand names, recipes and cooking methods	Objective measure
Disadvantages	Subjective reporting High participant burden Complex calculations Interview-based: requires trained interviewers which is cost and time consuming. Less used in epidemiological studies today.	Subjective reporting Misreporting Standard portion sizes Cognitive complex calculations Does not cover intake of all foods and beverages.	Subjective reporting Misreporting Several recalls required for estimation of habitual intake[189,190] If interview-based: requires trained interviewers which is cost and time consuming	Subjective reporting Several records for estimating usual intake[189,190] High participant burden Training of the participant Recording of many days may result in respondent fatigue and thereby change in dietary habits Cost and time consuming	Cost and time consuming Can be a burden for some participants, if having to book an appointment for a blood sample Only few recovery biomarkers available
Bias	Selection bias, information bias (interviewer, recall, intake, social desirability bias)	Selection bias, information bias (interviewer, recall, intake, social desirability bias)	Selection bias, information bias (interviewer, recall, intake, social desirability bias)	Selection bias, information bias (intake, social desirability bias)	Selection bias, information bias

3.4.1 The food frequency questionnaire

With the FFQ, participants are asked to report their usual long-term dietary intake over a specific period, usually months or a year. The FFQ consists of a long list of food items with close-ended frequency response categories typically ranging from never/rarely or monthly to several times a day. The list of food items can cover approximately the whole diet or specific dietary components of interests (short FFQs or screeners)[174,187,191]. In some FFQs, also known as quantitative or semi-quantitative FFQs, frequency intakes are combined with food portion sizes using either portion size photos or household measures [192]. After collection of FFQ data, daily intake of food groups, energy and nutrients are calculated based on reported frequencies and portion sizes (standard or specific) combined with a food composition database using a nutrient analysis software[174,192].

The FFQ has been widely applied in cohort studies for several reasons. First, the questionnaire measures usual long-term intake with enough precision to rank individuals according to dietary intake, which is sufficient in order to make risk predictions (risk ratio or odds ratio)[174]. Second, it is self-administered and today many are also web-based reducing the burden to participant as well as the back-end nutrient analysis[193]. Third, it has low cost compared to for instance interview-based instruments, which is important when assessing diet in populations of thousands of individuals[174,187]. The FFQ also have several limitations. It requires cognitively complex calculations, it can be difficult to recall foods consumed back in time (especially a year) and it does not cover all foods consumed[187].

3.4.2 The 24-hour dietary recall

The 24-HDR also relies on the participant's memory, but as the name indicates, the participant reports foods and beverages consumed over the past 24-hours [174,188]. With the 24-HDR, actual intake is measured. When only one 24-HDR per individual is obtained, average dietary intake at population level can be assessed. Assessment of the individual's usual dietary intake requires several recorded days per individual collected on non-consecutive days including both weekdays and weekend days [189,190]. The design of the 24-HDR is open-ended including detailed information about portion size, cooking method and sometimes also time and brand name of consumed item. Using a nutrient analysis software, daily intake of food groups, energy and nutrients are calculated based on reported intakes of foods and beverages and specified portion sizes combined with a food composition database[174,188]. The 24-HDR have traditionally been administered as a structured interview, where an interviewer (face-to-face or by telephone) asks about all foods and beverages consumed during the past 24-hours. The interviewer probes for amounts, cooking method and commonly forgotten foods [174,188].

The 24-HDR has mostly been applied in intervention studies and dietary surveys. Fortunately, due to the advancement of technology, web-based 24-HDRs with incorporated error checks and probes, have made it feasible and cost-effective to collect 24-HDRs in large cohort studies as an alternative to the FFQ[194-197]. Compared with the FFQ, the recall period of a 24-HDR is much shorter and does not require complex calculations, which makes it a more accurate method and

to some extend less prone to recall bias, though social desirability bias will still be present[188]. Presumably, since the 24-HDR is open-ended it may have larger variability in dietary intake than the FFQ, which has a fixed set of food items.

3.5 Evaluation of dietary assessment methods

All dietary assessment methods are subject to measurement errors which may result in attenuating or obscuring of associations between dietary exposure and disease outcome. It is therefore essential to evaluate how accurately the dietary assessment instrument measures true dietary intake. However, measurement of true dietary intake does not exist since there is currently no method without errors or gold standard. Instead, different methods with uncorrelated measurement errors can be used to "triangulate" the approximation of "true" intake[198].

3.5.1 Measurement errors in dietary assessment

Measurement error can be describes as the difference between the measured and the true value. Two main types of error exist: random and systematic errors[199]. Random error is defined as deviations scattered around the true mean intake and are caused by unpredictable variation between measurements such as day-to-day variation in dietary intake also referred to as within-individual random error[200]. Random errors tend to reduce the precision of the dietary measurement and consequently decrease of statistical power to detect a possible association between dietary exposure and outcome[201]. In diet-disease association studies, random error in the dietary exposure variable may cause attenuation of the correlation coefficient or the regression slope towards zero or the relative risk towards one[198]. Random errors may also, in some cases, lead to incorrect conclusions if the study involves estimating the proportion of a population being below or above a certain intake cut-off point, since random errors will lead to larger population variability[201]. Random errors can be mitigated by collecting repeated measures or increasing sample size[202]. In contrast, systematic error, is defined as consistent or proportional deviations from the true value in the same direction i.e., under- or overestimation of the true mean intake. Such error, also called bias, affects the accuracy of the measurement, and arises for instance when individuals with high true intake under-report or individuals with low true intakes over-report their dietary intake (intake-related bias). In addition, systematic error may also result from cultural or social desirability (person-specific bias)[203]. Systematic errors are considered more problematic since they may skew the measured population intake as well as attenuate or enlarge the association between a dietary exposure and disease outcome[201]. However, neither repeated measurements nor a larger sample size can reduce this type error[199]. **Figure 1** illustrates the effect of random and systematic errors on the precision and accuracy of the measured values.

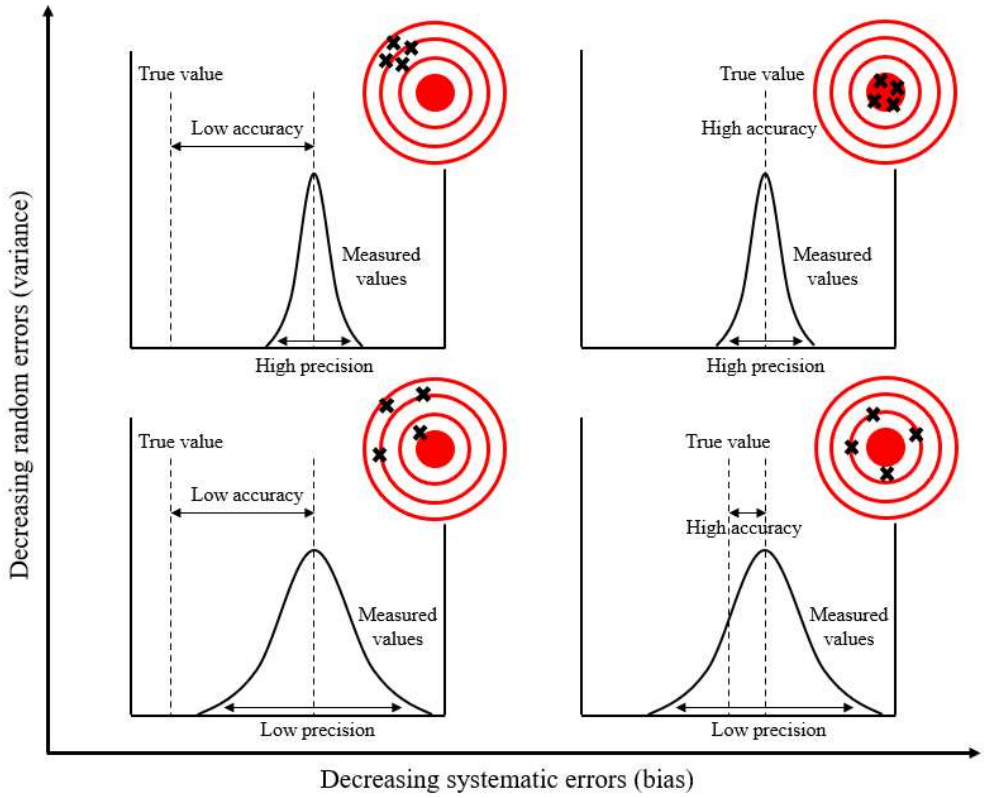

Figure 1. Illustration of effects of random and systematic errors on the precision and accuracy of the measured values.

3.5.2 Validity and reproducibility of dietary assessment methods

Validity describes the accuracy of a method i.e., to which degree the method measures what it is designed to measure. Absolute validity in dietary assessment is assessed by comparing the test method with a superior method representing true dietary intake without measurement errors[15]. However, as mentioned earlier, no method exists that can capture true dietary intake, with one exception perhaps being the measurement of energy intake by the DLW technique[204]. Direct observations could be used, as the best alternative, to obtain absolute validity, but this approach is difficult, time consuming and costly in practice and almost impossible if usual long-term dietary intake is the target[198]. Instead, relative validity is most often assessed. Here the test method is compared to a superior but imperfect method, also referred to as the reference method, which is considered to have a higher degree of validity[185]. In addition, the reference method should also have limited overlapping measurement errors with the test method. Since having similar measurement errors may result in high correlations between the test and reference method, which can falsely be interpreted as the test method having high accuracy[205]. Often when validating an FFQ, 24-HDRs or WFRs are used as reference methods. For both the 24-HDR and the WFR, the reported dietary intake is considered more accurate than the FFQ, due to the open-ended design as well as the level of detail regarding portion sizes and cooking methods[174]. Even though the 24-HDR has overlapping measurement errors with the FFQ i.e.,

being dependent on the participants' memory, the recalling period of the 24-HDR still differs from that of the FFQ since it is relatively short and does not include cognitive complex calculations.

Reproducibility describes to the precision of a method i.e., to which degree consistent measurements are obtained when for instance a dietary assessment method is repeated by the same person under the same circumstances[15]. The reproducibility (precision) can be affected by random within-individual variations but also non-random variability caused by factors such as season, change of diet or illness[206]. When evaluating the reproducibility of a self-report dietary assessment method, the time interval between administrations is important, though it can be difficult to determine which time interval is optimal. Administration of dietary assessments too close to each other in time may result in high reproducibility due to a training effect, whereas administration of the method assessments too far from each other in time may result in low reproducibility caused by true changes in dietary intake[207].

3.5.3 Statistical techniques to determine validity and reproducibility

Validity and reproducibility studies are conducted in order to identify possible systematic and random measurement errors. This is important when applying a new or an updated method or if a method is used in a different population than that it was originally designed for[15].

Several statistical tests are usually applied in order to determine different aspects of the validity and reproducibility of a dietary assessment method. These tests include correlation and different agreement analyses[191,208]. Correlation analysis is a common statistical method applied in validity and reproducibility studies[209]. It determines the strength of the relationship between two variables, in this regard between dietary intake variables measured by two different dietary assessment methods or by repeated administrations of the same method. The correlation coefficient ranges from -1 to 1 and a correlation coefficient close to one (negative or positive) would imply that the two variables are highly correlated. However, a correlation between two variables does not provide information about the differences between them and therefore a high correlation per se is not equated with having a high level of agreement[210].

Comparing group means of dietary intake measured by the test method and the reference method gives an indication of the agreement at group level i.e., the direction and magnitude of error. The ability to rank participants based on dietary intake is especially important in studies investigating diet-disease associations. Cross-classification refers to the degree of misclassification between the methods. Here participants are divided into categories of tertiles, quartiles or quintiles based on dietary intakes from the test and the reference method. The proportion of participants correctly classified into the same category or opposite category can then be determined, with the latter category resembling the proportion of participants being misclassified[209]. It has been proposed by Masson and colleagues that misclassification should not be higher than 10% and that the percentage of participants classified into the same tertile or quartile should be above 50%[208]. One should though have in mind that the percentage of agreement also contains chance agreement[209].

Bland-Altman analysis can be used to quantify the agreement between two measures across the entire range of intakes. With the Bland-Altman method it is possible to visually determine systematic differences (i.e., bias or as mentioned above, the mean difference at group level) between the test and the reference method and whether these differences are present across the range of dietary intakes. This is assessed by plotting the average of the methods on the x-axis and the difference between the two methods on the y-axis. Ideally, data points should be scattered equally below and above the mean difference throughout the range of dietary intakes[211,212]. In addition, it is possible to quantify the degree to which the test and the reference methods agree also referred to as limits of agreement (LOA). The LOA are calculated based on the mean differences and standard deviation (SD) between the two dietary intake measures, where 95% of the data points must be included within ± 2 SD of the mean difference[211,212]. The LOA can also be calculated non-parametrically. However, in order to fully determine whether the agreement between the two methods is acceptable or good depends on the objective of the particular study[191]. In other words, good agreement does not only mean that 95% of the data points lie within the LOA.

An additional statistical test in reproducibility studies is the intra-class correlation coefficient (ICC). The ICC is the ratio of the variance between subjects to the total variance (i.e., the variance between subjects plus the variance within subjects). The ICC ranges from 0-1 where an ICC close to 1, would indicate little within-subject variability and vice versa.

There are different forms of ICCs depending on the study purpose, design and type of measurement obtained[213]. The ICC for test-retest reproducibility refers to the variability in measurements obtained by the same instrument and from the same individuals. For instance, the variability in dietary intake between administrations of 24-HDRs as well as the variability in metabolite concentrations or gut microbiota abundance from repeated measures over time[18,214]. Other forms of ICC's refer to the inter-rater and intra-rater reliability i.e., where raters or interviewers may influence on the measurements. The ICC interrater reliability refers to the variation in measurements between two raters or interviewers obtained on the same group of individuals, whereas the ICC intra-rater reliability refers to the variation in measurements caused by one rater or interviewer obtained on different groups[213]

3.6 The human gut microbiota

The term gut microbiota refers to all microbes living in the gut including bacteria, bacteriophages, viruses, fungi and archaea, where bacteria are the most abundant microbe. Gut bacteria can be described at several taxonomic levels including phylum, class, order, family, genus, species, strain or clade level. Two phyla dominate the human gut microbiota; *Firmicutes* and *Bacteroidetes*[215,216]. At the genus level, the most abundant genera are *Bacteroides*, *Faecalibacterium* and *Prevotella,* though *Prevotella* is not prevalent in all humans[217]**.** The gut microbiota performs many important functions related to our health such as synthesis of vitamins and hormones, fermentation of non-digestible food components acting as a source of energy for colonocytes and the host, maintenance of intestinal epithelial barrier, regulation of immune homeostasis and protection against pathogens[218-221]. From studies of newborns and

infants, it has been shown that mode of birth is the first factor influencing the colonization of the almost sterile gut in terms of diversity, followed by other influential factors such as type of feeding as well as use of medication, particularly antibiotics[222]. This demonstrates that the gut microbiota is highly influenced by external factors already from the first years of life. Furthermore, studies of twins have revealed that external factors have greater impact on the gut microbiota than host genetics, although some bacteria are highly genetically controlled[223,224]. The gut microbiota composition may also differ with age, which have been shown between three age-groups: infants, adults and elder[225]. Furthermore, evidence on overall gut microbial sex-differences has so far been inconsistent[226-229], which could suggest that sex may only explain a small part of the total variation[229]. However, women may have higher diversity than men[230], which may be due to younger women having higher diversity[230]. Differences at the genus level have been found between women and men, where women have shown to be enriched with *Bifidobacterium*, *Ruminococcus*, and *Akkermansia* (and species *Akkermansia muciniphila*[230]) and men have shown to be enriched with *Prevotella*, *Megamonas*, *Fusobacterium*, and *Megasphaera*[228].

3.6.1 Gut microbiota variability and stability

During the past 20 years accumulating evidence has shown that alterations in the gut microbiota is associated with a number of different chronic diseases[231-234]. Though, to conduct meaningful investigations of the role of the gut microbiota in relation to disease outcomes gut microbes under investigation must have a low within-individual variation over time. So far, studies have consistently reported a larger between-individual variation compared to the within-individual variation at various taxonomic levels over time both short-term (days or months) and long-term (1-4 years)[16,18,235,236]. Some but few studies have further calculated the intra-class correlation coefficients (ICC) described in 3.5.3 as an indicator of gut microbiota stability. Five detected phyla (*Fusiobacteria*, *Proteobacteria*, *Bacteroidetes*, *Firmicutes*, *Actinobacteria*) have been reported with rather low ICC 0.00-0.44 over six months in three different study populations from Central and North America[17], where lower taxonomic levels have been reported with higher ICCs over one year (75% of 384 species, 80% of 80 genera, 66% of 82 families with ICCs >0.5)[18] in population from Sweden. Furthermore, different variability patterns have also been observed across species, where some species seem to be highly variable, others have a bimodal pattern and lastly some also seem to be rather stable[18,237]. However, more studies investigating gut microbial stability in different and larger populations are needed to get a better understanding of the stability of the gut microbiota especially of those genera, species and strains with low abundance. In **Table 2** is a summary of studies accessing the stability of the gut microbiota composition and abundance both short-term and long-term.

Table 2. Studies focusing on gut microbial taxonomic abundance stability and composition variation.

Study	Population	Sex	Age	Duration	Sequencing technique	Abundance stability and composition variation
Short-term						
Claesson et al. 2011[235]	Ireland (n=26)	F 50% M 50%	73y (65-88y)	3 months (2 samples)	16S rRNA gene sequencing	Differences in microbiota composition were larger between individuals than within individuals.
Mehta et al. 2018[16]	USA (n=308)	M	69y (52-82)	24-72 hour (n=308) (2 samples) 6 months (n=160) (2 samples)	Metagenomics sequencing	Differences in overall species composition was larger between individuals than within individuals for both short and intermediate periods. In a total of 146 species: 97% had an ICC >0.4 after 24-72h 87% had an ICC >0.4 after 6 months
Sinha et al. 2018[17]	Costa Rica (n=116) USA (n=44) USA (n=107)	F/M	Not reported 59y (SD±13)[227] 26y (18-40y)[238]	~6 months (2 sampling occasions) in each study	16S rRNA gene sequencing	For *Fusiobacteria*, *Proteobacteria, Bacteroidetes, Firmicutes, Actinobacteria* had: ICCs between 0.33-0.44 in the CRS ICCs between 0.03-0.51 for in NCI CRC ICCs between 0.00-0.39 in the HMP
Long-term						
Faith et al. 2013[239]	USA (n=33) (twins)	F	Not reported	Up to 5 years (2-13 samples)	LEA-Seq of Bacterial 16S rRNA genes	Approximately 70% of the same strains are remained after 1 year and 60% are remained after 5 years.
Chen et al. 2021[236]	The Netherlands (n=338)	F 44% M 56%	48.2y (18-80y)	4 years (2 samples)	Metagenomics sequencing	Differences in overall species composition was larger between individuals than within individuals (differences within-individuals were smaller for individuals with higher baseline alfa-diversity). 40% of the species showed no difference in relative abundance (63 out of 157).
Olsson et al. 2022[18]	Sweden (n=75)	F 52% M 48%	57.1y (SD ± 4.1)	1 year (4 samples)	Metagenomics sequencing 16S rRNA gene sequencing	Median gut microbiota relative abundance was larger between individuals than within individuals. ICCs >0.5 was 66% for families, 80% for genera, 75% species. Different variability patterns were observed: variable species, bimodal species and stable species. Enterotypes changed in 45% of the individuals during the study period.

Low Error Amplicon Sequencing (LEA-Seq), Intra-class correlation coefficient (ICC), 16S ribosomal ribonucleic acid (16S rRNA), Costa Rica study (CRS), National Cancer Institute colorectal cancer study (NCI CRC), Human microbiome project (HMP).

3.6.2 Effects of dietary patterns on gut microbiota

Several external factors have been associated with the gut microbiota composition and abundance in adults such as diet[30,240] and medication[241,242] and probably also physical activity[243] and smoking[244] among others. Of these factors, diet seems to have a major influence on the gut microbiota profile and composition. Long-term dietary habits have shown to be associated with differences in microbial profiles and composition. Different gut microbial community clusters i.e., enterotypes, have been reported in free-living individuals across populations. Three enterotypes have been identified, dominated by the genera *Bacteroides*, *Prevotella* and *Ruminococcus*/*Ruminococcaceae*[22,31,216,245], which seems to also be associated with different dietary patterns. The *Bacteroides* enterotype has been associated with high intake of animal protein, amino acids and saturated fat[30,246], where the enterotype *Prevotella* has oppositely been associated with higher intakes om carbohydrates, simples sugars[30], dietary fibers, fruits and eggs[246]. *Ruminococcus* enterotype has been associated with higher intakes of vegetables, nuts, legumes, seaweed and dietary fiber[31]. Differences in microbial abundance have also been reported between vegetarians and omnivores as well as with higher or lower adherence to the Mediterranean (MED), Healthy Nordic food index (HNFI), and PDIs[19-29] and listed in **Appendix 1**.

Highlights from **Appendix 1**.

- Higher abundance of ***Bacteroides*** and several *Bacteroides* spp. have been associated with diets of omnivores[20] and lower adherence to MED[247], conversely higher abundance of two *Bacteroides* spp., *cellulosilyticus* and *stercoris*, have been associated with higher adherence to the MED[25,29] and HNFI[28].

- Higher abundance of ***Parabacteroides***, *Parabacteroides spp.* ***Succinivibrio*** and *Succinivibrio* spp. have been reported in omnivores[20,22] and associated with lower adherence to MED[25].

- Higher abundance of ***Prevotella*** *and Prevotella copri* have been shown in vegetarians[20,22], where higher abundance of *Prevotella corporis* has been associated with lower adherence to MED[25].

- Higher abundance of ***Eubacterium*** *eligens* has been associated with higher adherence to MED[26,29].

- Higher abundance of ***Roseburia*** spp. and ***Ruminococcus*** spp. were associated with higher adherence to MED[25,26] and hPDI[26]. Conversely lower abundance of *Ruminococcus gnavus* and *torques* have been associated with higher adherence to MED[29].

- Lower abundance of ***Escherichia*** *coli* has been reported in vegetarians/vegans[19] and associated with higher adherence to MED[248], where *Escherichia Harmanii* has shown higher abundance in omnivores[20].

3.7 The role of diet and gut microbiota on cardiometabolic risk factors

Accumulating evidence have linked gut microbiota alterations with CMDs. Specifically, the last decade several observational studies have reported compositional changes in the gut microbiota of patients with CVD and T2D compared to controls or healthy individuals[10,11]. In patients with CHD and heart failure the relative abundance of *Bacteroidetes*[249,250], *Faecalibacterium*[251,252], *Ruminococcaceae*[251,253] have shown to be lower where *Enterococcus*[251,254], *Streptococcus* spp[252,254]. *Ruminococcus gnavus*[252,254] were higher compared with controls or healthy individuals. In a systematic review, including 18 observational studies, investigating the gut microbiota composition in patients with T2D (and prediabetes), a higher relative abundance of *Firmicutes*, *Lactobacillus* spp., *Escherichia* spp. *Streptococcus* spp. and a lower relative abundance of *Bacteroidetes*, *Clostridiales*, *Clostridium* spp. and *Faecalibacterium* spp. were reported in individuals with prediabetes or patients with T2D compared to controls[11]. Specific bacteria have also been associated with risk factors for cardiometabolic diseases. In a systematic review patients with obesity or metabolic disorders had in general higher relative abundance of *Prevotella*, *Blautia*, *Lactobacillus*, *Succinivibrio*, *Escherichia*, and *Fusobacterium* where higher relative abundance of *Akkermansia*, *Alistipes*, *Desulfovibrio, Bifidobacterium*, *Faecalibacterium*, *Oscillospira, Eubacterium*, *Odoribacter* were found in lean individuals[255]. Furthermore, in hypertensive individuals altered richness, diversity and composition have been reported compared to controls, where individuals with hypertension had in general lower relative abundance of *Coprococcus*, *Bacteroides* spp., *Roseburia* spp. and *Faecalibacterium* spp. compared to controls[256]. Chronic low-grade inflammatory markers have also been associated with features of the gut microbiota. For instance, lower microbiota diversity has been associated with higher levels of hs-CRP. In addition, the relative abundance of *Faecalibacterium*, *Bifidobacterium*, *Prevotella* and *Ruminococcus* were inversely related to hs-CRP and IL-6[257].

The link between the gut microbiota and development of CMDs is thought to be via the production of metabolites from gut microbial activities[258-260]. These metabolites are produced from both gut-host interactions and from gut interactions with external sources particularly diet i.e., undigestible carbohydrates, proteins and to a lesser extend fats. Gut-derived metabolites act as energy source for colonocytes but are also transported from the gut lumen over to the blood stream exerting different effects in various tissues and organs[261]. Thus, the food and nutrients that we consume play an important role in the gut microbiota composition and diversity as well as types and amounts of diet-derived microbial metabolites[261]. In fact, there is increasing evidence suggesting that the gut microbiota may act partly as a mediator in the relationship between diet and development of CMDs[259,262] (**Figure 3**).

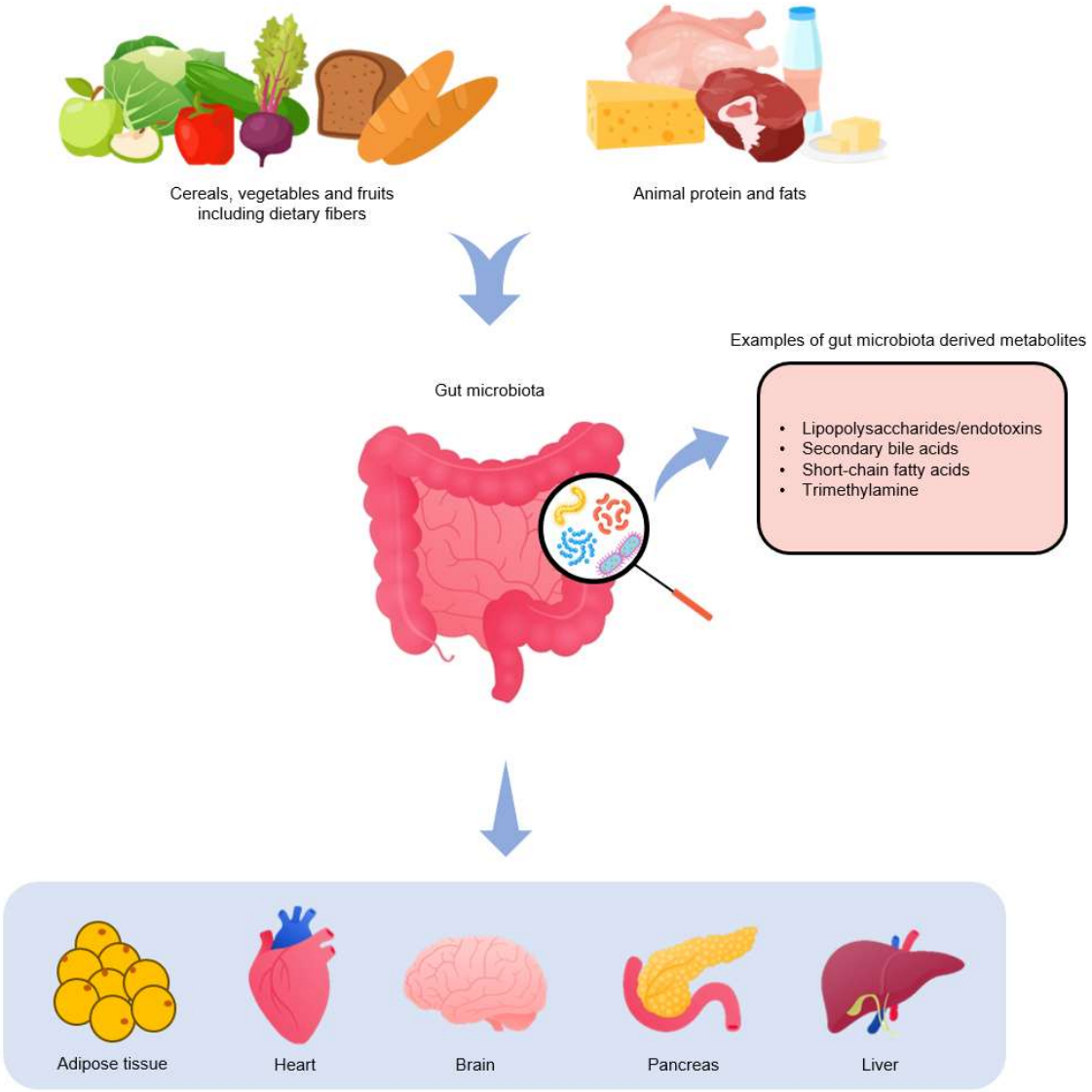

Figure 3. Diet and gut microbiota interactions and their influence on host health. Modified from Schroeder and Bäckhed[263].

Short-chain fatty acids

Several diet-derived microbial metabolites have been associated with host health. SCFAs are produced by bacterial fermentation in the gut from primarily undigested dietary fibers and resistant starch, where acetate, propionate and butyrate make up around 80% of all SCFAs. There are several SCFA producers such as *Prevotella* spp., *Ruminococcus* spp., *Bifidobatcerium* spp., *Akkermansia muciniphila*, *Ruminoccous bromii* and *Faecalibacterium prasunitzii* to name a few[258]. The potential effect of SCFAs on host health are various. For instance, butyrate is the main energy source for colonocytes and thereby an important substrate for maintenance of the intestinal epithelium[264]. In vitro, butyrate has shown to increase the integrity of the epithelial barrier through regulation of tight-junctions[265]. In adipose tissue SCFA have been suggested to be involved in regulation of the lipid metabolism by stimulating

adipogenesis (formation of adipocytes) and inhibiting lipolysis (breakdown of triglycerides)[266] as well as enhancing secretion of the satiety hormone leptin[267]. Besides leptin, SCFA may also induce the secretion of other appetite hormones, glucagon like peptide 1 (GLP-1) and peptide YY (PYY), from the gut[268]. Furthermore, SCFA may also play a role in regulation of glucose homeostasis by enhancing insulin release by the pancreas, increase insulin sensitivity and fatty acids oxidation and reduce lipid accumulation, gluconeogenesis in liver and skeletal muscles as well as reduce plasma glucose and cholesterol[269]. Lastly, SCFA may also be involved in inflammation by inducing the release of anti-inflammatory cytokines as well as inhibiting the release of pro-inflammatory cytokines[270]. However, the potential role of SCFA on health should be viewed in the light of that most of the research carried out have been cell or animal studies.

