The Role of Nutrition and Gut Microbiome in Type 2 Diabetes Risk

Zhangling Chen

The Role of Nutrition and Gut Microbiome in Type 2 Diabetes Risk

Zhangling Chen

The studies described in this thesis were performed within the Rotterdam Study and the Lifelines-Deep Study. We gratefully acknowledge the contributions of participants, research staff, data management, and health professionals of all studies.

Publication of this thesis was kindly supported by the Department of Epidemiology of Erasmus Medical Center and by Erasmus University Rotterdam. Additional financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

ISBN: 978-94-6332-579-0

Layout: Zhangling Chen and Loes Kema

Cover design: Loes Kema and Stevan Stojic

Print: GVO drukkers & vormgevers B.V.

The Role of Nutrition and Gut Microbiome in Type 2 Diabetes Risk

De rol van voeding en darmmicrobioom in type 2 diabetes risico

Thesis

to obtain the degree of Doctor from the

Erasmus University Rotterdam

by command of the rector magnificus

Prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Committee.

The public defense shall be held on

Tuesday 17th of December 2019 at 15:30 hours

by

Zhangling Chen

born in Guang'an, China

Erasmus University Rotterdam

Erasmus

Doctoral Committee

Promotor:	Prof.dr. M.A. Ikram
Other members:	Prof.dr. A.G. Uitterlinden
	Prof.dr. H. Boersma
	Dr. S.S. Soedamah-Muthu
Co-promotor:	Dr.ir. T. Voortman

Paranymphs

S. Lamballais

R. Zou

To my father

TABLE OF CONTENTS

MANUSCRIPTS BASED ON THE STUDIES DESCRIBED IN THIS THESIS

Chapter 2.1

Chen Z, Franco OH, Lamballais S, Ikram MA, Schoufour JD, Muka T, Voortman T. Associations of specific dietary protein with longitudinal insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. Clinical Nutrition. 2019. DOI: 10.1016/j.clnu.2019.01.021

Chapter 2.2

Chen Z, Glisic M, Song M, Aliahmad HA, Zhang X, Moumdjian AC, Gonzalez-Jaramillo V, Van der Schaft N, Bramer WM, Ikram MA, Voortman T. Dietary protein intake and all-cause and cause-specific mortality: results from the Rotterdam Study and a meta-analysis of prospective cohort studies (Under review).

Chapter 2.3

Chen Z*, Zuurmond MG*, Van der Schaft N, Nano J, Wijnhoven HAH, Ikram MA, Franco OH, Voortman T. Plant versus animal-based diets and insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. European Journal of Epidemiology. 2018;33(9):883-93.

Chapter 2.4

Chen Z, Schoufour JD, Rivadeneira F, Lamballais S, Ikram MA, Franco OH, Voortman T. Plant-based diet and adiposity over time in a middle-aged and elderly population: the Rotterdam Study. Epidemiology. 2019;30(2):303-10.

Chapter 3.1

Chen Z*, Radjabzadeh D*, Chen L*, Klurilshikov A, Ikram MA, Uitterlinden A, Zhernakova A, Fu J, Kraaij R, Voortman T. Gut microbiome, insulin resistance and type 2 diabetes: results from two large population-based studies (Manuscript).

Chapter 4.1

Chen Z, Radjabzadeh D, Ikram MA, Uitterlinden A, Kraaij R, Voortman T. Diet quality and gut microbiome: a large population-based study (Manuscript).

**denotes equal contribution*

Chapter 1

General Introduction

INTRODUCTION

Type 2 diabetes (T2D) is a common metabolic disease characterized by hyperglycemia. At present, more than 380 million people live with T2D.[1] T2D has been estimated as the sixth leading cause of death, largely attributable to high blood glucose and increased risks of cardiovascular diseases and other complications, which put a huge burden on health-care systems.[2] The epidemiology of T2D is influenced by multiple risk factors including multiple genetic, environmental, and behavioral factors (Table 1).[3] These multiple risk factors together fuel the development of T2D by possibly inducing pathophysiological defects in target organs or organ systems, such as insulin resistance in muscle and adipose tissue (Table 2).[1] In the process of development of T2D, there is a precursor condition referred to as prediabetes that is defined by blood glucose levels higher than normal, but not high enough yet to T2D thresholds.[4] Around 5–10% of people with prediabetes become diabetic every year, although the conversion rate varies with population characteristics and prediabetes definitions.[4]

Table 1. Examples of known risk factors for type 2 diabetes

Modifiable risk factors	Non-modifiable risk factors
Nutrition	Age
Physical inactivity	Sex
Sedentary behavior	Ethnicity
Overweight or obesity	History of gestational diabetes
Socioeconomics	Polycystic ovary syndrome
Components of the metabolic syndrome	Family history of diabetes
Cigarette smoking	Genetic predisposition, such as TCF7L2 gene
Inflammation	
Gut microbiome	
Some medications, such as beta-blockers, thiazides, and statins	

Table 2. Pathophysiological defects of type 2 diabetes

Organs/ organ systems	Pathophysiological defect
Pancreatic α and β cells	Loss of cell mass and function, impaired insulin secretion, dysregulated glucagon secretion, and increased glucagon concentration
Muscle and adipose tissue	Reduced peripheral glucose uptake, insulin resistance
Inflammation	Immune dysregulation
Liver	Increased hepatic glucose output
Kidney	Increased glucose reabsorption caused by of SGLT-2 receptors
Brain	Increased appetite, lack of satiety
Stomach or intestine	Increased rate of glucose absorption
Colon	Unbalanced gut microbiome

Despite increasing knowledge regarding risk factors for T2D, the incidence and prevalence of T2D continues to rise globally.[2] This calls for more effort to further address impact of risk factors on T2D. Nutrition, as a relatively easy modifiable risk factor, has attracted much attention, but much remains unclear.[5] Gut microbiome, a novel risk factor, has been suggested to play an important pathophysiological role in the development of T2D.[6, 7] As gut microbiome composition can be largely influenced by nutrition, and gut microbiome has been linked to T2D, gut microbiome has been proposed as a potential pathway through which nutrition may influence the development of T2D.[8] Therefore, further research on potential role of nutrition and gut microbiome in T2D risk can help provide new insight into etiology, mechanisms and thereby into the prevention, and therapy of T2D.

Nutrition and type 2 diabetes

To date, a large body of human studies has indicated the importance of nutrition in the prevention and management of T2D.[3, 5] Many studies have indicated that dietary macronutrients, such as carbohydrate, protein, and fat may affect T2D risk, which could differ by their specific subtypes.[5] Literature has also indicated that higher intake of certain foods, such as fruits, vegetables, and legumes, and lower intake of for example red and processed meat, are associated with lower T2D risk.[5] Although research on individual nutrients and food items is valuable, people generally do not consume isolated micronutrients or foods. Therefore, in addition to research on nutrients and foods, many researchers have paid much attention to dietary patterns. Evidence has indicated that adherence to some dietary patterns, such as a Mediterranean diet, the Dietary Approach to Stop Hypertension (DASH) diet, and plant-based diets, are associated with lower T2D risk.[9, 10] Overall, much effort and progress have been made in the nutrition research field for prevention of T2D. However, there are still a lot of inconsistencies in previous findings or limited data for certain topics. For example, although high long-term habitual animal protein intake has been consistently linked to higher T2D risk, the results for plant protein and T2D risk are mixed.[11] Furthermore, although associations for the Mediterranean diet and the DASH diet and T2D have been widely and consistently reported, data on plant-based diets are more limited.[12-14] Moreover, these topics have only been studied in certain specific populations, and diet habits are likely to vary according to sex, socioeconomic status, geographical location, ethnic group and culture, and vary over time, which calls for more nutrition research among diverse populations over time to further elucidate associations of nutrition with T2D.[15] Additionally, to better understand the role of nutrition in T2D risk and to identify targets for early prevention, it is reasonable to further explore associations of nutritional factors with risk factors and earlier stages of T2D, such as obesity, insulin resistance, and prediabetes, for which, to date less studies have been performed.

Gut microbiome and type 2 diabetes

The human gut microbiome is composed of bacteria, archaea, viruses and eukaryotic microbes that reside in and on our gut. These trillions of gut microorganisms reside in a complex ecosystem which operates as a "hidden organ" to influence our health and diseases.[16] New technologies, such as rapid nucleic acid sequencing, and advanced statistical technologies, have provided powerful tools to help our understanding of the gut microbiome. Recently, some studies have indicated that gut microbiome may play an important role in T2D.[6, 7, 17-19] For example, compared to non-diabetic participants, T2D patients have less alpha diversity in their gut microbiome composition.[20] Lean male donor fecal microbiota transplantation in males with metabolic syndrome resulted in a significant improvement in insulin sensitivity, along with an increased gut microbial diversity, including a distinct increase in butyrate-producing bacterial strains.[21] However, these previous studies had some limitations. They were limited by small sample size, by unclear inclusion and exclusion criteria of participants, and by their lack of control for important confounders, such as physical activity or social economic status.[6, 17-19] Furthermore, given most of these studies were conducted under trial conditions with a small number of specific participants, it is unclear whether these findings are applicable to real-world settings. Therefore, large population-based studies examining associations between gut microbiome composition and T2D risk with comprehensive adjustment for confounders are needed to further elucidate the role of gut microbiome in T2D risk in real-life settings.[17]

Nutrition and gut microbiome

Ongoing efforts have suggested that gut microbiome composition is modifiable and that it can be largely influenced by nutrition.[22, 23] However, these efforts have been mainly concentrated in researching the role of certain individual nutrients, such as fiber intake.[24] To date, few studies have examined the role of habitual overall diet in the gut microbiome composition in population-based settings.[25] To extend and update evidence on the role of diet in gut microbiome composition, well-conducted, large population-based studies considering key confounders, such as socioeconomic status, smoking and other lifestyle factors, are needed. Combined with ongoing research on gut microbiome and T2D risk, research on how nutrition affects gut microbiome could better help in developing strategies for prevention and treatment of T2D.

THIS THESIS

Objectives

The aim of this thesis was to study the role of nutrition and gut microbiome in T2D risk. To better unravel the role of nutrition and gut microbiome in T2D risk, I also included major risk factor and earlier stages of T2D, including adiposity, insulin resistance, and prediabetes (Figure 1). Therefore, the objectives were:

1. To examine associations of nutritional factors with adiposity, insulin resistance, prediabetes, T2D, and mortality.

2. To investigate associations between gut microbiome composition with insulin resistance and T2D.

3. To examine associations between nutritional factors and gut microbiome composition.

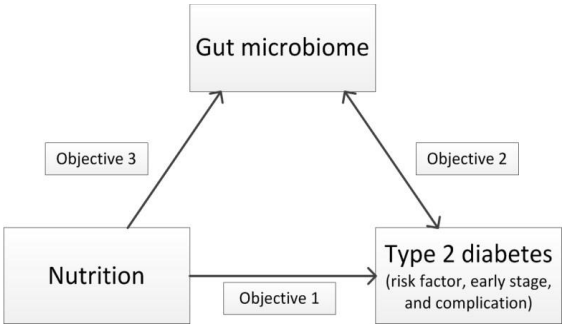


Figure 1. Overview of objectives of this thesis

Study design

Studies presented in this thesis were mainly carried out in the Rotterdam Study. These analyses were extended with analyses in the Lifelines-Deep Study and with a systematic review of existing literature.

The Rotterdam Study

The Rotterdam Study is a large ongoing population-based prospective cohort study among individuals aged ≥ 45 years in Ommoord district of Rotterdam, the Netherlands. The rationale and design of the Rotterdam Study are described in detail elsewhere.[26] Briefly, so far, a total of 14926 individuals of Ommoord district have been included in the three sub-cohorts of the study. The first sub-cohort,

Rotterdam Study-I (RS-I), was launched in 1990 and recruited 7983 inhabitants of the Ommoord district aged 55 years or older; the second sub-cohort, Rotterdam Study-II (RS-II), started in 2000 and included 3011 inhabitants of the Ommoord district aged 55 years or above; the third sub-cohort, Rotterdam Study-III (RS-III) started in 2006 by recruiting 3932 inhabitants in the study district with age 45 years or above. Upon entering the study, the participants underwent home-structured interviews and a series of examinations in our research center every 3-5 year. The Rotterdam Study has been approved by the Medical Ethics Committee according to the Wet Bevolkingsonderzoek: ERGO (Population Study Act: Rotterdam Study), executed by the Ministry of Health, Welfare and Sports of The Netherlands. All participants gave informed consent.

The Lifelines-Deep Study

The Lifelines-Deep Study is a sub-cohort of the Lifelines Cohort Study, a population-based cohort including all age groups living in the three provinces in the northern region of the Netherlands: Groningen, Friesland and Drenthe.[27] From 2006 through 2013, over 167000 individuals registered in the Lifelines Cohort Study. These participants received follow-up examinations every 5 years. From April to August 2013, 1539 Lifelines participants aged ≥ 18 years were invited to participate in the Lifelines-Deep Study. In the Lifelines-Deep Study, additional examinations were performed, including collection of fecal samples for gut microbiome composition. The Lifelines-Deep Study was approved by the ethics committee of the University Medical Centre Groningen. All participants signed an informed consent prior to enrolment.[28]

Systematic review and Meta-analysis

For Chapter 2.2, I conducted a systematic review and meta-analysis to include and pool results from several prospective cohorts. For the systematic review and meta-analysis, we performed extensive literature searches in several databases, including Medline via Ovid, EMBASE, Web of Science Core Collection, Cochrane CENTRAL via Wiley, PubMed and Google Scholar. No limits were set on language or year of publication. In order to identify additional relevant articles, the reference lists of the included studies and reviews were screened as well. We screened eligible articles and extracted data from individual studies by two independent reviewers. Finally, we pooled data from individual studies including the Rotterdam Study using a random-effects meta-analysis model.[29]

THESIS OUTLINE

Subsequent to this general introduction (Chapter 1), Chapter 2 of this thesis mainly focuses on the role of nutrition in T2D. Chapter 2.1 describes dietary protein intake in relation to insulin resistance, and risk of prediabetes and T2D in the Rotterdam Study. Chapter 2.2 demonstrates dietary protein

intake linked to risk of all-cause mortality and cause-specific mortality in the Rotterdam Study and a meta-analysis of prospective cohort studies. Chapter 2.3 and 2.4 focus on the associations between a plant-based diet with insulin resistance, risk of prediabetes and T2D (Chapter 2.3), and adiposity over time (Chapter 2.4) in the Rotterdam Study. Chapter 3 investigates the associations between gut microbiome composition and insulin resistance and risk of T2D in the Rotterdam Study and the Lifelines-Deep Study. Chapter 4 describes the association between diet quality and components of diet quality with gut microbiome composition in the Rotterdam Study. Chapter 5 provides an overview of the main findings from all studies in this thesis. Furthermore, in this chapter, I discuss the implications of our findings, methodological considerations of the studies and directions of future research.

REFERENCES

1. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. The Lancet. 2017;389(10085):2239-51.
2. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. 2016;388(10053):1545-602.
3. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nature Reviews Endocrinology. 2018;14(2):88.
4. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. The Lancet. 2012;379(9833):2279-90.
5. Ley SH, Hamdy O, Mohan V, Hu FB. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. The Lancet. 2014;383(9933):1999-2007.
6. Hartstra AV, Bouter KEC, Bäckhed F, Nieuwdorp M. Insights into the role of the microbiome in obesity and type 2 diabetes. Diabetes care. 2015;38(1):159-65.
7. Komaroff AL. The microbiome and risk for obesity and diabetes. Jama. 2017;317(4):355-6.
8. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. Human nutrition, the gut microbiome and the immune system. Nature. 2011;474(7351):327.
9. InterAct C. Mediterranean diet and type 2 diabetes risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) study: the InterAct project. Diabetes care. 2011;34(9):1913-8.
10. De Koning L, Chiuve SE, Fung TT, Willett WC, Rimm EB, Hu FB. Diet-quality scores and the risk of type 2 diabetes in men. Diabetes care. 2011;34(5):1150-6.
11. Shang X, Scott D, Hodge AM, English DR, Giles GG, Ebeling PR, et al. Dietary protein intake and risk of type 2 diabetes: results from the Melbourne Collaborative Cohort Study and a meta-analysis of prospective studies. The American Journal of Clinical Nutrition. 2016:ajcn140954.
12. Chen Z, Zuurmond MG, van der Schaft N, Nano J, Wijnhoven HAH, Ikram MA, et al. Plant versus animal based diets and insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. European journal of epidemiology. 2018;33(9):883-93.
13. 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 Med. 2016;13(6):e1002039.
14. Chen G-C, Koh W-P, Neelakantan N, Yuan J-M, Qin L-Q, van Dam RM. Diet Quality Indices and Risk of Type 2 Diabetes Mellitus: The Singapore Chinese Health Study. American journal of epidemiology. 2018;187(12):2651-61.
15. Hu FB, Willett WC. Current and Future Landscape of Nutritional Epidemiologic Research. Jama. 2018.
16. Guinane CM, Cotter PD. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. Therapeutic advances in gastroenterology. 2013;6(4):295-308.
17. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nature Reviews Microbiology. 2018:1.

18. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature. 2015;528(7581):262.
19. Larsen N, Vogensen FK, Van Den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PloS one. 2010;5(2):e9085.
20. Honda K, Littman DR. The microbiome in infectious disease and inflammation. Annual review of immunology. 2012;30:759-95.
21. Vrieze A, Van Nood E, Holleman F, Salojärvi J, Kootte RS, Bartelsman JFWM, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. Gastroenterology. 2012;143(4):913-6. e7.
22. Xu Z, Knight R. Dietary effects on human gut microbiome diversity. British Journal of Nutrition. 2015;113(S1):S1-S5.
23. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. Nature medicine. 2018;24(4):392.
24. 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(6):705-15.
25. 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(12):1645-55.
26. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. European journal of epidemiology. 2017;32(9):807-50.
27. Scholtens S, Smidt N, Swertz MA, Bakker SJL, Dotinga A, Vonk JM, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. International journal of epidemiology. 2014;44(4):1172-80.
28. Tigchelaar EF, Zhernakova A, Dekens JAM, Hermes G, Baranska A, Mujagic Z, et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. BMJ open. 2015;5(8):e006772.
29. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. Bmj. 2011;342:d549.