Less is known about the effect of amino acid-derived metabolites i.e., fermentation of proteins. Excess proteins are fermented in the gut into several different metabolites such as SCFA, branched-chain fatty acids, amines, ammonia, hydrogen sulfide, indoles and phenols[271]. These metabolites may exhibit adverse effects on host health and have been associated with kidney disease, steatosis and insulin resistance. Oppositely, an amino acids-derived metabolite (from tryptophan), indole propionic acid, has been associated with lower prevalence of T2D[261].

Trimethylamine N-oxide

TMAO is a gut-derived metabolite, which have been associated with increased risk of CVD incidents and CVD mortality[272]. TMAO can be obtained directly (high in some fish) or produced by fermentation of diet ingested choline, L-carnitine or betaine. TMA is then transported via the blood to the liver where it is further transformed via oxidation into the proatherogenic TMAO[273]. In a recent study, *Lachnoclostridium* has been associated with atherosclerotic patients and the species *Lachnoclostridium saccharolyticum* have shown to convert choline to TMA in vitro and enhance atherosclerosis in mice[274]. Accumulating evidence shows that TMAO is involved in atherosclerosis development and progression by playing a role in foam cell formation, endothelial dysfunction and plaque instability[275].

Secondary bile acids

When a fatty meal is consumed bile acids are released to aid the digestion and absorption of lipids as well as absorption of cholesterol and fat-soluble vitamins in the small intestine. These bile acids can be transformed into secondary bile acids by different gut microbes (*Clostridium*, *Enterococcus*, *Bifidobacterium*, *Lactobacillus* and *Bacteroides*). Some of these secondary bile acids are not reabsorbed but instead excreted in the stool. This will in turn enhance bile acid neo-synthesis in the liver and consequently results in loss of low-density lipoprotein[276].

Lipopolysaccharides

Lipopolysaccharides (LPS) are bacteria surface molecules, which are produced by most Gram-negative bacteria in the gut, for instance *Escherichia coli*. LPS has been associated with inflammation and elevated levels of LPS in blood (though not endotoxemia) has been seen following a high-fat meal. However, whether a high-fat diet consumed over a long period will results in elevated LPS levels and further unbeneficial health effects still needs to be elucidated[277].

4. STUDY DESIGN AND METHODS

4.1 Hypotheses and research strategies

Validation of dietary data and stability of the gut microbiota are important to ensure useful measurements for subsequent investigations of diet, the gut microbiota, and their role in disease development. The hypotheses of this thesis were therefore that dietary data are reliably assessed and that a certain proportion of the gut microbiota are stable over time. In the light of reliable diet and gut microbial data it was further hypothesized that dietary patterns are associated with differences in gut microbiota composition and that the association between diet and risk factors for CMDs are partly mediated trough the gut microbiota indicated by the *Prevotella*-to-*Bacteroides* (P/B) ratio.

More specifically the following five hypothesis were addressed in the thesis:

- The FFQ is acceptable in ranking of individuals according to energy and nutrient intakes (*paper I)* and is likewise reproducible after one year.
- The DQS is a good indicator of overall dietary quality, and a high adherence is associated with lower levels of CMD risk factors (*paper II*)
- A substantial part of the gut microbiota is relatively stable over one year (*paper III*)
- Dietary patterns are associated with differences in gut microbiota composition (*paper III*)
- The gut microbiota is partly a mediator of the association between diet and CMD risk factors (*paper IV*)

To investigate these hypotheses the DCH-NG MAX sub-cohort was established. The DCH-NG MAX included participants from whom biological samples (blood, urine, stool, saliva), anthropometric and body composition measurements as well as information about their diet and lifestyle at baseline, 6 months and 12 months were collected. Paper I, III, IV were based on data collected from the DCH-NG MAX sub-cohort, whereas paper II was based on data from the DCH-NG cohort.

4.2 The Diet, Cancer and Health – Next Generations cohort and MAX sub-cohort

4.2.1 DCH-NG

In 2015-2019 the Danish DCH-NG cohort was established with the purpose to constitute as a resource for transgenerational research of the role of genetic, environmental, behavioural and socioeconomic factors and the microbiome as well as their complex interactions[278]. The DCH-NG cohort is an extension of the Diet, Cancer and Health cohort[279] i.e., biological children, their spouses as well as grandchildren of the DCH participants were invited to participate in the study. In total, 183,764 descendants were invited by postal letter and 22% completed the study requirements resulting in 39,554 men and women from 18 to 79 years of age. All included participants provided informed consent either electronically or on paper. Participation in the study required completion of two comprehensive self-administered web-based questionnaires, an FFQ and a lifestyle questionnaire (LSQ), and a 30-minute clinical examination in one of the two study centers in Copenhagen or Aarhus. At the clinical examination urine, blood, saliva and stool samples were collected as well as measurements of height, weight, waist- and hip-circumference, body composition from bioimpedance (e.g., fat-free mass, fat mass, visceral fat mass, skeletal muscle mass) and BP[278] (**Figure 4**).

4.2.2 DCH-NG MAX

The DCH-NG MAX study is a validation sub-cohort of the DCH-NG cohort. The primary aim of the DCH-NG MAX study was to allow evaluation and validation of metabolomics, metagenomics, genetic analyses and questionnaires as well as to conduct explorative investigations of the association between genetics, microbiota, lifestyle and the molecular phenotype at baseline and over one year. The DCH-NG MAX study was established in 2017-2019. All participants visiting the Copenhagen study center for the clinical examination during August 2017 to January 2018 were invited to join the study and a total of 720 participants were enrolled. In this period study personal informed the participants about the validation study after completion of the baseline visit. The participants willing to participate in the study were further informed about the study requirements and signed the informed consent directly after. The inclusion criteria were completion of the DCH-NG FFQ, LSQ and clinical examination at baseline and the exclusion criteria was being pregnant. Additional diet measurements were included in the DCH-NG MAX study i.e., participants were also required to fill out two self-administered web-based 24-HDRs. Participants had to report all intakes of food and beverages consumed the day before the clinical examination (first recall) and all intakes of food and beverages consumed the day of the clinical examination (second recall). Participant were followed up at 6 months and 12 months where they were required to complete the DCH-NG FFQ, LSQ, 2x24-HDRs and visit the study center for the clinical examination[278,280](**Figure 4**).

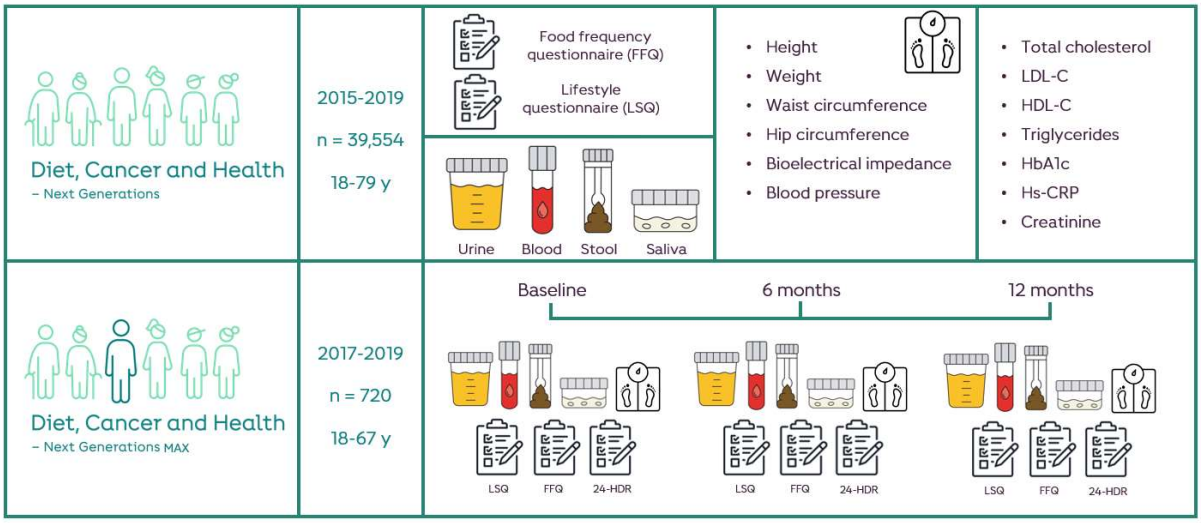

Figure 4. Overview of the data collection in Diet, Cancer and Health – Next Generations (DCH-NG) cohort and the DCH-NG MAX sub-cohort.

4.3 Study populations and design

4.3.1 Overview of study populations in paper I- IV

Participants included in paper I, III and IV were from the DCH-NG MAX sub-cohort, whereas participants included in paper II were a sub-sample from the DCH-NG cohort. Paper III and IV are based on the same population. A more detailed description of participants can be found in each paper, but a brief overview of samples size, age and sex in each paper are provided in **Table 3**.

Table 3. Brief overview of participant characteristics across paper I-IV.

	Food frequency questionnaire validation and reproducibility (Paper I)	**Dietary quality score validation and associations with cardiometabolic risk factors (Paper II)**	**Gut microbiota variation and associations with dietary patterns (Paper III)**	**Direct and indirect effect of diet on cardiometabolic risk factors (Paper IV)**
Population	DCH-NG MAX	DCH-NG	DCH-NG MAX	DCH-NG MAX
n	415	450	444	439
Age (years) Median (p25-p75)	50 (40-54)	49 (33-60)	49 (36-54)	F: 48 (30-53) M: 49 (40-54)
Sex (F/M %)	55 / 45	51 / 49	55 / 45	54 / 46

4.3.2 FFQ validity and reproducibility

In paper I the relative validity and reproducibility of the DCH-NG 376-item FFQ were assessed, and the study population was comprised of participants from the DCH-NG MAX. For the validity assessment the FFQ administered at baseline was compared with a mean of 3x24-HDRs for intakes of energy and nutrients. Therefore, participants who had completed the FFQ at baseline as well as the 24-HDRs at baseline, 6 months and 12 months were included. 12 participants were afterwards excluded since they only had 24-HDRs registered on weekend days, and it was assumed that the intake of foods and beverages was not representative of a usual diet. In total, 289 participants were included in the validity analysis. For the reproducibility assessment, the FFQ administered at baseline and at 12 months were compared for intakes of energy, nutrients and major food groups. Participants were included in the reproducibility analysis if they had complete FFQ at baseline and at 12 months, which resulted in a total of 415 participants.

4.3.3 DQS validity and associations with CMD risk factors

In paper II the DQS was validated against the DCH-NG 376-item FFQ and adherence to the DQS was associated with risk factors for CMDs. The study population was based on a sub-sample from the DCH-NG cohort, which was drawn during the main data collection. Besides their participation in the DCH-NG, the participants of the sub-sample had to fill out a 23-item FFQ (web-based), which was used to calculate the DQS. I.e., a total of 2,556 participants enrolled in the DCH-NG cohort from August 10, 2015 till April 16, 2016 were eligible if they had an e-mail address in the cohort database, anthropometric and body composition measurements, analysed blood samples from the clinical examination as well as complete FFQ and LSQ. From the 2,556 participants a random sample of 598 participants were included in the study containing five age groups (18-29, 30-39, 40-49, 50-59, >60 years) with approximately equal proportions of women and men in each age group. The sample size was selected based on recommended sample sizes in validation studies and with an expected participation rate equal to that of the main study. The 598 participants had to complete the 23-item FFQ within 14 days. 450 participants completed the 23-item FFQ and were included in the statistical analyses.

4.3.4 Diet, gut microbiota and their interaction with cardiometabolic risk factors

In paper III, the temporal gut microbiota variation at the genus level during one year was assessed. Furthermore, the association between adherence to different dietary indexes and their constituting food groups with the gut microbiota at the genus level was investigated. The dietary indexes included were the HNFI, relative Mediterranean diet score (rMED), PDI, hPDI, uPDI and pro-veg. For assessing the temporal variation of the gut microbiota over one year, participants with analysed stool samples from baseline, 6 months and 12 months were included. A stool sample was excluded if the participant had taken antibiotics <3 months prior to the

sampling (99 samples), which resulted in a total of 214 participants. For investigating the association between adherence to dietary indexes and the gut microbiota, participants with at least one analysed stool sample and at least two 24-HDRs were included, resulting in a total of 444 participants. For those participants that had two or three analysed stool samples, microbiota data were averaged, hence all participants would have one gut microbiota measure. Dietary index scores were calculated based on an average of two or three 24-HDRs per participant.

In paper IV the direct and indirect effect of diet on risk factors for CMDs were investigated using a structural equation model (SEM). The inclusion criteria were the same as for paper III described above, i.e., 444 participants with data on diet and gut microbiota. Five participants were afterwards excluded due to lack of bioelectrical impedance measures due to maybe being pregnant. A total of 439 participants were included in the analyses. A mean of two or three 24-HDRs was used to calculate g/day for each food component included in the HNFI and the rMED. The following risk factors for CMD were used as outcomes: BMI (kg/m^2), WC (cm), fat mass (%), SBP, diastolic blood pressure (DBP), total cholesterol (CHO), HDL-C, LDL-C, triglycerides (TG), HbA1c (mmol/l) and hs-CRP. Participants with missing outcomes were handled by full information maximum likelihood (FIML). FIML uses a likelihood function that accounts for all available data, including missing data information, without the need for imputation or case deletion.

4.4 Assessment of exposures and covariates

4.4.1 Daily intakes of foods, energy and nutrients

DCH-NG FFQ

The DCH-NG FFQ is a web-based semi-quantitative questionnaire with the purpose to measure habitual dietary intake in adults during one year. The FFQ was constructed from former questionnaires the paper-based FFQ from the DCH cohort and the web-based FFQ from the Danish National Birth cohort[281,282]. A thorough revision of the previous FFQs was made during 2014-2015, whereafter the web-based DCH-NG FFQ was made. We made modifications to the already existing questions and extended the food list, to better reflect food consumed at the time. From the DCH-NG FFQ, reported frequency consumption of each food item was multiplied with gender specific standard portion sizes. Portion sizes were mainly from EPIC-SOFT[283], and some were from the National Food Institute (DTU FOOD)[284] and the Danish National Birth Cohort[285]. Afterwards all food items were matched with a standard recipe. Ingredients (foods) from the recipes were linked to the Danish food composition table (Frida Food, version 4, 2019)[286]. Lastly, calculations of daily intakes of energy, nutrients and food groups were performed with FoodCalc version 1.3[287]. In paper 1 and II, daily intake of energy, nutrients and foods were used. Food items from the FFQ included in food groups used in paper I and II are listed in **Appendix 2** and **3**, respectively. In paper I the DCH-NG FFQ is referred to as the FFQ, whereas in paper II the FFQ is referred as the 376-item FFQ to differentiate it from the 23-item FFQ screener.

23-item FFQ

Daily and weekly intakes of food groups were calculated from the 23-item FFQ. Frequency consumption of each food item in the FFQ were multiplied with standard portion sizes from the National Food Institute (DTU FOOD)[284]. Food items from the 23-item FFQ included in food groups used in paper II are listed in **Appendix 3**.

24-HDRs

24-hour intake of foods and beverages were obtained using the online dietary assessment tool myfood24. Daily intakes of energy, nutrients and foods were calculated directly in myfood24[194]. Thus, each food item was multiplied by the specific portion size chosen by the participant (from portion size images, household measures or exact amounts) and linked to its corresponding food composition table from Denmark (Frida Food, version 2, 2017)[288], Sweden (Livsmedelverket, 2017)[289] or England (McCance and Widdowson's version 6 and 7, myfood24 branded UK food composition database)[290-292]. Food items from the 24-HDRs included in the food groups in each diet index used in paper III and IV are listed in **Appendix 4**.

4.4.2 Diet index scores

Diet scores were included in paper II, III and IV. An overview of indexes included in each paper are found in **Table 4**.

Table 4. Brief overview of diet index scores included in paper II-IV.

	Dietary quality score validation and associations with cardiometabolic risk factors (Paper II)	**Gut microbiota variation and associations with dietary patterns (Paper III)**	**Direct and indirect effect of diet on cardiometabolic risk factors (Paper IV)**
DQS	X		
HNFI		X	X*
rMED		X	X*
PDI		X	
hPDI		X	
uPDI		X	
pro-veg		X	

Dietary quality score (DQS), healthy Nordic food index (HNFI), relative Mediterranean diet score (rMED), plant-based diet index (PDI), healthy plant-based index (hPDI), unhealthy plant-based index (uPDI), provegetarian index (pro-veg). *No score was used but g/day for each food component included in the HNFI and rMED.

Dietary quality score

The DQS was validated and used as dietary exposure in paper II. The DQS was based on foods chosen *a priori* from nutritional and health aspects in relation to CVD. Four food groups were included: vegetables, fruit, fish and fats. A three-point scoring system was made for each group (0-2 points) to group individuals into three categories: having healthy, average or unhealthy dietary habits. The reported intakes to calculate the DQS were obtained from the 23-item FFQ. This 23-item FFQ was, in turn, shortened from the original FFQ (48 items) used for calculation

of the DQS[152]. The cut-off for a high score was based on recommended intakes from the Danish official dietary guidelines for vegetables, fruits and fish. A high score (2 points) was given for vegetables ≥5-7 servings/week, for fruits ≥3 pieces/day and for fish ≥200 g/week. For fat a high score was given for no use of spread on bread and no use of fat or use of olive oil for cooking. A low score (0 points) was given for vegetables ≤ 2 servings/week, for fruits ≤ 2 pieces/week and for non-consumers of fish. For fat a low score was given for using only saturated fat for cooking and spread. The score ranged from 0-8, which afterwards was converted to a score ranging from 1-9 and referred to as the 9-classed score[152,280].

Healthy Nordic Food Index

The HNFI was used as dietary exposure in paper III and IV. The HNFI was based on foods from the Nordic region chosen *a priori* with expected health-promoting effects. Six food groups were included in the score: fish, root vegetables, cabbage, apples and pears, whole grain oats and whole grain rye. The score was calculated based on an average of reported intakes from two or three 24-HDRs. For each food group 1 point was given for intakes above the sex-specific median and 0 points was given for intakes below the sex-specific median. The score ranged from 0-6 points[154].

Relative Mediterranean Diet score

The rMED was used as dietary exposures in paper III and IV. The rMED diet is a simplification of the original Mediterranean diet score. Nine food groups were included in the score: vegetables, fruit, legumes, fish, cereals, olive oil, alcohol, meat and dairy products. The score was calculated based on an average of reported intakes from two or three 24-HDRs. Each food group was divided into intake tertiles and for vegetables, fruit, legumes, fish and cereals a score of 0-2 were given to the first, second and third tertiles. The score was reversed for meat and dairy. For olive oil non-consumers were given 0 points, 1 point for intakes below the median and 2 points for intake equal to or above the median. For alcohol, the scoring was dichotomous where moderate alcohol consumption was given the highest score i.e., a reported alcohol intake of 5-25g/d for women and 10-50g/d for men were given 2 points. Those below or above these ranges were given 0 points. The score ranged from 0-18 points. Each food group was also adjusted for energy intake calculated as grams per 1,000 kcal[149].

Provegetarian and Plant-based Diet Indexes

The pro-veg, PDI, hPDI and uPDI were used as dietary exposure in paper III. Though, the indexes are named plant-based and provegetarian, the indexes also include animal-based foods. The pro-veg was constructed *a priori* based on studies about plant-based foods and CVD risk and mortality. Twelve food groups were incorporated in the score: vegetables, fruit, legumes, cereals, potatoes, nuts, vegetable oils, animal fat, dairy, egg, fish and seafood, meat. The score was calculated based on an average of reported intakes from two or three 24-HDRs. Each food group were divided in sex-specific quintiles. For intakes of plant-based food groups, a score of 1-5 were given to the first, second, third, fourth and fifth quintiles. For intakes of animal-based food groups the score was reversed. The score ranged from 12-60 points[150].

The PDI is similar to the pro-veg and constructed *a priori* based on studies about plant-based foods and risk of T2D. In addition, *a priori* healthy and unhealthy version of the PDI were

based on evidence from less healthy plant-based foods and risk of T2D. Eighteen food groups were included in all three PDI's. The healthy plant-based food groups encompassed: whole grains, vegetables, fruit, legumes, nuts, vegetable oil, tea and coffee. The unhealthy plant-based food groups encompassed: fruit juices, refined grains, potatoes, sugar-sweetened beverages, sweets and desserts. The animal-based food groups encompassed: eggs, dairy, fish and seafood, poultry, animal fat, unprocessed red meat, processed red meat and miscellaneous animal-based foods. The score was calculated based on an average of reported intakes from two to three 24-HDRs. Each food group was divided in sex-specific quintiles and given a score from 1-5. The score ranges from 18-90[151]. In the PDI, all plant-based foods were scored positively and for animal-based foods the score was reversed. In the hPDI, healthy plant-based foods were scored positively and both for unhealthy plant-based foods and animal-based foods the score was reversed. In the uPDI unhealthy plant-based foods were scored positively and both for healthy plant-based foods and animal-based foods the score was reversed. The differences in the scoring of each food groups in the PDI, hPDI and uPDI respectively are described in **Table 5**.

Since the food groups included in the rMED were energy-adjusted we decided to energy adjust all food groups within each index. Thus, each food group (except for alcohol in rMED) was energy adjusted by the density method (food g/1000 kcal). The scoring system for each index is shown in **Table 5**. Food items from the 24-HDRs were used to calculate the scores. Food items from the 24-HDRs included in food groups in each index are listed in **Appendix 4**.

Table 5. Scoring systems for the diet indexes, modified from paper III.

Food index[a]	No. food groups	Ranking	Scoring	Theoretical range
DQS	4	Frequency: high (2 points), medium (1 point), low/zero intakes (0 points). Fat was based on a summary of two subgroups (fat on bread and fat for cooking) and scoring was item specific. See details under 4.4.2.	High intake positive for vegetables, fruit and fish.	0-8 (1-9)[d]
HNFI[b,c]	6	Below (0 points) or above (1 point) the median	High intake positive for all food groups	0-6
rMED[b,c]	9	Tertile ranking (0-2 points) For alcohol, 2 points were assigned females with intakes of 5–25 g/day and males with intakes of 10–50 g/day). 0 points for above and below the ranges For olive oil, non-consumers (0 points), below (1 point) and above (2 points) the median	Positive: vegetables, legumes, fruits and nuts, cereals, fish and seafood, olive oil, and moderate alcohol consumption) Reverse: meat and dairy products	0-18
PDI[b,c]	18	Quintile ranking (1-5 points)	Positive: whole grains, fruits, vegetables, nuts, legumes, vegetable oils, tea and coffee, fruit and vegetable juices, refined grains, potatoes, sugar sweetened beverages, sweet and desserts Reverse: animal fat, dairy, egg, fish and seafood, meat, miscellaneous animal-based foods	18-90
hPDI[b,c]	18	Quintile ranking (1-5 points)	Positive: whole grains, fruits, vegetables, nuts, legumes, vegetable oils, tea, and coffee Reverse: fruit and vegetable juices, refined grains, potatoes, sugar sweetened beverages, sweet and desserts, animal fat, dairy, egg, fish and seafood, meat, miscellaneous animal-based foods	18-90
uPDI[b,c]	18	Quintile ranking (1-5 points)	Positive: fruit and vegetable juices, refined grains, potatoes, sugar sweetened beverages, sweet and desserts Reverse: whole grains, fruits, vegetables, nuts, legumes, vegetable oils, tea and coffee, animal fat, dairy, egg, fish and seafood, meat, miscellaneous animal-based foods	18-90
pro-veg[b,c]	12	Quintile ranking (1-5 points)	Positive: cereals, fruits, vegetables, nuts, legumes, vegetable oils, potatoes Reverse: animal fat, dairy, egg, fish and seafood, meat	12-60

[a] Dietary Quality Score (DQS), Healthy Nordic Food Index (HNFI), relative Mediterranean Diet score (rMED), Plant-based Diet Index (PDI), healthy Plant-based Diet Index (hPDI), unhealthy Plant-based Diet Index (uPDI), pro-vegetarian diet index (pro-veg).

[b] Each food group (apart from alcohol and olive oil in the rMED) were calculated as grams per 1,000 kcal.

[c] Positive indicates that higher intakes received higher scores. Reverse indicates that higher intakes received lower scores.

[d] The original score ranged from 0 to 8 points, which was afterwards converted to a score ranging from 1 to 9 and referred to as the 9-classed score.

4.4.3 Assessment of covariates

Information about education, physical activity and smoking habits were self-reported and obtained from the LSQ at baseline[278]. An overview of included covariates in each paper are shown in **Table 6**.

Age and sex

From the Danish Civil Registration System, the personal identification number (CPR) on all participants were obtained, which holds information about date of birth and sex. Age was calculated from date of the clinical assessment at baseline and date of birth.

Smoking habits

Smoking habits was based on questions regarding cigarette, cigar, cheroot, pipe and e-cigarette smoking and grouped into never, former and current smokers.

Physical activity

In papers I, III and IV physical activity was based on hours of sports per week calculated from questions regarding participation in exercise and sports activities. In paper I and III physical activity was only used in descriptive analyses. In paper II, physical activity was based on leisure time activities from questions regarding gardening, household chores, sedentary activities, grocery shopping etc. and was calculated as metabolic equivalent of task hours (MET-hours) per week.

Education

In papers I, III and IV, education was only used in descriptive analyses. Educational level was classified according to highest attained education grouped into five levels according to the DISCED-15[293] classification of completed educations and recommendations from the Danish Cancer Institute about classification of educational attainment where higher education included three levels. Highest attained education was grouped as follows: basic school, vocational training, higher (2-3 years), higher (3-4 years) and higher (+4 years). In paper II, education was classified similar to the previous study validating the DQS[152]. Highest attained education was grouped in three levels and a group with no education. Thus, highest attained education was grouped as follows: No education, <2 (basic school), 2-4 (vocational training, higher 2-3 years, higher 3-4 years), >4 (Higher +4 years).

Antibiotics

Use of antibiotics was self-reported and participants registered use of antibiotics within the past year in connection with stool sampling as well as at the clinical examination.

4.5 Gut microbiota

Participants in the DCH-MAX sub-cohort were instructed to collect stool samples in connection with each clinical examination (at baseline, 6 months and 12 months). Sampling took place at the study center or at home using a stool sampling kit. Stool samples were collected in tubes without preservatives and stored at -20°C in home freezer prior to return of the samples (within three days, transported in a cooling bag with cooling elements). Upon return, stool samples were stored at -80°C at the study center until further analysis. After completion of the data collection, stool samples were freeze-dried at the University of Copenhagen and analysed by 16S rRNA gene amplicon sequencing at the Swedish University of Agricultural Sciences, Uppsala. Here DNA was extracted from stool material and the V3-V4 region of the 16S rRNA gene was amplified by PCR using the primers 341F and 806R. In total, 1329 samples from 619 individuals were analysed. Afterwards, sequencing data were processed using the pipeline DADA2 within QIIME2 to generate amplicon sequence variants (ASVs)[294]. To assign taxonomy ASVs were matched to the Silva rRNA reference database[295,296]. Taxa was assigned from phylum to the genus level and the relative abundance of each taxon was calculated. A comprehensive description of stool sampling, freeze-drying, 16S rRNA gene amplicon sequencing, taxonomic assignment is found in paper III.

4.6 Risk factors for cardiometabolic diseases

Blood pressure

Using an automatic BP monitor (Omron M-10 IT/Omron HB-1300), systolic and diastolic BP (mmHg) was measured three times on the left arm after 5 min of rest with 1-2 min of rest in between measurements. In paper II an average of all three BP measurements was used, whereas in paper IV, an average of the last two BP measurements was used.

Body composition

For all body composition measurements participants were barefoot and in underwear or light clothing. Height (cm) was measured with a wireless stadiometer (Seca 264) and weight (kg) was measured with a body composition analyser (Seca mBCA515/514). BMI was calculated by dividing weight (kg) by the square of height (m). WC was measured twice midway between the lower rib margin and iliac crest (cm). A third WC measure was taken if the difference between the first and the second measurement was greater than 1 cm. An average of the two WC measurements or if a third measurement was taken an average of the two measurements closest to each other were used. Total body fat mass (kg) and percentage as well as visceral fat (VF) (liter) were measured by bioelectric impedance (Seca mBCA515/514).

Upfront analysis

CHO (mmol/l), LDL-C (mmol/l), HDL-C (mmol/l) and TG (mmol/l) were measured by enzymatic colorimetric techniques from lithium heparin plasma. HbA1c (mmol/mol) was measured by turbidimetric inhibition immuno assay (TINIA) from full blood (EDTA). Hs-CRP (mg/l) was measured by an immunoturbidimetric assay from lithium heparin plasma. Analysis

of CHO, LDL-C, HDL-C, TG, HbA1c and hs-CRP were all performed on a Cobas 6000 analyser at the Danish National Biobank.

4.7 Statistical analyses

An overview of the statistical analyses performed in paper I-IV are shown in **Table 6**.

4.7.1 FFQ relative validity and reproducibility (Paper I)

The following statistical tests were used to evaluate the validity ($FFQ_{baseline}$ compared with the mean of 3x24-HDRs) and the reproducibility ($FFQ_{baseline}$ compared with $FFQ_{12months)}$ of the FFQ for reported intakes of energy, nutrients and food groups. Nutrient intakes were energy adjusted by the density and residual method. Nutrient densities and nutrient residuals were skewed, therefore log and double-log transformation was applied to improve normality. The non-parametric Wilcoxon signed-rank test was used to test for systematic differences in intakes. Agreement was assessed using the Bland-Altman method[211,297], where median bias and 95% LOA were calculated with a non-parametric approach using the lower and upper 2.5 percentiles of the differences. The ability of the FFQ to rank individuals according to the level of intake was assessed by calculating the proportion of participants classified into the same, adjacent, opposite and extreme opposite quartiles (cross-classification). Lastly correlations were assessed by Spearman correlation analysis with 95% confidence intervals determined by the bootstrap method.

4.7.2 DQS validity and associations with CMD risk factors (Paper II)

Spearman correlation was used to assess the correlation between reported intakes of major food groups from the 376-item FFQ and 23-item FFQ. Reported intakes of energy, nutrients and food groups from the 376-item FFQ were used to evaluate the degree to which the DQS (calculated from the 23-item FFQ) reflected dietary quality. Linear trend across the categories was evaluated by modelling the score as both a continuous and a categorical variable and testing for model reduction. Test for trend was calculated using the 9-classed score and a high DQS (7-9 points) was used as reference. Spearman correlation coefficient was used to assess the correlation between the DQS and intakes of food groups and nutrients (from the 376-item FFQ). Linear regression model adjusted for sex and age and multiple linear regression model adjusted for sex, age, education, physical activity and smoking habits were used to assess the association between the DQS and Risk factors for CMDs.

4.7.3 Gut microbiota variation and associations with dietary patterns (Paper III)

The ICC, using two-way mixed effects, single model[213], was calculated to access the ratio between the intra-individual variance and inter-individual variance over the total variance of the relative abundance at the genus level. Hierarchical clustering (HC) with Ward´s method was used to identify gut microbiota community subgroups at the genus level, which were visualized using principal component analysis (PCA). Genera associated with each gut microbiota community subgroup was identified by orthogonal partial least squares discriminant analysis (OPLS-DA). Linear regression models were used to assess the association between 1) adherence to diet indexes and gut microbial community subgroups, 2) adherence to diet indexes and genera, 3) food groups and genera. All linear regression analyses were adjusted for age and sex. False discovery rate (FDR) was used in order to correct for multiple testing.

4.7.4 The role of diet and gut microbiota on CMD risk factors (Paper IV)

Structural equation modelling (SEM) was used in order to assess the direct and indirect effect, of the healthy Nordic and Mediterranean diet on CMD risk factors, through the gut microbiota. SEM is a statistical technique that can be used to analyse complex relationships among variables by integrating aspects of factor analysis and multiple regression. A key feature is that it enables the study of unobserved latent variables. A hypothesized model based on current knowledge about the relationship between diet, gut microbiota and CMD risk factors was established. Diet and adiposity displayed latent variables, where the gut microbiota and the remaining risk factors were observed variables. The gut microbiota was defined as the *Prevotella*-to-*Bacteroides* (P/B) ratio. The food components of the HNFI and rMED were transformed by square root to improve normality. Cube root for HDL and log2 for the remaining CMD risk factors. Separate analyses were performed for each sex and models were adjusted for age, smoking status, physical activity, meat intake and alcohol. To improve the fit of the hypothesised model forward search adjusted for multiple testing using the Holm-Bonferroni method was used. In the forward search, significant relations are added to the model sequentially until no other relations are significant. In a sensitivity analysis crude models adjusted for only age were performed. An additional sensitivity analysis was made in order to explore the differential effect of the P/B ratio between diet and adiposity. This was not possible to incorporate in the SEM, due to limitations with the R-packages used. Therefore, the total effect of diet on adiposity were compared between two sub-populations divided by their P/B ratio level, defines as below or above the median.

Table 6. Statistical analyses performed in paper I-IV.

Title	Statistical method	Test	Energy adjustment	Covariates	Multiple testing
Relative Validity and Reproducibility of a Web-Based Semi-Quantitative Food Frequency Questionnaire in the Danish Diet, Cancer, and Health—Next Generations MAX Study (**Paper I**)	Wilcoxon signed-rank test Bland-Altman method[211,297] Cross-classification Spearman correlation	▪ Energy and nutrient intakes from the $FFQ_{baseline}$ and mean of 3x24-HDRs ▪ Energy, nutrient, food group intakes from the $FFQ_{baseline}$ and $FFQ_{12months}$	Density method[298]: nutrient g/total energy Residual method[298]		
An updated validation of the Dietary Quality Score: associations with risk factors for cardiometabolic diseases in a Danish population (**Paper II**)	Spearman correlation	▪ Food group intakes from the 376-item FFQ and the 23-item FFQ ▪ Energy, nutrient and food group intakes from the 376-item FFQ and the DQS			
	Linear regression models	▪ Independent variables: intake of energy, nutrients, and food groups from the 376-item FFQ, risk factors for CMDs ▪ Dependent variable: DQS ▪ Simple model ▪ Multivariate model		Age, sex, physical activity, smoking habits, education	
Variation in gut microbiota during one year and ecological traits associated with six dietary patterns: Findings from the Danish Diet, Cancer, and Health – Next Generations MAX study (**Paper III**)	Intra-class correlation coefficients (ICC)[213]	▪ Two-way mixed effects, single model ▪ Variability within individuals (three measurements per individual)			
	Linear regression models	▪ Independent variables: diet indexes, food groups ▪ Dependent variable: gut microbiota ▪ Simple model ▪ Multivariate model	Density method: food g/1000 kcal	Age, sex	FDR
The interplay between diet and gut microbiota and associations with risk factors for cardiometabolic diseases in the Danish Diet, Cancer, and Health – Next Generations MAX study (**Paper IV**)	Structural equation model (SEM)[299]	Analyses were performed separately for each sex. ▪ Laten variables: diet (HNFI or rMED), adiposity (BMI, WC, RFM), lipidemia (CHO, LDL-C, HDL-C, TG) ▪ Observed variables: gut microbiota (P/B ratio), SBP, DBP, HbA1c, hs-CRP ▪ Crude model adjusted for age only ▪ Model adjusted for covariates ▪ Goodness-of-fit		Age, alcohol, meat intake, physical activity, smoking habits	Holm-Bonferroni in the forward search

Cardiometabolic diseases (CMD), food frequency questionnaure (FFQ), 24-hour dietary recalls (24-HDRs), Dietary Quality Score (DQS), Healhty Nordic Food Index (HNFI), relative Mediterranean diet index (rMED), body mass index (BMI), waist circumrence (WC), relative fat mass (RFM), total cholesterol (CHO), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG); systolic blood pressure (SBP), diastolic blood pressure (DBP), glycated hemoglobin (HbA1c), high sensitivity C-reactive protein (hs-CRP), *Prevotella*-to-*Bacteroides* (P/B) ratio, false discovery rate (FDR).

5.1 Relative validity and reproducibility of the DCH-NG food frequency questionnaire

Higher absolute median intakes of total energy and nutrients were reported with the $FFQ_{baseline}$ compared to the 24-HDRs (**Table 7**). Significant differences in total energy intake at group level were found between instruments and a low correlation coefficient (0.26, 95% CI: 0.15,0.36) but acceptable classification of participants' energy intake into the same or adjacent quartile (70%). There were significant differences at group level for most nutrients adjusted for energy intake by the density method (**Table 8**). However, improved accordance was observed when nutrients were adjusted for energy intake by the residual method i.e., no differences were observed between the instruments' nutrient residuals (**Table 9**).

Table 7. Median daily (p25–p75) absolute intakes of energy and selected nutrients from $FFQ_{baseline}$ and the 24-HDRs (n=289).

Energy and nutrients	FFQ Baseline		24-HDRs	
	Median	(p25–p75)	Median	(p25–p75)
Energy, kJ	10,832	(9136–13,269)	8491	(7200–10,099)
Protein, g	114	(94–139)	85	(69–103)
Total fat, g	96	(77–118)	82	(64–100)
SFA, g	30	(24–40)	26	(20–33)
MUFA, g	35	(28–44)	29	(22–36)
PUFA, g	15	(12–19)	14	(10–18)
EPA, g	0.16	(0.09–0.27)	0.03	(0.00–0.16)
DHA, g	0.25	(0.15–0.40)	0.08	(0.01–0.27)
Cholesterol, mg	372	(286–475)	217	(148–303)
Carbohydrate, g	281	(227–345)	215	(178–263)
Total sugar, g	122	(91–156)	62	(48–88)
Fibre, g	33	(25–42)	21	(16–27)
Alcohol, g	10	(3.7–18)	5.5	(0.00–15)

Food frequency questionnaire (FFQ), 24-hour dietary recalls (24-HDRs), saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), polyunsaturated fatty acids (PUFA), eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA), 25th percentile (p25), 75th percentile (p75). Table adapted from paper I.

Table 8. Energy and selected density-adjusted nutrient intakes from the FFQbaseline compared with the 24-HDRs for median bias, Spearman's correlation coefficient, Bland–Altman LOA, and cross-classification (n=289).

Energy and Density Adjusted Nutrients[a]	Median Bias[c]	*p*-Value[d]	Correlation Coefficient[h] (95% CI)	Bland–Altman LOA[e]		Cross-Classification (%)			
				Lower	Upper	Same Quartile	Adjacent Quartile	Opposite Quartile	Extreme Opposite Quartile
Energy, kJ	25	<0.0001	0.26 (0.15–0.36)	−32	189	30	40	24	6
Protein, g/kJ	1.06	<0.0001	0.50 (0.40–0.58)	−23	49	40	41	15	4
Total fat, g/kJ	0.92	<0.0001	0.43 (0.34–0.52)	−31	27	35	44	17	5
SFA, g/kJ	0.95	0.0006	0.46 (0.36–0.55)	−40	57	39	38	20	3
MUFA, g/kJ	0.98	0.0678	0.37 (0.25–0.47)	−38	63	35	40	19	7
PUFA, g/kJ	0.87	<0.0001	0.37 (0.26–0.47)	−51	61	31	46	18	6
EPA, g/kJ[b]	0	<0.0001	0.20 (0.07–0.31)	-[f]	-[f]	28	42	22	8
DHA, g/kJ[b]	0	<0.0001	0.28 (0.16–0.39)	-[f]	-[f]	29	43	20	7
Cholesterol, g/kJ	1.34	<0.0001	0.37 (0.26–0.48)	-[f]	-[f]	37	38	20	6
Carbohydrate, g/kJ	1.02	0.0037	0.50 (0.39–0.59)	−20	43	39	40	18	3
Total sugar, g/kJ	1.50	<0.0001	0.51 (0.41–0.60)	-[f]	-[f]	42	38	16	4
Fibre, g/kJ	1.21	<0.0001	0.57 (0.48–0.65)	−31	112	42	42	13	3
Alcohol, g/kJ[b]	0.01	0.9708	0.53 (0.44–0.62)	-[f]	-[f]	-[g]	-[g]	-[g]	-[g]

Food frequency questionnaire (FFQ), 24-hour dietary recalls (24-HDRs), saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), polyunsaturated fatty acids (PUFA), eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA), 25th percentile (p25), 75th percentile (p75). [a]Based on log-transformed density intakes. [b]Based on raw intake density intakes. [c]Median bias is reported as a percentage for log-transformed density intakes and unit difference for raw density intakes. [d]p-value, the test of difference in intake between (log[a]) FFQbaseline and (log[a]) mean of three 24-HDRs using Wilcoxon signed-rank test. [e]Bland–Altman limits of agreement (LOAs) are reported as a percentage difference. [f]Bland–Altman limits of agreement (LOAs) are not reported as LOAs depend on the level of the nutrient. [g]Alcohol had a large proportion of non-consumers in the 24-HDRs. [h]Spearman's correlation coefficient and 95% confidence intervals (CI). Table adapted from paper I.

Table 9. Selected nutrient residual intakes from the FFQbaseline compared with the 24-HDRs for Wilcoxon sing-rank test, Spearman's correlation coefficient, Bland-Altman (LOA) and cross-classification (n=289).

Nutrient residuals[a]	p-value[c]	Correlation Coefficient[h]		Bland-Altman LOA[d]		Cross-classification (%)			
		Crude (95% CI)	Energy-adjusted (95% CI)	Lower	Upper	Same quartile	Adjacent quartile	Opposite quartile	Extreme opposite quartile
Protein, g	0.9484	0.26 (0.14-0.37)	0.48 (0.38-0.56)	-29	42	41	38	17	4
Total fat, g	0.6999	0.26 (0.14-0.38)	0.44 (0.34-0.52)	-26	35	33	46	17	4
SFA, g	0.8726	0.32 (0.21-0.42)	0.45 (0.35-0.54)	-38	63	39	37	21	3
MUFA, g	0.7015	0.23 (0.10-0.34)	0.37 (0.27-0.47)	-38	63	36	38	20	6
PUFA, g	0.7704	0.22 (0.11-0.33)	0.37 (0.26-0.47)	-43	83	30	47	17	6
EPA, g[b]	0.2053	0.20 (0.09-0.31)	0.21 (0.10-0.31)	-[e]	-[e]	32	37	24	7
DHA, g[b]	0.1810	0.26 (0.15-0.37)	0.26 (0.15-0.37)	-[e]	-[e]	29	42	21	8
Cholesterol, mg	0.3378	0.31 (0.20-0.42)	0.37 (0.25-0.48)	-[e]	-[e]	35	39	19	6
Carbohydrate, g	0.4371	0.35 (0.25-0.45)	0.49 (0.39-0.58)	-23	37	40	37	19	4
Total sugar, g	0.4413	0.46 (0.35-0.55)	0.50 (0.39-0.59)	-[e]	-[e]	41	38	18	3
Fibre, g	0.8489	0.43 (0.32-0.53)	0.53 (0.43-0.62)	-44	82	41	40	16	3
Alcohol, g[b]	0.6808	0.55 (0.46-0.63)	0.46 (0.35-0.55)	-[e]	-[e]	47	33	15	5

Food frequency questionnaire (FFQ), 24-hour dietary recalls (24-HDRs), saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), polyunsaturated fatty acids (PUFA), eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA), 25th percentile (p25), 75th percentile (p75). [a]Based on double log-transformed energy adjusted nutrient intake by the residual method. [b]EPA, DHA, alcohol, vitamin K, thiamine, vitamin B6 and folate are based on crude nutrient intake and afterwards energy-adjusted by the residual method. [c]p-value , test of difference in intake between FFQbaseline and the mean of 3x24-HDRs by Wilcoxon signed-rank test. [d]Bland-Altman limits of agreement (LOA) are reported as percentage difference. [e]Bland-Altman limits of agreement (LOA) are not reported as LOA depend on the level of the nutrient. [h]Spearman's correlation coefficient and 95% confidence intervals (CI). Table adapted from paper I.

The discrepancy observed for total energy and nutrients between the $FFQ_{baseline}$ and the 24-HDRs may be related to certain aspects of the study design and the instruments themself. First, having a long list of food items in FFQs has been suggested to lead to overestimation of food intakes[15], and the present $FFQ_{baseline}$ consisted of 376-items. Conversely, FFQs with a long list of food items (>200) have been reported to better rank individuals according to nutrient intakes[300]. Second, there may have been too few 24-HDRs. Seven days have been recommended to adequately estimate total energy and even more for specific nutrient intakes[189]. However, this is not often achieved in validity studies. Third, the participants may have forgotten, or omitted foods consumed while completing the 24-HDRs, since it is an open-ended tool with no preset list of food items and no interviewer assistance. Fourth, the participants may also have had difficulties in remembering food portion sizes accurately. Despite the larger intake estimated by the $FFQ_{baseline}$ the absolute intake of energy and macronutrients are within the range of intakes estimated with several other web-based FFQs with varying number of food items (**Table 10**). To evaluate total energy intake more accurately, the FFQ could have been compared with an objective method, such as the DLW technique, but that is cumbersome and has not been performed as part of this validity assessment.

Table 10. A simple overview of reported intakes of total energy and macronutrients comparing the DCH-NG FFQ with other web-based FFQs.

	DCH-NG FFQ in MAX	**Other web-based FFQs**[301-313]
No. of items in the FFQ	376	44-279
Energy intake, kJ	10,832	7017-12,343
Fat, g	96	62-110
Protein, g	114	70-122
Carbohydrates, g	281	190-384

Diet, Cancer and Health – Next Generations (DCH-NG) cohort, Food frequency questionnaire (FFQ), 24-hour dietary recalls (24-HDRs).

Energy-adjusted correlation coefficients ranged from 0.18-0.58 and classification of participants' intake into the same or adjacent intake quartile ranged from 69%-86% for both nutrient densities and nutrient residuals (**Table 8 and 9**). These results are in line with results from previous validation studies[301-313], though caution should be made when comparing correlations coefficients or rankings across studies, since such estimates are population dependent (age, sex, population size, number of food items, portion sizes, recall period, number of 24-HDRs or WFRs). Moreover, Bland-Altman LOA were estimated for intake of total energy, nutrient densities and nutrient residuals but was not possible for most micronutrients since the LOA was dependent on the level of the nutrient. Bland–Altman plots for total energy and macronutrient densities (back-transformed) showed a bias with an increased dispersion with increased mean discernible as a funnel shape in the plots **Figure 5**. Validity of food group intakes was not assessed, due to too many zero intakes in the data from the 24-HDRs, although the sample size was considered adequate according to recommendations of including 200-300 subjects with three dietary records per subject[15]. One explanation may be that several food items

such as fish, poultry and red meat are not eaten on a regularly basis and therefore more recording days are needed.

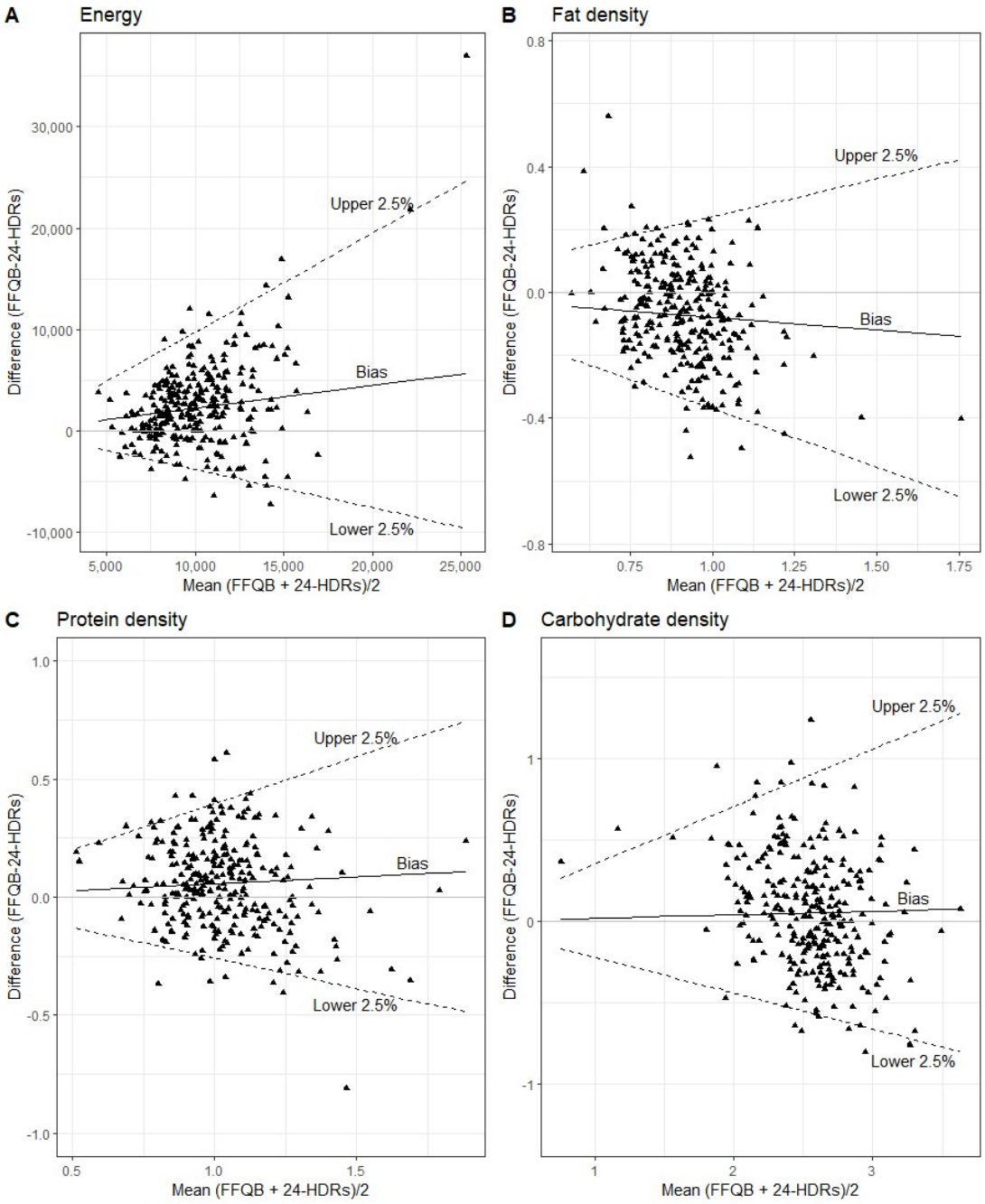

Figure 5. Bland–Altman plots comparing the food frequency questionnaire at baseline (FFQB) with the 24-hour dietary recalls (24-HDRs) for (A) energy (kJ); (B) fat density (g/kJ); (C) protein density (g/kJ); (D) carbohydrate density (g/kJ) (back-transformed). The solid line illustrates the median difference, and the dotted lines illustrate the upper and lower 2.5% percentiles. Figure adapted from Paper I.

Higher absolute median intakes of total energy, nutrients and foods groups were reported with the $FFQ_{baseline}$ compared with the $FFQ_{12months.}$ Significant differences were shown for total energy, some nutrient densities and the majority of food groups and again no significant differences were shown for nutrient residuals. The correlation coefficients were in general high. For total energy the coefficient was 0.67 (95% CI:0.61,0.73) and for nutrients densities and nutrient residuals it ranged from 0.52-0.88. Classification of participants' intake into the same or adjacent quartile was 88% for total energy and ranged from 80%-93% for nutrient densities and nutrient residuals. Furthermore, food group correlation coefficients were likewise high, ranging from 0.60-0.88 and classification of participants' intake into the same or adjacent quartile ranged from 82%-96% (**Table 11**). A reduction in dietary intake between first and second administration of an FFQ has been shown previously[301-304,309,311,314] and suggested to be due to a learning effect[301,309,314] or possibly fatigue. The length of time between the first and second administration may affect the results, since a short time interval may lead to higher correlations or agreement due to recall of previous reported intakes, where a long interval may result in low correlations or agreement also due to real changes in dietary intakes[315]. In summary, the $FFQ_{baseline}$ was found to be acceptable in ranking of participants' intake of total energy and energy-adjusted nutrients in comparison with the 24-HDRs. Moreover, the $FFQ_{baseline}$ showed satisfactory ranking of total energy, energy-adjusted nutrients and food group intakes in comparison with the $FFQ_{12months.}$

Table 11. Food group intakes from the FFQbaseline compared with the FFQ12 months for median bias, Spearman's correlation coefficient, Bland–Altman LOAs and cross-classification (n=415).

Food Groups [a]	Median Bias[b]	*p*-Value[d]	Correlation Coefficient[g]	Bland–Altman LOA[e]		Cross-Classification (%)			
			(95% CI)	Lower	Upper	Same Quartile	Adjacent Quartile	Opposite Quartile	Extreme Opposite Quartile
Fruits, g	−12.26	<0.0001	0.74 (0.68–0.79)	−66	123	51	41	6	1
Vegetables, g	−8.79	<0.0001	0.73 (0.67–0.77)	−54	95	52	39	9	1
Potatoes, g	−11.11	0.0003	0.69 (0.62–0.74)	−68	132	51	39	9	1
Legumes, g	- [c]	0.8210	0.75 (0.70–0.80)	- [c]	- [c]	53	36	11	1
Whole grains, g[f]	−9.93	0.0004	0.62 (0.55–0.69)	−74	228	46	40	11	3
Eggs, g	−10.44	<0.0001	0.61 (0.53–0.67)	−72	147	48	36	13	2
Poultry, g	- [c]	<0.0001	0.60 (0.52–0.66)	- [c]	- [c]	46	36	16	2
Red meat, g	- [c]	<0.0001	0.76 (0.70–0.81)	- [c]	- [c]	56	36	7	1
Processed red meat, g	- [c]	<0.0001	0.78 (0.73–0.82)	- [c]	- [c]	57	35	7	0
Fast food, g	- [c]	0.1363	0.75 (0.69–0.80)	- [c]	- [c]	55	37	7	1
Fish and seafood, g	- [c]	<0.0001	0.75 (0.70–0.80)	- [c]	- [c]	53	38	7	1
Dairy products, g	−7.4	0.0005	0.76 (0.70–0.80)	−83	281	57	35	7	1
Fermented dairy products, g	−12.36	<0.0001	0.72 (0.66–0.77)	−78	204	54	37	8	1
Fat products, g	−8.05	<0.0001	0.64 (0.57–0.70)	−67	81	45	42	11	2
Soft drinks, g	- [c]	0.0907	0.80 (0.75–0.84)	- [c]	- [c]	63	31	5	2
Coffee, g	- [c]	0.0025	0.78 (0.73–0.83)	- [c]	- [c]	60	33	7	0
Tea, g	- [c]	<0.0001	0.88 (0.84–0.90)	- [c]	- [c]	69	27	3	0

Food frequency questionnaire (FFQ). [a]Based on log-transformed intakes. [b]Median bias is reported as a percentage. [c]Percentage median bias and LOAs are not reported due to zero-intake. [d]p-value, the test of difference in intake between (log^a) $FFQ_{baseline}$ and (log^a) $FFQ_{12\ months}$ using the Wilcoxon signed-rank test. [e]Bland–Altman limits of agreement (LOAs) are reported as a percentage difference. [f]Estimated whole-grain intake from whole-grain products. [g]Spearman's correlation coefficient and 95% confidence intervals (CI). Table adapted from paper I.

5.2 Validity of the dietary quality score and associated cardiometabolic risk factors

The correlation coefficient of major food groups (fish, red meat, vegetables and fruits) between the 23-item FFQ and the 376-item FFQ ranged from 0.31-0.61. A higher DQS associated with a higher intake of fruits, vegetables, fish, fibre, vitamin and minerals and a lower intake of saturated fat. The correlation coefficient between the DQS and dietary intake from the FFQ ranged from 0.10-0.46 (**Table 12**).

Table 12. Intake of energy, nutrients and daily foods based on the 376-item FFQ by DQS category.

Energy, nutrients and foods	Healthy dietary habits (7-9 points) Median (P25, P75)	Average dietary habits (4-6 points) Median (P25, P75)	Unhealthy dietary habits (1-3 points) Median (P25, P75)	P-values for trend	Spearman´s correlation coefficients[a]
N	103	310	37		
Total energy (kJ)	9168 (8759, 11387)	7698 (6075, 9794)	6812 (5565, 8912)	<0.0001	0.25
Fibre (g)	31 (24, 36)	22 (17, 29)	16 (12, 23)	<0.0001	0.39
Saturated fat (E %)	9 (8, 11)	10 (9, 12)	11 (9, 12)	<0.0001	-0.25
Unsaturated fat (E %)	17 (14, 19)	16 (14, 18)	15 (14, 17)	0.0343	0.10
Vitamin B6 (mg)	2 (1.4, 2.0)	1 (1.1, 1.7)	1 (0.9, 1.4)	<0.0001	0.37
Vitamin B12 (mg)	6 (5, 8)	5 (4, 7)	5 (3, 7)	<0.0001	0.24
Vitamin E (mg)	10 (8, 13)	7 (5, 10)	5 (4, 7)	<0.0001	0.37
Vitamin C (mg)	122 (89, 153)	77 (60, 116)	53 (41, 76)	<0.0001	0.40
Vitamin D (mg)	3 (3, 4)	3 (2, 3)	2 (1, 2)	<0.0001	0.39
Vitamin K (mg)	135 (81, 183)	76 (54, 110)	44 (29, 68)	<0.0001	0.45
Calcium (mg)	1411 (1127, 1698)	1203 (841, 1671)	1093 (698, 1633)	<0.0001	0.19
Magnesium (mg)	512 (410, 598)	396 (313, 530)	326 (262, 413)	<0.0001	0.33
Selenium (µg)	55 (43, 67)	43 (33, 57)	36 (27, 44)	<0.0001	0.36
Iron (mg)	13 (11, 16)	10 (8, 13)	8 (6, 10)	<0.0001	0.39
Fruits (g)	83 (17, 227)	36 (11, 87)	9 (5, 18)	<0.0001	0.26
Vegetables (g)	187 (137, 268)	116 (82, 161)	69 (41, 104)	<0.0001	0.45
Fish (g)	34 (23, 44)	20 (10, 30)	9 (3, 15)	<0.0001	0.46

Food frequency questionnaire (FFQ), dietary quality score (DQS), [a]Spearman's correlation coefficient analyses are made using the 9-classed score. Table adapted from paper II.

Correlation coefficients were similar to those of the first DQS validation by Toft et al. which ranged from 0.05 to 0.55. However, caution should be made when comparing correlation coefficients, since these estimates are dependent of the population under study. Other studies have also reported satisfactory assessment of dietary quality based on short FFQs or screeners[316-318]. In fact, some studies also reported similar results between scores or indexes in relation to dietary quality whether calculated from a short or a long-FFQ[319-321], although it is important to highlight that short FFQs or screeners are not suitable for estimating habitual intake with high resolution or for estimating total energy intake.