Chapter 2

Nutrition and Type 2 Diabetes

Chapter 2.1

Dietary protein and type 2 diabetes

Chen Z, Franco OH, Lamballais S, Ikram MA, Schoufour JD, Muka T, Voortman T. Associations of specific dietary protein with longitudinal insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. Clinical Nutrition. 2019. DOI: 10.1016/j.clnu.2019.01.021

ABSTRACT

Background: High protein intake has been linked to increased type 2 diabetes (T2D) risk. However, if this association differs by protein from specific food sources, and if a habitual high protein intake affects insulin resistance and prediabetes risk are largely unknown.

Objectives: We aimed to investigate associations between protein intake from different food sources with longitudinal insulin resistance, and risk of prediabetes and T2D.

Methods: Our analyses included 6822 participants aged ≥45 years without diabetes at baseline in three sub-cohorts of the prospective population-based Rotterdam Study. We measured protein intake at baseline using food-frequency questionnaires. Data on longitudinal homeostatic model assessment of insulin resistance (HOMA-IR), and incidence of prediabetes and T2D were available from 1993-2014.

Results: During follow-up, we documented 931 prediabetes cases and 643 T2D cases. After adjusting for sociodemographic, lifestyle, and dietary factors, higher total protein intake was associated with higher longitudinal HOMA-IR and with higher risk of prediabetes and T2D (per 5% increment in energy from protein at the expense of carbohydrate, for HOMA-IR: β=0.10, (95%CI 0.07, 0.12); for prediabetes: HR=1.34 (1.24 1.44); for T2D: HR=1.37 (1.26, 1.49)). These associations were mainly driven by total animal protein (for HOMA-IR: 0.10 (0.07, 0.12); for prediabetes: 1.35 (1.24, 1.45); for T2D: 1.37 (1.26, 1.49)). The harmful associations of total animal protein were contributed to by protein from meat, fish, and dairy (e.g. for HOMA-IR: protein from meat, 0.13 (0.10, 0.17); from fish, 0.08 (0.03, 0.13); from dairy, 0.04 (0.0003, 0.08)). After additional adjustment for longitudinal waist circumference, associations of total protein and total animal protein with longitudinal HOMA-IR and prediabetes risk were attenuated but remained statistically significant. Total plant protein, as well as protein from legumes and nuts, from grains, from potatoes, or from fruits and vegetables, was not associated with any of the outcomes.

Conclusions: Higher intake of animal protein, from meat, dairy and fish food sources, is associated with higher longitudinal insulin resistance and risk of prediabetes and T2D, which may be partly mediated by obesity over time. Furthermore, plant protein from different sources is not related to insulin resistance, and risk of prediabetes and T2D. Our findings highlight the importance of specific protein food sources and that habitual high animal protein intake may already in early stages be harmful in the development of T2D.

INTRODUCTION

Diet is considered an important component of a healthy lifestyle in the prevention of type 2 diabetes (T2D).[1] One of the dietary factors of interest is protein. Short-term trials have reported beneficial effects of energy-restricted high-protein diet on obesity,[2] an important diabetes risk factor, due to increased satiety and energy expenditure.[3] However, several mechanistic and epidemiological studies have indicated that high levels of certain amino acids metabolized from dietary protein intake, such as branched-chain and aromatic amino acids, adversely affect glucose metabolism and insulin resistance.[4-6] Also, a recent review of eleven prospective cohort studies, reported that overall, higher habitual total protein intake is associated with higher T2D risk.[7] Most studies in the review observed that this positive association was mainly driven by total animal protein,[7] whereas evidence for plant protein is mixed.[7-9]

However, the effect of habitual protein intake on insulin resistance and prediabetes is unknown. T2D has a long asymptomatic continuous physiological process, preceded by insulin resistance, a core defect of the pathogenesis of T2D,[10] and by prediabetes, an early risk stage of T2D.[11] Although previous studies reported associations of protein intake with ultimate T2D risk, pathophysiological mechanisms behind these different earlier risk stages are not completely consistent;[12] thus effects of protein intake on insulin resistance and prediabetes risk might not be the same as for effects on T2D risk. To infer causal relations, longitudinal studies that seek to identify associations of protein intake with insulin resistance and prediabetes are warranted. However, to our knowledge, no studies have directly examined the associations between protein intake with longitudinal insulin resistance and prediabetes risk. Furthermore, almost all previous studies have investigated associations for intake of total protein, total animal protein and total plant protein, but not of protein from more specific food sources, especially for plant protein sources, for which evidence is very inconsistent.[7, 13]

Therefore, we aimed to investigate the associations between protein intake from different food sources in an iso-energetic diet, with longitudinal insulin resistance and risk of prediabetes and T2D in a large Dutch population-based study.

METHODS

Study population

The current study was embedded within the Rotterdam Study (RS), a population-based cohort study including people aged ≥ 45 years living in the Ommoord District of Rotterdam, the Netherlands. Details on the design of the Rotterdam Study are described elsewhere.[14] The cohort consisted of three sub-cohorts. The baseline examination of the first sub-cohort (RS-I) was done in 1989-93 among

participants aged 55 years and over (n= 7983). In 2000-01, the study was extended with a second sub-cohort (RS-II) of new individuals who had aged to 55 years or moved into the study area after 1990 (n=3011). In 2006, a third sub-cohort (RS-III) with new individuals was recruited and included inhabitants aged 45 years and older (n=3932). Follow-up examinations were performed every 3-5 years in each sub-cohort. The study has been approved by the Medical Ethics Committee of Erasmus University Medical Center and all participants gave informed consent.

Population for current analyses

We used data from all three sub-cohorts (Supplemental Figure 1). For 6822 participants who were free of diabetes at baseline (RS-I-1: n=2976, RS-II-1: n=1418, and RS-III-1: n=2428), we had dietary data available at baseline. For the analyses of insulin resistance, from this group (n=6822) we excluded participants without data on homeostatic model assessment of insulin resistance (HOMA-IR) at baseline and follow-up, resulting in 6657 participants (RS-I: n=2899, RS-II: n=1396, RS-III: n=2362). For the analyses of prediabetes risk, from the group (n=6822) we excluded participants with prediabetes at baseline or without follow-up data of prediabetes, resulting in 5795 participants (RS-I: n=2492, RS-II: n=1152, RS-III: n=2151). Finally, for the analyses of T2D, we excluded participants without follow-up data of T2D still from the 6822 participants, resulting in 6813 participants (RS-I: n=2976, RS-II: n=1414, RS-III: n=2423). Data on the outcomes were available from 1993 to 2014.

Assessment of protein intake

At the baseline visits of RS-I and RS-II, food intake data were obtained using a semi-quantitative 170-item food-frequency questionnaire (FFQ). For dietary assessment at baseline in RS-III (2006-08) and for the follow-up measurements in RS-I (RS-I-5, 2009-11) and RS-II (RS-II-3, 2011-12), a semi-quantitative 389-item FFQ was used. Both FFQs were previously validated for nutrient intakes against other dietary assessment methods, for which results have been described elsewhere.[15-17] Food intake data were converted to energy and nutrient intake using the Dutch Food Composition tables 1993, 2001, 2006, and 2011 (NEVO). Intakes of protein and other macronutrients were expressed as percentage of total energy. Data on protein intake at baseline were analyzed in main analyses, and data on protein intake at follow-up in RS-I and RS-II were analyzed in sensitivity analyses.

Assessment of insulin resistance

Fasting blood was drawn at the research center at two time points in each sub-cohort (at RS-I-3 (1997-99) and I-5 (2009-11), at RS-II-1 (2000-01) and II-3 (2011-12), and at RS-III-1 (2006-08) and III-2 (2012-14)). Glucose levels were measured using the glucose hexokinase method. Serum insulin levels were measured on a Roche Modular Analytics E170 analyzer. The HOMA-IR was calculated using the formula: fasting insulin (mU/L) × fasting glucose (mmol/L)/22.5.

Prevalence and incidence of type 2 diabetes and prediabetes

At baseline and during follow-up, information from general practitioners, structured home interviews, pharmacy dispensing records, and follow-up examinations in our research center, was used to identify incident T2D and prediabetes cases. We defined T2D and prediabetes according to WHO guidelines.[18] Participants who fulfilled at least one of the following criteria were diagnosed as incident T2D: 1. Fasting blood glucose concentration of 7.0 mmol/L or higher; 2. Non-fasting blood glucose concentration of 11.1 mmol/L or higher; 3. The use of blood glucose-lowering medications. Prediabetes was defined as having fasting blood glucose between 6.0 and 7.0 mmol/L or non-fasting blood glucose between 7.7 and 11.1 mmol/L. All potential incident T2D and prediabetes were independently identified by two study physicians. In case of disagreement, consensus was sought by consulting endocrinologists. These cases were monitored until 2012.

Assessment of covariates

Information on smoking, medication use, and education levels, was obtained during home interviews at baseline. Waist circumference (WC) was measured at the research center at baseline (RS-I-1 (1989-93), RS-II-1 (2000-01), RS-III-1 (2006-08)) and follow-up period (RS-I-3 (1997-99) and RS-I-5 (2009-11); RS-II-2 (2004-05) and RS-II-3 (2011-12); RS-III-2 (2012-14)). WC was measured at the level midway between the lower rib margin and the iliac crest with the participant in a standing position. Physical activity was assessed with an adapted version of the Zutphen Physical Activity Questionnaire at RS-I-3 and RS-II-1, and with the LASA Physical Activity Questionnaire at RS-III-1. Physical activities were further weighted by their intensity with Metabolic Equivalent of Task (MET), obtained from the 2011 version of the Compendium of Physical Activities. Overall dietary quality was taken into account according to the Dutch Guidelines for a Healthy Diet, for which a sum score for adherence to these dietary guidelines (0-14) was calculated from the FFQ data.[17]

Hypertension at baseline was defined using the following criteria: systolic blood pressure ≥ 140 mmHg; or/and diastolic blood pressure ≥ 90 mmHg; or use of blood pressure-lowering medication. Cardiovascular disease (CVD) at baseline and during follow-up was defined as having a medical record of myocardial infarction, coronary artery bypass surgery, or percutaneous transluminal coronary angioplasty.[19] Serum cholesterol and triacylglycerol were measured at baseline with an automated enzymatic procedure. Information on family history of diabetes was available at RS-I-1 and RS-II-1 and was defined as having at least one parent or sibling with T2D.

Data analysis

Descriptive data were expressed as mean (SD), median (25th percentile–75th percentile), or in percentages. To better approximate normal data distributions, we used natural log-transformed values

for HOMA-IR. We used linear mixed models with a random-effects structure including a random intercept and slope (for time of repeated measurements of HOMA-IR) to analyze the associations between protein intake and longitudinal HOMA-IR. We used Cox proportional hazard models to analyze the associations between protein intake and risk of prediabetes and T2D. Because effects of macronutrient intake cannot be separated from the effects of overall energy intake or intake of other macronutrients, we modelled macronutrient substitution effects.

For all models we used multivariable nutrient density substitution models for protein intake at the expense of carbohydrate.[20] As no indications for non-linear associations for the main models were found, all primary analyses were performed using models assuming linearity. All macronutrients were entered in the models per 5 energy percent (E%).[20, 21] We first examined the associations for total protein intake at the expense of carbohydrate by including total protein intake, total energy, total fat intake, and alcohol, in the model (i.e. all macronutrients except for carbohydrate). Subsequently we examined the associations for total animal and total plant protein intake at the expense of carbohydrate in similar models, for which we mutually adjusted for total animal and total plant protein. The coefficients in these models indicated change in outcomes (e.g. average change in longitudinal HOMA-IR over time) by replacement of carbohydrate by other nutrients. The effect estimate for protein in this model could therefore be interpreted as a 5 E% higher protein intake at the expense of an isocaloric amount of carbohydrates. For all main analyses, model 1 included intake of protein, total fat, total energy, alcohol, baseline age, and sex, and for the analyses of longitudinal HOMA-IR we additionally adjusted for the time of repeated measurements of HOMA-IR; model 2 was additionally adjusted for smoking status, educational levels, diet quality score, physical activity, and family history of diabetes. In model 3, we additionally adjusted for longitudinal WC to account for the potential effect of obesity over time in the potential associations between protein intake with longitudinal insulin resistance, and risk of prediabetes and T2D. Especially in model 3, to account for the potential effect of obesity over time in the associations with risk of prediabetes and T2D, we used novel joint models combing linear mixed model with a random-effects structure including a random intercept and slope (for time of repeated measurements of WC) and cox proportional hazard model. The cox model part of the joint model was the same as model 2 and the linear mixed model part of joint model, in which the outcomes were the repeated measurements of WC before the onset of prediabetes or T2D, was additionally adjusted for the time of longitudinal WC measurements, and the interactions between protein and the time of these repeated measurements.[22]

Effect modification was examined by including interactions of intake of total protein, total animal protein, and total plant protein with age, sex in model 2, and longitudinal WC (continuous data) in model 3. In case of significant interactions ($p<0.05$), the analyses would be stratified. We performed several sensitivity analyses based on model 2 to test whether the associations of total protein, total animal protein, and total plant protein with outcomes were robust. First, we replaced total fat by

carbohydrate, to examine whether it made a difference if protein was consumed at the expense of fat; and we split total fat in subgroups of fatty acids (saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), polyunsaturated fatty acids (PUFA), trans-unsaturated fatty acids (Trans-fat) at the expense of carbohydrate to control the effect of subgroups of fatty acids. Second, we additionally added cholesterol, hypertension, triglycerides, and CVD history, since these factors could mediate associations. Third, we excluded participants who developed CVD during follow-up to exclude the possibility of significant change of diet and lifestyle. Fourth, we examined the associations of total protein, total animal protein, and total plant protein with longitudinal HOMA-IR using data of HOMA-IR that was measured before onset of T2D at follow-up. Last, we further adjusted for protein intake 20 years after baseline in RS-I and 10 years after baseline in RS-II to examine whether our main results were robust after incorporating potential effect of dietary intake at follow-up among the participants with these data available in RS-I and RS-II.

To further explore whether the associations of animal protein and plant protein differ by more specific food sources, we further examined the associations of isocaloric replacement of carbohydrate with protein from meat, protein from dairy, protein from fish, protein from legumes and nuts, protein from potato, protein from grains, and protein from fruits and vegetables with longitudinal insulin resistance, and risk of prediabetes and T2D. In this modelling approach, the percentage of energy intake from protein from meat, dairy, fish, legumes and nuts, potatoes, grains, and fruits and vegetables were simultaneously included in one model, with adjustment for total energy, alcohol, SFA, MUFA, PUFA, Trans-fat, fiber, age, sex, smoking status, educational levels, diet quality score, physical activity, and family history of diabetes for analyses of risk of prediabetes and T2D. The time of longitudinal HOMA-IR measurements was additionally included in the multivariate model for analysis of longitudinal insulin resistance.

All analyses were performed separately for RS-I, RS-II, and RS-III, and the results were pooled using fixed-effects meta-analysis. To adjust for potential bias associated with missing data (Supplemental Table 1), a multiple imputation procedure (n=10) was used for missing data of covariates. Statistical procedures were performed with the use of SPSS statistical software, version 21.0 (IBM Corp, Armonk, NY) and R version 3.3.1 (The R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Baseline characteristics

Characteristics and dietary intakes of the study population are presented in Table 1. Average total protein intake in our population was 85.8 ± 25.1 g/day, this corresponded to 1.20 ± 0.3 g/kg BW/day, which is higher than the recommended intake of 0.8 g/kg BW/day.[23] In our study population, most protein came from animal sources (total animal protein: 53.6 ± 19.0 g/day). Mean percentage of energy intake from total protein was 16.3%, from total animal protein was 10.3%, and from total plant sources was 6.0%. Main animal protein sources were meat and dairy; the main plant protein source was grains. During a median 5.7 years of follow-up, we documented 931 prediabetes cases among 5795 participants. During a median 7.2 years of follow-up, we documented 643 T2D cases among 6813 participants.

Intake of total protein, total animal protein, and total plant protein with insulin resistance, risk of prediabetes and T2D

After multivariable adjustment (Model 2), higher intake of total protein and of total animal protein was associated with higher longitudinal insulin resistance (for per 5 E% higher total protein at the expense of carbohydrate, β=0.10 (95%CI 0.07, 0.12)); for total animal protein at the expense of carbohydrate, β=0.10 (95%CI 0.07, 0.12)) (Table 2). After additional adjustment for longitudinal WC (Model 3), the estimates were attenuated but still statistically significant.

In line with this, higher total protein intake was associated with higher prediabetes risk (hazard ratio (HR) 1.34 (95%CI 1.24, 1.44)), also mainly driven by total animal protein (1.35 (95%CI 1.24, 1.45)) not by total plant protein (Table 3). After additional adjustment for longitudinal WC (model 3), these estimates were slightly attenuated but still statistically significant.

Similarly, higher intake of total protein and total animal, but not total plant protein, was associated with higher T2D risk (HR for total protein 1.37 (95%CI 1.26, 1.49)) (Table 4). After additional adjustment for longitudinal WC, the associations attenuated and no longer statistically significant (e.g. for total protein intake: 1.12 (95%CI 0.96, 1.30)).

Table 1. Baseline characteristics of study participants

	RS-I (n=2976)	RS-II (n=1418)	RS-III (n=2428)	Pooled (n=6822)
Age at dietary assessment (years)	79.1±4.6	72.1±4.9	56.8±6.4	65.4±11.3
Sex, male (%)	40.4	45.0	40.5	41.4
Waist circumference (cm)	88.8±10.8	93.7±26.8	92.5±22.6	91.2±19.7
Physical activity (MET-hours/week)				
- Zutphen Physical Activity Questionnaire	80.7 (55.4-116.3)	77.3 (52.9-105.0)	NA	\
- LASA Physical Activity Questionnaire	NA	NA	42·9 (17·7-82·8)	\
Education level (%)				
- Primary	15.7	7.1	9.5	11.7
- Lower	43.7	45.2	34.4	40.6
- Intermediate	30.2	28.4	27.4	28.8
- Higher	10.0	17.9	28.4	18.2
Smoking (%)				
- Never	28.9	28.9	31.9	32.0
- Ever	47.6	47.6	43.5	44.7
- Current	23.0	23.0	24.4	22.8
Dietary intake				
Energy intake (kcal/day)	1985±505	2155±569	2337±865	2146±685
Carbohydrate intake (g/day)	207.1 (172.7-249.3)	213.7(170.7-263.0)	249.1 (198.7-309.6)	216.1 (123.5-322.9)
Carbohydrate intake (E%)	43.5 (39.1-47.8)	41.7 (36.7-46.1)	45.0 (40.9-49.5)	43.7 (39.1-85.8)
Total fat intake (g/day)	80.2±27.6	88.4±25.9	85.8±49.1	83.9±36.6
Total fat intake (E%)	34.2±6.3	34.6±5.7	29.9±6.6	32.8±6.6
Alcohol intake (g/day)	4.5 (0.35-15.3)	8.7(0.7-22.6)	8.1 (1.4-19.7)	6.5 (0-44.1)
Alcohol intake (E%)	1.6 (0.14, 5.5)	2.8 (0.27-7.1)	2.6 (0.43, 6.1)	2.0 (0.18, 6.2)
Total protein intake (g/day)	81.8±19.4	88.8±25.7	89.0±29.84	85.8±25.13
Total protein intake (E%)	16.8±3.0	16.6±2.9	15.6±2.6	16.3±2.9

Table 1. Baseline characteristics of study participants (Continued)

	RS-I (n=2976)	RS-II (n=1418)	RS-III (n=2428)	Pooled (n=6822)
Total animal protein intake (g/day)	53.1±15.6	57.8±22.0	51.6±20.8	53.6±19.0
Total animal protein intake (E%)	11.0±2.9	10.9±3.3	9.1±2.8	10.3±3.1
- Protein from meat (g/day)	20.7 (15.5, 26.9)	23.7 (17.5, 31.8)	20.8(14.8, 28.6)	21.3 (15.5, 28.4)
- Protein from meat (E%)	4.3 (3.2, 5.5)	4.5 (3.2, 6.1)	3.8 (2.7, 5.0)	4.1 (3.0, 5.5)
- Protein from dairy (g/day)	24.8 (18.3, 32.0)	23.7 (16.1, 32.1)	17.6 (11.8, 24.6)	22.0 (15.0, 29.9)
- Protein from dairy (E%)	5.1 (3.8, 6.7)	4.5 (3.2, 6.0)	3.2 (2.2, 4.3)	4.3 (2.9, 5.8)
- Protein from fish (g/day)	2.0 (0, 4.8)	1.4 (0.2, 3.8)	4.7 (2.4, 8.1)	2.9 (0.6, 5.7)
- Protein from fish (E%)	0.4 (0, 0.9)	0.3 (0.04, 0.7)	0.9 (0.4, 1.4)	0.6 (0.1, 1.1)
Total plant protein intake (g/day)	28.7±8.5	30.9±11.3	37.4±14.9	32.3±12.3
Total plant protein intake (E%)	5.8±1.2	5.7±1.3	6.4±1.3	6.04±1.32
- Protein from grains (g/day)	13.3 (10.5, 16.7)	16.2 (12.0, 21.3)	18.3 (13.6, 24.5)	15.3 (11.5, 20.2)
- Protein from grains (E%)	2.8 (2.2, 3.3)	3.2 (2.5, 3.9)	3.3 (2.6, 4.1)	3.0 (2.4, 3.7)
- Protein from legumes and nuts (g/day)	2.2 (1.2, 4.6)	2.3 (0.7, 5.0)	3.4 (1.7, 6.7)	2.6 (1.2, 5.3)
- Protein from legumes and nuts (E%)	0.5 (0.3, 0.9)	0.4 (0.1, 0.9)	0.6 (0.3,1.1)	0.5 (0.3, 1.0)
- Protein from potato (g/day)	2.8 (1.9, 3.9)	2.5 (1.6, 3.4)	1.9 (1.1, 3.1)	2.4 (1.5, 3.4)
- Protein from potato (E%)	0.6 (0.4, 0.8)	0.5 (0.3, 0.7)	0.4 (0.2, 0.5)	0.5 (0.3, 0.7)
- Protein from vegetables and fruits (g/day)	2.3 (1.7, 2.9)	5.0 (3.6, 6.5)	10.6 (6.6, 15.2)	4.0 (2.3, 8.2)
- Protein from vegetables and fruits (E%)	0.5 (0.3, 0.6)	1.0 (0.7, 1.2)	1.9 (1.3, 2.7)	0.8 (0.5, 1.5)

Values are percentages for categorical variables, mean ± SD for continuous variables with a normal distribution, or median (25th percentile–75th percentile) for continuous variables with a skewed distribution; on the basis of unimputed data.

Abbreviations: RS, Rotterdam Study; SD, standard deviation; E%, percentage of total energy intake; MET, metabolic equivalent of task; NA, not available

Table 2. Associations of protein intake with longitudinal insulin resistance

	RS-I (n=2899)			RS-II (n=1396)			RS-III (n=2362)			Pooled results (n=6657)		
HOMA-IR	**β**	**95%CI**	**P value**	**β**	**95%CI**	**P value**	**β**	**95%CI**	**P value**	**β**	**95%CI**	**P value**
Total protein (per 5 E%)												
Model 1	0.06	0.02, 0.10	0.002	0.09	0.03, 0.14	0.001	0.11	0.06, 0.15	0.001	0.08	0.06, 0.11	0.001
Model 2	0.08	0.04, 0.12	0.001	0.09	0.04, 0.14	0.001	0.13	0.08, 0.17	0.001	0.10	0.07, 0.12	0.001
Model 3	0.001	-0.04, 0.04	0.96	0.07	0.02, 0.12	0.004	0.10	0.06, 0.15	0.001	0.05	0.03, 0.07	0.001
Total animal protein (per 5 E%)												
Model 1	0.07	0.03, 0.11	0.001	0.08	0.03, 0.13	0.003	0.11	0.07, 0.16	0.001	0.09	0.06, 0.11	0.001
Model 2	0.08	0.04, 0.12	0.001	0.08	0.03, 0.13	0.002	0.13	0.08, 0.17	0.001	0.10	0.07, 0.12	0.001
Model 3	0.005	-0.03, 0.04	0.92	0.07	0.02, 0.12	0.001	0.10	0.06, 0.15	0.001	0.05	0.03, 0.07	0.001
Total plant protein (per 5 E%)												
Model 1	-0.04	-0.14, 0.05	0.38	-0.06	-0.20, 0.08	0.40	-0.03	-0.13, 0.07	0.51	-0.04	-0.10, 0.02	0.19
Model 2	0.01	-0.09, 0.12	0.78	-0.02	-0.18, 0.12	0.71	0.03	-0.07, 0.13	0.55	0.01	-0.05, 0.08	0.69
Model 3	-0.06	-0.15, 0.03	0.18	-0.03	-0.17, 0.12	0.71	0.03	-0.06, 0.13	0.50	-0.02	-0.08, 0.04	0.53

Values are β coefficients and 95% confidence intervals (CI) from linear mixed models for the difference of intake of per 5 E% protein in HOMA-IR at the expense of carbohydrate.

Model 1 is adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex and time point of HOMA-IR measurement.

Model 2 is additionally adjusted for smoking status (never, past, or current), education level (primary, lower, intermediate, or higher), diet quality score, physical activity (METh/wk), use of blood glucose-lowering medications in follow-up, and family history of diabetes (for RS-I and RS-II only).

Model 3 is adjusted for all covariates in model 2 and additionally for longitudinal WC.

Abbreviations: WC, waist circumference; HOMA-IR, homeostatic model assessment of insulin resistance; CI, confidence interval; RS, Rotterdam Study; E%, percentage of total energy intake

Table 3. Associations of protein intake with risk of prediabetes

	RS-I (n=2492)			RS-II (n=1152)			RS-III (n=2151)			Pooled results (n=5795)		
Prediabetes	**HR**	**95%CI**	**P value**	**HR**	**95%CI**	**P value**	**HR**	**95%CI**	**P value**	**HR**	**95%CI**	**P value**
Total protein (per 5 E%)												
Model 1	1.31	1.12, 1.54	0.001	1.55	1.27, 1.91	0.001	0.55	0.28, 1.07	0.08	1.35	1.20, 1.53	0.001
Model 2	1.32	1.21, 1.43	0.001	1.51	1.23, 1.84	0.001	0.70	0.35, 1.42	0.32	1.34	1.24, 1.44	0.001
Model 3	1.22	1.03, 1.46	0.02	1.46	1.17, 1.82	0.001	1.05	0.74, 1.49	0.78	1.27	1.12, 1.45	0.003
Total animal protein (per 5 E%)												
Model 1	1.32	1.12, 1.55	0.001	1.58	1.28, 1.96	0.001	1.19	0.87, 1.62	0.28	1.38	1.22, 1.55	0.001
Model 2	1.32	1.21, 1.44	0.001	1.57	1.27, 1.93	0.001	1.22	0.89, 1.68	0.21	1.35	1.24, 1.45	0.001
Model 3	1.17	0.95, 1.46	0.48	1.61	1.28, 2.02	0.001	1.06	0.75, 1.50	0.71	1.28	1.13, 1.46	0.001
Total plant protein (per 5 E%)												
Model 1	1.17	0.77, 1.76	0.46	1.99	1.07, 3.70	0.03	0.64	0.31, 1.32	0.23	1.20	0.88, 1.63	0.26
Model 2	1.13	0.91, 1.41	0.53	2.52	1.32, 4.83	0.01	0.86	0.40, 1.85	0.69	1.19	0.98, 1.46	0.08
Model 3	1.39	0.90, 2.14	0.13	1.12	0.59, 2.11	0.10	0.73	0.32, 1.61	0.43	1.18	0.86, 1.63	0.32

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5 E% protein intake difference with risk of prediabetes at the expense of carbohydrate.

Model 1 is adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age and sex.

Model 2 is additionally adjusted for smoking status (never, past, or current), education level (primary, lower, intermediate, or higher), diet quality score, physical activity (METh/wk) and family history of diabetes (for RS-I and RS-II only).

Model 3 is adjusted for all covariates in model 2 and additionally for time point of WC measurements, longitudinal WC, and the interaction between protein intake and longitudinal WC.

Abbreviations: WC, waist circumference; CI, confidence interval, HR, hazard ratio; RS, Rotterdam Study; E%, percentage of total energy intake

Table 4. Associations of protein intake with risk of type 2 diabetes

	RS-I (n=2976)			RS-II (n=1414)			RS-III (n=2423)			Pooled results (n=6813)		
Type 2 diabetes	**HR**	**95%CI**	**P value**	**HR**	**95%CI**	**P value**	**HR**	**95%CI**	**P value**	**HR**	**95%CI**	**P value**
Total protein (per 5 E%)												
Model 1	1.27	1.05, 1.54	0.01	1.33	1.03, 1.72	0.03	1.95	1.33, 2.87	0.001	1.37	1.18, 1.58	0.001
Model 2	1.26	1.14, 1.39	0.02	1.29	1.00, 1.67	0.04	1.96	1.61, 2.39	0.001	1.37	1.26, 1.49	0.001
Model 3	1.04	0.85, 1.27	0.70	1.09	0.83, 1.44	0.53	1.62	1.05, 2.50	0.03	1.12	0.96, 1.30	0.16
Total animal protein (per 5 E%)												
Model 1	1.28	1.06, 1.54	0.01	1.34	1.02, 1.74	0.03	1.94	1.32, 2.84	0.001	1.37	1.19, 1.58	0.001
Model 2	1.26	1.14, 1.39	0.02	1.32	1.02, 1.72	0.03	1.95	1.60, 2.37	0.001	1.37	1.26, 1.49	0.001
Model 3	1.03	0.84, 1.26	0.78	1.11	0.83, 1.18	0.48	1.41	0.72, 2.76	0.32	1.11	0.94, 1.28	0.21
Total plant protein (per 5 E%)												
Model 1	1.05	0.64, 1.73	0.85	1.44	0.67, 3.11	0.35	1.54	0.62, 3.82	0.36	1.21	0.83, 1.77	0.32
Model 2	0.93	0.71, 1.23	0.80	1.84	0.82, 4.14	0.14	1.67	1.02, 2.74	0.03	1.12	0.87, 1.40	0.35
Model 3	0.95	0.55, 1.64	0.86	1.94	0.88, 4.27	0.92	1.18	0.24, 5.70	0.84	1.20	0.79, 1.80	0.38

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5 E% protein intake difference with risk of type 2 diabetes at the expense of carbohydrate.

Model 1 is adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age and sex.

Model 2 is additionally adjusted for smoking status (never, past, or current), education level (primary, lower, intermediate, or higher), diet quality score, physical activity (METh/wk) and family history of diabetes (for RS-I and RS-II only).

Model 3 is adjusted for all covariates in model 2 and additionally for time point of WC measurements, longitudinal WC, and the interaction between protein intake and longitudinal WC.

Abbreviations: WC, waist circumference; CI, confidence interval, HR, hazard ratio; RS, Rotterdam Study; E%, percentage of total energy intake

Table 5. Associations of protein from specific food sources with longitudinal insulin resistance, and risk of prediabetes and type 2 diabetes

Protein intake (per 5 E%)	Insulin resistance (n=6657)		Prediabetes (n=5795)		Type 2 diabetes (n=6813)	
	β (95%CI)	P value	HR (95%CI)	P value	HR (95%CI)	P value
Protein intake from animal food sources						
-Meat	0.13 (0.10, 0.17)	<0.0001	1.54 (1.31, 1.80)	<0.0001	1.40 (1.12, 1.75)	0.002
-Fish	0.08 (0.03, 0.13)	0.001	1.31 (1.03, 1.65)	0.02	1.65 (1.30, 2.10)	<0.0001
-Dairy	0.04 (0.0003, 0.08)	0.01	1.26 (1.06, 1.49)	0.01	1.23 (1.00, 1.49)	0.04
Protein from plant food sources						
-Legumes & nuts	-0.06 (-0.15, 0.03)	0.21	1.00 (0.63, 1.62)	0.99	0.73 (0.40, 1.32)	0.30
-Grains	0.03 (-0.05, 0.12)	0.45	1.51 (0.96, 2.36)	0.07	1.68 (0.94, 3.00)	0.08
-Potatoes	0.16 (-0.12, 0.43)	0.27	1.25 (0.34, 4.53)	0.74	0.59 (0.12, 2.75)	0.50
-Fruits & Vegetables	-0.08 (-0.03, 0.19)	0.16	0.90 (0.40, 2.03)	0.79	1.39 (0.53, 3.60)	0.51

Effect estimates are pooled results of regression coefficients (β) for HOMA-IR or hazard ratios (HRs) for incidence of prediabetes or type 2 diabetes with their 95%-confidence intervals (95%CIs) in RS-I, RS-II, and RS-III, using fixed-effects meta-analysis. The multivariable models simultaneously include energy of protein from meat, from dairy, from fish, from legumes and nuts, from grains, from potato, and from fruits and vegetables, adjusted for SFA intake (5 E%), MUFA intake (5 E%), PUFA intake (5 E%), Trans-fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), fiber (grams), age, sex, smoking status (never, past, or current), education level (primary, lower, intermediate, or higher), diet quality score, physical activity (METh/wk) and family history of diabetes (for RS-I and RS-II only).

Abbreviations: HOMA-IR, homeostatic model assessment of insulin resistance; CI, confidence interval, HR, hazard ratio; RS, Rotterdam Study; E%, percentage of total energy intake, SFA, saturated fat acids intake; MUFA, monounsaturated fat acids intake; PUFA, polyunsaturated fat acids intake; Trans-fat, trans-unsaturated fat acids intake.

Intake of protein from various food sources with insulin resistance, risk of prediabetes and T2D

We further examined the associations of protein from more specific animal and plant food sources with longitudinal insulin resistance, and risk of prediabetes and T2D. In multivariable models, higher intakes of protein from meat, from fish, and from dairy were all associated with higher longitudinal insulin resistance, and risk of prediabetes and T2D (e.g. for HOMA-IR: protein from meat, 0.13 (95%CI 0.10, 0.17), from fish, 0.08 (95%CI 0.03, 0.13), and from dairy, 0.04 (95%CI 0.0003, 0.08); and for prediabetes risk: protein from meat, 1.54 (95%CI 1.31, 1.80), from fish, 1.31 (95%CI 1.03, 1.65), from dairy, 1.26 (95%CI 1.06, 1.49)). Protein from legumes and nuts, protein from grains, protein from potato, and protein from fruits and vegetables, were not associated with longitudinal insulin resistance, and risk of prediabetes and T2D (Table 5).

Sensitivity analyses

The associations between total protein, total animal protein, and total plant protein with longitudinal insulin resistance and risk of prediabetes and T2D did not differ by age, sex, or longitudinal WC (all interactions $p>0.05$). In all sensitivity analyses, including for example the additional adjustments for cholesterol, hypertension, triglycerides and CVD history; and modelling replacement of protein at the expense of fat instead of carbohydrate, the estimates were similar and remained statistically significant (Supplemental Tables 2-17).

DISCUSSION

In this large population-based prospective cohort, higher habitual protein intake, mainly from animal food sources, was associated with a persistently higher insulin resistance over time and a higher risk of prediabetes and T2D. Obesity over time appeared to be a mediator in these associations. We observed that protein from meat, from fish, and from dairy all contributed to these associations. Intake of total plant protein and protein from legumes and nuts, protein from potatoes, protein from grains or protein from fruits and vegetables were not associated with longitudinal insulin resistance, risk of prediabetes or T2D.