Furthermore, a higher DQS was also associated with cardiometabolic risk factors including lower levels of absolute fat mass (AFM), RFM, VF, WC, LDL-C and hs-CRP as well as higher levels of HDL-C (**Table 13 and 14**). Studies investigating associations between adherence to other dietary quality scores (based on short FFQs) and CMD factors have shown inconclusive results: some have shown associations between a higher score and lower levels of BMI and WC

as well as lower odds having overweight, obese, having hypertension and high TG[318,322] whereas others have not found associations between a higher score and CMD risk factors[316,317]. The reason for this discrepancy between studies, may be due to population differences including health status, covariates and actual levels of the consumed foods but also the fact that the dietary quality scores or indexes investigated in the different studies varied substantially in terms of included food groups. Lastly, only few studies have assessed the association between dietary quality scores based on short FFQs and CMD risk factors, which makes a general conclusion difficult.

Table 13. Associations between the dietary quality score and BP and anthropometric risk factors for cardiometabolic diseases.

	Healthy dietary habits[e]	**Average dietary habits[e]**	**Unhealthy dietary habits[e]**	**P-values for trend[c]**
	(7-9 points)	(4-6 points)	(1-3 points)	
n	103	310	37	
SBP (mm Hg)				
Mean (s.d.)	123.0 (15.9)	123.4 (17.8)	128.0 (15.5)	
Simpel model[a], ß (95% CI)	0	-0.28 (-3.59-3.04)	4.10 (-1.51-9.71)	0.7514
Multivariate model[b], ß (95% CI)	0	-0.14 (-3.52-3.24)	3.34 (-2.57-9.25)	0.9566
DBP (mm Hg)				
Mean (s.d.)	81.8 (9.9)	81.9 (11.0)	86.2 (10.3)	
Simpel model[a], ß (95% CI)	0	-0.10 (-2.30-2.10)	4.50 (0.78-8.23)	0.1765
Multivariate model[b], ß (95% CI)	0	-0.14 (-2.37-2.10)	4.13 (0.22-8.03)	0.2852
WC (cm)				
Mean (s.d.)	85. 6 (11.7)	88.6 (12.2)	93.8 (18.1)	
Simpel model[a], ß (95% CI)	0	2.49 (0.12-4.85)	7.53 (3.52-11.53)	0.0007
Multivariate model[b], ß (95% CI)	0	2.00 (-0.33-4.33)	5.27 (1.19-9.35)	0.0161
BMI(kg/m2)				
Mean (s.d.)	24.5 (3.9)	25.1 (4.0)	26.8 (6.2)	
Simpel model[a], ß (95% CI)	0	0.53 (-0.36-1.42)	2.32 (0.81-3.82)	0.0171
Multivariate model[b], ß (95% CI)	0	0.35 (-0.51-0.21)	1.31 (-0.19-2.82)	0.1856
VF (liter)[d]				
Mean (s.d.)	1.42 (1.36)	1.84 (1.57)	2.64 (2.42)	
Simpel model[a], ß (95% CI)	0	0.34 (0.05-0.63)	1.24 (0.63-1.62)	<0.0001
Multivariate model[b], ß (95% CI)	0	0.31 (0.03-0.60)	0.81 (0.31-1.30)	0.0003
AFM (kg)				
Mean (s.d.)	20.4 (9.6)	22.4 (9.4)	25.5 (13.3)	
Simpel model[a], ß (95% CI)	0	2.18 (0.16-4.20)	6.39 (2.97-9.81)	0.0006
Multivariate model[b], ß (95% CI)	0	1.73 (-0.24-3.69)	4.15 (0.71-7.58)	0.0106
RFM (%)				
Mean (s.d.)	27.4 (10.0)	29.0 (9.0)	29.5 (9.2)	
Simpel model[a], ß (95% CI)	0	2.06 (0.56-3.57)	4.70 (2.16-7.24)	0.0002
Multivariate model[b], ß (95% CI)	0	1.78 (0.28-3.27)	3.48 (0.86-6.10)	0.0030

Systolic blood pressure (SBP), diastolic blood pressure (DBP), waist circumference (WC), body mass index (BMI), visceral fat (VF), absolute fat mass (AFM), relative fat mass (RFM). [a]Simple linear regression models adjusted for sex and age. [b]Multivariate linear regression models adjusted for sex, age, physical activity, smoking and education. [c]Made using the 9-classed score. [d]Six subjects excluded from the analyses due to missing values. [e]The dietary quality score categories are based on the 23-item FFQ. Table adapted from paper II.

Table 14. Associations between the dietary quality score and biological risk factors for cardiometabolic diseases.

	Healthy dietary habits[e]	Average dietary habits[e]	Unhealthy dietary habits[e]	P-values for trend[d]
	(7-9 points)	(4-6 points)	(1-3 points)	
n	103	310	37	
CHO (mmol/l)				
Mean (s.d.)	4.99 (1.01)	5.08 (0.95)	5.25 (1.12)	
Simpel model[a], ß (95% CI)	0	0.08 (-0.11-0.28)	0.35 (0.01-0.68)	0.0279
Multivariate model[b], ß (95% CI)	0	0.06 (-0.14-0.27)	0.28 (-0.08-0.63)	0.0565
LDL-C (mmol/l)				
Mean (s.d.)	3.06 (0.92)	3.18 (0.87)	3.42 (1.16)	
Simpel model[a], ß (95% CI)	0	0.11 (-0.08-0.31)	0.39 (0.07-0.71)	0.0104
Multivariate model[b], ß (95% CI)	0	0.11 (-0.09-0.30)	0.35 (0.01-0.69)	0.0133
HDL-C (mmol/l)				
Mean (s.d.)	1.70 (0.45)	1.60 (0.43)	1.50 (0.39)	
Simpel model[a], ß (95% CI)	0	-0.08 (-0.17-0.01)	-0.12 (-0.27-0.03)	0.0231
Multivariate model[b], ß (95% CI)	0	-0-08 (-0.17-0.01)	-0.11 (-0.27-0.05)	0.0379
TG[d] (log mmol/l)				
Mean (s.d.)	0.07 (0.52)	0.18 (0.52)	0.19 (0.52)	
Simpel model[a], ß (95% CI)	0	0.09 (-0.02-0.20)c	0.09 (-0.09-0.28)c	0.0773
Multivariate model[b], ß (95% CI)	0	0.08 (-0.03-0.19)c	0.04 (-0.16-0.23)c	0.2194
HbA1c (mmol/mol)				
Mean (s.d.)	34.1 (4.0)	34.2 (4.8)	35.4 (6.3)	
Simpel model[a], ß (95% CI)	0	0.06 (-0.92-1.03)	1.51 (-0.14-3.17)	0.1227
Multivariate model[b], ß (95% CI)	0	-0.18 (-1.16-0.81)	0.92 (-0.81-2.64)	0.4762
Hs-CRP (mg/L)				
Mean (s.d.)	0.92 (1.01)	1.60 (2.61)	1.70 (1.85)	
Simpel model[a], ß (95% CI)	0	0.70 (0.19-1.20)	0.93 (0.08-1.79)	0.0051
Multivariate model[b], ß (95% CI)	0	0.63 (0.12-1.14)	0.63 (-0.26-1.52)	0.0449

Total cholesterol (CHO), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), glycated hemoglobin (HbA1c), high sensitivity C-reactive protein (hs-CRP). [a]Simple linear regression models adjusted for sex and age. [b]Multivariate linear regression models adjusted for sex, age, physical activity, smoking and education. [c]Data are natural logarithmically transformed. [d]Made using the 9-classed score. [e]The dietary quality score categories are based on the 23-item FFQ. Table adapted from paper II.

In summary, based on our findings, the DQS appears to be a valid tool to assess dietary quality and it showed clear associations with CMD risk factors and is therefore suitable for estimating dietary quality in large populations, for use in clinical settings or in studies where diet is not the main interest.

5.3 Gut microbiota variability and associations with different dietary patterns

Out of the 234 genera, 91 (39%) had an ICC>0.5 i.e., higher between than within-individual variability. For 100 genera (43%) the ICC<0.5, indicated higher within individual variability, whereas for 43 genera (18%) it was not possible to calculate an ICC. This was due to problems with convergence of the models that were used for estimation of the ICC for some of the genera. Few studies have measured the ICC of gut microbes as mentioned in section 3.6.1 and comparisons are difficult due to differences in the taxonomic levels investigated and the populations under study. One study, which comprised three different populations reported low ICCs at the phylum level[17]. Another study reported strains with ICC>0.4 as having good reproducibility, however without reporting the actual ICC values, which again limits comparison[16]. A recent study from Sweden estimated the ICC based on 4 samples per person in 75 healthy individuals and showed good reproducibility at family (66% ICC>0.5), genus (80% ICC>0.5) and strain-levels (75% ICC>0.5)[18]. The reason for a higher microbe stability in this study could be due to differences in preprocessing, in particular having a higher prevalence threshold. Also, different ICC versions exist, with different interpretations. However, this information was not available for any of the mentioned studies. A core microbiota of 40 genera present in at least 95% of the samples was also found. Interestingly, 12 out of the 40 genera had an ICC<0.5, which means that even though a microbe is highly prevalent it does not necessarily mean that it is also stable. Overall, from our study a certain proportion of the gut microbiota at the genus level was indicated as stable and these microbes may be used in investigations of diet and gut microbiota in relation to disease outcome. Further, the association between those gut microbial genera with an ICC>0.5 and plant-based dietary patterns i.e., the HNFI, rMED, PDI, hPDI, uPDI and pro-veg were explored. Three gut microbiota community subgroups were observed, in turn enriched by *Bacteroides*, *Prevotella*-9 and *Ruminococcaceae* (**Figure 6A, B**) corresponding to previously identified enterotypes reported[22,31,216,245]. As anticipated, the *Prevotella*-9 and *Ruminococcaceae-groups* were associated with higher adherence to the healthy plant-based dietary patterns (HNFI, rMED, PDI, hPDI, pro-veg) compared to the *Bacteroides* group (**Figure 6C**).

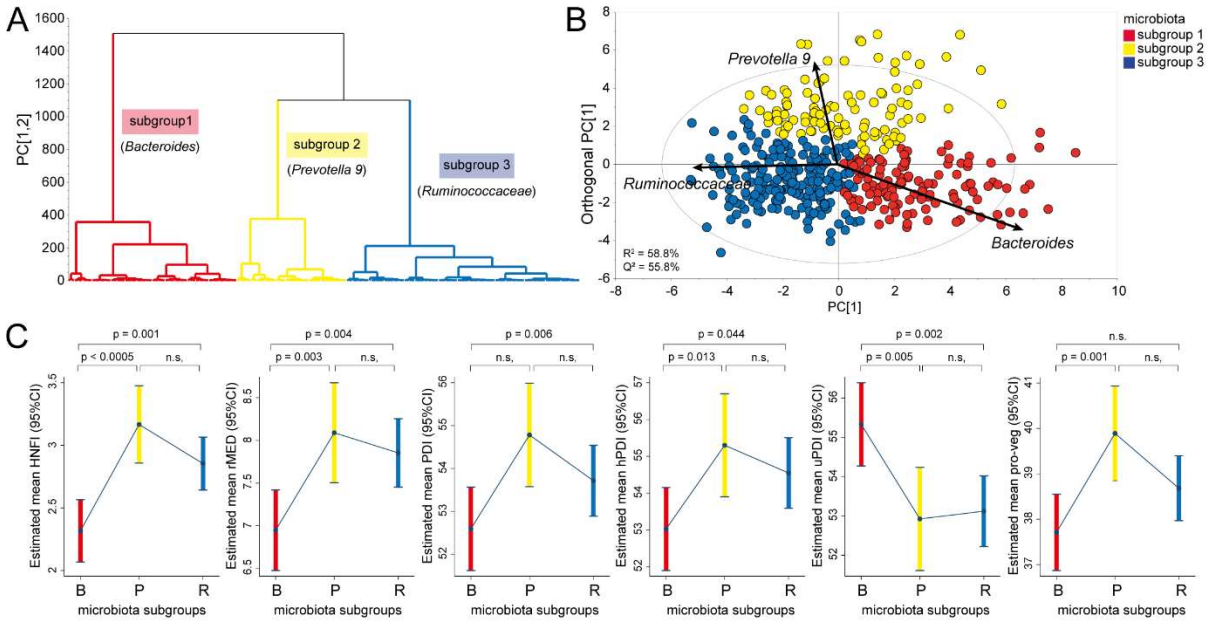

Figure 6. Diet indexes associated with microbiota community subgroups. Gut microbiota were log transformed, centered and scaled to unit variance by genus. **A**: Hierarchical clustering of gut microbiota at the genus level defined three subgroups. **B**: OPLS-DA model separated the three subgroups dominated by *Bacteroides* (subgroup 1), *Prevotella* 9 (subgroup 2), and *Ruminococcaceae* (subgroup 3) (R2 = 58.8%, Q2 = 55.8%, p < 0.0001). **C**: Predicted Healthy Nordic Food Index (HNFI), relative Mediterranean diet score (rMED), Plant-based Diet Index (PDI), healthy Plant-based Diet Index (hPDI), unhealthy Plant-based Diet Index (uPDI), and pro-vegetarian index (pro-veg) for each bacterial community subgroup B (*Bacteroides*-group), P (*Prevotella*-9-group), and R (*Ruminococcaceae*-group). Estimated mean values were obtained from linear models adjusted for age and sex. A p-value below 0.05 was considered significant.

Additionally, higher dietary index scores were significantly associated with 22 genera (**Figure 7**). As expected, for the uPDI, the direction of the association was reversed compared to the other dietary indexes. Several of those genera showing a higher abundance with higher diet index score have been reported to have fiber fermenting properties such as *Ruminoccoccae* and *Coprococcus*[323]. These results are in line with other studies. A higher abundance of *Ruminococcaceae* spp., *Coprococcus* spp., *Ruminococcus* spp. have previously been associated with higher adherence to the Mediterranean diet and hPDI. In addition, higher abundance of *Bacteroides* spp. and *Parabacteroides* were shown to be associated with omnivores diet as well as low adherence to the Mediterranean diet[20,24,25]. In contrast to the literature, where higher adherence to the Mediterranean diet has been associated with higher abundance of *Flavinofractor* spp., *Oscillibacter* spp. and *Erysipelatoclostridium* spp.[25], we found a lower abundance though at the genus level. Furthermore, the association between all individual food components of each diet index and the 22 genera was also explored. Fruits, vegetables, whole grains/cereals, and nuts were most strongly and consistently associated with the 22 genera (**Figure 8**).

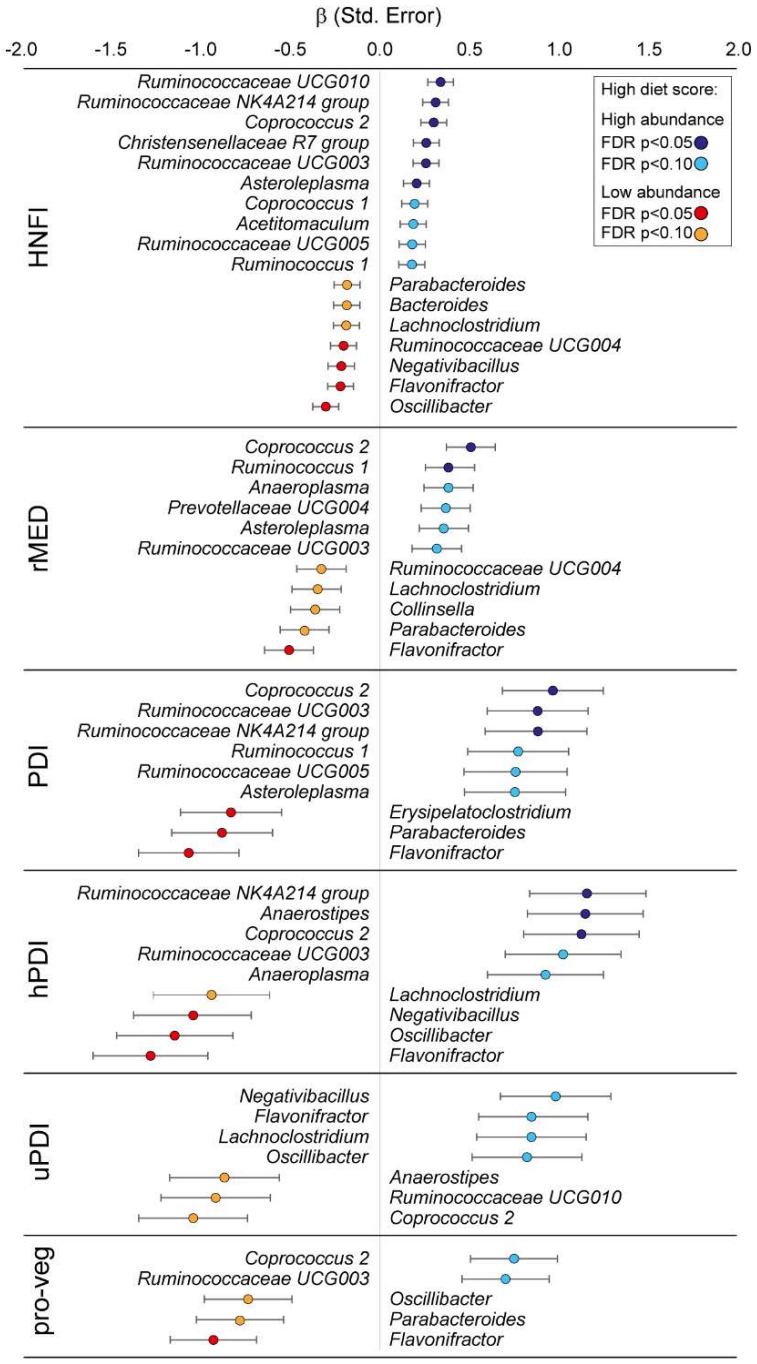

Figure 7. Diet indexes associated with microbiota at the genus level, examined by linear regression (for n=91 genera log-transformed and scaled to unit variance; mean ± SD, 0.0 ± 1.0). All models were adjusted for sex and age. Presented are beta-coefficients (β) and Standard error of the mean (Std. Error) for the genera for each diet index: Healthy Nordic Food Index (HNFI), relative Mediterranean diet score (rMED), Plant-based Diet Index (PDI), healthy Plant-based Diet Index (hPDI), unhealthy Plant-based Diet Index (uPDI), and pro-vegetarian index (pro-veg). Blue bars indicate a higher relative abundance, and red bars indicate a lower relative abundance of the indicated genus associated with a higher diet index score. The dark-colored bars indicate those genera with an false discovery rate (FDR) adjusted p-value <0.05 and light-colored bars indicate those genera with an FDR adjusted p-value <0.1.

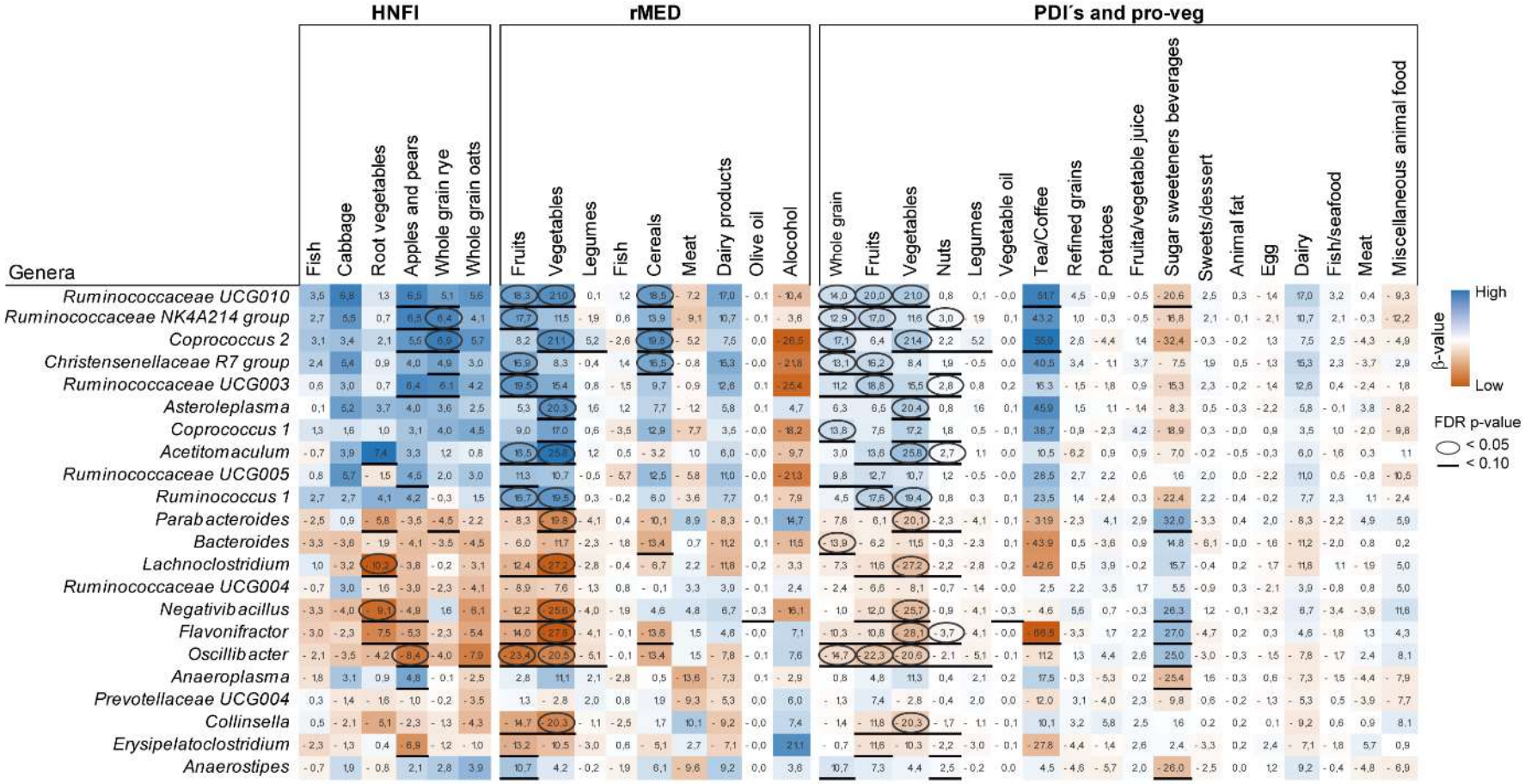

Figure 8. Associations between individual food groups and diet index-associated genera (n = 22). Associations between food groups g/d (for each diet index) and log-transformed genera scaled to unit variance were evaluated with sex- and age-adjusted linear regression analyses. Presented are beta-coefficients (β). Reported intake of each food group associated with a significantly higher (blue) or lower (brown) relative abundance of a genus is marked. Genera with false discovery rate (FDR) <0.1 and 0.05 were marked with a black line or black circle, respectively.

In summary, 39% of the gut microbiota was found to have an ICC>0.5. Gut microbial subgroups were also identified, consistent with already reported enterotypes i.e., *Bacteroides*, *Prevotella* and *Ruminococcus*. As anticipated, the *Prevotella*-9 and *Ruminococcaceae*-groups were associated with higher adherence to the healthy plant-based indexes (HNFI, rMED, PDI, hPDI, pro-veg) compared to the *Bacteroides*-group. Additionally, higher adherence to the healthy plant-based indexes were found to be associated with the relative abundance of 22 genera, for instance higher relative abundance of fiber-fermenting genera. These 22 genera also associated most strongly with intakes of vegetables, fruit, cereal/whole grains and nuts.

5.4 Direct and indirect effects of healthy Nordic and Mediterranean diet patterns, mediated by gut microbiota, on cardiometabolic risk factors

The final models with the healthy Nordic diet are shown in **Figure 9** and **Figure 10** with standardized estimates for women and men. Models for the Mediterranean diet in women and men can be found in Supplementary Materials in paper IV. No indirect effect of the healthy Nordic or Mediterranean diets on risk factors for CMD, mediated by the P/B ratio in women or men were found (**Table 15**). In addition, sensitivity analysis showed no significant differences between the effect of diets on adiposity in those having a high P/B ratio compared to those having a low P/B ratio (see supplementary materials in paper IV).

A few intervention studies have investigated the interaction between diet and gut microbiota on health. An observational study of 300 men where carriers of *Prevotella copri* had higher risk of myocardial infarction and that non-carriers with high adherence to the Mediterranean diet had a lower risk of myocardial infarction[29]. Moreover, three small intervention studies have reported differential effects of the P/B ratio on glucose metabolism and weight loss. In the study by Christensen et al., healthy individuals with overweight and with high P/B ratio lost more weight than individuals with low P/B ratio consuming a diet high in fibre and/or rich in whole grains[12]. In the study by Hjort et al., individuals with overweight or obese and with high P/B ratio lost more weight and body fat compared to individuals with low P/B ratio, independently of diet[13]. In addition, in another small intervention study, responders to a breakfast of barley kernel bread (improvements in blood glucose and insulin levels) were shown to have a higher P/B ratio compared to non-responders[14]. Furthermore, transplantation of fecal material to mice, also showed lower blood glucose levels, higher *Prevotella* in mice colonised with the microbiota from responders compared to non-responders[14].

Importantly, the current study is a cross-sectional study including 439 individuals with self-reported diets based on a limited number of 24-HDRs and gut microbiota data based on 16S rRNA gene sequencing, which is less sensitive and specific than quantitative polymerase chain reaction (qPCR) (used in some of the studies). Also, we used averages of the exposure, mediator and outcomes in the model, which may disturb the time order and probably also the effect of diet on CMD risk factors, mediated via the P/B ratio. Moreover, with the SEM, it is likely that this complex multidimentional problem has been oversimplified.

Expectedly, an association was found between adherence to the healthy Nordic and Mediterranean diets and lower levels of adiposity in women. For men this association was only observed for the healthy Nordic diet (**Table 15**). These results are in comparison with other studies investigating the association between a high adherence to the Mediterranean diet or a Nordic diet and obesity as well as other CMD risk factors, with a majority of studies investigating the Mediterranean diet. Overall, a high adherence to the Mediterranean diet or a Nordic diet were associated with lower risk of overweight and obesity or BMI[9,324,325]. Furthermore, in the current study adherence to the healthy Nordic and Mediterranean diets were also associated with lower levels of lipidemia and hs-CRP, this effect was partly indirect mediated by adiposity in women. Again, this was only observed in men with adherence to the healthy Nordic diet. (**Table 15**). Both elevated levels of blood lipids and CRP have previously been associated with obesity[59,326].

In summary, based on these findings the effect of a healthy Nordic or Mediterranean diet on CMD risk factors does not seem to be mediated through the P/B ratio. Though a these diets have an effect on CMD risk factors, some were mediated by adiposity.

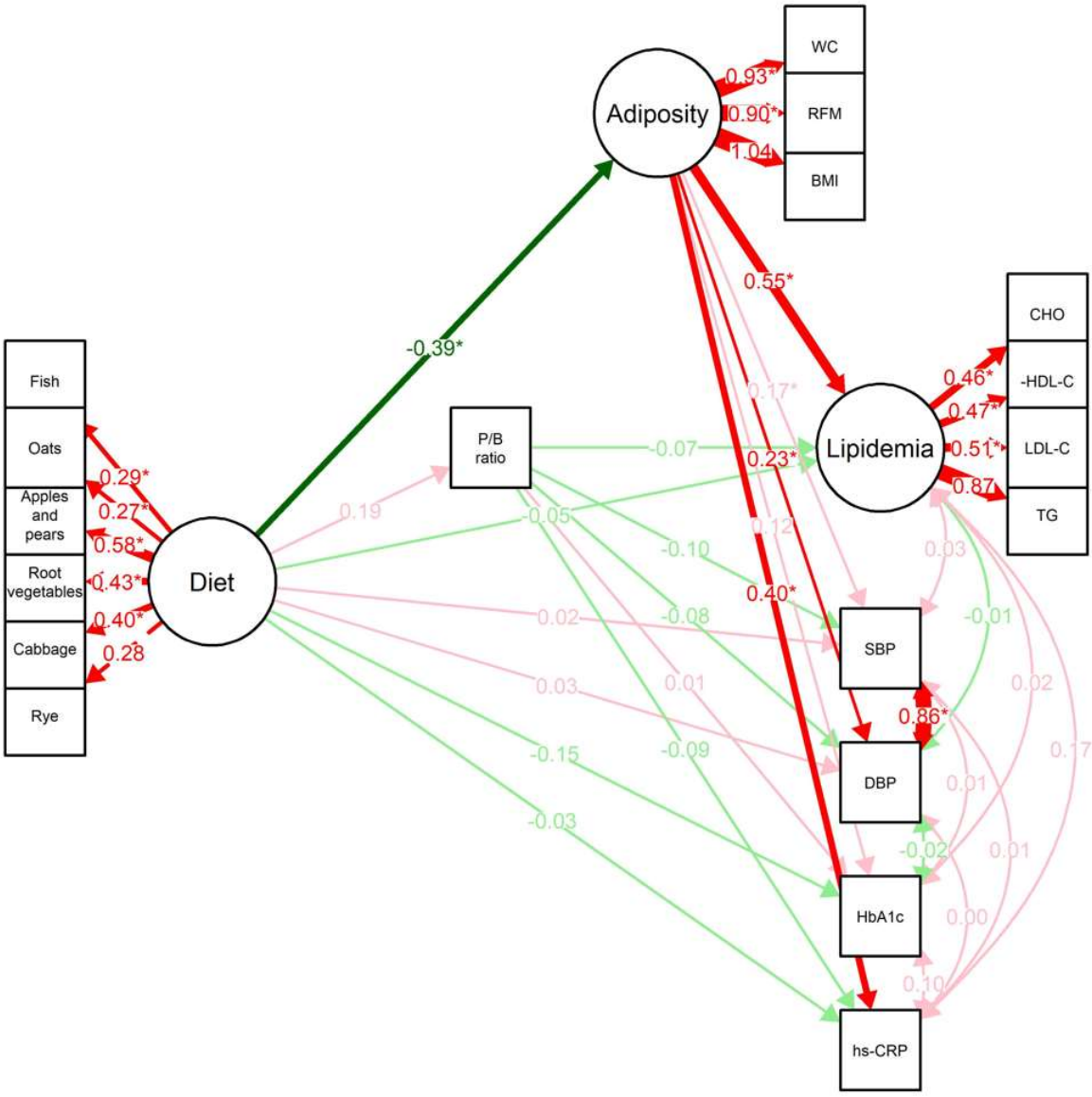

Figure 9. The final model for women with standardized estimates of the relationship between the healthy Nordic diet, *Prevotella*-to-*Bacteroides* (P/B) ratio, and risk factors for cardiometabolic diseases (CMD), adjusted for age, smoking habits, physical activity, alcohol, and meat intakes. The goodness of fit indices: X2 = 379, df = 199, p-value <0.001, RMSEA = 0.065 (0.065, 0.075), CFI = 0.940, SRMR = 0.076. The color and thickness of the arrows indicate the sign and effect size of the estimate: green color indicates a reduction, red color indicates an increase, the thicker the arrow, the larger the effect. An (*) signifies statistical significance (p-value <0.05). Adiposity is defined by body mass index (BMI), waist circumference (WC) and relative fat mass (RFM); lipidemia is defined by total cholesterol (CHO), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG); systolic blood pressure (SBP), diastolic blood pressure (DBP), glycated hemoglobin (HbA1c), high sensitivity C-reactive protein (hs-CRP).