Our results for T2D risk support those of previous observational studies that found positive associations between total protein and total animal protein and T2D risk.[7] Furthermore, we extended this evidence by also reporting associations of higher total protein and total animal protein with higher longitudinal insulin resistance and risk of prediabetes. Some small previous human studies also indicated harmful effects of high protein diets on insulin resistance when the intake was prolonged.[24] More importantly, our study further extended the evidence in this field by examining the associations

of protein from more detailed animal food sources with longitudinal insulin resistance, and risk of prediabetes and T2D in long-term follow-up. We observed independent associations of higher intake of protein from meat, from fish, and from dairy with higher insulin resistance, and risk of prediabetes and T2D, indicating that the observed associations were not mainly driven by protein from a particular animal food source. Few previous studies have observed associations between animal protein from more specific animal food sources with insulin resistance, and risk of prediabetes and T2D.[8, 13] Similar to our findings, van Nielen et al. observed that the association of animal protein with T2D risk was not explained by protein from a particular animal food source by examining whether the association of animal protein substantially changed when excluding protein from meat, fish, or dairy from total animal protein.[8] In contrast, a recent study by Heli et al. reported null associations of protein from meat, fish or dairy with T2D risk.[13] The discrepancy could be explained by the lower animal protein intakes in the study of Heli et al[13] as compared to animal protein intakes in van Nielen et al.'s and our studies.

The potential mechanisms behind the associations of habitual high long-term animal protein with the development of T2D, are largely unknown. One mechanism could involve effects of specific amino acids metabolized mainly from animal protein. Dietary branched-chain and aromatic amino acids are mainly derived from animal-sourced protein, and they can lead to phosphorylation of the insulin receptor by activating mammalian target of rapamycin (mTOR), thereby undermining the normal regulation of glucose and insulin levels.[4, 24, 25] Previous studies also proposed other possible mechanisms. For example, co-occurrence of nutrients in animal protein-rich foods, such as heme iron, saturated fat, nitrites, and advanced glycation end products might attribute to the positive associations.[9, 26] However, we observed that the associations of animal protein intake were independent of other macronutrients and overall diet quality, in which red and processed meat, rich in heme iron, saturated fat, nitrites, and advanced glycation end products, is an important component, suggesting specific effect of animal protein. In contrast, some short-term randomized trials reported beneficial effects of high protein diets (mean protein contents were 1.25 ± 0.17 g/kg BW/day in these randomized control trials) on obesity, a main risk factor of T2D, which may be explained by induced satiety and energy expenditure due to short-term high protein intake.[3] On short-term, high protein intake may increase gluconeogenesis and cause a high ketogenic state, which contribute to increased satiety and energy expenditure.[3, 27, 28] This increased satiety and energy expenditure can result in a lower energy intake and a negative energy balance, and thereby promote weight loss and weight maintenance.[29]

Our results were also in line with most of previous studies reporting null associations of total plant protein.[7] In contrast, the studies in American populations reported a modest inverse association with T2D risk.[9] Moreover, we extended this evidence by further exploring the associations of plant protein from more specific food sources, including protein from legumes and nuts, grains, potatoes, fruits and

vegetables, for which we observed null associations. Little previous evidence for protein from more specific plant food sources and T2D risk was available and was mainly limited to classification of total plant protein as protein from grains and non-grains only.[13] For example, Virtanen et al. reported null associations between protein from grains or from non-grains with T2D risk, but did not examine the associations for protein from more specific plant food sources.[13]

Finally, we also observed that after additional adjustment for longitudinal WC, the associations of total and animal protein intake with insulin resistance, and risk of prediabetes and T2D were attenuated, although still statistically significant for insulin resistance, and risk of prediabetes. This suggests that obesity seems to be a mediator in the associations. Previous studies[7-9] reported that obesity could be a mediator but were limited by adjusting for only baseline obesity. Because obesity is a factor of an overall unhealthy lifestyle, obesity could be both intermediate and confounder in the associations. However, in our analyses we adjusted for main indicators of lifestyle, such as physical activity and overall diet quality, before correcting for longitudinal measures of obesity, therefore, the attenuation of the associations by additional adjustment for obesity is more likely to be explained by the mediation role of obesity.

Strengths and limitations

Our study has several strengths. First, our study is the first study that directly examined associations of protein intake with longitudinal insulin resistance and prediabetes risk in large population. Studying these early risk stages help minimize reverse causation and understand how protein intake influence the development of T2D. We found that the associations of protein intake with longitudinal insulin resistance and risk of prediabetes and T2D were consistent, which indicates that it could be beneficial to limit habitual high animal protein intake already for early stages in the development of T2D. Second, to our knowledge, this is also the first prospective study to use longitudinal obesity as a time-varying covariate in linear mixed model and Joint model to examine the role of obesity in the development of T2D, namely, we have accounted for the effect of longitudinal obesity in the associations between protein intake and the development of T2D. Third, our study comprehensively examined these associations for protein from more specific food sources instead of protein from total animal protein and plant protein, which adds literature into this field and may facilitate public health recommendations. Fourth, we adjusted for a wide range of potential confounders, including many lifestyle and dietary variables, which is important, especially when studying a single nutrient. Last, our results were robust to various sensitivity analyses, including additional adjustment for a broad range of other cardiovascular risk factors, and different macronutrient substitution effects.

There are several limitations we should consider. First, we only used data on protein intake at baseline in main analyses, which may not represent long-time protein intake. Therefore, analysis of data on

repeated measurements of dietary protein intake over time would be preferable. However, the exclusion of participants who were likely to change their diet during follow-up, such as participants with cardiovascular diseases at baseline or follow-up, and the additional adjustment for protein intake 20 years after baseline in RS-I and 10 years after baseline in RS-II, did not change the results; furthermore, estimates were similar in three sub-cohorts with different follow-up. Combined, the results from these sensitivity analyses indicated that our conclusions were robust. Second, as our current study was conducted within an observational population-based cohort study among general population, the variation in protein intake was not that large. A larger variation would be preferable to explore the role of plant protein intake in the development of T2D risk. However, several previous cohort studies reported similar amount of variation of plant protein intake, and also observed associations between plant protein intake and T2D risk. This indicates that the amount of variation of plant protein among our participants would have been sufficient to pick up associations with these outcomes. Third, misclassification of protein intake could have occurred. However, given the prospective study design, this measurement error was likely to be non-differential, which would have attenuated observed associations. Fourth, although we adjusted for many potential confounders, the possibility of residual confounding cannot be dismissed, for example, through the meal distribution of protein and energy. Finally, our results may not be generalizable to people of other age or race and need further replication.

Conclusions

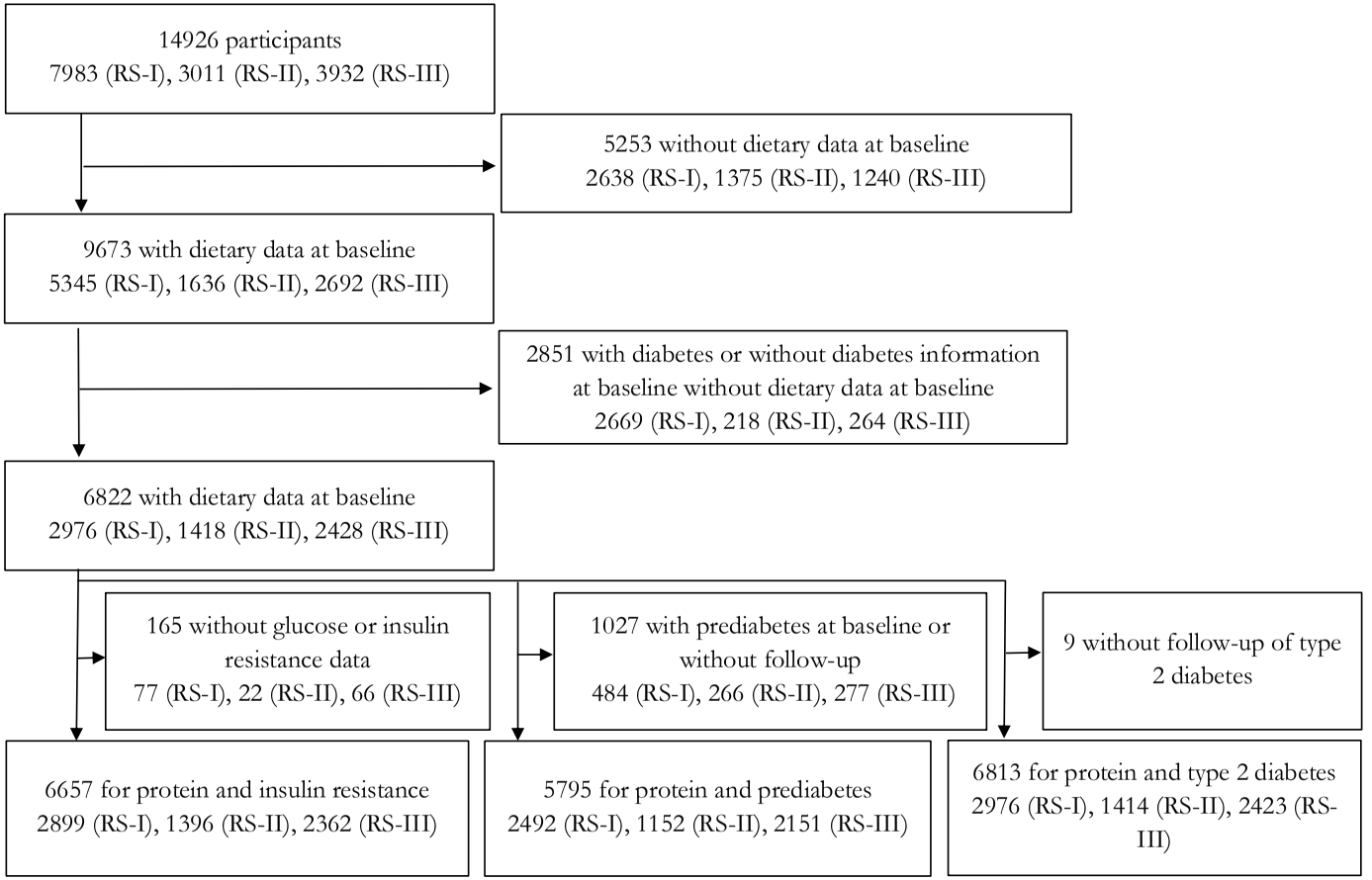
In this large prospective cohort study, higher intake of total protein and total animal protein was associated with higher longitudinal insulin resistance and risk of prediabetes and T2D. Obesity over time appeared to partly mediate these associations. Protein from meat, fish and dairy all contributed to these associations. Intake of protein from legumes and nuts, grains, potatoes, or vegetables and fruits was not associated with insulin resistance, risk of prediabetes or T2D. Our findings indicate the importance of protein sources and that limiting high intake of protein from animal food sources may be beneficial in preventing development of T2D.

REFERENCES

1. Hu FB. Globalization of Diabetes The role of diet, lifestyle, and genes. Diabetes care. 2011;34(6):1249-57.
2. Wycherley TP, Moran LJ, Clifton PM, Noakes M, Brinkworth GD. Effects of energy-restricted high-protein, low-fat compared with standard-protein, low-fat diets: a meta-analysis of randomized controlled trials. The American journal of clinical nutrition. 2012:ajcn. 044321.
3. Westerterp-Plantenga MS, Lemmens SG, Westerterp KR. Dietary protein–its role in satiety, energetics, weight loss and health. British journal of nutrition. 2012;108(S2):S105-S12.
4. Tremblay F, Marette A. Amino acid and insulin signaling via the mTOR/p70 S6 kinase pathway A negative feedback mechanism leading to insulin resistance in skeletal muscle cells. Journal of Biological Chemistry. 2001;276(41):38052-60.
5. Lu J, Xie G, Jia W, Jia W. Insulin resistance and the metabolism of branched-chain amino acids. Frontiers of medicine. 2013;7(1):53-9.
6. Würtz P, Soininen P, Kangas AJ, Rönnemaa T, Lehtimäki T, Kähönen M, et al. Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. Diabetes care. 2013;36(3):648-55.
7. Shang X, Scott D, Hodge AM, English DR, Giles GG, Ebeling PR, et al. Dietary protein intake and risk of type 2 diabetes: results from the Melbourne Collaborative Cohort Study and a meta-analysis of prospective studies. The American Journal of Clinical Nutrition. 2016:ajcn140954.
8. van Nielen M, Feskens EJ, Mensink M, Sluijs I, Molina E, Amiano P, et al. Dietary protein intake and incidence of type 2 diabetes in Europe: the EPIC-InterAct Case-Cohort Study. Diabetes Care. 2014;37(7):1854-62.
9. Malik VS, Li Y, Tobias DK, Pan A, Hu FB. Dietary Protein Intake and Risk of Type 2 Diabetes in US Men and Women. American journal of epidemiology. 2016:kwv268.
10. Muoio DM, Newgard CB. Molecular and metabolic mechanisms of insulin resistance and β-cell failure in type 2 diabetes. Nature reviews Molecular cell biology. 2008;9(3):193-205.
11. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. The Lancet. 2012;379(9833):2279-90.
12. Stumvoll M, Goldstein BJ, van Haeften TW. Pathogenesis of type 2 diabetes. Endocrine research. 2007;32(1-2):19-37.
13. Virtanen HEK, Koskinen TT, Voutilainen S, Mursu J, Tuomainen T-P, Kokko P, et al. Intake of different dietary proteins and risk of type 2 diabetes in men: the Kuopio Ischaemic Heart Disease Risk Factor Study. British Journal of Nutrition. 2017;117(6):882-93.
14. Hofman A, Brusselle GG, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2016 objectives and design update. European journal of epidemiology. 2015;30(8):661-708.
15. Klipstein-Grobusch Kd, Den Breeijen JH, Goldbohm RA, Geleijnse JM, Hofman A, Grobbee DE, et al. Dietary assessment in the elderly: validation of a semiquantitative food frequency questionnaire. European Journal of Clinical Nutrition. 1998;52(8):588-96.

16. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. The American journal of clinical nutrition. 1993;58(4):489-96.
17. Voortman T, Kiefte-de Jong JC, Ikram MA, Stricker BH, van Rooij FJA, Lahousse L, et al. Adherence to the 2015 Dutch dietary guidelines and risk of non-communicable diseases and mortality in the Rotterdam Study. Eur J Epidemiol. 2017.
18. Organization WH. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WH. 2006.
19. Leening MJG, Ferket BS, Steyerberg EW, Kavousi M, Deckers JW, Nieboer D, et al. Sex differences in lifetime risk and first manifestation of cardiovascular disease: prospective population based cohort study. Bmj. 2014;349:g5992.
20. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. The American journal of clinical nutrition. 1997;65(4):1220S-8S.
21. Hu FB, Stampfer MJ, Rimm E, Ascherio A, Rosner BA, Spiegelman D, et al. Dietary fat and coronary heart disease: a comparison of approaches for adjusting for total energy intake and modeling repeated dietary measurements. American journal of epidemiology. 1999;149(6):531-40.
22. Rizopoulos D. Joint models for longitudinal and time-to-event data: With applications in R: Chapman and Hall/CRC; 2012.
23. World Health Organization (WHO). Protein and Amino Acid Requirements in Human Nutrition: report of a joint FAO/WHO/UNU expert consultation. Geneva, Switzerland: 2002.
24. Rietman A, Schwarz J, Tomé D, Kok FJ, Mensink M.? European journal of clinical nutrition. 2014;68(9):973.
25. Laplante M, Sabatini DM. mTOR signaling in growth control and disease. Cell. 2012;149(2):274-93.
26. Pan A, Sun Q, Bernstein AM, Schulze MB, Manson JE, Willett WC, et al. Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. The American journal of clinical nutrition. 2011;94(4):1088-96.
27. Johnstone AM, Horgan GW, Murison SD, Bremner DM, Lobley GE. Effects of a high-protein ketogenic diet on hunger, appetite, and weight loss in obese men feeding ad libitum–. The American journal of clinical nutrition. 2008;87(1):44-55.
28. Veldhorst MAB, Westerterp-Plantenga MS, Westerterp KR. Gluconeogenesis and energy expenditure after a high-protein, carbohydrate-free diet–. The American journal of clinical nutrition. 2009;90(3):519-26.
29. Goran MI. Energy metabolism and obesity. Medical Clinics of North America. 2000;84(2):347-62.

SUPPLEMENTAL MATERIAL



Supplemental Figure 1. Participants selection

Supplemental Table 1. Number and percentages of missing values of covariates in all participants in different analyses before imputation

	Education level	Smoking status	Physical activity
Analyses of HOMA-IR (n=6657)	40 (0.60%)	32 (0.48%)	239(3.6%)
Analyses of prediabetes (n=5795)	32 (0.55%)	25 (0.43%)	237(4.1%)
Analyses of type 2 diabetes (n=6813)	41 (0.60%)	33 (0.48%)	272 (4.0%)

For all covariates of all analyses, there were only missing values on education level, smoking status and physical activity before imputation.

Supplemental Table 2. Associations of protein intake at the expense of total fat intake with longitudinal insulin resistance

(Per 5 E%)	RS-I (n=2899)			RS-II (n=1396)			RS-III (n=2362)			Pooled results (n=6657)		
	β	95%CI	P value	β	95%CI	P value	β	95%CI	P value	β	95%CI	P value
Total protein	0.07	0.03, 0.11	0.001	0.05	-0.01, 0.10	0.11	-0.04	-0.12, 0.07	0.50	0.05	0.02, 0.08	0.001
Animal protein	0.07	0.03, 0.12	0.001	0.05	-0.01, 0.11	0.08	0.13	0.08, 0.18	0.001	0.09	0.06, 0.11	0.001
Plant protein	0.03	-0.02, 0.08	0.53	-0.07	-0.22, 0.07	0.31	0.07	-0.04, 0.17	0.23	0.03	-0.02, 0.07	0.20

Values are β coefficient and 95% confidence intervals (CI) for the difference of intake of per 5E% protein in HOMA-IR at the expense of total fat. All models are multivariate models corresponding to model 2 in the main analyses, adjusted for carbohydrate intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, time point of HOMA-IR measurement, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous), and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 3. Associations of protein intake at the expense of total fat intake with risk of prediabetes

(Per 5 E%)	RS-I (n=2492)			RS-II (n=1152)			RS-III (n=2151)			Pooled results (n=5795)		
	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value
Total protein	1.23	1.13, 1.34	0.01	1.42	1.13, 1.77	0.002	0.72	0.36, 1.45	0.36	1.35	1.28, 1.42	0.001
Animal protein	1.24	1.14, 1.35	0.01	1.41	1.23, 1.77	0.003	1.16	0.84, 1.60	0.36	1.27	1.18, 1.36	0.001
Plant protein	1.10	0.89, 1.36	0.67	2.01	1.01, 3.77	0.03	0.82	0.39, 1.72	0.59	1.14	0.94, 1.38	0.20

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5E% protein intake difference with risk of prediabetes at the expense of total fat. All models are multivariate models corresponding to model 2 in the main analyses, adjusted for carbohydrate intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous) and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 4. Associations of protein intake at the expense of total fat intake with risk of type 2 diabetes

(Per 5 E%)	RS-I (n=2976)			RS-II (n=1414)			RS-III (n=2423)			Pooled results (n=6813)		
	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value
Total protein	1.19	0.98, 1.44	0.09	1.36	1.03, 1.80	0.03	1.75	1.22, 2.51	0.01	1.31	1.14, 1.52	0.001
Animal protein	1.21	0.99, 1.47	0.06	1.36	1.03, 1.80	0.03	1.74	1.18, 2.56	0.01	1.33	1.11, 1.60	0.002
Plant protein	0.91	0.55, 1.52	0.72	1.44	0.66, 3.12	0.36	1.47	0.91, 2.37	0.43	1.21	0.88, 1.67	0.23

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5E% protein intake difference with risk of type 2 diabetes at the expense of total fat. All models are multivariate models corresponding to model 2 in the main analyses, adjusted for carbohydrate intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous) and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 5. Associations of protein intake with longitudinal insulin resistance, adjusted for fatty acid subgroups

(Per 5 E%)	RS-I (n=2899)			RS-II (n=1396)			RS-III (n=2362)			Pooled results (n=6657)		
	β	95%CI	P value	β	95%CI	P value	β	95%CI	P value	β	95%CI	P value
Total protein	0.13	0.07, 0.19	0.001	0.13	0.06, 0.20	0.001	0.15	0.10, 0.20	0.001	0.14	0.10, 0.17	0.001
Animal protein	0.13	0.06, 0.20	0.001	0.13	0.06, 0.20	0.001	0.16	0.10, 0.21	0.001	0.14	0.11, 0.17	0.001
Plant protein	0.09	-0.07, 0.25	0.28	0.22	0.02, 0.42	0.03	0.04	-0.07, 0.16	0.47	0.09	0.00, 0.17	0.04

Values are based on multivariable linear regression models and reflect differences (95% CI) in HOMA-IR per 5 E% protein intake at the expense of carbohydrate. All models are multivariate models corresponding to model 2 in the main analyses, adjusted for saturated fat acids intake (5 E%), polyunsaturated fatty acids intake (5 E%), monounsaturated fat acids intake (5 E%), trans fatty acids intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, time point of HOMA-IR measurement, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous), and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 6. Associations of protein intake with risk of prediabetes, adjusted for fatty acid subgroups

(Per 5 E%)	RS-I (n=2492)			RS-II (n=1152)			RS-III (n=2151)			Pooled results (n=5795)		
	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value
Total protein	1.34	1.23, 1.47	0.001	1.55	1.25, 1.92	0.001	0.73	0.34, 1.54	0.40	1.36	1.25, 1.47	0.001
Animal protein	1.34	1.01, 1.77	0.04	1.57	1.26, 1.96	0.001	1.27	0.91, 1.77	0.17	1.43	1.23, 1.67	0.001
Plant protein	1.52	0.76, 3.05	0.24	2.35	1.18, 4.69	0.02	0.88	0.40, 1.96	0.76	1.53	1.00, 2.31	0.05

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5E% protein intake difference with risk of prediabetes at the expense of carbohydrate. All models are multivariate model corresponding to model 2 in the main analyses, adjusted for saturated fat acids intake (5 E%), polyunsaturated fatty acids intake (5 E%), monounsaturated fat acids intake (5 E%), trans fatty acids intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous) and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 7. Associations of protein intake with risk of type 2 diabetes after adjustment for fatty acid subgroups

(Per 5 E%)	RS-I (n=2976)			RS-II (n=1414)			RS-III (n=2423)			Pooled results (n=6813)		
	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value
Total protein	1.31	1.18, 1.46	0.01	1.33	1.01, 1.74	0.04	2.06	1.39, 3.06	0.001	1.35	1.22, 1.48	0.001
Animal protein	1.31	1.18, 1.46	0.01	1.34	1.01, 1.77	0.04	2.06	1.41, 2.99	0.001	1.35	1.23, 1.49	0.001
Plant protein	0.98	0.74, 1.31	0.95	1.92	0.81, 4.58	0.14	1.79	0.70, 4.58	0.25	1.09	0.84, 1.42	0.51

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5E% protein intake difference with risk of type 2 diabetes at the expense of carbohydrate. All models are multivariate model corresponding to model 2 in the main analyses, adjusted saturated fat acids intake (5 E%), polyunsaturated fatty acids intake (5 E%), monounsaturated fat acids intake (5 E%), trans fatty acids intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous) and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 8. Associations of protein intake with longitudinal insulin resistance, additionally adjusted for cardiovascular risk factors

(Per 5 E%)	RS-I (n=2899)			RS-II (n=1396)			RS-III (n=2362)			Pooled results (n=6657)		
	β	95%CI	P value	β	95%CI	P value	β	95%CI	P value	β	95%CI	P value
Total protein	0.08	0.04, 0.12	0.001	0.07	0.01, 0.12	0.02	0.09	0.04, 0.13	0.001	0.08	0.05, 0.11	0.001
Animal protein	0.08	0.04, 0.13	0.001	0.06	0.01, 0.12	0.02	0.08	0.04, 0.13	0.001	0.08	0.05 0.11	0.001
Plant protein	0.04	-0.06, 0.15	0.43	-0.02	-0.16, 0.13	0.83	0.07	0.01, 0.14	0.023	0.05	0.00, 0.10	0.05

Values are β coefficient and 95% confidence intervals (CI) for the difference of intake of per 5E% protein in HOMA-IR at the expense of carbohydrate. All models are multivariate model corresponding to model 2 in the main analyses, adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, time point of HOMA-IR measurement, smoking status(never, past or current), education level (primary, lower, intermediate or higher), diet quality score, time of physical activity (continuous), TAG (continuous), hypertension (yes or no), cholesterol(continuous), CVD history ((yes or no), and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 9. Associations of protein intake with risk of prediabetes, additionally adjusted for cardiovascular risk factors

(Per 5 E%)	RS-I (n=2492)			RS-II (n=1152)			RS-III (n=2151)			Pooled results (n=5795)		
	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value
Total protein	1.31	1.11, 1.55	0.001	1.35	1.08, 0.70	0.01	1.18	0.83, 1.68	0.35	1.31	1.16, 1.49	0.001
Animal protein	1.30	1.00, 1.69	0.06	1.38	1.09, 1.74	0.01	1.20	0.85, 1.69	0.31	1.31	1.12, 1.53	0.001
Plant protein	1.46	0.76, 2.81	0.25	1.93	0.96, 3.87	0.07	0.95	0.59, 1.53	0.83	1.25	0.90, 1.75	0.19

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5E% protein intake difference with risk of prediabetes at the expense of carbohydrate. All models are multivariate model corresponding to model 2 in the main analyses, adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous), TAG (continuous), hypertension (yes or no), cholesterol (continuous), CVD history ((yes or no), and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 10. Associations of protein intake with risk of type 2 diabetes, additionally adjusted for cardiovascular risk factors

(Per 5 E%)	RS-I (n=2976)			RS-II (n=1414)			RS-III (n=2423)			Pooled results (n=6813)		
	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value
Total protein	1.23	1.01, 1.49	0.04	1.29	0.98, 1.71	0.07	1.85	1.19, 2.86	0.01	1.31	1.13, 1.52	0.001
Animal protein	1.25	1.03, 1.51	0.03	1.33	1.00, 1.77	0.05	1.85	1.19, 2.85	0.01	1.27	1.09, 1.49	0.003
Plant protein	0.84	0.50, 1.43	0.52	2.10	0.9, 4.85	0.08	1.69	0.92, 3.11	0.09	1.27	0.88, 1.82	0.20

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5en% protein intake difference with risk of type 2 diabetes at the expense of carbohydrate. All models are multivariate models corresponding to model 2 in the main analyses, adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous), TAG (continuous), hypertension (yes or no), cholesterol (continuous), CVD history ((yes or no), and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 11. Associations of protein intake with longitudinal insulin resistance, excluding patients with incidence of CVD in follow-up

(Per 5 E%)	RS-I (n=2576)			RS-II (n=1333)			RS-III (n=2145)			Pooled results (n=6234)		
	β	95%CI	P value	β	95%CI	P value	β	95%CI	P value	β	95%CI	P value
Total protein	0.08	0.04, 0.13	0.001	0.07	0.01, 0.13	0.02	0.10	0.05, 0.15	0.001	0.08	0.06, 0.11	0.001
Animal protein	0.08	0.04, 0.13	0.001	0.07	0.01, 0.13	0.03	0.07	0.02, 0.13	0.01	0.08	0.05, 0.11	0.001
Plant protein	0.03	-0.08, 0.14	0.63	-0.05	-0.21, 0.11	0.51	0.06	-0.01, 0.13	0.03	0.04	-0.02, 0.10	0.17

Values are based on multivariable linear regression models and reflect differences (95% CI) in HOMA-IR per 5E% protein intake at the expense of carbohydrate. All models are multivariate model corresponding to model 2 in the main analyses, adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, time point of HOMA-IR measurement, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous), and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 12. Associations of protein intake with prediabetes, excluding patients with incidence of CVD in follow-up

(Per 5 E%)	RS-I (n=2167)			RS-II (n=1058)			RS-III (n=2118)			Pooled results (n=5343)		
	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value
Total protein	1.37	1.14, 1.64	0.001	1.51	1.21, 1.88	0.001	1.11	0.78, 1.58	0.57	1.38	1.21, 1.57	0.001
Animal protein	1.37	1.15, 1.65	0.001	1.54	1.22, 1.93	0.001	0.91	0.63, 1.31	0.62	1.35	1.18, 1.54	0.001
Plant protein	0.99	0.62, 1.60	0.97	2.16	1.06, 4.37	0.03	0.67	0.41, 1.09	0.11	0.93	0.68, 1.28	0.67

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5 E% protein intake difference with risk of prediabetes at the expense of carbohydrate. All models are Multivariate model corresponding to model 2 in the main analyses, adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous) and family history of diabetes (yes or no).

Supplemental Table 13. Associations of protein intake with risk of type 2 diabetes, excluding patients with incidence of CVD in follow-up

(Per 5 E%)	RS-I (n=2586)			RS-II (n=1299)			RS-III (n=2378)			Pooled results (n=6263)		
	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value	HR	95%CI	P value
Total protein	1.25	1.01, 1.55	0.04	1.40	1.07, 1.83	0.01	1.89	1.23, 2.92	0.004	1.37	1.17, 1.60	0.001
Animal protein	1.27	1.02, 1.57	0.03	1.44	1.09, 1.91	0.01	1.88	1.22, 2.90	0.004	1.39	1.19, 1.63	0.001
Plant protein	0.73	0.41, 1.29	0.27	2.25	0.95, 5.35	0.70	1.59	0.87, 2.93	0.13	1.22	0.84, 1.77	0.31

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5 E% protein intake difference with risk of type 2 diabetes at the expense of carbohydrate. All models are multivariate model corresponding to model 2 in the main analyses, adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous) and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 14. Associations of protein intake with longitudinal insulin resistance, censored HOMA-IR data before onset of type 2 diabetes

(Per 5 E%)	RS-I (n=2893)			RS-II (n=1383)			RS-III (n=2252)		
	β	95%CI	P value	β	95%CI	P value	β	95%CI	P value
Total protein	0.08	0.04, 0.12	0.0001	0.07	0.02, 0.12	0.005	0.11	0.06, 0.15	0.0001
Animal protein	0.08	0.04, 0.12	0.0001	0.07	0.02, 0.12	0.01	0.11	0.06, 0.16	<0.0001
Plant protein	0.02	-0.09, 0.12	0.74	-0.04	-0.19, 0.11	0.61	0.05	-0.06, 0.16	0.37

(Per 5 E%)	Pooled results (n=6528)		
	β	95%CI	P value
Total protein	0.09	0.06, 0.11	<0.0001
Animal protein	0.09	0.06, 0.11	<0.0001
Plant protein	0.02	-0.05, 0.08	0.61

Values are β coefficient and 95% confidence intervals (CI) for the difference of intake of per 5 E% protein in HOMA-IR, censored HOMA-IR data before onset of type 2 diabetes at the expense of carbohydrate. All models are multivariate models corresponding to model 2 in the main analyses, adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), age, sex, time point of HOMA-IR measurement, smoking status(never, past or current), education level (primary, lower, intermediate or higher), diet quality score, time of physical activity (continuous), and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 15. Associations of protein intake with longitudinal insulin resistance additionally adjusted for protein intake in follow-up

(Per 5 E%)	RS-I (n=987)			RS-II (n=826)		
	β	95%CI	P value	β	95%CI	P value
Total protein	0.08	0.04, 0.13	0.001	0.24	0.02, 0.47	0.001
Animal protein	0.08	0.04, 0.13	0.001	0.12	0.05, 0.19	0.001
Plant protein	0.03	-0.08, 0.14	0.63	0.24	0.03, 0.45	0.02

Values are based on multivariable linear mixed regression models and reflect differences (95% CI) in insulin resistance (HOMA-IR) per 5 E% protein intake at the expense of carbohydrate.

All models are multivariate models corresponding to model 2 in the main analyses, adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), protein intake 20 years after baseline in RS-I or 10 years after baseline in RS-II, time point of HOMA-IR measurement, age, sex, smoking status(never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous), and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 16. Associations of protein intake with risk of prediabetes additionally adjusted for protein intake in follow-up

(Per 5 E%)	RS-I (n=913)			RS-II (n=721)		
	HR	95%CI	P value	HR	95%CI	P value
Total protein	1.23	1.01, 1.55	0.04	1.31	1.06, 1.72	0.01
Animal protein	1.22	1.02, 1.57	0.02	1.38	1.06, 1.69	0.01
Plant protein	0.70	0.40, 1.22	0.23	2.22	0.90, 5.35	0.68

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5 E% protein intake difference with risk of prediabetes at the expense of carbohydrate. All models are multivariate models corresponding to model 2 in the main analyses, adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), protein intake 20 years after baseline in RS-I or 10 years after baseline in RS-II, age, sex, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous) and family history of diabetes (for RS-I and RS-II only).

Supplemental Table 17. Associations of protein intake with risk of type 2 diabetes, additionally adjusted for protein intake in follow-up

(Per 5 E%)	RS-I (n=1076)			RS-II (n=855)		
	HR	95%CI	P value	HR	95%CI	P value
Total protein	1.24	1.01, 1.55	0.04	1.40	1.07, 1.83	0.01
Animal protein	1.27	1.02, 1.57	0.03	1.44	1.09, 1.91	0.01
Plant protein	0.73	0.41, 1.29	0.27	2.25	0.95, 5.35	0.72

Values are hazard ratios (HRs) and 95% confidence intervals (CI) for per 5 E% protein intake difference with risk of type 2 diabetes at the expense of carbohydrate.

All models are multivariate models corresponding to model 2 in the main analyses, adjusted for total fat intake (5 E%), total energy intake (kcal/d), alcohol intake (5 E%), protein intake 20 years after baseline in RS-I or 10 years after baseline in RS-II, age, sex, smoking status (never, past or current), education level (primary, lower, intermediate or higher), diet quality score, physical activity (continuous) and family history of diabetes (for RS-I and RS-II only).

Chapter 2.2

Dietary protein and mortality

Chen Z, Glisic M, Song M, Aliahmad HA, Zhang X, Moumdjian AC, Gonzalez-Jaramillo V, Van der Schaft N, Bramer WM, Ikram MA, Voortman T. Dietary protein intake and all-cause and cause-specific mortality: results from the Rotterdam Study and a meta-analysis of prospective cohort studies.

ABSTRACT

Background: Short-term high protein diets appear to improve cardiometabolic risk profile, but long-term high protein intake has been associated with higher cardiometabolic diseases risk. Evidence for associations between protein intake with mortality is inconsistent.

Objectives: We aimed to examine associations of dietary protein from different sources with all-cause and cause-specific mortality.

Methods: We followed 7786 participants from three sub-cohorts of the Rotterdam Study, a population-based cohort in the Netherlands. Dietary data were collected using food-frequency questionnaires at baseline (1989-93, 2000-01, 2006-08). Deaths were followed until 2018. Associations were examined using Cox regression. Additionally, we performed a highest versus lowest meta-analysis and a dose-response meta-analysis to summarize results from the Rotterdam Study and previous prospective cohorts.

Results: During a median follow-up of 13.0 years, 3589 deaths were documented in the Rotterdam Study. In this cohort, after multivariable adjustment, higher total protein intake was associated with higher all-cause mortality (e.g. highest versus lowest quartile of total protein intake as percentage of energy (Q4 versus Q1), HR=1.12 (1.01, 1.25)); mainly explained by higher animal protein and CVD mortality (Q4 versus Q1, CVD mortality: 1.28 (1.03, 1.60)). Plant protein intake was not associated with all-cause or cause-specific mortality. These findings for total and animal protein intake were corroborated in a meta-analysis of eleven prospective cohort studies including the Rotterdam Study (total 64306 deaths among 350452 participants): higher total protein intake was associated with higher all-cause mortality (pooled RR for highest versus lowest quantile 1.05 (1.01, 1.10)), and for dose-response per 5 energy percent (E%) increment, 1.02 (1.004, 1.04)); again mainly driven by an association between animal protein and CVD mortality (highest versus lowest, 1.09 (1.01, 1.18), per 5 E% increment, 1.05 (1.02, 1.09)). Furthermore, in the meta-analysis a higher plant protein intake was associated with lower all-cause and CVD mortality (e.g. for all-cause mortality, highest versus lowest, 0.93 (0.87, 0.99), per 5 E% increment, 0.87 (0.78, 0.98); for CVD mortality, highest versus lowest 0.86 (0.73, 1.00)).