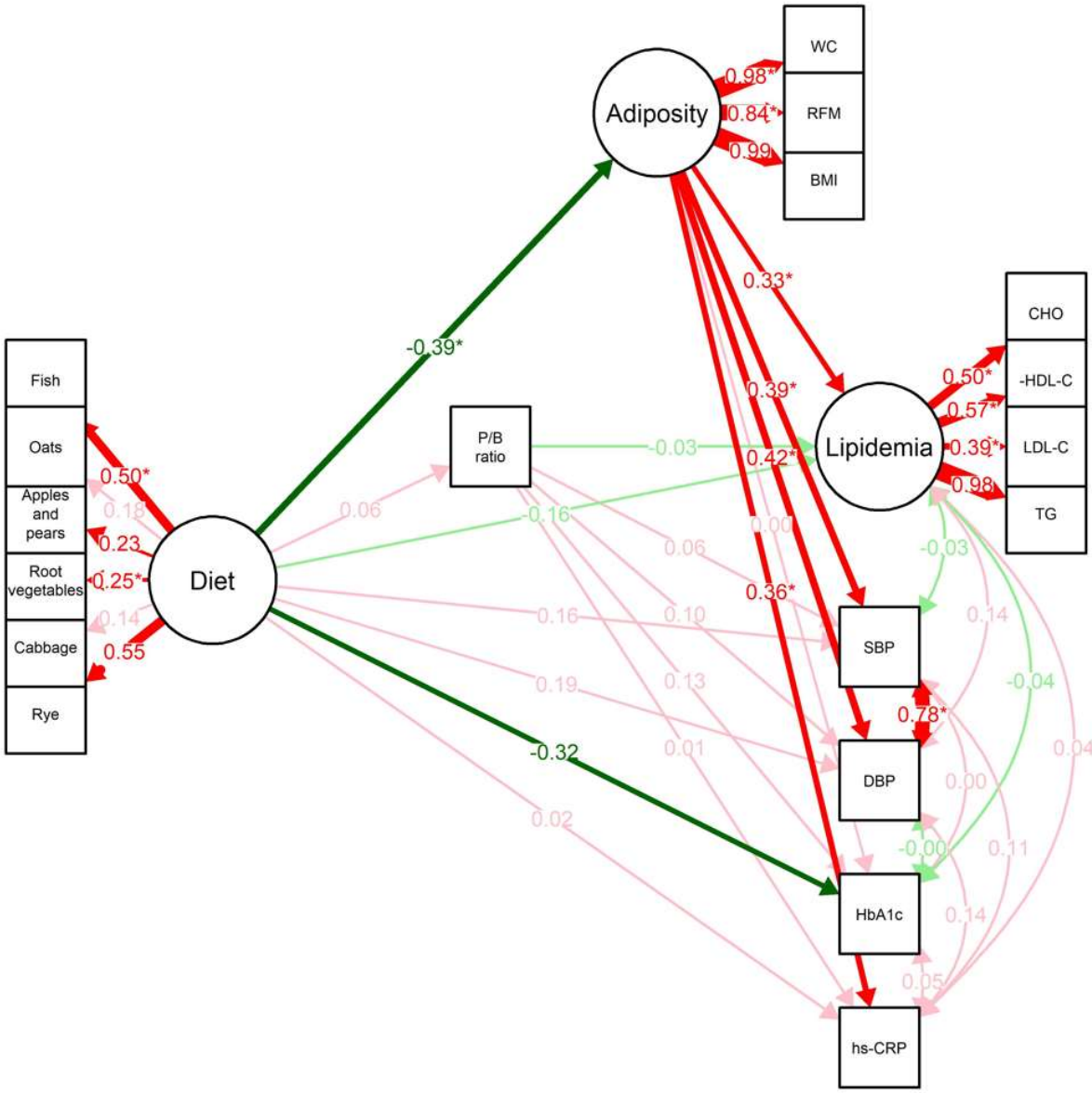

Figure 10. The final model for men with standardized estimates of the relationship between the healthy Nordic diet, *Prevotella*-to-*Bacteroides* (P/B) ratio, and risk factors for cardiometabolic diseases (CMD), adjusted for age, smoking habits, physical activity, alcohol, and meat intakes. The goodness of fit indices: X2 = 343, df = 199, p-value <0.001, RMSEA = 0.060 (0.049, 0.070), CFI = 0.931, SRMR = 0.071. The color and thickness of the arrows indicate the sign and effect size of the estimate: green color indicates a reduction, red color indicates an increase, the thicker the arrow, the larger the effect. An (*) signifies statistical significance (p-value <0.05). Adiposity is defined by body mass index (BMI), waist circumference (WC) and relative fat mass (RFM); lipidemia is defined by total cholesterol (CHO), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG); systolic blood pressure (SBP), diastolic blood pressure (DBP), glycated hemoglobin (HbA1c), high sensitivity C-reactive protein (hs-CRP).

Table 15. Unstandardized estimates and 95% confidence intervals for women and men showing the direct, indirect, and total effect of the healthy Nordic and Mediterranean diets on risk factors for CMD. The models were adjusted for age, smoking habits, and physical activity. In addition, the model with healthy Nordic diet was adjusted for alcohol and meat intake.

From	To	Direct effect	Indirect effect via P/B ratio	Indirect effect via adiposity	Total effect
Women					
Healthy Nordic diet	Adiposity	**-5.3 (-9.4, -1.0)**	-0.0 (-0.3, 0.3)		**-5.3 (-9.4, -1.0)**
	Lipidemia	-1.6 (-8.3, 5.6)	-0.4 (-1.3, 0.5)	**-6.7 (-12.2, -0.9)**	-8.6 (-16.5, 0.1)
	SBP	0.3 (-2.2, 2.8)	-0.2 (-0.5, 0.1)	-0.7 (-1.6, 0.2)	-0.7 (-2.9, 1.6)
	DBP	0.3 (-2.0, 2.7)	-0.2 (-0.5, 0.1)	-0.9 (-2.0, 0.1)	-0.8 (-2.9, 1.4)
	HbA1c	-1.7 (-4.4, 1.2)	0.0 (-0.2, 0.3)	-0.5 (-1.2, 0.2)	-2.1 (-4.7, 0.5)
	Hs-CRP	-2.8 (-22.3, 21.4)	-1.7 (-4.7, 1.4)	**-14.4 (-25.6, -1.5)**	-18.2 (-36.4, 5.1)
Men					
Healthy Nordic diet	Adiposity	**-2.3 (-4.0, -0.5)**	0.1 (-0.2, 0.4)		**-2.1 (-3.7, -0.5)**
	Lipidemia	-3.2 (-8.1, 2.1)	-0.0 (-0.2, 0.1)	**-2.5 (-4.5, -0.5)**	**-5.6 (-10.8, -0.2)**
	SBP	0.6 (-0.4, 1.6)	0.0 (-0.0, 0.1)	**-0.6 (-1.1, -0.03)**	0.0 (-0.8, 1.0)
	DBP	0.8 (-0.3, 1.9)	0.0 (-0.1, 0.1)	**-0.7 (-1.3, -0.03)**	0.1 (-0.8, 1.1)
	HbA1c	-1.8 (-3.5, 0.0)	0.0 (-0.1, 0.2)	-0.0 (-0.4, 0.4)	**-1.7 (-3.2, -0.2)**
	Hs-CRP	0.8 (-7.4, 9.8)	0.0 (-0.3, 0.3)	**-4.8 (-9.1, -0.3)**	-4.0 (-11.2, 3.8)
Women					
Mediterranean diet	Adiposity	**-6.2 (-10.3, -1.8)**	-0.0 (-0.2, 0.1)		**-6.2 (-10.3, -1.8)**
	Lipidemia	-3.6 (-10.3, 3.6)	-0.2 (-0.7, 0.3)	**-7.6 (-13.2, -1.6)**	**-11.0 (-19.6, -1.5)**
	SBP	-2.2 (-4.9, 0.6)	-0.1 (-0.3, 0.1)	-0.6 (-1.5, 0.3)	**-2.8 (-5.4, -0.1)**
	DBP	-2.0 (-4.5, 0.6)	-0.1 (-0.2, 0.1)	-0.9 (-1.8, 0.1)	**-2.9 (-5.5, -0.2)**
	HbA1c	-1.6 (-4.0, 0.9)	0.0 (-0.1, 0.1)	-0.7 (-1.6, 0.2)	-2.2 (-4.6, 0.2)
	Hs-CRP	-17.9 (-36.2, 5.7)	-0.7 (-2.6, 1.2)	**-15.0 (-25.8, -2.5)**	**-30.7 (-49.4, -5.1)**
Men					
Mediterranean diet	Adiposity	-7.5 (-16.4, 2.3)	0.1 (-0.4, 0.7)		-7.4 (-16.2, 2.3)
	Lipidemia	-11.1 (-27.8, 9.4)	0.1 (-0.6, 0.7)	-8.8 (-19.3, 3.0)	-18.9 (-39.0, 7.7)
	SBP	-1.4 (-5.1, 2.4)	-0.0 (-0.3, 0.3)	-1.4 (-3.4, 0.6)	-2.9 (-7.2, 1.7)
	DBP	-0.7 (-4.0, 2.7)	-0.0 (-0.5, 0.4)	-1.8 (-4.1, 0.6)	-2.5 (-6.5, 1.6)
	HbA1c	-5.1 (-11.8, 2.1)	-0.0 (-0.5, 0.4)	-0.2 (-1.7, 1.2)	-5.3 (-12.1, 1.9)
	Hs-CRP	6.4 (-20.6, 42.6)	0.0 (-0.4, 0.5)	-16.8 (-34.7, 5.9)	-11.5 (-34.6, 19.7)

Cardiometabolic diseases (CMD); adiposity is defined by body mass index (BMI), waist circumference (WC) and relative fat mass (RFM); lipidemia is defined by total cholesterol (CHO), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG); systolic blood pressure (SBP), diastolic blood pressure (DBP), glycated hemoglobin (HbA1c), high sensitivity C-reactive protein (hs-CRP), *Prevotella*-to-*Bacteroides* (P/B) ratio. A one-unit increase in the latent variable healthy Nordic diet (Mediterranean diet) corresponds to a one-unit increase in whole grain rye (cereal) on the square root scale. The effect of diet is represented as % change in the CMD risk factors. For the latent variables adiposity and lipidemia the estimate corresponds to % change in BMI and TG respectively. Numbers in bold indicate that the effect is statistically significant.

6. CONCLUDING DISCUSSION

In this thesis, observational studies from the Diet Cancer and Health – Next Generations (DCH-NG) cohort and the DCH-NG MAX sub-cohort were used to validate the DCH-NG food frequency questionnaire (FFQ) and the dietary quality score (DQS) as well as to assess the temporal gut microbiota variability in a Danish population. In addition, investigations of the association between different dietary patterns and the gut microbiota as well as their interplay in relation to CMD risk factors were explored.

It was found that the assessment of the relative validity and reproducibility of the web-based DCH-NG FFQ adds to the large body of evidence that the FFQ can rank individuals by their dietary intake and that the FFQ is suitable in epidemiological studies that investigate diet in relation to disease outcomes (Paper I). The DCH-NG FFQ is web-based and updated from two previous FFQs; the paper-based version from the DCH cohort and the web-based FFQ from the Danish National Birth Cohort[281,282]. The transformation of the paper-based FFQs into web-based FFQs have resulted in several improvements and advantages[327]. An increase in response rate and less missing food items have been reported in adults completing both a paper-based and web-based FFQ[328]. This is most likely due to the general use of technology by a great proportion of the adult population and possibly a preference towards the web-based format. However, this may be different in elderly. The participants also have the possibility to skip items that they do not consume and as already mentioned incorporation of error checks minimises missing data. A large upgrade to the back-end nutrient analysis has also been made with the web-based versions, since raw data is immediately available upon completion of the FFQ and the estimations of energy and nutrient intakes are thereby faster[193]. However, a study comparing a paper-based and web-based FFQ, have not found large differences in the validity of dietary intake[329]. Even though technology has made improvements with the web-based FFQ, the main errors of self-reported dietary assessment methods still prevail. It is likely that the response rate in the DCH-NG and DCH-NG MAX would be lower if questionnaires had to be filled out by pencil and mailed back. To overcome some of the challenges of self-reported dietary intake, calibration of dietary intakes with objective measures such as established recovery biomarkers or concentration biomarkers could be made[330]. The purpose of calibration is to calculate correction factors which can then be used to adjust relative risk estimates. So far, a limited number of biomarkers exist and still lacks for many food groups. Overall, technology has led to considerable improvements to the FFQ, though recall and social desirability bias still exist. It is hard to eliminate these errors, but further effort should be used to calibrate dietary intake from self-reported instruments with dietary biomarkers. Also, further development of combined dietary assessments methods with different sources of measurement errors is warranted.

The DQS, based on a short FFQ, was found to be a valid tool to assess dietary quality (Paper II). Since the DQS does not require complex nutrient analyses, but only a 23-item FFQ, it is eligible for monitoring the quality of diet in large populations. The DQS could also be used in clinical settings or in studies where diet is not the main focus. When dietary intake is assessed, it is always a compromise between the cost (both data collection and processing of dietary data),

level of detail and burden to the participants. Different dietary assessment methods provide dietary data with different levels of accuracy. Short FFQs or screeners for instance can provide enough information about dietary intake to estimate overall quality of diet with minor participant burden. The complex cognitive calculations which are present for more comprehensive FFQs are believed to be reduced to some extent. The limitation with such a short questionnaire is the lack of reliable assessment of total energy intake and intake of foods and beverages[174]. The short 23-item FFQ used to calculate the DQS in paper II is an example of such questionnaire[153]. The DQS is a simple tool with the purpose to evaluate quality of dietary habits and has been used as a tool to monitor the quality of dietary habits in the Danish population since 2010[331,332]. A central point is that the coverage of the target population is presumed to be higher with a short FFQ than if a more comprehensive dietary assessment method would be used. This is of utmost importance when surveillance of large populations is in focus. Notably, the purpose of short FFQs is not to estimate total dietary intake, but to provide information about the prevalence of overall quality of dietary habits which could be valuable information to for instance local and national authorities to initiate activities for health promotion.

A large proportion of the gut microbiota seem to have yet unexplored random variations. When investigating the role of specific gut microbes in relation to health outcomes, the probability of discovering underlying associations in population-base studies will be higher for bacteria with little within-subject variation over time. Study the variation and estimating its within- and between subject components in relation to each other will help to design subsequent endpoint studies with regards to sample size. Only a few studies on the temporal variability of the human gut microbiota are available in even fewer population groups[16-18], despite the importance of such studies to evaluate how well determination of a specific bacterial taxa in a single fecal sample would reflect the long-term level of that bacteria. Novel data on temporal variability of the gut microbiota at the genus level in a Danish population was provided in Paper III to complement the current scarce data available. It was found that a large proportion of microbes have substantial temporal variability. This will have implications for endpoint studies and the dimensioning of such studies. Bacteria that vary over time within an individual will be difficult to assess in relation to disease outcomes, because a single determination in a sample will not capture the underlying average presence of the bacterium which is related to disease risk. In addition, the use of 16S rRNA gene amplicon sequencing is a commonly used method for analysing the composition and relative abundance of gut microbes, including in the DCH-NG MAX study. However, the sensitivity of this technique is limited to identification at the genus level. Yet, species within the same genus have shown to exhibit distinct properties[296] and the use of more comprehensive assessment of the gut microbiota in terms of identification of species and strains as well as functional pathways are therefore further needed.

In paper III, evidence was provided that adherence to healthy plant-based patterns differs according to known gut microbial sub-group i.e., *Bacteroides*, *Prevotella*-9 and *Ruminococcaceae*-groups. It was also found that healthy plant-based patterns and food constituents such as fruits, vegetables, wholegrain/cereals and nuts were associated with the relative abundance of specific genera, in particular, a higher abundance of genera with fiber-

fermenting properties. This adds to the existing evidence that diet is associated with gut microbiota community subgroups and certain genera, particularly plant-based foods. However, our investigation provides no support of P/B ratio being of relevance for the associations between diet and cardiometabolic health. Neither could we find solid explanations for the discrepancies with literature. The question of whether the gut microbiota may be involved in modulation and/or mediation of the effect of diet on CMD risk factors needs to be further investigated taking into account more complex dimensions of the gut microbiota with addition of metabolites formed from diet and microbiota interactions. Also, to get a deeper understanding of this relationship, species and strains should be further investigated since similar microbial species can for instance have distinct functions[333]. The gut microbiota was based on a small piece of a stool sample and constitute a surrogate marker of the microbial content of the gut. Even though bacteria are found from the oral cavity down to the colon, most of our bacteria resides in the gut. A stool sample is a feasible material to collect in order to study the gut microbiota, but it may not represent the conditions in the whole gut for instance in the intestine[334]. In addition, 16S rRNA gene sequencing was used, which allows for taxonomic classification down to genus level. In order to reliably assess species or strains, full genome sequencing or qPCR are needed. With full genome sequencing it is also possible to investigate functional pathways, which can provide important information in unravelling the association between diet and the gut microbiota. A challenge with the full genome sequencing, is that it also requires substantially more complex pre-processing of data.

6.1 Methodological considerations

6.1.1 Selection bias

Selection bias can arise if the relation between an exposure and an outcome differs for individuals participating in the study compared to individuals invited but do not participate[335]. The main focus of the DCH-NG cohort was the family design i.e., invited participants constituted descendants from the previous DCH cohort. In total, 22% of the invited descendants completed the study requirements. The DCH-NG participants and non-participants differed according to several factors. Participants compared to non-participants were more likely to be women, middle aged and married, to have shorter distance to the study center, higher income and educational level as well as having medium to high skilled occupation and managing responsibility[278]. To compare characteristics of participants from the main cohort and the DCH-NG MAX sub-cohort, participants included in the FFQ validation (paper I) seem to have a similar distribution of sex and age. For highest attained education, there was a larger proportion being in the highest category and a smaller proportion being in the vocational training category in DCH-NG MAX compared to the main cohort. In addition, in the DCH-NG MAX sub-cohort only participants visiting the Copenhagen study center were invited to participate. Whether there may be regional differences in DCH-NG cohort has not yet been investigated and therefore whether participants in the DCH-NG MAX may differ from the general DCH-NG cohort is difficult to discuss at this point. However, participants that were family related were also included to mimic the design of the main cohort. In addition, an equal number of men and

women with different age and fasting strata were included, to have the possibility to investigate the influence of sex, age and fasting status in metabolomics or other biological measurements, but this should not give rise to a selected group. In paper II, a random sub-sample of the DCH-NG enrolled during 2015 to 2016, were included to validate the DQS and associate adherence to the score with CMD risk factors. The DQS is a tool currently used to monitor dietary quality in the Danish population[336].In paper II, there was a smaller proportion of participants with unhealthy dietary habits (both females and males 8%) compared to the proportion from the National survey from 2017 (females 12%, males 20%). Our population seem to be healthier than the general population in terms of dietary habits. However, I do not believe that this will lead to differences in association between dietary quality and CMD risk factors but more like an underestimation.

6.1.2 Information bias in the context of self-reported dietary intake

Information bias arises when the exposure, outcome or confounders are measured inaccurately or misclassified. In self-reported dietary intake different kinds of information bias may be present such as recall bias and intake-related bias.

Recall bias

Recall bias have often been related to dietary intake in case-control studies, due to differences in recalling of dietary intake or other information between cases and control. FFQs are also subject to recall bias especially when the recall period is long such as one year[193]. In the DCH-NG FFQ the recall period is one year and therefore recall bias may to some extent exist. Whether this can explain the discrepancy between the dietary intake reported with the FFQ compared to the 24-HDRs is difficult to conclude. The 24-HDR is also subject to recall bias, but maybe to a lesser extent than the FFQ since the recall period is rather short.

Intake related bias

This type of bias is usually related to the pressure to adhere to a certain dietary pattern due to social or cultural norms also known as social-desirability bias. For instance, individuals with high intake of presumable unhealthy foods may underreport their intake while individuals with low intake of presumable healthy foods may overreport their intake. This type of error most likely will result in attenuation of the slope in regression analysis[337]. Long-term dietary intake reported with an FFQ are thought to be influence by social desirability and perhaps to a larger extend than short-term dietary intake reported with 24-HDRs due to differences in cognitive memory[338]. In the validation of the FFQ (paper I) higher levels of energy and nutrient intakes were found compared with intakes from the 24-HDRs, since it was not possible to compare food group intakes between the FFQ and the 24-HDRs it is difficult to discuss any potential over- or underestimation of healthy or unhealthy foods. In general, I would assume the presence of social desirability bias in some degree.

6.1.3 Confounding

Confounding is another type of bias that can distort an association when the effect of an exposure on an outcome is confused by another factor or several other factors. A confounder is both associated with the exposure and the outcome but does not lie on the causal pathway[335]. In paper II, potential confounders c earlier literature. In the simple linear regression model, adjustments for sex and age were made, where in the multivariate linear regression model we adjusted further for physical activity and education level as well as smoking habits. For some of the CMD risk factors, the estimates changed between the simple and adjusted model, which could indicate confounding. In paper III, we were more hesitant to avoid over-adjustment. Adjustments were only made for age and sex and therefore there may be other confounding factors such as physical activity, BMI or other types of medication. It has been recommended to adjust for confounding variables in microbiota and disease association studies, in order to reduce false gut microbiota-disease associations[339]. In paper IV, the analyses were separate by sex (due to effect modification, see below). In the crude models, age was adjusted for. In the adjusted models, for rMED further adjustments for physical activity and smoking, whereas for HNFI further adjustments for alcohol and meat were made. There were differences between estimates for the simple and fully adjusted models, which again could indicate confounding. Still, there may be other confounding factors that have not been considered.

7. CONCLUSIONS

- A 376-item FFQ was evaluated. Higher absolute intakes of total energy and nutrients were reported with the FFQ compared to the 24-HDRs. For the relative validity, ranking of individuals dietary intake according to total energy, nutrient densities and nutrient residuals were acceptable. For the reproducibility, correlation coefficients and ranking of individuals according to energy, energy-adjusted nutrients and food group intakes were satisfactory. Overall, the FFQ's ability to rank individuals according to their dietary intake was considered satisfactory and may therefore be used in epidemiological studies with diet as the exposure. Thus, further studies are needed to validate intake of food groups from the FFQ.

- The correlation coefficients of major food groups between the 23-item FFQ and 376-item FFQ was acceptable. The DQS is a good indicator of overall dietary quality in the investigated population of Danish men and women. It could successfully be used to rank individuals into groups of having healthy, average or unhealthy dietary habits. Adherence to the DQS was significantly associated with CMD risk factors including lower levels of WC, AFM, RFM, VF, LDL-C and hs-CRP and higher HDL-C levels. Thus, the DQS is a simple and easy tool suitable for evaluating dietary quality in large populations.

- A substantial part (39%) of the gut microbiota at genus level had moderate to good stability (ICC>0.5). These genera therefore may be studied in relation to disease outcomes in prospective studies with acceptable precision. Based on genera with ICC>0.5 gut microbial subgroups (enterotypes) were identified: *Bacteroides*, *Prevotella*-9 and *Ruminoccoccae*-groups. *Prevotella*-9 and *Ruminococcaceae*-groups were associated with higher adherence to the healthy plant-based indexes (HNFI, rMED, PDI, hPDI, pro-veg) compared to the *Bacteroides*-group. Dietary patterns (HNFI, rMED, and PDI, hPDI and uPDI) were also associated with 22 specific genera. Higher adherence to the healthy plant-based indexes associated with higher relative abundance of genera with known fibre fermenting properties. Furthermore, vegetables, fruit, whole grains/cereal and nuts were most strongly associated with these genera.

- The effect of adherence the healthy Nordic or Mediterranean diets on CMD risk factors did not appear to be mediated by the *Prevotella*-to-*Bacteroides* ratio. However, adherence to these healthy diets were associated with lower levels of CMD risk factors, in particular lower levels of adiposity as well as lower levels lipidemia and hs-CRP, mediated by adiposity.

8. FUTURE PERSPECTIVES

- Though web-based FFQs have advantages over the paper-based version, measurement errors still exist. Studies where several dietary assessment tools are combined could be an approach in future studies in order to improve dietary data. For instance, the combination of the FFQ with 24-HDRs. Thereby foods that are not consumed regularly will be assessed together with more accurate portion sizes.

- Currently, the DQS is used to assess and monitor the quality of dietary habits in the Danish population. To our knowledge, this score has not been validated outside of Denmark. The score is based on consumption of vegetables, fruit, fish and fat which are not related to specific cultural or regional foods. However, the 23-item FFQ may be and therefore it would be interesting to first investigate its validity in other Nordic countries in order to evaluate if this simple score could be used as a more universal score to monitor quality of dietary habits.

- More studies investigating the stability of genera, species and strains in other countries are warranted in order to get a more comprehensive evaluation and understanding of the temporal variability of specific gut microbes. Three gut microbial community subgroups were identified, also referred to as enterotypes in the literature. Further assessment of the stability of these subgroups, would be interesting to explore, since these are not discrete groups.

- Further studies investigating the effect of diet on CMD risk factors where more complex dimensions of the gut microbiota are included are needed. Since it is the gut derived metabolites and not the microbiota itself that exerts the main health effects these studies should be combined with metabolomics data.

9. ACKNOWLEDGEMENTS

I would like to give a special thank you to all the participants in the DCH-NG cohort and the sub-cohort MAX as well as former study personal. Your contributions to science are invaluable and the foundation of my thesis.

Rikard Landberg (main supervisor at Chalmers), you are an inspiration both as a researcher and as a people's manager. Thank you for continues positivity, motivation and guidance through my PhD project. Thank you for believing in me and having trust in me to take on this PhD project with all the twist and turns it has involved. I truly appreciate your encouragement throughout the years - I have learned a lot from you.

Jytte Halkjær (co-supervisor at DCI), thank you for all our good and sometime tough times with establishing the DCH-NG. I am still proud of what we were able to accomplish in our little team. Your door is always open both for scientific discussions but also for cozy talks, I appreciate and value that. Thank you for your help, guidance and support throughout my whole PhD time.

Carl Brunius (co-supervisor at Chalmers), I am thankful for all your help and support during my PhD project. You are always positive, and I have learned a lot from your enlightening scientific inputs and our discussions. I also appreciate your help in the finalization process of my papers and the thesis.

Susanne Bügel (supervisor, Copenhagen University), thank you for support and guidance during my PhD studies and for making me feel welcome at University of Copenhagen.

Anne Tjønneland and Anja Olsen, thank you for giving me the opportunity to be part of the team behind the DCH-NG back 2014 and for support during my PhD. Thank you for always having open doors to your offices and for putting aside time for my questions over the years. Also, thank for creating a work environment where freedom with responsibility is a corner stone.

Susanne Rosthøj, thank you for challenging and broaden my statistical knowledge. Your excitement for delving into new statistical models is contagious and I have enjoyed our many discussions, and I am grateful for all your help. Thank you for taking the time for my endless questions. Mia Klinten Grand, I have enjoyed our collaboration the last months of my PhD project. Thank you for helping me with my last paper and I look forward to finalizing this work together.

Kristina E.T. Pedersen, (office mate) I sincerely enjoyed our time working together establishing the DCH-NG cohort with Jytte, Anja and Anne. What a journey it has been. You are always positive, ready to help out or give a hug if needed, thank you for that and for support, good times in the office, and great trips over the years.

Julie L.M. Andersen, (office mate) and Rikke L. Nielsen (former colleague), thank you for taking over the data collection of the DCH-NG MAX study while I was on maternity leave, you

did a tremendous job. Julie, thank you for great discussions and laughs in our office and help along the way with DCH-NG data cleaning. Rikke, also thank you for assisting me with developing the Danish food database for myfood24, I would not have made it without your help and valuable dietician knowledge.

Anders and Ingegerd, it has been a pleasure working with you on our two papers. I am grateful for all your help, scientific knowledge and discussions as well as contributions to the papers. I have learned a lot from you during the last year of my PhD studies.

Torben Hansen and Johan Dicksved thank you for contributing with freeze drying and analysis of stool samples and microbiota scientific inputs.

Katja Boll and Peter F. Hendriksen what a privilege it has been to have you as data managers. Katja thank you for your work and help with the dietary analysis, delivering datasets and for being there in urgent matters. Peter, thank you for your help with data processing and in urgent matters.

Nick Martinussen (former colleague) for being our data manager throughout planning and data collection of the DCH-NG cohort and MAX study. Thank you for always having time, and for our exhausting long diet-analysis meetings which were mainly doable due to our weekly cinnamon roll tasting ritual.

Thank you to former and current colleagues from the Diet, Cancer and Health group at the Danish Cancer Institute. I could not have wished for better colleagues or place to work.

Colleagues from Chalmers University of Technology, former and current it has been a pleasure being part of the Nutrition and Food Sciences, especially at my two months stay and visits from time to time. I could easily have spent more time in your department.

And not to forget my dear mother who have helped out in busy periods during my PhD studies and for endless support.

Last, but not least thank you to my incredible husband Stavros for inexhaustible support, encouragement and push when needed and to our three wonderful boys Hektor, Atlas and Emil for showing me what is important in life♥

10. REFERENCES

1. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Reference Life Table. *Institute for Health Metrics and Evaluation* (IHME).
2. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet.* 2023;402:203-234.
3. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. *Journal of the American College of Cardiology.* 2020;76:2982-3021.
4. Vasudevan AR, Ballantyne CM. Cardiometabolic risk assessment: an approach to the prevention of cardiovascular disease and diabetes mellitus. *Clinical cornerstone.* 2005;7:7-16.
5. Guo F, Moellering DR, Garvey WT. The progression of cardiometabolic disease: validation of a new cardiometabolic disease staging system applicable to obesity. *Obesity.* 2014;22:110-118.
6. de Waard A-KM, Hollander M, Korevaar JC, Nielen MMJ, Carlsson AC, Lionis C, et al. Selective prevention of cardiometabolic diseases: activities and attitudes of general practitioners across Europe. *European Journal of Public Health.* 2018;29:88-93.
7. Faasse S, Braun H, Vos M. The role of NAFLD in cardiometabolic disease: an update. *F1000Research.* 2018;7:170.
8. Rao GHR. Cardiometabolic diseases: a global perspective. *Journal of Cardiology & Cardiovascular Therapy*. 2018.
9. Huo R, Du T, Xu Y, Xu W, Chen X, Sun K, et al. Effects of Mediterranean-style diet on glycemic control, weight loss and cardiovascular risk factors among type 2 diabetes individuals: a meta-analysis. *European Journal of Clinical Nutrition.* 2015;69:1200-1208.
10. Trøseid M, Andersen G, Broch K, Hov JR. The gut microbiome in coronary artery disease and heart failure: Current knowledge and future directions. *EBioMedicine.* 2020;52.
11. Letchumanan G, Abdullah N, Marlini M, Baharom N, Lawley B, Omar MR, et al. Gut Microbiota Composition in Prediabetes and Newly Diagnosed Type 2 Diabetes: A Systematic Review of Observational Studies. *Frontiers in cellular and infection microbiology.* 2022;12.
12. Christensen L, Vuholm S, Roager HM, Nielsen DS, Krych L, Kristensen M, et al. Prevotella Abundance Predicts Weight Loss Success in Healthy, Overweight Adults Consuming a Whole-Grain Diet Ad Libitum: A Post Hoc Analysis of a 6-Wk Randomized Controlled Trial. *The Journal of nutrition.* 2019;149:2174-2181.
13. Hjorth MF, Blædel T, Bendtsen LQ, Lorenzen JK, Holm JB, Kiilerich P, et al. Prevotella-to-Bacteroides ratio predicts body weight and fat loss success on 24-week diets varying in macronutrient composition and dietary fiber: results from a post-hoc analysis. *International Journal of Obesity.* 2019;43:149-157.
14. Kovatcheva-Datchary P, Nilsson A, Akrami R, Lee YS, De Vadder F, Arora T, et al. Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. *Cell metabolism.* 2015;22:971-982.
15. Willett W, Lenart E. Reproducibility and Validity of Food Frequency Questionnaires. *Nutritional Epidemiology.* 3rd ed. Oxford University Press; 2012:96–141.

16. Mehta RS, Abu-Ali GS, Drew DA, Lloyd-Price J, Subramanian A, Lochhead P, et al. Stability of the human faecal microbiome in a cohort of adult men. *Nature microbiology.* 2018;3:347-355.
17. Sinha R, Goedert JJ, Vogtmann E, Hua X, Porras C, Hayes R, et al. Quantification of Human Microbiome Stability Over 6 Months: Implications for Epidemiologic Studies. *American Journal of Epidemiology.* 2018;187:1282-1290.
18. Olsson LM, Boulund F, Nilsson S, Khan MT, Gummesson A, Fagerberg L, et al. Dynamics of the normal gut microbiota: A longitudinal one-year population study in Sweden. *Cell host & microbe.* 2022;30:726-739.e723.
19. Zimmer J, Lange B, Frick JS, Sauer H, Zimmermann K, Schwiertz A, et al. A vegan or vegetarian diet substantially alters the human colonic faecal microbiota. *European Journal of Clinical Nutrition.* 2012;66:53-60.
20. Ruengsomwong S, La-Ongkham O, Jiang J, Wannissorn B, Nakayama J, Nitisinprasert S. Microbial Community of Healthy Thai Vegetarians and Non-Vegetarians, Their Core Gut Microbiota, and Pathogen Risk. *Journal of microbiology and biotechnology.* 2016;26:1723-1735.
21. Kabeerdoss J, Shobana Devi R, Regina Mary R, Ramakrishna BS. Faecal microbiota composition in vegetarians: comparison with omnivores in a cohort of young women in southern India. *British Journal of Nutrition.* 2012;108:953-957.
22. de Moraes AC, Fernandes GR, da Silva IT, Almeida-Pititto B, Gomes EP, Pereira AD, et al. Enterotype May Drive the Dietary-Associated Cardiometabolic Risk Factors. *Frontiers in cellular and infection microbiology.* 2017;7:47.
23. Mitsou EK, Kakali A, Antonopoulou S, Mountzouris KC, Yannakoulia M, Panagiotakos DB, et al. Adherence to the Mediterranean diet is associated with the gut microbiota pattern and gastrointestinal characteristics in an adult population. *British Journal of Nutrition.* 2017;117:1645-1655.
24. Garcia-Mantrana I, Selma-Royo M, Alcantara C, Collado MC. Shifts on Gut Microbiota Associated to Mediterranean Diet Adherence and Specific Dietary Intakes on General Adult Population. *Frontiers in microbiology.* 2018;9:890.
25. Rosés C, Cuevas-Sierra A, Quintana S, Riezu-Boj JI, Martínez JA, Milagro FI, et al. Gut Microbiota Bacterial Species Associated with Mediterranean Diet-Related Food Groups in a Northern Spanish Population. *Nutrients.* 2021;13.
26. Peters BA, Xing J, Chen GC, Usyk M, Wang Z, McClain AC, et al. Healthy dietary patterns are associated with the gut microbiome in the Hispanic Community Health Study/Study of Latinos. *The American journal of clinical nutrition.* 2023;117:540-552.
27. Miao Z, Du W, Xiao C, Su C, Gou W, Shen L, et al. Gut microbiota signatures of long-term and short-term plant-based dietary pattern and cardiometabolic health: a prospective cohort study. *BMC medicine.* 2022;20:204.
28. Gaundal L, Myhrstad MCW, Rud I, Gjøvaag T, Byfuglien MG, Retterstøl K, et al. Gut microbiota is associated with dietary intake and metabolic markers in healthy individuals. *Food & nutrition research.* 2022;66.
29. Wang DD, Nguyen LH, Li Y, Yan Y, Ma W, Rinott E, et al. The gut microbiome modulates the protective association between a Mediterranean diet and cardiometabolic disease risk. *Nature medicine.* 2021;27:333-343.
30. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science.* 2011;334:105-108.
31. Noh H, Jang HH, Kim G, Zouiouich S, Cho SY, Kim HJ, et al. Taxonomic Composition and Diversity of the Gut Microbiota in Relation to Habitual Dietary Intake in Korean Adults. *Nutrients.* 2021;13.

32. Lim MY, Rho M, Song Y-M, Lee K, Sung J, Ko G. Stability of Gut Enterotypes in Korean Monozygotic Twins and Their Association with Biomarkers and Diet. *Scientific reports.* 2014;4:7348.
33. WHO. *Cardiovascular diseases (CVDs) [Fact sheet].* 2021.
34. Rafieian-Kopaei M, Setorki M, Doudi M, Baradaran A, Nasri H. Atherosclerosis: process, indicators, risk factors and new hopes. *International journal of preventive medicine.* 2014;5:927-946.
35. Committee ADAPP. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2022. *Diabetes Care.* 2021;45:17-38.
36. Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. *Lancet.* 2017;389:1238-1252.
37. Riazi K, Azhari H, Charette JH, Underwood FE, King JA, Afshar EE, et al. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. *The lancet Gastroenterology & hepatology.* 2022;7:851-861.
38. Einarson TR, Acs A, Ludwig C, Panton UH. Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007-2017. *Cardiovascular diabetology.* 2018;17:83.
39. Alicic RZ, Rooney MT, Tuttle KR. Diabetic Kidney Disease: Challenges, Progress, and Possibilities. *Clinical journal of the American Society of Nephrology.* 2017;12:2032-2045.
40. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, Jafar TH, Heerspink HJ, Mann JF, et al. Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. *Lancet.* 2013;382:339-352.
41. Mantovani A, Byrne CD, Bonora E, Targher G. Nonalcoholic Fatty Liver Disease and Risk of Incident Type 2 Diabetes: A Meta-analysis. *Diabetes Care.* 2018;41:372-382.
42. Przybyszewski EM, Targher G, Roden M, Corey KE. Nonalcoholic fatty liver disease and cardiovascular disease. *Clinical Liver Disease*. 2021;17:19-22.
43. Rodgers JL, Jones J, Bolleddu SI, Vanthenapalli S, Rodgers LE, Shah K, et al. Cardiovascular Risks Associated with Gender and Aging. *Journal of cardiovascular development and disease.* 2019;6.
44. Yoshida Y, Chen Z, Fonseca VA, Mauvais-Jarvis F. Sex Differences in Cardiovascular Risk Associated With Prediabetes and Undiagnosed Diabetes. *American journal of preventive medicine.* 2023.
45. Gao Z, Chen Z, Sun A, Deng X. Gender differences in cardiovascular disease. *Medicine in Novel Technology and Devices.* 2019;4.
46. Kautzky-Willer A, Leutner M, Harreiter J. Sex differences in type 2 diabetes. *Diabetologia.* 2023;66:986-1002.
47. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nature Reviews Nephrology.* 2020;16:377-390.
48. Hajar R. Genetics in Cardiovascular Disease. *Heart views : the official journal of the Gulf Heart Association.* 2020;21:55-56.
49. Micha R, Peñalvo JL, Cudhea F, Imamura F, Rehm CD, Mozaffarian D. Association Between Dietary Factors and Mortality From Heart Disease, Stroke, and Type 2 Diabetes in the United States. *Journal of the American Medical Association.* 2017;317:912-924.
50. Lavie CJ, Ozemek C, Carbone S, Katzmarzyk PT, Blair SN. Sedentary Behavior, Exercise, and Cardiovascular Health. *Circulation research.* 2019;124:799-815.
51. Clark AM, DesMeules M, Luo W, Duncan AS, Wielgosz A. Socioeconomic status and cardiovascular disease: risks and implications for care. *Nature Reviews Cardiology.* 2009;6:712-722.

52. Organisation WH. Obesity and overrweight [Fact sheet]. 2021.
53. Bosomworth NJ. Normal-weight central obesity: Unique hazard of the toxic waist. *Canadian family physician Medecin de famille canadien.* 2019;65:399-408.
54. Hruby A, Hu FB. The Epidemiology of Obesity: A Big Picture. *PharmacoEconomics.* 2015;33:673-689.
55. Sowers JR. Obesity as a cardiovascular risk factor. *American Journal of Medicine.* 2003;115:37-41.
56. Rhee EJ, Cho JH, Kwon H, Park SE, Jung JH, Han KD, et al. Association between abdominal obesity and increased risk for the development of hypertension regardless of physical activity: A nationwide population-based study. *Journal of clinical hypertension.* 2018;20:1417-1426.
57. Shariq OA, McKenzie TJ. Obesity-related hypertension: a review of pathophysiology, management, and the role of metabolic surgery. *Gland surgery.* 2020;9:80-93.
58. Bays HE. "Sick fat," metabolic disease, and atherosclerosis. *The American journal of medicine.* 2009;122:S26-37.
59. Shabana, Shahid SU, Sarwar S. The abnormal lipid profile in obesity and coronary heart disease (CHD) in Pakistani subjects. *Lipids in health and disease.* 2020;19:73.
60. Wondmkun YT. Obesity, Insulin Resistance, and Type 2 Diabetes: Associations and Therapeutic Implications. *Diabetes, metabolic syndrome and obesity : targets and therapy.* 2020;13:3611-3616.
61. Janssen J. Hyperinsulinemia and Its Pivotal Role in Aging, Obesity, Type 2 Diabetes, Cardiovascular Disease and Cancer. *International journal of molecular sciences.* 2021;22.
62. Zatterale F, Longo M, Naderi J, Raciti GA, Desiderio A, Miele C, et al. Chronic Adipose Tissue Inflammation Linking Obesity to Insulin Resistance and Type 2 Diabetes. *Frontiers in physiology.* 2019;10:1607.
63. Nour M, Lutze SA, Grech A, Allman-Farinelli M. The Relationship between Vegetable Intake and Weight Outcomes: A Systematic Review of Cohort Studies. *Nutrients.* 2018;10.
64. Ledoux TA, Hingle MD, Baranowski T. Relationship of fruit and vegetable intake with adiposity: a systematic review. *Obesity reviews : an official journal of the International Association for the Study of Obesity.* 2011;12:143-150.
65. Maki KC, Palacios OM, Koecher K, Sawicki CM, Livingston KA, Bell M, et al. The Relationship between Whole Grain Intake and Body Weight: Results of Meta-Analyses of Observational Studies and Randomized Controlled Trials. *Nutrients.* 2019;11.
66. González-Morales R, Canto-Osorio F, Stern D, Sánchez-Romero LM, Torres-Ibarra L, Hernández-López R, et al. Soft drink intake is associated with weight gain, regardless of physical activity levels: the health workers cohort study. *The international journal of behavioral nutrition and physical activity.* 2020;17:60.
67. Malik VS, Hu FB. The role of sugar-sweetened beverages in the global epidemics of obesity and chronic diseases. *Nature reviews Endocrinology.* 2022;18:205-218.
68. Cox CE. Role of Physical Activity for Weight Loss and Weight Maintenance. *Diabetes spectrum : a publication of the American Diabetes Association.* 2017;30:157-160.
69. Rapsomaniki E, Timmis A, George J, Pujades-Rodriguez M, Shah AD, Denaxas S, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people. *Lancet.* 2014;383:1899-1911.
70. American Heart Association. High Blood Pressure. The Facts About HBP. What is high blood pressure? 2023; https://www.heart.org/en/health-topics/high-blood-pressure/the-facts-about-high-blood-pressure/what-is-high-blood-pressure. Accessed 08.15.2023.

71. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension. *European heart journal.* 2018;39:3021-3104.
72. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet.* 2002;360:1903-1913.
73. Whelton PK, Carey RM, Aronow WS, Casey DE, Jr., Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension.* 2018;71:1269-1324.
74. Diseases NIoDaDaK. High Blood Pressure & Kidney Disease. 2020; https://www.niddk.nih.gov/health-information/kidney-disease/high-blood-pressure. Accessed 08.17.2023.
75. Rao MV, Qiu Y, Wang C, Bakris G. Hypertension and CKD: Kidney Early Evaluation Program (KEEP) and National Health and Nutrition Examination Survey (NHANES), 1999-2004. *American journal of kidney diseases : the official journal of the National Kidney Foundation.* 2008;51:30-37.
76. Emdin CA, Anderson SG, Woodward M, Rahimi K. Usual Blood Pressure and Risk of New-Onset Diabetes: Evidence From 4.1 Million Adults and a Meta-Analysis of Prospective Studies. *Journal of the American College of Cardiology.* 2015;66:1552-1562.
77. Oikonomou D, Georgiopoulos G, Katsi V, Kourek C, Tsioufis C, Alexopoulou A, et al. Non-alcoholic fatty liver disease and hypertension: coprevalent or correlated? *European journal of gastroenterology & hepatology.* 2018;30:979-985.
78. Sun D, Zhou T, Heianza Y, Li X, Fan M, Fonseca VA, et al. Type 2 Diabetes and Hypertension. *Circulation research.* 2019;124:930-937.
79. Slama M, Susic D, Frohlich ED. Prevention of hypertension. *Current opinion in cardiology.* 2002;17:531-536.
80. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. *Nature reviews Nephrology.* 2020;16:223-237.
81. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *European heart journal.* 2020;41:111-188.
82. Jung E, Kong SY, Ro YS, Ryu HH, Shin SD. Serum Cholesterol Levels and Risk of Cardiovascular Death: A Systematic Review and a Dose-Response Meta-Analysis of Prospective Cohort Studies. *International journal of environmental research and public health.* 2022;19.
83. Thomas F, Bean K, Guize L, Quentzel S, Argyriadis P, Benetos A. Combined effects of systolic blood pressure and serum cholesterol on cardiovascular mortality in young (<55 years) men and women. *European heart journal.* 2002;23:528-535.
84. Libby P, Buring JE, Badimon L, Hansson GK, Deanfield J, Bittencourt MS, et al. Atherosclerosis. *Nature reviews Disease primers.* 2019;5:56.
85. Martin A, Lang S, Goeser T, Demir M, Steffen HM, Kasper P. Management of Dyslipidemia in Patients with Non-Alcoholic Fatty Liver Disease. *Current atherosclerosis reports.* 2022;24:533-546.
86. Mooradian AD. Dyslipidemia in type 2 diabetes mellitus. *Nature clinical practice Endocrinology & metabolism.* 2009;5:150-159.

87. Riccardi G, Vaccaro O, Costabile G, Rivellese AA. How Well Can We Control Dyslipidemias Through Lifestyle Modifications? *Current cardiology reports.* 2016;18:66.
88. Cefalu WT. Insulin resistance: cellular and clinical concepts. *Experimental biology and medicine.* 2001;226:13-26.
89. Giri B, Dey S, Das T, Sarkar M, Banerjee J, Dash SK. Chronic hyperglycemia mediated physiological alteration and metabolic distortion leads to organ dysfunction, infection, cancer progression and other pathophysiological consequences: An update on glucose toxicity. *Biomedicine & Pharmacotherapy.* 2018;107:306-328.
90. de Luca C, Olefsky JM. Inflammation and insulin resistance. *FEBS letters.* 2008;582:97-105.
91. Ward PA. Acute and Chronic Inflammation. In: *Fundamentals of Inflammation.* Cambridge University Press; 2010:1-16.
92. Liu Y, Guan S, Xu H, Zhang N, Huang M, Liu Z. Inflammation biomarkers are associated with the incidence of cardiovascular disease: a meta-analysis. *Frontiers in Cardiovascular Medicine*. 2023;10.
93. Wang X, Bao W, Liu J, Ouyang YY, Wang D, Rong S, et al. Inflammatory markers and risk of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care.* 2013;36:166-175.
94. Alfaddagh A, Martin SS, Leucker TM, Michos ED, Blaha MJ, Lowenstein CJ, et al. Inflammation and cardiovascular disease: From mechanisms to therapeutics. *American journal of preventive cardiology.* 2020;4.
95. Shimobayashi M, Albert V, Woelnerhanssen B, Frei IC, Weissenberger D, Meyer-Gerspach AC, et al. Insulin resistance causes inflammation in adipose tissue. *The Journal of clinical investigation.* 2018;128:1538-1550.
96. Hotamisligil GS, Arner P, Caro JF, Atkinson RL, Spiegelman BM. Increased adipose tissue expression of tumor necrosis factor-alpha in human obesity and insulin resistance. *The Journal of clinical investigation.* 1995;95:2409-2415.
97. Hotamisligil GS, Peraldi P, Budavari A, Ellis R, White MF, Spiegelman BM. IRS-1-mediated inhibition of insulin receptor tyrosine kinase activity in TNF-alpha- and obesity-induced insulin resistance. *Science.* 1996;271:665-668.
98. Libby P, Ridker PM, Maseri AJC. Inflammation and atherosclerosis. 2002;105:1135-1143.
99. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* 2019;393:1958-1972.
100. Mozaffarian D, Rosenberg I, Uauy R. History of modern nutrition science-implications for current research, dietary guidelines, and food policy. *British Medical Journal (Clinical Research Edition).* 2018;361:2392.
101. Slavin JL, Lloyd B. Health benefits of fruits and vegetables. *Advances in nutrition.* 2012;3:506-516.
102. Taha Gökmen Ü, Ayşe Nur S, Onur Ç, Funda Pınar Ç. Role of Vegetables in Human Nutrition and Disease Prevention. In: *Vegetables.* IntechOpen; 2018:Ch. 2.
103. Lampe JW. Health effects of vegetables and fruit: assessing mechanisms of action in human experimental studies. *The American journal of clinical nutrition.* 1999;70:475-490.
104. Slavin J. Whole grains and human health. *Nutrition research reviews.* 2004;17:99-110.
105. Aune D, Keum N, Giovannucci E, Fadnes LT, Boffetta P, Greenwood DC, et al. Whole grain consumption and risk of cardiovascular disease, cancer, and all cause and cause specific mortality: systematic review and dose-response meta-analysis of prospective studies. *British Medical Journal (Clinical Research Edition)* . 2016;353:2716.

106. Bechthold A, Boeing H, Schwedhelm C, Hoffmann G, Knüppel S, Iqbal K, et al. Food groups and risk of coronary heart disease, stroke and heart failure: A systematic review and dose-response meta-analysis of prospective studies. *Critical reviews in food science and nutrition.* 2019;59:1071-1090.
107. Ghanbari-Gohari F, Mousavi SM, Esmaillzadeh A. Consumption of whole grains and risk of type 2 diabetes: A comprehensive systematic review and dose-response meta-analysis of prospective cohort studies. *Food science & nutrition.* 2022;10:1950-1960.
108. Topping DL, Clifton PM. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiological reviews.* 2001;81:1031-1064.
109. Makki K, Deehan EC, Walter J, Bäckhed F. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell host & microbe.* 2018;23:705-715.
110. He J, Zhang P, Shen L, Niu L, Tan Y, Chen L, et al. Short-Chain Fatty Acids and Their Association with Signalling Pathways in Inflammation, Glucose and Lipid Metabolism. *International journal of molecular sciences.* 2020;21.
111. Giosuè A, Calabrese I, Vitale M, Riccardi G, Vaccaro O. Consumption of Dairy Foods and Cardiovascular Disease: A Systematic Review. *Nutrients.* 2022;14.
112. Feng Y, Zhao Y, Liu J, Huang Z, Yang X, Qin P, et al. Consumption of Dairy Products and the Risk of Overweight or Obesity, Hypertension, and Type 2 Diabetes Mellitus: A Dose-Response Meta-Analysis and Systematic Review of Cohort Studies. *Advances in nutrition.* 2022;13:2165-2179.
113. González S, Fernández-Navarro T, Arboleya S, de los Reyes-Gavilán CG, Salazar N, Gueimonde M. Fermented Dairy Foods: Impact on Intestinal Microbiota and Health-Linked Biomarkers. *Frontiers in Microbiology*. 2019;10.
114. Skåre JU, Brantsæter AL, Frøyland L, Hemre GI, Knutsen HK, Lillegaard ITL, et al. Benefit-risk assessment of fish and fish products in the Norwegian diet–an update. Opinion of the Scientific Steering Committee of the Norwegian Scientific Committee for Food Safety. 2014.
115. Giosuè A, Calabrese I, Lupoli R, Riccardi G, Vaccaro O, Vitale M. Relations between the Consumption of Fatty or Lean Fish and Risk of Cardiovascular Disease and All-Cause Mortality: A Systematic Review and Meta-Analysis. *Advances in nutrition.* 2022;13:1554-1565.
116. Zhang Y, Zhuang P, He W, Chen JN, Wang WQ, Freedman ND, et al. Association of fish and long-chain omega-3 fatty acids intakes with total and cause-specific mortality: prospective analysis of 421 309 individuals. *Journal of internal medicine.* 2018;284:399-417.
117. Barbaresko J, Lang A, Szczerba E, Baechle C, Beckhaus J, Schwingshackl L, et al. Dietary Factors and All-Cause Mortality in Individuals With Type 2 Diabetes: A Systematic Review and Meta-analysis of Prospective Observational Studies. *Diabetes Care.* 2023;46:469-477.
118. Pastorino S, Bishop T, Sharp SJ, Pearce M, Akbaraly T, Barbieri NB, et al. Heterogeneity of Associations between Total and Types of Fish Intake and the Incidence of Type 2 Diabetes: Federated Meta-Analysis of 28 Prospective Studies Including 956,122 Participants. *Nutrients.* 2021;13.
119. Innes JK, Calder PC. Marine Omega-3 (N-3) Fatty Acids for Cardiovascular Health: An Update for 2020. *International journal of molecular sciences.* 2020;21.
120. Hidayat K, Yang J, Zhang Z, Chen GC, Qin LQ, Eggersdorfer M, et al. Effect of omega-3 long-chain polyunsaturated fatty acid supplementation on heart rate: a meta-analysis of randomized controlled trials. *European Journal of Clinical Nutrition.* 2018;72:805-817.

121. Miller PE, Van Elswyk M, Alexander DD. Long-chain omega-3 fatty acids eicosapentaenoic acid and docosahexaenoic acid and blood pressure: a meta-analysis of randomized controlled trials. *American journal of hypertension.* 2014;27:885-896.
122. AbuMweis S, Jew S, Tayyem R, Agraib L. Eicosapentaenoic acid and docosahexaenoic acid containing supplements modulate risk factors for cardiovascular disease: a meta-analysis of randomised placebo-control human clinical trials. *Journal of human nutrition and dietetics : the official journal of the British Dietetic Association.* 2018;31:67-84.
123. Gao LG, Cao J, Mao QX, Lu XC, Zhou XL, Fan L. Influence of omega-3 polyunsaturated fatty acid-supplementation on platelet aggregation in humans: a meta-analysis of randomized controlled trials. *Atherosclerosis.* 2013;226:328-334.
124. Mozaffarian D, Rimm EB. Fish Intake, Contaminants, and Human HealthEvaluating the Risks and the Benefits. *Journal of the American Medical Association.* 2006;296:1885-1899.
125. Shi W, Huang X, Schooling CM, Zhao JV. Red meat consumption, cardiovascular diseases, and diabetes: a systematic review and meta-analysis. *European heart journal.* 2023;44:2626-2635.
126. Zhang X, Liang S, Chen X, Yang J, Zhou Y, Du L, et al. Red/processed meat consumption and non-cancer-related outcomes in humans: umbrella review. *British Journal of Nutrition.* 2023;130:484-494.
127. Williams PJN, Dietetics. Nutritional composition of red meat. 2007;64:S113-S119.
128. Kotopoulou S, Zampelas A, Magriplis E. Risk Assessment of Nitrite and Nitrate Intake from Processed Meat Products: Results from the Hellenic National Nutrition and Health Survey (HNNHS). *International journal of environmental research and public health.* 2022;19.
129. Siri-Tarino PW, Sun Q, Hu FB, Krauss RM. Saturated fat, carbohydrate, and cardiovascular disease. *The American journal of clinical nutrition.* 2010;91:502-509.
130. Jackson SL, Cogswell ME, Zhao L, Terry AL, Wang CY, Wright J, et al. Association Between Urinary Sodium and Potassium Excretion and Blood Pressure Among Adults in the United States: National Health and Nutrition Examination Survey, 2014. *Circulation.* 2018;137:237-246.
131. Micha R, Michas G, Mozaffarian D. Unprocessed red and processed meats and risk of coronary artery disease and type 2 diabetes--an updated review of the evidence. *Current atherosclerosis reports.* 2012;14:515-524.
132. Tang WHW, Li XS, Wu Y, Wang Z, Khaw KT, Wareham NJ, et al. Plasma trimethylamine N-oxide (TMAO) levels predict future risk of coronary artery disease in apparently healthy individuals in the EPIC-Norfolk prospective population study. *American heart journal.* 2021;236:80-86.
133. Wang DD, Hu FB. Dietary Fat and Risk of Cardiovascular Disease: Recent Controversies and Advances. *Annual Review of Nutrition.* 2017;37:423-446.
134. Comission E. Trans fats in foods a new regulation for EU consuemrs. Factsheet (2019). https://food.ec.europa.eu/system/files/2019-04/fs_labelling-nutrition_transfats_factsheet-2019.pdf. Accessed 04.10.2023.
135. Boateng L, Ansong R, Owusu WB, Steiner-Asiedu M. Coconut oil and palm oil's role in nutrition, health and national development: A review. *Ghana medical journal.* 2016;50:189-196.
136. Peou S, Milliard-Hasting B, Shah SA. Impact of avocado-enriched diets on plasma lipoproteins: A meta-analysis. *Journal of Clinical Lipidology.* 2016;10:161-171.
137. Santos LP, Gigante DP, Delpino FM, Maciel AP, Bielemann RM. Sugar sweetened beverages intake and risk of obesity and cardiometabolic diseases in longitudinal

studies: A systematic review and meta-analysis with 1.5 million individuals. *Clinical nutrition ESPEN.* 2022;51:128-142.
138. Fernández-Solà J. The Effects of Ethanol on the Heart: Alcoholic Cardiomyopathy. *Nutrients.* 2020;12.
139. Ronksley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *British Medical Journal (Clinical research ed).* 2011;342:671.
140. Larsson SC, Wallin A, Wolk A. Alcohol consumption and risk of heart failure: Meta-analysis of 13 prospective studies. *Clinical nutrition.* 2018;37:1247-1251.
141. Li X-H, Yu F-f, Zhou Y-H, He J. Association between alcohol consumption and the risk of incident type 2 diabetes: a systematic review and dose-response meta-analysis1. *The American Journal of Clinical Nutrition.* 2016;103:818-829.
142. Lankester J, Zanetti D, Ingelsson E, Assimes TL. Alcohol use and cardiometabolic risk in the UK Biobank: A Mendelian randomization study. *PLoS One.* 2021;16.
143. Willett W, Sampson LJNe. Foods and nutrients. 2013;40:17.
144. Schulze MB, Martínez-González MA, Fung TT, Lichtenstein AH, Forouhi NG. Food based dietary patterns and chronic disease prevention. 2018;361:2396.
145. Zhao J, Li Z, Gao Q, Zhao H, Chen S, Huang L, et al. A review of statistical methods for dietary pattern analysis. *Nutrition journal.* 2021;20:37.
146. Krebs-Smith SM, Subar AF, Reedy J. Examining Dietary Patterns in Relation to Chronic Disease: Matching Measures and Methods to Questions of Interest. *Circulation.* 2015;132:790-793.
147. Heidemann C, Schulze MB, Franco OH, van Dam RM, Mantzoros CS, Hu FB. Dietary patterns and risk of mortality from cardiovascular disease, cancer, and all causes in a prospective cohort of women. *Circulation.* 2008;118:230-237.
148. Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HAB, Kuczynski KJ, et al. Update of the Healthy Eating Index: HEI-2010. *Journal of the Academy of Nutrition and Dietetics.* 2013;113:569-580.
149. Buckland G, González CA, Agudo A, Vilardell M, Berenguer A, Amiano P, et al. Adherence to the Mediterranean diet and risk of coronary heart disease in the Spanish EPIC Cohort Study. *American Journal of Epidemiology.* 2009;170:1518-1529.
150. Martínez-González MA, Sánchez-Tainta A, Corella D, Salas-Salvadó J, Ros E, Arós F, et al. A provegetarian food pattern and reduction in total mortality in the Prevención con Dieta Mediterránea (PREDIMED) study. *The American journal of clinical nutrition.* 2014;100 Suppl 1:320-328.
151. Satija A, Bhupathiraju SN, Rimm EB, Spiegelman D, Chiuve SE, Borgi L, et al. Plant-Based Dietary Patterns and Incidence of Type 2 Diabetes in US Men and Women: Results from Three Prospective Cohort Studies. *PLoS medicine.* 2016;13.
152. Toft U, Kristoffersen LH, Lau C, Borch-Johnsen K, Jørgensen T. The Dietary Quality Score: validation and association with cardiovascular risk factors: the Inter99 study. *European Journal of Clinical Nutrition.* 2007;61:270-278.
153. Rostgaard-Hansen A, Lau C, Halkjær J, Olsen A, Toft U. An updated validation of the Dietary Quality Score: associations with risk factors for cardiometabolic diseases in a Danish population. *European Journal of Nutrition.* 2023:1-10.
154. Olsen A, Egeberg R, Halkjær J, Christensen J, Overvad K, Tjønneland A. Healthy aspects of the Nordic diet are related to lower total mortality. *The Journal of nutrition.* 2011;141:639-644.
155. Hansen CP, Overvad K, Kyrø C, Olsen A, Tjønneland A, Johnsen SP, et al. Adherence to a Healthy Nordic Diet and Risk of Stroke: A Danish Cohort Study. *Stroke.* 2017;48:259-264.