Conclusions: Evidence from prospective cohort studies to date suggests that total protein intake is positively associated with all-cause mortality, mainly driven by a harmful association of animal protein with CVD mortality. Plant protein intake is inversely associated with all-cause and CVD mortality. Our findings support current dietary recommendations to increase intake of plant protein in place of animal protein.

INTRODUCTION

Defining the role of dietary protein intake in health has been a long-standing research topic of interest and remains a high priority in nutrition research. Although protein delivers amino acids that are crucial for various body functions, protein intake in the general population tends to be much higher than required.[1] Short-term randomized clinical trials have indicated that a higher protein intake replacing carbohydrate favors weight management and improves blood lipid and lipoprotein profiles and glycemic regulation.[2-4] These beneficial effects on cardiometabolic risk profile have been shown to be partly dependent on weight loss and possibly owing to the enhanced postprandial satiety and energy expenditure when replacing carbohydrate with protein.[5] However, several prospective observational studies have reported that long-term high intake of total and animal protein is associated with higher risk of type 2 diabetes[6] and cardiovascular diseases (CVD).[7]

Recently, to further explore the role of dietary protein intake in overall health, several previous studies have examined the associations between protein intake and all-cause and cause-specific mortality, but with apparently inconsistent results.[8-16] For example, Song et al. reported that higher animal protein intake was associated with higher CVD mortality risk, while higher plant protein intake was associated with lower risk of all-cause and CVD mortality.[12] In contrast, Kelemen et al. reported null associations of total and animal protein with risk of all-cause and CVD mortality, but beneficial association of plant protein with CVD mortality.[9] Tharrey et al. observed that higher animal protein intake was associated with higher CVD mortality, while plant protein intake was not associated with CVD mortality.[15]

Therefore, we aimed to investigate the associations of total, animal, and plant protein intake with all-cause and cause-specific mortality in the Rotterdam Study. Furthermore, to clarify the currently mixed evidence from previous studies, we also systematically reviewed and meta-analyzed our findings with those from previous prospective studies to evaluate the association of dietary protein intake with mortality.

METHODS

The current study consisted of two stages. First, we analyzed the associations of protein intake with all-cause and cause-specific mortality in the Rotterdam study. Second, we conducted a systematic review and meta-analysis by combining the new results from the Rotterdam Study with results from previous prospective cohort studies.

Methods in the Rotterdam Study

Study design and population in the Rotterdam Study

The first stage of this study was conducted within three sub-cohorts of the Rotterdam study (RS), a large prospective cohort study of participants aged 45 years and above in Rotterdam, the Netherlands.[17] Briefly, the first sub-cohort of the Rotterdam Study (RS-I) was initiated in the period of 1989-93 by recruiting participants aged ≥ 55 years from the district of Ommoord (n=7983). In 2000-01, the study was extended with a second sub-cohort (RS-II) including new individuals who had become 55 years of age or moved into the study area (n=3011). In 2006-08, a third sub-cohort (RS-III) was started of new individuals aged 45 years and older (n=3932). Until the end of 2008, 14926 participants were contained in the three sub-cohorts at baseline. We collected information every 3-5 years through interviews for which we visited the participants at their homes, through questionnaires which the participants returned, and through examinations in our dedicated research center which is in the Ommoord district. In the home interviews, we collected information such as education level, smoking status, medical history, and income. At the examination center, we mainly focused on examinations of imaging (of heart, blood vessels, eyes, skeleton, and later brain) and on collecting biospecimens that enabled further in-depth molecular and genetic analysis. The Rotterdam Study has been approved by the institutional review board (Medical Ethics Committee) of Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. The approval has been renewed every 5 years. All participants gave informed consent.[17]

For the current analysis within the Rotterdam Study, of the 14926 participants who participated at baseline, we had reliable dietary data available for 9701. Reasons for absence of valid dietary data were: not having received dietary assessment because of logistic reasons; living in a resident home for elderly, or suspected dementia; not having completed the dietary assessment; or unreliable dietary intake according to a trained nutritionist or an estimated energy intake of <500 or >5000 kcal/day).[18,19] From these 9701 participants, we further excluded 1914 participants with CVD, diabetes, or cancer at baseline, and 1 participant without follow-up data on mortality, leaving 7786 participants for the main analyses (Supplemental Figure 1).

Dietary assessment

Dietary intake in the Rotterdam Study was assessed at baseline in all three sub-cohorts using a semi-quantitative food questionnaire (FFQ) as described in more details elsewhere.[19,20] Briefly, we used an FFQ with 170 food items to assess dietary intake at baseline of RS-I (1989-93) and at baseline of RS-II (2000-01). This 170-item FFQ was validated against fifteen 24-hour food records and four 24-hour urinary urea excretion samples which were collected on non-consecutive days over a period of a year in a subsample of the Rotterdam Study (n=80), as described in detail elsewhere[21]: Adjusted Pearson's

correlation intake against the food records were 0.66 for total protein intake and 0.59 for plant protein intake; and Spearman's correlation for protein intake against urinary urea was 0.67. A 389-item FFQ was used to assess dietary intake at baseline of RS-III (2006-08). This 389-item FFQ was previously validated in two other Dutch populations using a 9-day dietary record[22] and a 4-week dietary history[23]: Pearson's correlations for intakes of different nutrients varied from 0.40 to 0.86. Food intake data were converted to energy and nutrient intake based on the Dutch Food Composition tables.

Ascertainment of death

Information on vital status of the participants was obtained from clinical follow-up data collection and from municipal records. General practitioners reported events of interest by means of a computerized system or notified new events annually. Trained research assistants subsequently collected information from medical records at the general practitioners' offices, hospitals and nursing homes. Two research physicians independently identified the events according to the International Classification of Diseases, Tenth revision (ICD-10). Afterwards, a senior physician reviewed all coded events. Cause-specific mortality was recoded according to the ICD-10 codes (CVD cause: F01, I05-99 (non-stroke CVD cause: I05-51, 70-99, stroke cause: F01, I60-69); cancer cause: C01-97). Coded information on all-cause mortality was available until May 2018 and coded information on cause-specific mortality was available until January 2015.

Covariates

Information on age, sex, smoking status, and education level was obtained from questionnaires at baseline. Information on physical activity was obtained using the adapted version of the Zutphen Physical Activity Questionnaire at the third visit of RS-I (1997-99) and at baseline of RS-II, and the LASA Physical Activity Questionnaire at baseline of RS-III. Physical activities were weighted according to intensity with Metabolic Equivalent of Task (MET), from the Compendium of Physical Activities version 2011. To account for differences between the two questionnaires, questionnaire-specific z-scores of MET-hours per week were calculated. We measured body weight and height at baseline in our research center and body mass index (BMI) was calculated. A previously defined diet quality score was calculated to reflect adherence to Dutch dietary guidelines as described in detail elsewhere.[20] Briefly, this was a sum-score of the adherence to guidelines for 14 individual foods items, including: vegetables (≥200 g/day), fruit (≥200 g/day), whole-grains (≥90 g/day), legumes (≥135 g/week), nuts (≥15 g/day), dairy (≥350 g/day), fish (≥100 g/week), tea (≥450 mL/day), ratio whole-grains: total grains (≥50%), ratio unsaturated fats and oils: total fats (≥50%), red and processed meat (<300 g/week), sugar-containing beverages (≤150 mL/day), alcohol (≤10 g/day) and salt (≤6 g/day). We scored every participant as adhering to this item ('yes' scored as 1) or not adhering to the item ('no' scored as 0). The total diet quality score theoretically ranged from 0 (no adherence) through 14

(full adherence). We previously reported that a higher score was associated with a lower premature mortality risk and a lower risk of developing some of the chronic diseases on which the guidelines were based.[20] Information on CVD, diabetes, and cancers was obtained from general practitioners, pharmacies' datasets, Nationwide Medical Register, or follow-up examinations in our research center.

Data analyses

We expressed intake of dietary protein and other macronutrients as a percentage of total energy consumption. Baseline characteristics of the Rotterdam population are presented for the whole group and by quartiles of total, animal, or plant protein intake. Trends of these characteristics across quartiles of protein intake were examined by using means and linear regression for continuous variables, or chi-square tests for categorical variables. After confirming that the assumption for proportional hazards was met on the basis of Schoenfeld residuals,[24] we used Cox proportional hazard models to analyze associations of dietary protein intake with all-cause and cause-specific mortality. Because effects of macronutrient intake cannot be separated from the effects of overall energy intake or intake of other macronutrients, we modelled macronutrient substitution effects. For our main models, we used multivariable nutrient density substitution models for protein intake at the expense of carbohydrate. For this aim, models were used with adjustment for total energy intake and percentage of energy from subtypes of fats (saturated fat (SFA), monounsaturated fat (MUFA), polyunsaturated fat (PUFA), and trans-fat (TSF)), and from alcohol.[25] Therefore, coefficients from these models were interpreted as the estimated effects of substituting a certain percentage of energy from protein intake for equivalent energy from carbohydrate intake. We estimated hazard ratios (HRs) and their 95% confidence intervals (CIs) of mortality by comparing participants in each quartile of protein intake as percentage of energy with those in the lowest quartile. To quantify a linear trend, we assigned the median within each quartile and modeled this variable continuously. Furthermore, we also modelled dietary protein intake as continuous variable and estimated HRs and 95%CIs per 5 energy percent (E%) increment from protein at the expense of carbohydrate.

For all main analyses, we included intake of protein, SFA, MUFA, PUFA, TSF, total energy, alcohol, baseline age, sex, and RS-cohort in model 1; we additionally adjusted for smoking status, education level, overall diet quality score, fiber intake, physical activity, and BMI in model 2. For analysis of animal and plant protein intake, mutual adjustment for plant and animal protein was performed.

Sensitivity analyses

We conducted a series of sensitivity analyses to test robustness of our main results. First, we replaced fat intake by carbohydrate in the main models (model 2), to examine whether it made a difference if dietary protein was consumed at the expense of fat instead of carbohydrate. Second, we examined if the associations of animal protein intake with all-cause and cause-specific mortality differed by protein

from specific animal food sources, such as meat, dairy, fish, and eggs at expense of carbohydrate. In this modelling approach, the percentages of energy intake from protein from meat, dairy, fish, and eggs were simultaneously included in one model, with adjustment for plant protein and all other covariates in model 2. Third, we examined the interaction effect of total, animal, or plant protein with age, sex, BMI, or physical activity by including their interaction terms in model 2, to explore whether the associations of protein intake and mortality differed by these factors. Last, to minimize reverse causality bias, we excluding the participants who died within the first 2 years of follow-up in the Rotterdam study.

We performed all analyses based on combined data from RS-I, RS-II, and RS-III. All variables included in analyses were used to predict missingness patterns. Missing values on covariates were assumed to be missing at random and accounted for using multiple imputations (m=10 imputations).[26] Supplemental Table 1 shows the percentage of missing data for covariates in the Rotterdam Study. Statistical procedures were performed with the use of R version 3.3.1 (The R Foundation for Statistical Computing, Vienna, Austria).

Methods for the systematic review and meta-analysis

The systematic review was conducted using a predefined protocol and reported in accordance with the PRISMA and MOOSE guidelines.[27, 28] Medline (Ovid), Embase.com, and the Cochrane Central Register of Controlled Trials were searched from inception until August 27, 2019 (date last searched), with assistance of an experienced biomedical information specialist. The detailed search strategy is shown in Supplemental Table 2. Two independent reviewers conducted an initial screening of all titles or abstracts and then evaluated all potentially relevant articles based on full text reviews. Eligible studies were included if they (i) were observational studies with a longitudinal design (i.e., prospective cohort); and (ii) had assessed the variance of estimates of the association between dietary protein intake (total, animal and/or plant protein) with all-cause mortality and/or cause-specific mortality in a general population (i.e., populations that were not selected based on pre-existing health conditions). We contacted the investigators for relevant data if their studies were potentially eligible for this study. We extracted the following characteristics from the included studies: first author, cohort name, country, publication year, age at entry, sex, sample size, duration of follow-up, assessment of dietary protein intake, ascertainment of outcomes, the most adjusted association estimates and corresponding measures of variation, and variables that were entered into the multivariable model as potential confounders. In case of multiple publications from the same study, we included the most up-to date or comprehensive information. We used the nine-star Newcastle–Ottawa Scale (NOS) to assess study quality on the basis of selection of three domains: selection of participants (population representativeness), comparability (adjustment for confounders), and ascertainment of outcomes of interest. Nine points on the NOS reflects the highest study quality.[29]

Data synthesis and analysis

We conducted highest versus lowest and dose-response meta-analyses, using the most adjusted association estimation from each original study. For the main meta-analysis, we estimated pooled RRs for highest versus lowest quantile of protein intake using random-effects models.[30] Heterogeneity was determined using the Cochrane χ^2 statistic and the I^2 statistic.[31] We additionally conducted dose-response meta-analyses for all-cause, CVD, and cancer mortality, using a generalized least-squares regression approach.[32] In estimating dose-response trends, several approximations across categories of dietary protein intake were applied: the midpoint or mean value of the amount of dietary protein intake, distributions of deaths and person years, HR and 95% CI. When sufficient data ($n \geq 5$) studies[33] contributed to a dose-response meta-analysis, non-linearity was explored using restricted cubic splines with three knots (10%, 50%, and 90%).[34] A Wald-type test was used to test statistical significance of non-linearity.[34]

In the main meta-analyses comparing quantiles, we conducted subgroup analyses stratified by geographical study location. As sensitivity analysis, we examined the influence of individual studies on the overall risk estimates comparing quantiles by recalculating the pooled estimates after excluding one study at a time. As a second set of sensitivity analyses, we additionally incorporated studies reporting estimations not expressed in E% in the dose-response meta-analysis, for which we could only approximate protein intake in E%.[15] Additionally, publication bias was evaluated through a funnel plot[35] and Egger's test.[36, 37] We used STATA release 12 (Stata Corp, College Station, Texas) for all highest versus lowest meta-analyses. The dose-response meta-analysis was conducted with "dosresmeta" package[34] in R version 3.3.1 (The R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Results in the Rotterdam Study

Characteristics of population

For the 7786 participants of the Rotterdam Study in our current analysis, mean age at baseline was 63.7 ± 8.7 years, and 60.8% of all participants were female. Furthermore, average total protein intake was 85.8 ± 25.1 g/day (16.4% ± 2.3% of total energy), this corresponded to 1.20 ± 0.3 g/kg BW/day, which is higher than the recommended intake of 0.8 g/kg BW/day (34). Most protein came from animal sources (53.6 ± 19.0 g/day, and 10.3% ± 2.5% of total energy). Compared to participants in the lowest quartile of animal protein intake, those in the highest quartile of animal protein had a slightly higher BMI, were more often lower educated, and more likely to smoke. In contrast, compared to the

participants in the lowest quartile of plant protein intake, those in the highest quartile of plant protein intake had higher overall diet quality, and were more often highly educated, and less likely to smoke (Table 1, Supplemental Table 3).

Associations of protein intake with all-cause mortality and cause-specific mortality

During a median follow-up of 13.0 years (25th-75th percentile, 8.3-19.1 years, data on all-cause mortality available until May 2018), we documented 3589 deaths. During a median follow-up of 12.8 years (25th-75th percentile, 8.2-19.0 years, data on cause-specific mortality available until January 2015), we documented 877 CVD deaths (of which, 594 non-stroke CVD deaths and 283 stroke deaths), 896 cancer deaths, and 1289 deaths due to other causes (which consisted of various specific causes, all with relatively small numbers).

As shown in Table 2, in multivariable models (Model 2), higher total protein intake was associated with higher risk of all-cause mortality, CVD mortality, and non-stroke CVD mortality (e.g. for all-cause mortality, the highest quartile versus the lowest quartile of total protein intake as percentage of energy (Q4 versus Q1), HR: 1.12, 95%CI (1.01, 1.25); per 5 E% increment in total protein, HR: 1.09, 95%CI (1.02, 1.17); and for CVD mortality, per 5 E% HR: 1.20, 95%CI (1.05, 1.37)). These associations were mainly explained by animal protein intake (Table 3) (e.g. Q4 versus Q1, for all-cause mortality, 1.18 (1.05, 1.31), for CVD mortality, 1.28 (1.03, 1.60); per 5 E% increment, for all-cause mortality, 1.20 (1.05, 1.37); for CVD mortality, 1.19 (1.04, 1.37). Total, or animal protein intake was not associated with stroke mortality, cancer mortality, and other mortality. Besides, plant protein intake was not associated with all-cause and cause-specific mortality (Tables 2-4).

Sensitivity analyses

In the Rotterdam Study, we observed similar results for protein intake and all-cause and cause-specific mortality when at the expense of fat instead of carbohydrate (Supplemental Table 4). We also observed that higher intake of protein from meat or from dairy was associated higher risk of all-cause mortality or CVD mortality (Supplemental Table 5). We observed no interaction effects of protein intake with age, BMI, or physical activity, but we did observe a significant interaction effect of animal protein intake with sex for all-cause mortality (p value for the interaction term = 0.02). Specifically, the association between animal protein intake and all-cause mortality was stronger in male participants (Q4 versus Q1: 1.42 (1.20, 1.68), per 5 E% increment: 1.42 (1.13, 1.77)); while the association was null in female participants (Q4 versus Q1: 1.01 (0.87, 1.17), per 5 E% increment: 1.04 (0.90, 1.21)). Last, the results were similar after excluding deaths cases within the first two years of follow-up (Supplemental Table 6).