156. Vitale M, Masulli M, Calabrese I, Rivellese AA, Bonora E, Signorini S, et al. Impact of a Mediterranean Dietary Pattern and Its Components on Cardiovascular Risk Factors, Glucose Control, and Body Weight in People with Type 2 Diabetes: A Real-Life Study. *Nutrients.* 2018;10.
157. Schramm S, Bramming M, Davidsen M, Rosendahl H, Tolstrup J. Sygdomsbyrden i Danmark:-risikofaktorer. 2023.
158. Keys A, Menotti A, Karvonen MJ, Aravanis C, Blackburn H, Buzina R, et al. The diet and 15-year death rate in the seven countries study. *American Journal of Epidemiology.* 1986;124:903-915.
159. Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, Gnardellis C, Lagiou P, Polychronopoulos E, et al. Diet and overall survival in elderly people. *British Medical Journal (Clinical research ed).* 1995;311:1457-1460.
160. Bach A, Serra-Majem L, Carrasco JL, Roman B, Ngo J, Bertomeu I, et al. The use of indexes evaluating the adherence to the Mediterranean diet in epidemiological studies: a review. *Public Health Nutrition.* 2006;9:132-146.
161. Serra-Majem L, Trichopoulou A, de la Cruz JN, Cervera P, Álvarez AG, La Vecchia C, et al. Does the definition of the Mediterranean diet need to be updated? *Public Health Nutrition.* 2004;7:927-929.
162. Rosato V, Temple NJ, La Vecchia C, Castellan G, Tavani A, Guercio V. Mediterranean diet and cardiovascular disease: a systematic review and meta-analysis of observational studies. *European Journal of Nutrition.* 2019;58:173-191.
163. Estruch R, Ros E, Salas-Salvadó J, Covas MI, Corella D, Arós F, et al. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts. *The New England journal of medicine.* 2018;378:e34.
164. Salas-Salvadó J, Bulló M, Babio N, Martínez-González M, Ibarrola-Jurado N, Basora J, et al. Reduction in the incidence of type 2 diabetes with the Mediterranean diet: results of the PREDIMED-Reus nutrition intervention randomized trial. *Diabetes Care.* 2011;34:14-19.
165. Salas-Salvadó J, Fernández-Ballart J, Ros E, Martínez-González M-A, Fitó M, Estruch R, et al. Effect of a Mediterranean Diet Supplemented With Nuts on Metabolic Syndrome Status: One-Year Results of the PREDIMED Randomized Trial. *Archives of internal medicine.* 2008;168:2449-2458.
166. Mithril C, Dragsted LO, Meyer C, Tetens I, Biltoft-Jensen A, Astrup A. Dietary composition and nutrient content of the New Nordic Diet. *Public Health Nutrition.* 2013;16:777-785.
167. Adamsson V, Reumark A, Cederholm T, Vessby B, Risérus U, Johansson G. What is a healthy Nordic diet? Foods and nutrients in the NORDIET study. *Food & nutrition research.* 2012;56.
168. Kanerva N, Kaartinen NE, Schwab U, Lahti-Koski M, Männistö S. The Baltic Sea Diet Score: a tool for assessing healthy eating in Nordic countries. *Public Health Nutrition.* 2014;17:1697-1705.
169. Uusitupa M, Hermansen K, Savolainen MJ, Schwab U, Kolehmainen M, Brader L, et al. Effects of an isocaloric healthy Nordic diet on insulin sensitivity, lipid profile and inflammation markers in metabolic syndrome - a randomized study (SYSDIET). *Journal of internal medicine.* 2013;274:52-66.
170. Massara P, Zurbau A, Glenn AJ, Chiavaroli L, Khan TA, Viguiliouk E, et al. Nordic dietary patterns and cardiometabolic outcomes: a systematic review and meta-analysis of prospective cohort studies and randomised controlled trials. *Diabetologia.* 2022;65:2011-2031.

171. Huang T, Yang B, Zheng J, Li G, Wahlqvist ML, Li D. Cardiovascular disease mortality and cancer incidence in vegetarians: a meta-analysis and systematic review. *Annals of nutrition & metabolism.* 2012;60:233-240.
172. Kim H, Lee K, Rebholz CM, Kim J. Plant-based diets and incident metabolic syndrome: Results from a South Korean prospective cohort study. *PLoS medicine.* 2020;17:e1003371.
173. Margetts BM, Nelson M. Measuring dietary exposure in nutritional epidemiological studies. *Nutrition research reviews.* 1995;8:165-178.
174. Thompson FE, Subar AF. Dietary assessment methodology. In: *Nutrition in the Prevention and Treatment of Disease*. Elsevier Inc. 2017:5-48.
175. Michels KB. Nutritional epidemiology--past, present, future. *International journal of epidemiology.* 2003;32:486-488.
176. Willet W. Nature of variation in diet. In: *Nutritional Epidemiology.* 3rd ed. Oxford University Press; 2012:33-49.
177. Hebert JR, Clemow L, Pbert L, Ockene IS, Ockene JK. Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. *International journal of epidemiology.* 1995;24:389-398.
178. Hebert JR, Ma Y, Clemow L, Ockene IS, Saperia G, Stanek EJ, III, et al. Gender Differences in Social Desirability and Social Approval Bias in Dietary Self-report. *American Journal of Epidemiology.* 1997;146:1046-1055.
179. Lafay L, Mennen L, Basdevant A, Charles MA, Borys JM, Eschwège E, et al. Does energy intake underreporting involve all kinds of food or only specific food items? Results from the Fleurbaix Laventie Ville Santé (FLVS) study. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity.* 2000;24:1500-1506.
180. Picó C, Serra F, Rodríguez AM, Keijer J, Palou A. Biomarkers of Nutrition and Health: New Tools for New Approaches. *Nutrients*. 2019;11:1092.
181. Corella D, Ordovás JM. Biomarkers: background, classification and guidelines for applications in nutritional epidemiology. *Nutricion hospitalaria.* 2015;31.
182. Cross AJ, Lampe JW, Rock CL, Boushey CJ. Biomarkers and their use in nutrition intervention. In: *Nutrition in the prevention and treatment of disease.* Elsevier inc. 2017:217-234.
183. Woodside JV, Draper J, Lloyd A, McKinley MC. Use of biomarkers to assess fruit and vegetable intake. *The Proceedings of the Nutrition Society.* 2017;76:308-315.
184. Lampe JW, Huang Y, Neuhouser ML, Tinker LF, Song X, Schoeller DA, et al. Dietary biomarker evaluation in a controlled feeding study in women from the Women's Health Initiative cohort. *The American journal of clinical nutrition.* 2017;105:466-475.
185. Gibson RS. Validity in dietary assessment methods. In: *Principles of nutritional assessment.* Oxford university press; 2005:149-196.
186. Guasch-Ferré M, Bhupathiraju SN, Hu FB. Use of Metabolomics in Improving Assessment of Dietary Intake. *Clinical Chemistry.* 2018;64:82-98.
187. Willett W. Food Frequency Methods. In: *Nutritional Epidemiology.* 3rd ed. Oxford University Press; 2012:70–95.
188. Baranowski T. 24-Hour Recall and Diet Record Methods. In: *Nutritional Epidemiology.* 3rd ed. Oxford University Press; 2012:49–69.
189. Nelson M, Black AE, Morris JA, Cole TJ. Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. *The American journal of clinical nutrition.* 1989;50:155-167.

190. Ma Y, Olendzki BC, Pagoto SL, Hurley TG, Magner RP, Ockene IS, et al. Number of 24-hour diet recalls needed to estimate energy intake. *Annals of epidemiology.* 2009;19:553-559.
191. Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires - a review. *Public Health Nutr.* 2002;5:567-587.
192. Pérez Rodrigo C, Aranceta J, Salvador G, Varela-Moreiras G. Food frequency questionnaires. *Nutricion hospitalaria.* 2015;31 Suppl 3:49-56.
193. Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: current state of the art and future prospects. *F1000Research.* 2017;6:926.
194. Carter MC, Albar SA, Morris MA, Mulla UZ, Hancock N, Evans CE, et al. Development of a UK Online 24-h Dietary Assessment Tool: myfood24. *Nutrients.* 2015;7:4016-4032.
195. Subar AF, Kirkpatrick SI, Mittl B, Zimmerman TP, Thompson FE, Bingley C, et al. The Automated Self-Administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. *Journal of the Academy of Nutrition and Dietetics.* 2012;112:1134-1137.
196. Foster E, Lee C, Imamura F, Hollidge SE, Westgate KL, Venables MC, et al. Validity and reliability of an online self-report 24-h dietary recall method (Intake24): a doubly labelled water study and repeated-measures analysis. *Journal of nutritional science.* 2019;8:e29.
197. Liu B, Young H, Crowe FL, Benson VS, Spencer EA, Key TJ, et al. Development and evaluation of the Oxford WebQ, a low-cost, web-based method for assessment of previous 24 h dietary intakes in large-scale prospective studies. *Public Health Nutrition.* 2011;14:1998-2005.
198. Willett W. Correction for the effects of measurement error. In: *Nutritional Epidemiology.* 3rd ed. Oxford University Press; 2012:287-304.
199. Gibson RS. Measurement errors in dietary assessment In: *Principles of nutritional assessment.* Oxford university press; 2005:105-128.
200. Bennett DA, Landry D, Little J, Minelli C. Systematic review of statistical approaches to quantify, or correct for, measurement error in a continuous exposure in nutritional epidemiology. *BMC medical research methodology.* 2017;17:146.
201. Dietary Assessment Primer, Effects of Measurement Error. . https://dietassessmentprimer.cancer.gov/concepts/error/error-effects.html. Accessed 08.04.2023.
202. Gibson RS. Introduction. In: *Principles of nutritional assessment.* Oxford university press; 2005:1-19.
203. Dietary Assessment Primer, Types of Measurement Error. https://dietassessmentprimer.cancer.gov/concepts/error/error-effects.html. Accessed 08.04.2023.
204. Speakman JR, Yamada Y, Sagayama H, Berman ESF, Ainslie PN, Andersen LF, et al. A standard calculation methodology for human doubly labeled water studies. *Cell reports Medicine.* 2021;2:100203.
205. Margetts BM, Nelson M. *Design concepts in nutritional epidemiology.* OUP Oxford; 1997.
206. Gibson RS. Reproducibility in dietary assessment. In: *Principles of nutritional assessment.* Oxford university press; 2005:129-148.
207. Tsubono Y, Nishino Y, Fukao A, Hisamichi S, Tsugane S. Temporal change in the reproducibility of a self-administered food frequency questionnaire. *American Journal of Epidemiology.* 1995;142:1231-1235.

208. Masson LF, McNeill G, Tomany JO, Simpson JA, Peace HS, Wei L, et al. Statistical approaches for assessing the relative validity of a food-frequency questionnaire: use of correlation coefficients and the kappa statistic. *Public Health Nutrition.* 2003;6:313-321.
209. Lombard MJ, Steyn NP, Charlton KE, Senekal M. Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. *Nutrition journal.* 2015;14:40.
210. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in clinical research.* 2017;8:187-191.
211. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical methods in medical research.* 1999;8:135-160.
212. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-310.
213. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine.* 2016;15:155-163.
214. Breier M, Wahl S, Prehn C, Fugmann M, Ferrari U, Weise M, et al. Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples. *PLoS One.* 2014;9:e89728.
215. Lynch SV, Pedersen O. The Human Intestinal Microbiome in Health and Disease. *The New England journal of medicine.* 2016;375:2369-2379.
216. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. *Nature.* 2011;473:174-180.
217. Piquer-Esteban S, Ruiz-Ruiz S, Arnau V, Diaz W, Moya A. Exploring the universal healthy human gut microbiota around the World. *Computational and structural biotechnology journal.* 2022;20:421-433.
218. Uebanso T, Shimohata T, Mawatari K, Takahashi A. Functional Roles of B-Vitamins in the Gut and Gut Microbiome. *Molecular Nutrition & Food Research*. 2020;64.
219. Martin AM, Sun EW, Rogers GB, Keating DJ. The Influence of the Gut Microbiome on Host Metabolism Through the Regulation of Gut Hormone Release. *Frontiers in physiology.* 2019;10:428.
220. Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, et al. Gut microbiota functions: metabolism of nutrients and other food components. *European Journal of Nutrition.* 2018;57:1-24.
221. Wu HJ, Wu E. The role of gut microbiota in immune homeostasis and autoimmunity. *Gut microbes.* 2012;3:4-14.
222. Suárez-Martínez C, Santaella-Pascual M, Yagüe-Guirao G, Martínez-Graciá C. Infant gut microbiota colonization: influence of prenatal and postnatal factors, focusing on diet. *Frontiers in microbiology.* 2023;14:1236254.
223. Vilchez-Vargas R, Skieceviciene J, Lehr K, Varkalaite G, Thon C, Urba M, et al. Gut microbial similarity in twins is driven by shared environment and aging. *EBioMedicine.* 2022;79:104011.
224. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature.* 2018;555:210-215.
225. Odamaki T, Kato K, Sugahara H, Hashikura N, Takahashi S, Xiao J-z, et al. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC Microbiology.* 2016;16:90.

226. de la Cuesta-Zuluaga J, Kelley ST, Chen Y, Escobar JS, Mueller NT, Ley RE, et al. Age- and Sex-Dependent Patterns of Gut Microbial Diversity in Human Adults. *mSystems.* 2019;4.
227. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. *Journal of the National Cancer Institute.* 2013;105:1907-1911.
228. Takagi T, Naito Y, Inoue R, Kashiwagi S, Uchiyama K, Mizushima K, et al. Differences in gut microbiota associated with age, sex, and stool consistency in healthy Japanese subjects. *Journal of gastroenterology.* 2019;54:53-63.
229. Kim YS, Unno T, Kim BY, Park MS. Sex Differences in Gut Microbiota. *The world journal of men's health.* 2020;38:48-60.
230. Sinha T, Vich Vila A, Garmaeva S, Jankipersadsing SA, Imhann F, Collij V, et al. Analysis of 1135 gut metagenomes identifies sex-specific resistome profiles. *Gut microbes.* 2019;10:358-366.
231. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature.* 2013;500:541-546.
232. Allin KH, Nielsen T, Pedersen O. Mechanisms in endocrinology: Gut microbiota in patients with type 2 diabetes mellitus. *European journal of endocrinology.* 2015;172:167-177.
233. Novakovic M, Rout A, Kingsley T, Kirchoff R, Singh A, Verma V, et al. Role of gut microbiota in cardiovascular diseases. *World journal of cardiology.* 2020;12:110-122.
234. Gao Z, Guo B, Gao R, Zhu Q, Qin H. Microbiota disbiosis is associated with colorectal cancer. *Frontiers in microbiology.* 2015;6:20.
235. Claesson MJ, Cusack S, O'Sullivan O, Greene-Diniz R, de Weerd H, Flannery E, et al. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proceedings of the National Academy of Sciences of the United States of America.* 2011;108.
236. Chen L, Wang D, Garmaeva S, Kurilshikov A, Vich Vila A, Gacesa R, et al. The long-term genetic stability and individual specificity of the human gut microbiome. *Cell.* 2021;184:2302-2315.
237. Roager HM, Licht TR, Poulsen SK, Larsen TM, Bahl MI. Microbial enterotypes, inferred by the prevotella-to-bacteroides ratio, remained stable during a 6-month randomized controlled diet intervention with the new nordic diet. *Applied and environmental microbiology.* 2014;80:1142-1149.
238. Aagaard K, Petrosino J, Keitel W, Watson M, Katancik J, Garcia N, et al. The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology.* 2013;27:1012-1022.
239. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et al. The long-term stability of the human gut microbiota. *Science.* 2013;341.
240. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature.* 2014;505:559-563.
241. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nature microbiology.* 2018;3:1255-1265.
242. Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Mannerås-Holm L, et al. Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. *Nature medicine.* 2017;23:850-858.
243. Aya V, Flórez A, Perez L, Ramírez JD. Association between physical activity and changes in intestinal microbiota composition: A systematic review. *PLoS One.* 2021;16.

244. Antinozzi M, Giffi M, Sini N, Gallè F, Valeriani F, De Vito C, et al. Cigarette Smoking and Human Gut Microbiota in Healthy Adults: A Systematic Review. *Biomedicines.* 2022;10.
245. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. *Science.* 2016;352:560-564.
246. Lim MY, Rho M, Song YM, Lee K, Sung J, Ko G. Stability of gut enterotypes in Korean monozygotic twins and their association with biomarkers and diet. *Scientific reports.* 2014;4:7348.
247. García-Montero C, Fraile-Martínez O, Gómez-Lahoz AM, Pekarek L, Castellanos AJ, Noguerales-Fraguas F, et al. Nutritional Components in Western Diet Versus Mediterranean Diet at the Gut Microbiota-Immune System Interplay. Implications for Health and Disease. *Nutrients.* 2021;13.
248. Mitsou EK, Kakali A, Antonopoulou S, Mountzouris KC, Yannakoulia M, Panagiotakos DB, et al. Adherence to the Mediterranean diet is associated with the gut microbiota pattern and gastrointestinal characteristics in an adult population. *British Journal of Nutrition.* 2017;117:1645-1655.
249. Emoto T, Yamashita T, Sasaki N, Hirota Y, Hayashi T, So A, et al. Analysis of Gut Microbiota in Coronary Artery Disease Patients: a Possible Link between Gut Microbiota and Coronary Artery Disease. *Journal of atherosclerosis and thrombosis.* 2016;23:908-921.
250. Cui L, Zhao T, Hu H, Zhang W, Hua X. Association Study of Gut Flora in Coronary Heart Disease through High-Throughput Sequencing. *BioMed research international.* 2017;2017:3796359.
251. Zhu Q, Gao R, Zhang Y, Pan D, Zhu Y, Zhang X, et al. Dysbiosis signatures of gut microbiota in coronary artery disease. *Physiological genomics.* 2018;50:893-903.
252. Cui X, Ye L, Li J, Jin L, Wang W, Li S, et al. Metagenomic and metabolomic analyses unveil dysbiosis of gut microbiota in chronic heart failure patients. *Scientific reports.* 2018;8:635.
253. Luedde M, Winkler T, Heinsen FA, Rühlemann MC, Spehlmann ME, Bajrovic A, et al. Heart failure is associated with depletion of core intestinal microbiota. *ESC heart failure.* 2017;4:282-290.
254. Jie Z, Xia H, Zhong S-L, Feng Q, Li S, Liang S, et al. The gut microbiome in atherosclerotic cardiovascular disease. *Nature Communications.* 2017;8:845.
255. Xu Z, Jiang W, Huang W, Lin Y, Chan FKL, Ng SC. Gut microbiota in patients with obesity and metabolic disorders - a systematic review. *Genes & nutrition.* 2022;17:2.
256. Naik SS, Ramphall S, Rijal S, Prakash V, Ekladios H, Mulayamkuzhiyil Saju J, et al. Association of Gut Microbial Dysbiosis and Hypertension: A Systematic Review. *Cureus.* 2022;14.
257. van den Munckhof ICL, Kurilshikov A, Ter Horst R, Riksen NP, Joosten LAB, Zhernakova A, et al. Role of gut microbiota in chronic low-grade inflammation as potential driver for atherosclerotic cardiovascular disease: a systematic review of human studies. *Obesity reviews : an official journal of the International Association for the Study of Obesity.* 2018;19:1719-1734.
258. Nogal A, Valdes AM, Menni C. The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. *Gut microbes.* 2021;13:1-24.
259. Warmbrunn MV, Herrema H, Aron-Wisnewsky J, Soeters MR, Van Raalte DH, Nieuwdorp M. Gut microbiota: a promising target against cardiometabolic diseases. *Expert review of endocrinology & metabolism.* 2020;15:13-27.

260. Chakaroun RM, Olsson LM, Bäckhed F. The potential of tailoring the gut microbiome to prevent and treat cardiometabolic disease. *Nature Reviews Cardiology.* 2023;20:217-235.
261. Roager HM, Dragsted LO. Diet-derived microbial metabolites in health and disease. *Nutrition Bulletin*. 2019;44:216-227.
262. Chakaroun RM, Olsson LM, Bäckhed F. The potential of tailoring the gut microbiome to prevent and treat cardiometabolic disease. *Nature Reviews Cardiology.* 2023;20:217-235.
263. Schroeder BO, Bäckhed F. Signals from the gut microbiota to distant organs in physiology and disease. *Nature medicine.* 2016;22:1079-1089.
264. Hague A, Butt AJ, Paraskeva C. The role of butyrate in human colonic epithelial cells: an energy source or inducer of differentiation and apoptosis? *Proceedings of the Nutrition Society.* 1996;55:937-943.
265. Peng L, Li ZR, Green RS, Holzman IR, Lin J. Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein kinase in Caco-2 cell monolayers. *The Journal of nutrition.* 2009;139:1619-1625.
266. Hong YH, Nishimura Y, Hishikawa D, Tsuzuki H, Miyahara H, Gotoh C, et al. Acetate and propionate short chain fatty acids stimulate adipogenesis via GPCR43. *Endocrinology.* 2005;146:5092-5099.
267. Xiong Y, Miyamoto N, Shibata K, Valasek MA, Motoike T, Kedzierski RM, et al. Short-chain fatty acids stimulate leptin production in adipocytes through the G protein-coupled receptor GPR41. *Proceedings of the National Academy of Sciences*. 2004;101:1045-1050.
268. Byrne CS, Chambers ES, Morrison DJ, Frost G. The role of short chain fatty acids in appetite regulation and energy homeostasis. *Int J Obes (Lond).* 2015;39:1331-1338.
269. Tang R, Li L. Modulation of Short-Chain Fatty Acids as Potential Therapy Method for Type 2 Diabetes Mellitus. *The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie medicale.* 2021;2021:6632266.
270. Li M, van Esch BCAM, Wagenaar GTM, Garssen J, Folkerts G, Henricks PAJ. Pro- and anti-inflammatory effects of short chain fatty acids on immune and endothelial cells. *European Journal of Pharmacology.* 2018;831:52-59.
271. Diether NE, Willing BP. Microbial Fermentation of Dietary Protein: An Important Factor in Diet–Microbe–Host Interaction. *Microorganisms.* 2019;7.
272. Sanchez-Gimenez R, Ahmed-Khodja W, Molina Y, Peiró OM, Bonet G, Carrasquer A, et al. Gut Microbiota-Derived Metabolites and Cardiovascular Disease Risk: A Systematic Review of Prospective Cohort Studies. *Nutrients.* 2022;14.
273. Zeisel SH, Warrier M. Trimethylamine N-Oxide, the Microbiome, and Heart and Kidney Disease. *Annual Review of Nutrition.* 2017;37:157-181.
274. Cai YY, Huang FQ, Lao X, Lu Y, Gao X, Alolga RN, et al. Integrated metagenomics identifies a crucial role for trimethylamine-producing Lachnoclostridium in promoting atherosclerosis. *NPJ biofilms and microbiomes.* 2022;8:11.
275. Zhu Y, Li Q, Jiang H. Gut microbiota in atherosclerosis: focus on trimethylamine N-oxide. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica.* 2020;128:353-366.
276. Pushpass R-AG, Alzoufairi S, Jackson KG, Lovegrove JA. Circulating bile acids as a link between the gut microbiota and cardiovascular health: impact of prebiotics, probiotics and polyphenol-rich foods. *Nutrition research reviews.* 2022;35:161-180.

277. Wassenaar TM, Zimmermann K. Lipopolysaccharides in Food, Food Supplements, and Probiotics: Should We be Worried? *European journal of microbiology & immunology.* 2018;8:63-69.
278. Petersen KEN, Halkjær J, Loft S, Tjønneland A, Olsen A. Cohort profile and representativeness of participants in the Diet, Cancer and Health-Next Generations cohort study. *European journal of epidemiology.* 2022;37:117-127.
279. Tjønneland A, Olsen A, Boll K, Stripp C, Christensen J, Engholm G, et al. Study design, exposure variables, and socioeconomic determinants of participation in Diet, Cancer and Health: a population-based prospective cohort study of 57,053 men and women in Denmark. *Scandinavian journal of public health.* 2007;35:432-441.
280. Rostgaard-Hansen AL, Rosthøj S, Brunius C, Olsen SF, Bjerregaard AA, Cade JE, et al. Relative Validity and Reproducibility of a Web-Based Semi-Quantitative Food Frequency Questionnaire in the Danish Diet, Cancer, and Health-Next Generations MAX Study. *Nutrients.* 2023;15.
281. Mikkelsen TB, Osler M, Olsen SF. Validity of protein, retinol, folic acid and n-3 fatty acid intakes estimated from the food-frequency questionnaire used in the Danish National Birth Cohort. *Public Health Nutrition.* 2006;9:771-778.
282. Tjønneland A, Overvad K, Haraldsdóttir J, Bang S, Ewertz M, Jensen OLEM. Validation of a Semiquantitative Food Frequency Questionnaire Developed in Denmark. *International journal of epidemiology.* 1991;20:906-912.
283. Slimani N, Deharveng G, Charrondière RU, van Kappel AL, Ocké MC, Welch A, et al. Structure of the standardized computerized 24-h diet recall interview used as reference method in the 22 centers participating in the EPIC project. *Computer methods and programs in biomedicine.* 1999;58:251-266.
284. Ygil KH. *Mål, vægt og portionsstørrelser på fødevarer.* Søborg: DTU Fødevareinstituttet, Afdeling for Ernæring; 2013.
285. Mikkelsen TB, Osler M, Olsen SF. Validity of protein, retinol, folic acid and n–3 fatty acid intakes estimated from the food-frequency questionnaire used in the Danish National Birth Cohort. *Public health nutrition.* 2006;9:771-778.
286. Fødevaredata (frida.fooddata.dk), version 4, 2019. Food, Technical University of Denmark; 2019. https://frida.fooddata.dk/.
287. Lauritsen J. FoodCalc 1.3. 1998.
288. Fødevaredata (frida.fooddata.dk), version 2, 2017. Food, Technical University of Denmark; 2017. https://frida.fooddata.dk/.
289. Livsmeddelverket. Swedish Food Agency; 2017. https://www.livsmedelsverket.se/.
290. McCance RA, Widdowson EM. *McCance and Widdowson's the Composition of Foods.* Royal Society of Chemistry; 2014.
291. Robertson CE. McCance and Widdowson's The Composition of Foods - Sixth Summary Edition. In. Vol 28. Oxford, UK: Blackwell Publishing Ltd; 2003:81-83.
292. Carter MC, Hancock N, Albar SA, Brown H, Greenwood DC, Hardie LJ, et al. Development of a New Branded UK Food Composition Database for an Online Dietary Assessment Tool. *Nutrients.* 2016;8.
293. [Uddannelsesklassifikation (DISCED-15), afsluttede uddannelser, v1:2023]. 2023; https://www.dst.dk/da/Statistik/dokumentation/nomenklaturer/disced15-audd. Accessed 14.09, 2023.
294. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods.* 2016;13:581-583.

295. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research.* 2013;41:D590-596.
296. Callahan B. DADA2 Pipeline Tutorial. 2023; https://benjjneb.github.io/dada2/training.html. Accessed 20.10.2023.
297. Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet.* 1986;327:307-310.
298. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *The American journal of clinical nutrition.* 1997;65:1220S-1228S; discussion 1229S-1231S.
299. Kline RB. *Principles and practice of structural equation modeling.* 4th ed. Guilford publications; 2015.
300. Molag ML, de Vries JHM, Ocké MC, Dagnelie PC, van den Brandt PA, Jansen MCJF, et al. Design Characteristics of Food Frequency Questionnaires in Relation to Their Validity. *American Journal of Epidemiology.* 2007;166:1468-1478.
301. Labonté M, Cyr A, Baril-Gravel L, Royer MM, Lamarche B. Validity and reproducibility of a web-based, self-administered food frequency questionnaire. *European Journal of Clinical Nutrition.* 2012;66:166-173.
302. Fallaize R, Forster H, Macready AL, Walsh MC, Mathers JC, Brennan L, et al. Online dietary intake estimation: reproducibility and validity of the Food4Me food frequency questionnaire against a 4-day weighed food record. *Journal of medical Internet research.* 2014;16:190.
303. Kristal AR, Kolar AS, Fisher JL, Plascak JJ, Stumbo PJ, Weiss R, et al. Evaluation of web-based, self-administered, graphical food frequency questionnaire. *Journal of the Academy of Nutrition and Dietetics.* 2014;114:613-621.
304. Alawadhi B, Fallaize R, Franco RZ, Hwang F, Lovegrove J. Web-Based Dietary Intake Estimation to Assess the Reproducibility and Relative Validity of the EatWellQ8 Food Frequency Questionnaire: Validation Study. *JMIR formative research.* 2021;5:e13591.
305. El Mesmoudi N, Al Dhaheri AS, Feehan J, Stojanovska L, Ali HI. Validation of a quantitative web-based food frequency questionnaire to assess dietary intake in the adult Emirati population. *PLoS One.* 2022;17.
306. Knudsen VK, Hatch EE, Cueto H, Tucker KL, Wise L, Christensen T, et al. Relative validity of a semi-quantitative, web-based FFQ used in the 'Snart Forældre' cohort - a Danish study of diet and fertility. *Public Health Nutrition.* 2016;19:1027-1034.
307. Kato E, Takachi R, Ishihara J, Ishii Y, Sasazuki S, Sawada N, et al. Online version of the self-administered food frequency questionnaire for the Japan Public Health Center-based Prospective Study for the Next Generation (JPHC-NEXT) protocol: Relative validity, usability, and comparison with a printed questionnaire. *Journal of epidemiology.* 2017;27:435-446.
308. Medin AC, Carlsen MH, Hambly C, Speakman JR, Strohmaier S, Andersen LF. The validity of a web-based FFQ assessed by doubly labelled water and multiple 24-h recalls. *British Journal of Nutrition.* 2017;118:1106-1117.
309. Affret A, El Fatouhi D, Dow C, Correia E, Boutron-Ruault MC, Fagherazzi G. Relative Validity and Reproducibility of a New 44-Item Diet and Food Frequency Questionnaire Among Adults: Online Assessment. *Journal of medical Internet research.* 2018;20:e227.
310. Vaz JS, Buffarini R, Schneider BC, Bielemann RM, Gonçalves H, Assunção MCF. Relative validity of a computer-based semi-quantitative FFQ for use in the Pelotas (Brazil) Birth Cohort Studies. *Public Health Nutrition.* 2021;24:34-42.

311. Azarias HGA, Marques-Rocha JL, Miranda A, Dos Santos LC, Gomes Domingos AL, Hermsdorff HHM, et al. Online Food Frequency Questionnaire From the Cohort of Universities of Minas Gerais (CUME Project, Brazil): Construction, Validity, and Reproducibility. *Frontiers in nutrition.* 2021;8.
312. Christensen SE, Möller E, Bonn SE, Ploner A, Bälter O, Lissner L, et al. Relative validity of micronutrient and fiber intake assessed with two new interactive meal- and Web-based food frequency questionnaires. *Journal of medical Internet research.* 2014;16:59.
313. Christensen SE, Möller E, Bonn SE, Ploner A, Wright A, Sjölander A, et al. Two new meal- and web-based interactive food frequency questionnaires: validation of energy and macronutrient intake. *Journal of medical Internet research.* 2013;15:109.
314. Carlsen MH, Andersen LF, Hjartåker A. Reproducibility and feasibility of an online self-administered food frequency questionnaire for use among adult Norwegians. *Food & nutrition research.* 2021;65.
315. Cui Q, Xia Y, Wu Q, Chang Q, Niu K, Zhao Y. A meta-analysis of the reproducibility of food frequency questionnaires in nutritional epidemiological studies. *International Journal of Behavioral Nutrition and Physical Activity.* 2021;18:12.
316. Haghighatdoost F, Sarrafzadegan N, Mohammadifard N, Sajjadi F, Maghroon M, Boshtam M, et al. Healthy Eating Index and Cardiovascular Risk Factors among Iranians. *Journal of the American College of Nutrition.* 2013;32:111-121.
317. van Lee L, Feskens EJ, Meijboom S, Hooft van Huysduynen EJ, van't Veer P, de Vries JH, et al. Evaluation of a screener to assess diet quality in the Netherlands. *British Journal of Nutrition.* 2016;115:517-526.
318. Masip G, Keski-Rahkonen A, Pietilainen KH, Kujala UM, Rottensteiner M, Vaisanen K, et al. Development of a Food-Based Diet Quality Score from a Short FFQ and Associations with Obesity Measures, Eating Styles and Nutrient Intakes in Finnish Twins. *Nutrients.* 2019;11.
319. Cleghorn CL, Harrison RA, Ransley JK, Wilkinson S, Thomas J, Cade JE. Can a dietary quality score derived from a short-form FFQ assess dietary quality in UK adult population surveys? *Public health nutrition.* 2016;19:2915-2923.
320. England CY, Thompson JL, Jago R, Cooper AR, Andrews RC. Development of a brief, reliable and valid diet assessment tool for impaired glucose tolerance and diabetes: the UK Diabetes and Diet Questionnaire. *Public Health Nutrition.* 2017;20:191-199.
321. Robinson SM, Jameson KA, Bloom I, Ntani G, Crozier SR, Syddall H, et al. Development of a Short Questionnaire to Assess Diet Quality among Older Community-Dwelling Adults. *The journal of nutrition, health & aging.* 2017;21:247-253.
322. Persson CE, Rothenberg E, Hansson PO, Welin C, Strandhagen E. Cardiovascular risk factors in relation to dietary patterns in 50-year-old men and women: a feasibility study of a short FFQ. *Public Health Nutrition.* 2019;22:645-653.
323. Zhu LB, Zhang YC, Huang HH, Lin J. Prospects for clinical applications of butyrate-producing bacteria. *World journal of clinical pediatrics.* 2021;10:84-92.
324. Lotfi K, Saneei P, Hajhashemy Z, Esmaillzadeh A. Adherence to the Mediterranean Diet, Five-Year Weight Change, and Risk of Overweight and Obesity: A Systematic Review and Dose–Response Meta-Analysis of Prospective Cohort Studies. *Advances in Nutrition.* 2022;13:152-166.
325. Skreden M, Hillesund ER, Wills AK, Brantsæter AL, Bere E, Øverby NC. Adherence to the New Nordic Diet during pregnancy and subsequent maternal weight development: a study conducted in the Norwegian Mother and Child Cohort Study (MoBa). *British Journal of Nutrition.* 2018;119:1286-1294.

326. Ellulu MS, Patimah I, Khaza'ai H, Rahmat A, Abed Y. Obesity and inflammation: the linking mechanism and the complications. *Archives of medical science : AMS.* 2017;13:851-863.
327. Illner AK, Freisling H, Boeing H, Huybrechts I, Crispim SP, Slimani N. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. *International journal of epidemiology.* 2012;41:1187-1203.
328. Zazpe I, Santiago S, De la Fuente-Arrillaga C, Nuñez-Córdoba JM, Bes-Rastrollo M, Martínez-González MA. Paper-Based Versus Web-Based Versions of Self-Administered Questionnaires, Including Food-Frequency Questionnaires: Prospective Cohort Study. *JMIR public health and surveillance.* 2019;5.
329. Lo Siou G, Csizmadi I, Boucher BA, Akawung AK, Whelan HK, Sharma M, et al. The Comparative Reliability and Feasibility of the Past-Year Canadian Diet History Questionnaire II: Comparison of the Paper and Web Versions. *Nutrients.* 2017;9.
330. Prentice RL, Vasan S, Tinker LF, Neuhouser ML, Navarro SL, Raftery D, et al. Metabolomics Biomarkers for Fatty Acid Intake and Biomarker-Calibrated Fatty Acid Associations with Chronic Disease Risk in Postmenopausal Women. *The Journal of Nutrition.* 2023;153:2663-2677.
331. Christensen AI, Davidsen M, Ekholm O, Hansen SE, Algren MH, Juel K. *Den Nationale Sundhedsprofil 2010-Hvordan har du det?* : Sundhedsstyrelsen; 2011.
332. Rosendahl H, Davidsen M, Møller SR, Román JEI, Kragelund K, Christensen AI, et al. Danskernes sundhed: Den nationale sundhedsprofil 2021.
333. Yeoh YK, Sun Y, Ip LYT, Wang L, Chan FKL, Miao Y, et al. Prevotella species in the human gut is primarily comprised of Prevotella copri, Prevotella stercorea and related lineages. *Scientific reports.* 2022;12:9055.
334. Vuik F, Dicksved J, Lam SY, Fuhler GM, van der Laan L, van de Winkel A, et al. Composition of the mucosa-associated microbiota along the entire gastrointestinal tract of human individuals. *United European gastroenterology journal.* 2019;7:897-907.
335. Rothman KJ GS. Precision and validity in epidemiologic studie In: *Modern Epidemiology.* Lippencott - Raven Publishers; 1998.
336. Jensen HAR, Davidsen M, Ekholm O, Christensen AI. Danskernes sundhed-Den nationale sundhedsprofil 2017. 2018.
337. Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, Riboli E, et al. Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutrition.* 2002;5:915-923.
338. Hébert JR. Social Desirability Trait: Biaser or Driver of Self-Reported Dietary Intake? *Journal of the Academy of Nutrition and Dietetics.* 2016;116:1895-1898.
339. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. *Nature.* 2020;587:448-454.

APPENDICES

APPENDIX 1

Overview of different dietary patterns and associated genera or species.

Dietary pattern	Adherence	Abundance	Genus or species	Reference[19-28]
Omnivores	NA	Only omnivores	***Acidaminococcus***	Ruengsomwong 2016
Mediterranean	Higher adherence	Lower abundance	***Acidaminococcus intestini***	Peters 2023
Healthy plant-based	Higher adherence	Lower abundance	***Acidaminococcus intestini***	Peters 2023
Vegetarian/vegan	NA	Only vegetarian	*Acinetobacter*	Ruengsomwong 2016
Omnivores	NA	Higher abundance	*Akkermansia*	Ruengsomwong 2016
Omnivores	NA	Higher abundance	***Alistipes***	Ruengsomwong 2016
Mediterranean	Lower adherence	Higher abundance	***Alistipes timonensis***	Róses 2021
Mediterranean	Higher adherence	Higher abundance	*Anaerostipes hadrus*	Peters 2023
Mediterranean	Lower adherence	Higher abundance	*Anaerotruncus colihominis*	Róses 2021
Healthy Nordic	Higher adherence	Lower abundance	*Bacilli*	Gaundal 2022
Mediterranean	Higher adherence	Higher abundance	*Bacterium LF-3*	Peters 2023
Healthy plant-based	Higher adherence	Higher abundance	*Bacterium LF-3*	Peters 2023
Vegetarian/vegan	NA	Lower abundance	***Bacteroides***	Zimmer 2012
Omnivores	NA	Higher abundance	***Bacteroides***	Ruengsomwong 2016
Mediterranean	Higher adherence	Higher abundance	***Bacteroides cellulosilyticus***	Róses 2021
Mediterranean	Higher adherence	Higher abundance	***Bacteroides cellulosilyticus***	Wang 2021
Omnivores	NA	Higher abundance	***Bacteroides dorei***	Ruengsomwong 2016
Mediterranean	Lower adherence	Higher abundance	***Bacteroides ovatus***	Garcia-Mantrana 2018
Healthy Nordic	Higher adherence	Higher abundance	***Bacteroides stercoris***	Gaundal 2022
Omnivores	NA	Higher abundance	***Bacteroides thethiotaomicron***	Ruengsomwong 2016
Omnivores	NA	Higher abundance	***Bacteroides uniformis***	Ruengsomwong 2016
Mediterranean	Lower adherence	Higher abundance	***Bacteroides uniformis***	Garcia-Mantrana 2018
Omnivores	NA	Higher abundance	***Bacteroides vulgatus***	Ruengsomwong 2016
Vegetarian/vegan	NA	Lower abundance	***Bifidobacterium***	Zimmer 2012
Mediterranean	Higher adherence	Higher abundance	***Bifidobacterium animalis***	Róses 2021
Mediterranean	Higher adherence	Lower abundance	***Bifidobacterium bifidum***	Peters 2023
Healthy plant-based	Higher adherence	Higher abundance	*Blautia*	Miao 2022
Vegetarian/vegan	NA	Only vegetarian	*Bulleidia*	Ruengsomwong 2016
Mediterranean	Lower adherence	Higher abundance	***Butyricicoccus pullicaecorum***	Róses 2021
Mediterranean	Higher adherence	Higher abundance	***Butyrivibrio crossotus***	Peters 2023
Healthy plant-based	Higher adherence	Higher abundance	***Butyrivibrio crossotus***	Peters 2023
Vegetarian/vegan	NA	Only vegetarian	*Caldimonas Elusimicrobium*	Ruengsomwong 2016
Mediterranean	Higher adherence	Higher abundance	*Candida albicans*	Mitsou 2017
Mediterranean	Higher adherence	Higher abundance	*Catabacter hongkongensis*	Róses 2021
Mediterranean	Higher adherence	Higher abundance	*Catenibacterium*	Garcia-Mantrana 2018
Plant-based	Higher adherence	Lower abundance	*Catenisphaera*	Miao 2022
Mediterranean	Lower adherence	Higher abundance	*Christensenella minuta*	Róses 2021
Mediterranean	Higher adherence	Lower abundance	***Clostridium***	Garcia-Mantrana 2018
Omnivores	NA	Higher abundance	***Clostridium clostridioforme***	Ruengsomwong 2016
Vegetarian/vegan	NA	Lower abundance	***Clostridium coccides***	Kabeerdoss 2012
Mediterranean	Higher adherence	Higher abundance	***Clostridium sp L2 50***	Wang 2021
Mediterranean	Higher adherence	Lower abundance	*Collinsella aerofaciens*	Wang 2021
Mediterranean	Higher adherence	Higher abundance	*Coprococcus eutactus*	Peters 2023
Healthy plant-based	Higher adherence	Lower abundance	*Dorea*	Miao 2022
Vegetarian/vegan	NA	Lower abundance	*Enterobacteriaceae*	Zimmer 2012
Mediterranean	Higher adherence	Higher abundance	*Erysipelatoclostridium ramosum*	Róses 2021
Vegetarian/vegan	NA	Lower abundance	***Escherichia coli***	Zimmer 2012
Mediterranean	Higher adherence	Lower abundance	***Escherichia coli***	Mitsou 2017
Omnivores	NA	Higher abundance	***Escherichia hermannii***	Ruengsomwong 2016
Healthy Nordic	Higher adherence	Lower abundance	***Eubacterium biforme***	Gaundal 2022
Mediterranean	Higher adherence	Higher abundance	***Eubacterium eligens***	Peters 2023
Healthy plant-based	Higher adherence	Higher abundance	***Eubacterium eligens***	Peters 2023
Mediterranean	Higher adherence	Higher abundance	***Eubacterium eligens***	Wang 2021
Mediterranean	Lower adherence	Higher abundance	***Eubacterium saphenum***	Róses 2021

Unhealthy plant-based	Higher adherence	Lower abundance	***Eubacterium xylanophilum group***	Miao 2022
Unhealthy plant-based	Higher adherence	Higher abundance	*Exiguobacterium*	Miao 2022
Unhealthy plant-based	Higher adherence	Higher abundance	*F0332*	Miao 2022
Mediterranean	Higher adherence	Higher abundance	*Faecalibacterium prausnitzii*	Wang 2021
Mediterranean	Lower adherence	Higher abundance	*Gordonibacter pamelaeae*	Róses 2021
Omnivores	NA	Higher abundance	*Halomonas*	De Moraes 2017
Vegetarian/vegan	NA	Higher abundance	*Klebsiella*	Ruengsomwong 2016
Vegetarian/vegan	NA	Higher abundance	*Klebsiella pneumoniae*	Ruengsomwong 2016
Mediterranean	Higher adherence	Higher abundance	***Lachnospiraceae bacterium TF01-11***	Peters 2023
Healthy plant-based	Higher adherence	Higher abundance	***Lachnospiraceae bacterium TF01-11***	Peters 2023
Healthy Nordic	Higher adherence	Lower abundance	*Lactobacillus spp.*	Gaundal 2022
Mediterranean	Higher adherence	Lower abundance	***Megasphaera massiliensis***	Peters 2023
Healthy plant-based	Higher adherence	Lower abundance	***Megasphaera massiliensis***	Peters 2023
Healthy plant-based	Higher adherence	Lower abundance	*Mogibacterium timidum*	Peters 2023
Mediterranean	Higher adherence	Higher abundance	*Oscillibacter valericigenes*	Róses 2021
Mediterranean	Higher adherence	Higher abundance	*Oscillospira (Flavonifractor) plautii*	Róses 2021
Mediterranean	Higher adherence	Higher abundance	*Papillibacter cinnamivorans*	Róses 2021
Omnivores	NA	Higher abundance	***Parabacteroides***	Ruengsomwong 2016
Omnivores	NA	Higher abundance	***Parabacteroides distasonis***	Ruengsomwong 2016
Mediterranean	Lower adherence	Higher abundance	***Parabacteroides goldsteinii***	Róses 2021
Mediterranean	Higher adherence	Higher abundance	***Paraprevotella clara***	Róses 2021
Omnivores	NA	Only omnivores	*Pediococcus*	Ruengsomwong 2016
Omnivores	NA	Only omnivores	*Peptoniphilus*	Ruengsomwong 2016
Plant-based	Higher adherence	Lower abundance	*Peptostreptococcus*	Miao 2022
Healthy plant-based	Higher adherence	Higher abundance	***Polynucleobacter***	Miao 2022
Unhealthy plant-based	Higher adherence	Lower abundance	***Polynucleobacter***	Miao 2022
Vegetarian/vegan	NA	Higher abundance	***Prevotella***	Ruengsomwong 2016
Vegetarian/vegan	NA	Higher abundance	***Prevotella***	De Moraes 2017
Vegetarian/vegan	NA	Higher abundance	***Prevotella copri***	Ruengsomwong 2016
Mediterranean	Lower adherence	Higher abundance	***Prevotella corporis***	Róses 2021
Vegetarian/vegan	NA	Lower abundance	***Roseburia***	Kabeerdoss 2012
Omnivores	NA	Lower abundance	***Roseburia***	De Moraes 2017
Mediterranean	Higher adherence	Higher abundance	***Roseburia faecis***	Róses 2021
Mediterranean	Higher adherence	Higher abundance	***Roseburia hominis***	Peters 2023
Mediterranean	Higher adherence	Higher abundance	***Roseburia intestinalis***	Peters 2023
Healthy plant-based	Higher adherence	Higher abundance	***Roseburia intestinalis***	Peters 2023
Healthy plant-based	Higher adherence	Higher abundance	*Ruminococcaceae UCG009*	Miao 2022
Mediterranean	Higher adherence	Higher abundance	***Ruminococcus bromii***	Róses 2021
Mediterranean	Higher adherence	Lower abundance	***Ruminococcus gnavus***	Wang 2021
Mediterranean	Higher adherence	Higher abundance	***Ruminococcus lactaris***	Peters 2023
Mediterranean	Higher adherence	Lower abundance	***Ruminococcus torques***	Wang 2021
Healthy Nordic	Higher adherence	Lower abundance	*Streptococcus salivarius spp. (thermophilus)*	Gaundal 2022
Omnivores	NA	Only omnivores	***Succinicibrio***	Ruengsomwong 2016
Omnivores	NA	Higher abundance	***Succinivibrio***	De Moraes 2017
Mediterranean	Lower adherence	Higher abundance	***Succinivibrio dextrinosolvens***	Róses 2021
Omnivores	NA	Only omnivores	*Turicibacter*	Ruengsomwong 2016
Unhealthy plant-based	Higher adherence	Higher abundance	*ZOR0006*	Miao 2022

APPENDIX 2

Food items from the DCH-NG 376-item FFQ in each food group used in paper I.

Food groups	**Food items from the 376-item FFQ**
Fruits	Lemon, lemon juice, orange, grapefruit, lime, tangerine, apple, pear, banana, peach, nectarine, kiwi fruit, plum, watermelon, honeydew melon, cantaloup melon, muskmelon, grape, mango, papaya, pineapple, pomegranate, kaki fruit, passion fruit, strawberry, blueberry, raspberry, blackberry, currant, cherry, gooseberry, rhubarb, canned fruit, dried fruit
Vegetables	Spinach, lettuce, cucumber, squash/zucchini, tomato, tomato (canned), tomato puree, tomato ketchup, aubergine, avocado, sweet pepper, olives, peas, green beans, sweetcorn, radish, beet, beet (canned), carrot, horse-radish, Jerusalem artichoke, parsnip, parsley root, swede, turnip, kohlrabi, celeriac, ginger root, cauliflower, brussels sprouts, broccoli, cabbage, kale, mushroom, onion, garlic, spring onion, asparagus, asparagus (canned), chives, parsley, leek, celery, basil
Potatoes	Potato, potato flour
Legumes	Chickpeas, beans, lentils
Eggs	Eggs, egg yolk, egg white
Poultry	Chicken, turkey, duck, goose, turkey minced, chicken sliced, turkey sliced, chicken nuggets
Red meat	Veal, beef, beef minced, pork, pork minced, lamb, meat balls, liver, heart
Processed red meat	Roast beef sliced, pork liver paste, liver pate, sausage, salami, smoked ham, ham sliced, veal sliced, bacon
Fast food	Pommes frites, pizza, burger, fried spring rolls
Fish and seafood	Cod, plaice, flounder, saithe, tuna, striped catfish, shrimp, oyster, crayfish, lobster, tuna (canned), shrimp (canned), caviar, rainbow trout, charr, garfish, cod roe, mussel, crab, cod roe (canned), herring, mackerel, salmon, halibut, mackerel (smoked, canned), herring (pickled), bucklingpaté, sardine (canned), salmon (smoked), halibut (smoked)
Dairy products	Skimmed milk, reduced fat milk, whole milk, chocolate milk, cream, icecream,
Fermented dairy products	Buttermilk, yoghurt, cheese, sour cream
Fat	Olive oil, rape seed oil, sunflower oil, thistle oil, grapeseed oil, corn oil, coconut oil, peanut oil, sesame oil, linseed oil, palm oil, butter, blended spread, margarine, pork lard, goose fat
Soft drinks	Lemonade/iced tea with and without sugar, soft drink with and without sugar
Coffee	Filter coffee, instant coffee, French press coffee, espresso, decaf or grain coffee, caffe latte, cappuccino
Tea	Green tea, white tea, black tea, rooibos tea, herbal tea, chai latte
Whole grain products[a]	Rolled oats, spelt flakes, rye flakes, oatmeal/wholegrain porridge, cereal products, rye bread, wholemeal bread/buns, crispbread

[a] Amount of whole grains were calculated from wholegrain products

APPENDIX 3

Food items from the 23-item FFQ and 376-item FFQ included in the food groups: fish, red meat, vegetables and fruits.

Food groups	Food items from the 23-item FFQ	Food items from the 376-item FFQ
Fish	Cold cuts – fish, meals with fish	Cod, plaice, flounder, saithe, tuna, striped catfish, shrimp, oyster, crayfish, lobster, tuna (canned), shrimp (canned), caviar, rainbow trout, charr, garfish, cod roe, mussel, crab, cod roe (canned), herring, mackerel, salmon, halibut, mackerel (smoked, canned), herring (pickled), bucklingpaté, sardine (canned), salmon (smoked), halibut (smoked)
Red meat	Cold cuts – meat, meals with beef, veal, pork or lamb	Veal, beef, beef minced, pork, pork minced, lamb, meat balls, liver, heart
Vegetables	Salads, other raw vegetables, cooked vegetables	Spinach, lettuce, cucumber, squash/zucchini, tomato, tomato (canned), tomato puree, tomato ketchup, tomato soup, aubergine, avocado, sweet pepper, olives, peas, green beans, sweetcorn, radish, beet, beet (canned), carrot, horse-radish, Jerusalem artichoke, parsnip, parsley root, swede, turnip, kohlrabi, celeriac, ginger root, cauliflower, brussels sprouts, broccoli, cabbage, kale, mushroom, onion, garlic, spring onion, asparagus, asparagus (canned), chives, parsley, leek, celery, basil
Fruits	Fruit	Lemon, lemon juice, orange, grapefruit, lime, apple, pear, banana, peach, nectarine, kiwi fruit, plum, watermelon, honeydew melon, cantaloup melon, muskmelon, grape, mango, papaya, pineapple, pomegranate, kaki fruit, passion fruit, strawberry, blueberry, raspberry, blackberry, currant, cherry, gooseberry, rhubarb

APPENDIX 4

Food items from the 24-HDRs included in food groups for each diet index in paper III and IV, modified from paper III.

Diet index	Food group	Food items from the 24-HDR
Healthy Nordic Food Index (HNFI)	Fish	Fish fatty fresh, fish patty processed, fish fatty dish, fish lean fresh, fish lean processed, fish lean dish
	Cabbage	Cabbages
	Root vegetables	Root vegetables
	Apples and pears	Apple raw/stewed/baked without sugar, pears raw/stewed without sugar, dried pear and apple, apple and pear stewed or backed with sugar
	Whole grain (oats)	Rolled oats, oat porridge made with water or milk
	Whole grain (rye)	Rye bread, rye crispbread, rye flour porridge, rye crushed grains or flakes
Relative Mediterranean Diet score (rMED)	Fruits (including nuts and seeds)	Citrus fruits, fruits dried, other fruits, fruiting vegs, fruits preserved, nuts and seeds, fruiting vegs
	Vegetables (excluding potatoes)	Cabbages, fruiting vegs, leafy vegs, mushrooms, onion garlic, other root vegs, sauces dips dressings, stalk vegs sprouts, vegetables dish, sauces dips dressings
	Legumes	Legumes, bouillon, fruiting vegs, sauces dips dressings, nuts and seeds, stalk vegs sprouts, vegetable dish, sauces dips dressings
	Fish (including fresh or frozen, excluding fish products and preserved fish	Fish lean fresh, fish fatty fresh, fish lean dish, fish fatty dish
	Cereals (unrefined and whole grain)	Cereals refined, cereals whole grains, vegetable dish
	Total meat (meat and meat products)	Poultry, poultry dish, poultry proc, processed meat, processed meat dish, red meat, red meat dish, vegetable dish
	Dairy products	Butter, dairy products fatty, dairy products lean, ice creams, sauces dips dressings
	Olive oil	Olive oil
	Alcohol	Wine, beer, spirits and brandy, alcopops
Plant-based Diet Index (PDI), healthy Plant-based Diet Index (hPDI), unhealthy Plant-based Diet Index (uPDI), Provegetarian index (pro-veg)[a]	Whole grains	Cereal wholegrains, cereal wholergrains
	Fruits	Citrus fruits, fruits dried, other fruits, fruits preserved
	Vegetables	Cabbages, fruiting vegs, leafy vegs, mushrooms, onion garlic, other root vegs, seaweed, stalk vegs sprouts, vegetable dish, sauces dips dressings
	Nuts	Nuts and seeds
	Legumes	Legumes, sauces dips dressings, stalk vegs sprouts, vegetable dish
	Vegetable oils	Vegetable oils
	Tea and coffee[b]	Coffee, tea
	Refined grains	Cereals refined, vegetable dish
	Potatoes	Crisps, potatoes, potatoes fatty, other root vegs
	Fruit and vegetable juices[b]	Fruit juices, fruit juice, vegetables juice
	Sugar sweetened beverages[b]	Soft drinks, soft drinks light, sugar jam syrups
	Sweets and desserts[b]	Cakes and biscuits, chocolate candy bars, confect non choc, desserts, sugar jam syrups
	Animal fat	Butter, margarines, other animal fat
	Egg	Eggs
	Dairy	Butter, dairy products fatty, dairy products lean, ice cream, sauces dips dressings
	Fish and seafood	Fish fatty fresh, fish fatty proc, fish lean fresh, fish lean proc
	Meat	Poultry, poultry proc, proc meat, red meat, red meat dish
	Miscellaneous animal-based foods[b]	Fish lean dish, fish fatty dish, poultry dish, proc meat dish, red meat dish, vegetable dish, eggs, sauces dips dressings

[a] PDI, hPDI, uPDI and pro-veg are not included in paper IV.
[b] Food groups not included in pro-veg.