Table 1. Characteristics of the Rotterdam Study population

		By extreme quartiles of total protein		By extreme quartiles of animal protein		By extreme quartiles of plant protein	
	(n=7786)	**Quartile 1 (n=1947) ≤14.4 E%**	**Quartile 4 (n=1947) >18.1 E%**	**Quartile 1 (n=1947) ≤8.4 E%**	**Quartile 4 (n=1947) >12.1 E%**	**Quartile 1 (n=1947) ≤5.2 E%**	**Quartile 4 (n=1947) >6.7 E%**
Age (years)	63.7 (8.7)	63.9 (9.4)	64.2 (8.2)	62.1 (9.1)	64.9 (8.2)*	66.2 (8.8)	60.6 (7.9)*
Sex (%)							
-Female	60.8	12.8	18.4*	12.9	17.9*	13.9	15.6*
-Male	39.2	12.7	6.6*	12.1	7.1*	11.1	9.4*
BMI (kg/m^2)	26.6 (3.9)	25.8 (3.6)	27.5 (4.3)*	25.8 (3.7)	27.4 (4.2)*	26.4 (3.7)	26.6 (4.1)
Smoking status (%)							
-Never	23.8	7.2	9.3	7.8	8.9	6.8	9.1
-Ever	42.3	11.0	10.1	11.4	9.8	9.6	11.2
-Current	23.8	6.7	5.4*	5.7	6.2*	8.5	4.7*
Education level (%)							
-Primary	15.3	3.6	4.1	3.1	4.5	4.6	3.1
-Low	41.1	10.0	10.9	9.5	10.9	10.8	9.6
-Intermediate	27.2	7.2	6.6	7.1	6.4	6.6	6.5
-High	15.8	4.1	3.2	5.2	3.1*	2.9	5.7*
Physical activity (MET-hours/week)							
-RS-I and II	80.4	74.3	83.9	76.7	83.0	75.9	84.8
	(55.7,113.2)	(47.8,105.6)	(59.8,117.5)*	(49.5,108.7)	(58.1,116.0)*	(47.6, 105.5)	(61.6, 121.0)*
-RS-III	43.0	41.0	39.4	42.8	38.0	38.0	48.0
	(17.7, 82.6)	(17.1,84.3)	(15.0, 56.4)	(18.0,84.1)	(15.0, 72.4)	(16.0, 72.4)	(18.7, 87.8)*
Total protein (g/day)	85.8 (25.1)	76.6 (21.2)	90.2 (25.4)*	78.9 (23.4)	90.4 (25.5)*	83.5 (25.0)	86.2 (24.3)*
Total protein (E%)	16.4 (2.3)	13.0 (1.3)	20.3 (2.1)*	13.4 (1.7)	20.0 (2.3)*	15.8 (3.2)	16.8 (2.9)*

Table 1. Characteristics of the Rotterdam Study population (Continued)

		By extreme quartiles of total protein		By extreme quartiles of animal protein		By extreme quartiles of plant protein	
	(n=7786)	**Quartile 1 (n=1947) ≤14.4 E%**	**Quartile 4 (n=1947) >18.1 E%**	**Quartile 1 (n=1947) ≤8.4 E%**	**Quartile 4 (n=1947) >12.1 E%**	**Quartile 1 (n=1947) ≤5.2 E%**	**Quartile 4 (n=1947) >6.7 E%**
-Animal protein (g/day)	53.6 (19.0)	42.9 (14.0)	63.0 (30.0)*	40.5 (13.3)	65.5 (20.9)*	59.4 (21.6)	46.2 (15.8)*
-Animal protein (E%)	10.3 (2.5)	7.3 (1.6)	14.2 (2.5)*	6.9 (1.3)	14.5 (2.2)*	11.3 (3.3)	9.1 (3.0)*
-Plant protein (g/day)	30.3 (8.9)	33.7 (12.1)	27.2 (9.4)*	38.4 (13.7)	25.0 (7.6)*	24.0 (7.1)	39.9 (13.5)*
-Plant protein (E%)	6.2 (1.5)	5.7 (1.3)	6.1 (1.5)*	6.5 (1.6)	5.6 (1.2)*	4.5 (0.6)	7.6 (1.1)*
Total fat (g/day)	82.7 (29.7)	93.5 (34.2)	68.9 (24.2)*	72.8 (27.0)	90.9 (35.1)*	91.2 (33.3)	76.5 (29.7)*
Total fat (E%)	35.3 (6.6)	35.3 (7.3)	34.5 (6.2)*	34.3 (7.3)	35.6 (6.5)*	38.1 (7.1)	32.3 (6.0)*
-SFA (g/day)	31.7 (12.3)	35.1 (13.6)	27.0 (10.7)*	29. 2 (12.5)	32.8 (13.4)*	37.1 (14.4)	26.8 (10.5)*
-SFA (E%)	13.6 (3.3)	13.3 (3.4)	13.5 (3.3)*	12.4 (3.3)	14.2 (3.4)*	15.5 (3.5)	11.5 (2.6)*
-MUFA (g/day)	27.8 (11.2)	31.8 (13.2)	23.0 (8.6)*	24.3 (13.7)	30.9 (9.4)*	30.9 (12.7)	25.9 (11.2)*
-MUFA (E%)	11.8 (2.8)	11.9 (3.2)	11.5 (2.5)*	11.5 (3.2)	11.9 (2.7)*	12.8 (3.2)	11.0 (2.7)*
-PUFA (g/day)	16.5 (7.8)	19.3 (9.3)	13.0 (6.0)*	19.8 (9.4)	13.1 (6.2)*	16.4 (8.5)	17.0 (8.1)*
-PUFA (E%)	7.0 (2.5)	7.2 (2.7)	6.5 (2.5)*	7.4 (2.7)	6.5 (2.5)*	6.8 (2.7)	7.2 (2.3)*
-TSF (g/day)	1.7 (1.2, 2.4)	1.9 (1.3, 2.9)	1.4 (1.0, 1.9)	1.6 (1.1, 2.4)	1.5 (1.1, 2.1)*	2.1 (0.25, 3.21)	1.27 (0.91, 1.80)*
-TSF (E%)	0.72 (0.52, 1.05)	0.72 (0.52, 1.10)	0.72 (0.54, 0.97)	0.61 (0.45, 0.92)	0.77 (0.59, 1.06)*	0.92 (0.65, 1.33)	0.56 (0.43, 0.74)*
Carbohydrate (g/day)	228.2 (76.6)	269.0 (86.5)	186.2 (56.2)*	275.0 (87.1)	183.9 (56.3)*	216.4 (77.7)	243.7 (80.9)*
Carbohydrate (E%)	43.5 (7.1)	45.5 (7.9)	41.9 (6.7)*	46.7 (7.6)	40.6 (6.8)*	40.1 (7.9)	46.6 (6.4)*
Diet quality score	6.7 (1.9)	6.3 (2.0)	7.2 (1.8)*	6. 8 (2.0)	6.9 (1.8)	5.7 (1.7)	7.7 (1.7)*

Table 1. Characteristics of the Rotterdam Study population (Continued)

		By extreme quartiles of total protein		By extreme quartiles of animal protein		By extreme quartiles of plant protein	
	(n=7786)	**Quartile 1 (n=1947) ≤14.4 E%**	**Quartile 4 (n=1947) >18.1 E%**	**Quartile 1 (n=1947) ≤8.4 E%**	**Quartile 4 (n=1947) >12.1 E%**	**Quartile 1 (n=1947) ≤5.2 E%**	**Quartile 4 (n=1947) >6.7 E%**
Fiber (gram)	19.5 (15.1, 26.6)	20.8 (15.4, 28.9)	17.7 (14.3, 22.6)*	24.6 (18.1, 33.9)	16.6 (13.6, 20.9)*	15.2 (12.2, 19.5)	26.2 (19.6, 35.2)*

Variables expressed as mean (SD), median (25th percentile–75th percentile), or percentage.

P-trend was assessed were tested with linear regression (continuous variables) or with chi-square test (categorical variables).

* indicates $P < 0.05$ for trend across quartiles.

Abbreviations: MET, metabolic equivalent of task; E%, energy percent; SFA, saturated fat acids, MUFA, monounsaturated fat acids, PUFA, polyunsaturated fat acids, TSF, trans fat acid

Table 2. Associations of total protein intake with all-cause and cause-specific mortality in the Rotterdam Study (comparison is isocaloric substitution for carbohydrate)

Total protein	HR (95% CI) per 5 E% increment n = 7786	Quartile 1 n=1947	Quartile 2 n=1946	Quartile 3 n=1946	Quartile 4 n=1947	P trend
Median intake (E%)	**16.2**	**13.3**	**15.3**	**17.0**	**19.7**	
All-cause mortality						
Number of deaths	n=3589	n=878	n=853	n=884	n=974	
Model 1	1.04 (0.97, 1.11)	1 (Reference)	1.00 (0.91, 1.10)	0.96 (0.87, 1.05)	1.04 (0.93, 1.15)	0.60
Model 2	1.09 (1.02, 1.17)	1 (Reference)	1.05 (0.95, 1.16)	1.02 (0.92, 1.13)	1.12 (1.01, 1.25)	0.06
CVD mortality						
Number of deaths	n=877	n=220	n=191	n=205	n=261	
Model 1	1.16 (1.01, 1.32)	1 (Reference)	0.95 (0.78, 1.16)	0.93 (0.76, 1.14)	1.16 (0.94, 1.42)	0.15
Model 2	1.20 (1.05, 1.37)	1 (Reference)	0.99 (0.81, 1.21)	0.99 (0.80, 1.21)	1.22 (0.99, 1.52)	0.06
Non-stroke CVD mortality						
Number of deaths	n=594	n=147	n=125	n=143	n=179	
Model 1	1.24 (1.06, 1.45)	1 (Reference)	0.93 (0.73, 1.18)	0.96 (0.76, 1.23)	1.19 (0.92, 1.53)	0.13
Model 2	1.27 (1.08, 1.49)	1 (Reference)	0.96 (0.75, 1.23)	1.02 (0.79, 1.31)	1.23 (0.95, 1.60)	0.04
Stroke mortality						
Number of deaths	n=283	n=73	n=66	n=62	n=82	
Model 1	0.99 (0.78, 1.24)	1 (Reference)	1.00 (0.71, 1.40)	0.85 (0.59, 1.21)	1.09 (0.76, 1.57)	0.86
Model 2	1.05 (0.83, 1.33)	1 (Reference)	1.04 (0.74, 1.47)	0.91 (0.64, 1.32)	1.19 (0.82, 1.74)	0.42
Cancer mortality						
Number of deaths	n=896	n=243	n=220	n=220	n=213	
Model 1	0.92 (0.81, 1.06)	1 (Reference)	0.92 (0.76, 1.10)	0.87 (0.71, 1.05)	0.84 (0.68, 1.04)	0.10
Model 2	0.94 (0.82, 1.08)	1 (Reference)	0.95 (0.78, 1.14)	0.89 (0.73, 1.08)	0.87 (0.70, 1.08)	0.18

Table 2. Associations of total protein intake with all-cause and cause-specific mortality in the Rotterdam Study (comparison is isocaloric substitution for carbohydrate) (Continued)

Total protein	HR (95% CI) per 5 E% increment n = 7786	Quartile 1 n=1947	Quartile 2 n=1946	Quartile 3 n=1946	Quartile 4 n=1947	P trend
Median intake (E%)	**16.2**	**13.3**	**15.3**	**17.0**	**19.7**	
Other mortality						
Number of deaths	n=1289	n=311	n=309	n=318	n=351	
Model 1	1.00 (0.90, 1.11)	1 (Reference)	1.04 (0.89, 1.22)	0.95 (0.80, 1.12)	1.02 (0.86, 1.22)	0.99
Model 2	1.09 (0.97, 1.22)	1 (Reference)	1.12 (0.95, 1.32)	1.06 (0.89, 1.25)	1.16 (0.97, 1.39)	0.17

Effect estimates are hazard ratios (HRs) and 95%-confidence intervals (95%CIs) derived from Cox proportional hazards regression models. Estimates are based on pooled results of imputed data.

Model 1: Age, sex, RS-cohort (RS-I, -II, and -III), intake of total energy, SFA (E%), MUFA (E%), PUFA (E%), TSFA (E%) and alcohol (E%).

Model 2: Model 1 + fiber, overall diet quality score, physical activity (z-score of metabolic equivalents of task-hours/week), education level (primary, lower, intermediate, and high), smoking status (never, ever, current), and BMI.

Abbreviations: SFA, saturated fat acids, MUFA, monounsaturated fat acids, PUFA, polyunsaturated fat acids, TSF, trans fat acids; BMI, body mass index.

Table 3. Associations of animal protein with all-cause and cause-specific mortality in the Rotterdam Study (comparison is isocaloric substitution for carbohydrate)

Animal protein	HR (95% CI) per 5 E% increment n = 7786	Quartile 1 n=1947	Quartile 2 n=1946	Quartile 3 n=1946	Quartile 4 n=1947	P trend
Median intake (E%)	**10.2**	**7.2**	**9.3**	**11.1**	**13.9**	
All-cause mortality						
Number of deaths	n=3589	n=692	n=887	n=970	n=1040	
Model 1	1.10 (0.96, 1.25)	1 (Reference)	1.07 (0.96, 1.18)	1.05 (0.95, 1.16)	1.13 (1.01, 1.26)	0.04
Model 2	1.20 (1.05, 1.37)	1 (Reference)	1.06 (0.95, 1.17)	1.08 (0.97, 1.20)	1.18 (1.05, 1.31)	0.003
CVD mortality						
Number of deaths	n=877	n=169	n=216	n=219	n=273	
Model 1	1.16 (1.02, 1.32)	1 (Reference)	1.08 (0.88, 1.32)	0,97 (0.79, 1.20)	1.24 (1.00, 1.54)	0.08
Model 2	1.19 (1.04, 1.37)	1 (Reference)	1.07 (0.87, 1.32)	1.01 (0.82, 1.25)	1.28 (1.03, 1.60)	0.03
Non-stroke CVD mortality						
Number of deaths	n=594	n=109	n=145	n=150	n=190	
Model 1	1.25 (1.07, 1.47)	1 (Reference)	1.11 (0.86, 1.43)	1.05 (0.81, 1.35)	1.36 (1.04, 1.77)	0.02
Model 2	1.27 (1.08, 1.49)	1 (Reference)	1.10 (0.85, 1.42)	1.06 (0.81, 1.37)	1.34 (1.03, 1.75)	0.03
Stroke mortality						
Number of deaths	n=283	n=60	n=71	n=69	n=83	
Model 1	0.98 (0.78, 1.24)	1 (Reference)	1.01 (0.71, 1.43)	0.84 (0.58, 1.21)	1.03 (0.71, 1.49)	0.99
Model 2	1.05 (0.83, 1.33)	1 (Reference)	1.01 (0.71, 1.44)	0.90 (0.62, 1.30)	1.12 (0.76, 1.64)	0.64
Cancer mortality						
Number of deaths	n=896	n=184	n=248	n=230	n=234	
Model 1	0.93 (0.82, 1.07)	1 (Reference)	1.09 (0.90, 1.32)	0.94 (0.77, 1.15)	0.97 (0.78, 1.21)	0.51
Model 2	0.95 (0.82, 1.08)	1 (Reference)	1.08 (0.89, 1.32)	0.94 (0.77, 1.16)	0.98 (0.78, 1.22)	0.54

Table 3. Associations of animal protein with all-cause and cause-specific mortality in the Rotterdam Study (comparison is isocaloric substitution for carbohydrate) (Continued)

Animal protein	HR (95% CI) per 5 E% increment n = 7786	Quartile 1 n=1947	Quartile 2 n=1946	Quartile 3 n=1946	Quartile 4 n=1947	P trend
Median intake (E%)	**10.2**	**7.2**	**9.3**	**11.1**	**13.9**	
Other mortality						
Number of deaths	n=1289	n=240	n=301	n=364	n=384	
Model 1	1.02 (0.92, 1.14)	1 (Reference)	1.02 (0.86, 1.21)	1.03 (0.87, 1.22)	1.09 (0.91,1.31)	0.22
Model 2	1.09 (0.97, 1.22)	1 (Reference)	1.03 (0.86, 1.22)	1.11 (0.93, 1.32)	1.21 (1.00, 1.46)	0.07

Effect estimates are hazard ratios (HRs) and 95%-confidence intervals (95%CIs) derived from Cox proportional hazards regression models. Estimates are based on pooled results of imputed data.

Model 1: Age, sex, and RS-cohort (RS-I, -II, and -III), total energy, SFA (E%), MUFA (E%), PUFA (E%), TSFA (E%), and alcohol (E%).

Model 2: Model 1 + fiber, overall diet quality, physical activity (z-score of metabolic equivalents of task-hours/week), education level (primary, lower, intermediate, and high), smoking status (never, ever, current), and BMI.

Abbreviations: SFA, saturated fat acids, MUFA, monounsaturated fat acids, PUFA, polyunsaturated fat acids, TSF, trans fat acids; BMI, body mass index.

Table 4. Associations of plant protein with all-cause and cause-specific mortality in the Rotterdam study (comparison is isocaloric substitution for carbohydrate)

Plant protein	HR (95% CI) per 5 E% increment n=7786	Quartile 1 n=1947	Quartile 2 n=1946	Quartile 3 n=1946	Quartile 4 n=1947	P trend
Median intake (E%)	**5.8**	**4.6**	**5.5**	**6.2**	**7.3**	
All-cause mortality						
Number of deaths	n=3,589	n=1,176	n=986	n=813	n=614	
Model 1	0.80 (0.67, 0.95)	1 (Reference)	0.85 (0.78, 0.93)	0.84 (0.76, 0.93)	0.90 (0.80, 1.01)	0.05
Model 2	1.09 (0.88, 1.35)	1 (Reference)	0.93 (0.85, 1.03)	0.94 (0.84, 1.04)	1.06 (0.92, 1.21)	0.53
CVD mortality						
Number of deaths	n=877	n=282	n=235	n=210	n=150	
Model 1	1.02 (0.72, 1.46)	1 (Reference)	0.88 (0.74, 1.06)	0.98 (0.80, 1.19)	1.03 (0.81, 1.30)	0.72
Model 2	1.28 (0.84,1.96)	1 (Reference)	0.96 (0.80, 1.17)	1.06 (0.86, 1.31)	1.19 (0.91, 1.57)	0.17
Non-stroke CVD mortality						
Number of deaths	n=594	n=190	n=155	n=148	n=101	
Model 1	1.03 (0.67, 1.59)	1 (Reference)	0.85 (0.68, 1.05)	0.99 (0.78, 1,25)	0.98 (0.74, 1.31)	0.91
Model 2	1.32 (0.79, 2.22)	1 (Reference)	0.92 (0.73, 1.17)	1.07 (0.83, 1.39)	1.16 (0.83, 1.62)	0.29
Stroke mortality						
Number of deaths	n=283	n=92	n=80	n=62	n=49	
Model 1	1.09 (0.55, 2.16)	1 (Reference)	0.96 (0.70, 1.31)	0.96 (0.68, 1.37)	1.13 (0.74,1.71)	0.64
Model 2	1.36 (0.61, 3.03)	1 (Reference)	1.05 (0.76, 1.46)	1.05 (0.72, 1.52)	1.27 (0.79, 2.04)	0.37
Cancer mortality						
Number of deaths	n=896	n=305	n=227	n=204	n=160	
Model 1	0.72 (0.48, 1.04)	1 (Reference)	0.76 (0.64, 0.91)	0.80 (0.66, 0.97)	0.82 (0.65, 1.03)	0.08
Model 2	0.84 (0.53, 1.32)	1 (Reference)	0.81 (0.68, 0.98)	0.85 (0.69, 1.04)	0.90 (0.69, 1.17)	0.44

Table 4. Associations of plant protein with all-cause and cause-specific mortality in the Rotterdam study (comparison is isocaloric substitution for carbohydrate) (Continued)

Plant protein	HR (95% CI) per 5 E% increment n=7786	Quartile 1 n=1947	Quartile 2 n=1946	Quartile 3 n=1946	Quartile 4 n=1947	P trend
Median intake (E%)	**5.8**	**4.6**	**5.5**	**6.2**	**7.3**	
Other mortality						
Number of deaths	n=1,289	n=441	n=392	n=261	n=195	
Model 1	0.59 (0.43, 0.83)	1 (Reference)	0.92 (0.80, 1.07)	0.75 (0.64, 0.89)	0.85 (0.69, 1.03)	0.01
Model 2	0.90 (0.62, 1.3)	1 (Reference)	1.04 (0.89, 1.21)	0.87 (0.73, 1.04)	1.06 (0.84, 1.34)	0.87

Effect estimates are hazard ratios (HRs) and 95%-confidence intervals (95%CIs) derived from Cox proportional hazards regression models. Estimates are based on pooled results of imputed data.

Model 1: Age, sex, and RS-cohort (RS-I, -II, and -III), total energy, SFA (E%), MUFA (E%), PUFA (E%), TSFA (E%), and alcohol (E%)

Model 2: Model 1 + fiber, overall diet quality, physical activity (z-score of metabolic equivalents of task-hours/week), education level (primary, lower, intermediate, and high), smoking status (never, ever, current), and BMI.

Abbreviations: SFA, saturated fat acids, MUFA, monounsaturated fat acids, PUFA, polyunsaturated fat acids, TSF, trans fat acids; BMI, body mass index.

Meta-analysis results of the Rotterdam Study and previous prospective cohort studies

Literature search and characteristics of studies

In the initial search, we identified 12152 potentially relevant unique citations. After screening and detailed full-text assessment, ten previously published articles were eligible for the systematic review.[8-15, 39, 40] Finally, ten previous studies were eligible for the meta-analysis,[8-15, 39, 40] resulting in a total of eleven prospective studies including the Rotterdam study, with a total number of 350452 participants and 64306 deaths (Supplemental Figure 2). The number of participants (from 1100 to 131342) and deaths (from 60 to 36115) varied widely across these eleven studies. Median duration of follow-up ranged from 12.0 to 28.0 years. Of the eleven studies, eight[9-12, 14, 15, 39] were conducted among North American and European populations (87% of total participants of this meta-analysis), in which the mean or median intake of total protein ranged from around 70 through 93 grams/day, mainly from animal protein intake with a mean or median ranging from around 54 through 65 grams/day. Three studies were conducted within Japanese populations, with a total of 82171participants.[8, 13, 40] Detailed characteristics and quality assessment of these studies have been summarized in Table 5 and Supplemental Table 7. Overall, all the eleven studies were medium to high quality.

Meta-analyzed associations for protein intake and all-cause and cause-specific mortality

Highest versus lowest meta-analysis

Nine studies[8-14, 40] including the Rotterdam Study presented associations of comparing the highest with the lowest categories of protein intake with mortality, and thus were summarized into the highest versus lowest meta-analysis. Figure 1 shows the results of the highest versus lowest meta-analysis. Of the nine studies, six examined associations[9, 11, 12, 14, 40] for total protein intake with all-cause mortality (59841 all-cause deaths among 247863 participants). Comparing the highest quantile of total protein intake with the lowest quantile, the pooled RR was 1.05, 95%CI (1.01, 1.10), I^2 = 9.8%, $P_{heterogeneity}$ = 0.35 for all-cause mortality. Five studies[9, 11, 12, 40] examined associations for total protein and CVD mortality (14704 CVD deaths among 245222 participants), with a pooled estimate of 1.08, 95%CI (0.98, 1.20), I^2 = 20.4%, $P_{heterogeneity}$ = 0.29. Six studies[9-12, 40] examined associations for total protein and cancer mortality; and two studies[12] on other mortality. For both these outcomes, pooled RRs were null (Figure 1A). For animal protein intake, five studies reported associations with all-cause mortality,[9, 12, 14, 40] CVD mortality,[8, 9, 12, 40] or cancer mortality,[9, 10, 12, 40] and two studies[12] with other mortality (Figure 1B). While null pooled associations were observed for all-cause, cancer, and other mortality, a significant pooled RR was observed for CVD mortality: 1.09, 95%CI (1.01, 1.18), I^2 = 0.0%, $P_{heterogeneity}$ = 0.43. For plant protein intake, similar studies were included with a pooled RR of 0.93, 95%CI (0.87, 0.99), I^2 = 38.7%, $P_{heterogeneity}$ = 0.16 for all-cause mortality, and 0.86 (0.73, 1.00), I^2= 48.2%, $P_{heterogeneity}$

= 0.09 for CVD mortality. We observed null associations for plant protein and cancer mortality and other mortality (Figure 1C).

Dose-response meta-analysis

We performed dose-response meta-analyses based on six studies[11-14, 40] (Supplemental Table 8), from which sufficient data could be extracted to estimate dose-response estimates. In these studies, the median animal protein intake ranged from 4.3 E% through 20.0 E%, and plant protein from 2.6 E% through 8.4 E%. We found no evidence for non-linear associations (Wald test: $p>0.05$). In line with the highest versus lowest meta-analysis, we observed a positive linear association between total protein intake and all-cause mortality (per 5 E% increment, 1.02 (1.004, 1.04), I^2= 37.9%, $P_{heterogeneity}$ = 0.17), mainly driven by animal protein intake and CVD mortality (Per 5 E% increment, 1.05 (1.02, 1.09), I^2= 31.2%, $P_{heterogeneity}$ = 0.23)) (Supplemental Table 8, and Figure 2A-B). Furthermore, we observed an inverse linear association between plant protein intake with all-cause mortality (per 5 E% increment, 0.87 (0.78, 0.98), I^2= 40.0%, $P_{heterogeneity}$ = 0.17) (Supplemental Table 8, Figure 2C). We observed no dose-response associations for the other examined associations (Supplemental Table 8).

Subgroup and sensitivity meta-analysis

We observed that several meta-analysis results were modified by geographical study location (Supplemental Table 9). For total protein and all-cause mortality and for animal protein and CVD mortality, positive associations were observed in North American and European populations, whereas null associations were observed in Japanese populations. For plant protein, inverse associations with all-cause and CVD mortality were only observed in North American and Japanese populations, but not in European populations (Supplemental Table 9). For the sensitivity analyses, as shown in Supplemental Table 10, most of the pooled associations were similar after excluding one study at each turn; and thus, were not driven by one individual study. For plant protein and CVD mortality, excluding the Rotterdam Study substantially reduced the heterogeneity. Supplemental Table 11 shows the results of the second set of sensitivity analysis in which we included two additional studies in the dose-response meta-analysis that did not report associations for protein in E% but rather in g/day[15] or in SD.[39] After incorporating results from the study by Bates et al.[39] for total protein intake with all-cause mortality and CVD mortality, the pooled dose-response association between total protein intake and all-cause mortality was null, but with high heterogeneity (I^2= 87.8%, $P_{heterogeneity}$=0.004) (Supplemental Table 11). Estimates for animal and plant protein were not available in this study. After incorporating results from the study by Tharrey et al.[15] for animal and plant protein and CVD mortality, the results remained similar (e.g. for animal protein and CVD mortality: per 5 E% increment, 1.08 (1.01, 1.16)) (Supplemental Table 11). The appearance of funnel plots was symmetrical for all analyses, and Egger's test results were not significant (Supplemental Figure 3), suggesting no publication bias.

Table 5. Characteristics of the prospective studies included in the systematic review and meta-analysis

Authors	Study cohort/populations	Country	Baseline age (years)	Female (%)	Follow-up (years)
Sauvaget et al[8]	The Adult Health Study	Japan	35-89	100	14
Kelemen et al[9]	The Iowa Women's Health Study	US	55-69	100	16.4
Smit et al[10]	The Puerto Rico Heart Health Program	Puerto Rico	45-64	0	12
[1]Bates et al[39]	The community-living population of mainland Britain	UK	76·7	50.2	14
Levine et al[11]	NHANES III	US	64.8	55.4	13.1
Song et al[12]	Nurses' Health Study & Health Professional Follow-up study	US	49	64.7	27.0
[1]Tharrey et al[15]	The Adventist Health Study 2	US and Canada	>25	NA	9.4
Kurihara et al[13]	The National Integrated Project for Prospective Observation of Non-communicable Disease	Japan	52.6	58.4	13.9
Virtanen et al[14]	The Kuopio Ischaemic Heart Disease Risk Factor Study	Finland	52.7-53.7	0	22.31
Budhathoki et al[40]	Japan Public Health Center–based Prospective Cohort Study	Japan	45 to 74	54.5	18
[2]Chen et al et al	The Rotterdam Study	Netherlands	63.5	60.8	13.0

Table 5. Characteristics of the prospective studies included in the systematic review and meta-analysis (Continued)

Authors	Number of deaths			Level of adjustment	NOS[3]
	All deaths	CVD deaths	Cancer deaths		
Sauvaget et al[8]	NA	60	NA	++	7
Kelemen et al[9]	3978	739	1676	+++	8
Smit et al[10]	NA	NA	167	++	8
[1]Bates et al[39]	749	199	Na	+	6
Levine et al[11]	2553	1212	638	+++	8
Song et al[12]	36115	8851	13159	+++	9
[1]Tharrey et al[15]	NA	2276	NA	+++	9
Kurihara et al[13]	1,213	354	NA	+++	9
Virtanen et al[14]	1225	618	347	+++	9
Budhathoki et al[40]	12381	3,025	5055	+++	9
[2]Chen et al et al	3589	877	896	+++	9

Level of adjustment: +, minimally adjusted (typically adjusted for age, sex, and CVD confounders but not for other nutritional factors); ++, adjusted for other macronutrients and/or other nutritional factors; +++, adjusted for animal and plant protein intake.

[1] indicates inclusion only in the systematic review (and in a sensitivity meta-analysis), not in the main meta-analysis because of different format of estimates.

[2]The current study

[3]NOS score, Newcastle–Ottawa Scale score with a theoretical range from zero to nine with higher scores reflecting higher study design quality

Abbreviations: CVD death, cardiovascular diseases death; NOS, Newcastle–Ottawa Scale, score with a theoretical range from zero to nine with higher scores reflecting higher study design quality; NA, not available

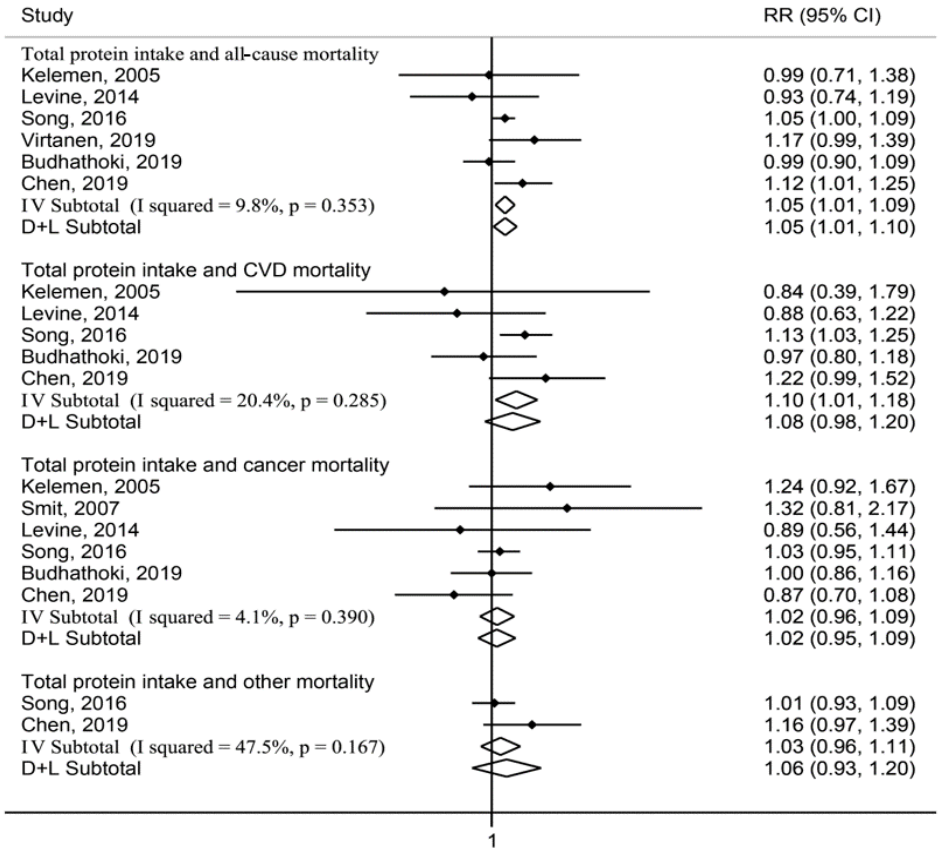


Figure 1 A. Total protein and mortality

Solid dots denote individual HRs, horizontal lines demote individual 95% CIs, open diamonds correspond to the pooled RRs including the 95% CIs, p values denote $P_{heterogeneity}$ values, I-V Subtotal denotes fixed-effects analysis, and D+L Subtotal denotes random-effects analysis. Abbreviations: CVD mortality, cardiovascular mortality, RR, relative risk; CI, confidential interval

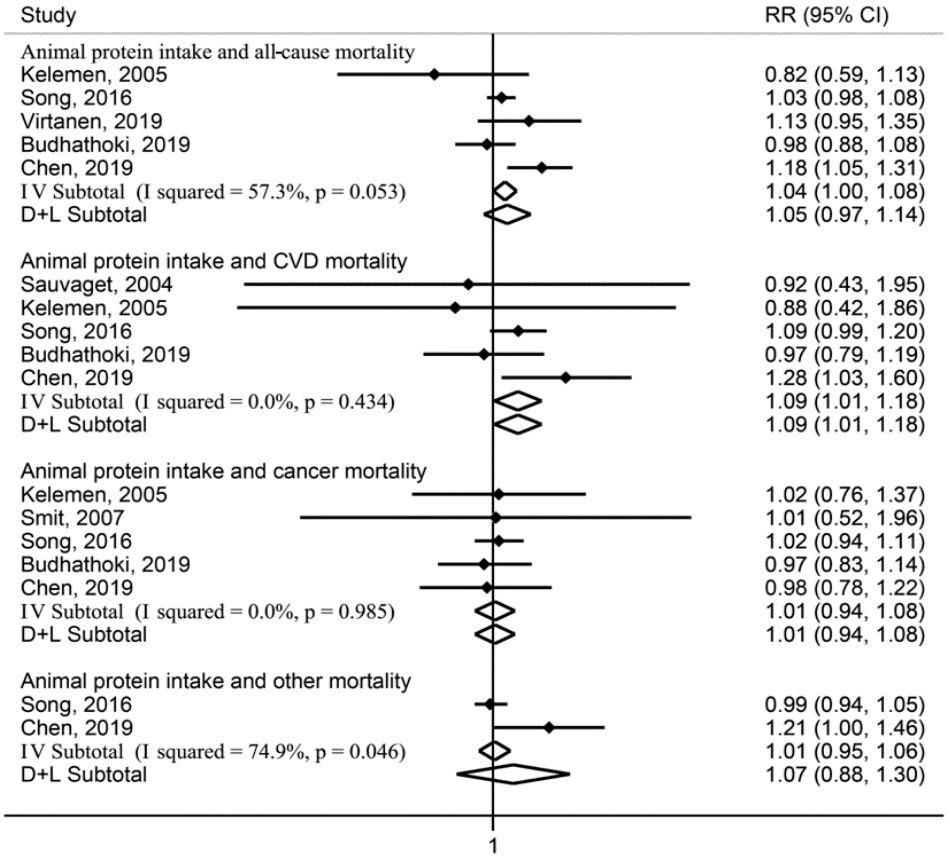


Figure 1 B. Animal protein and mortality

Solid dots denote individual HRs, horizontal lines demote individual 95% CIs, open diamonds correspond to the pooled RRs including the 95% CIs, p values denote $P_{heterogeneity}$ values, I-V Subtotal denotes fixed-effects analysis, and D+L Subtotal denotes random-effects analysis. Abbreviations: CVD mortality, cardiovascular mortality, RR, relative risk; CI, confidential interval

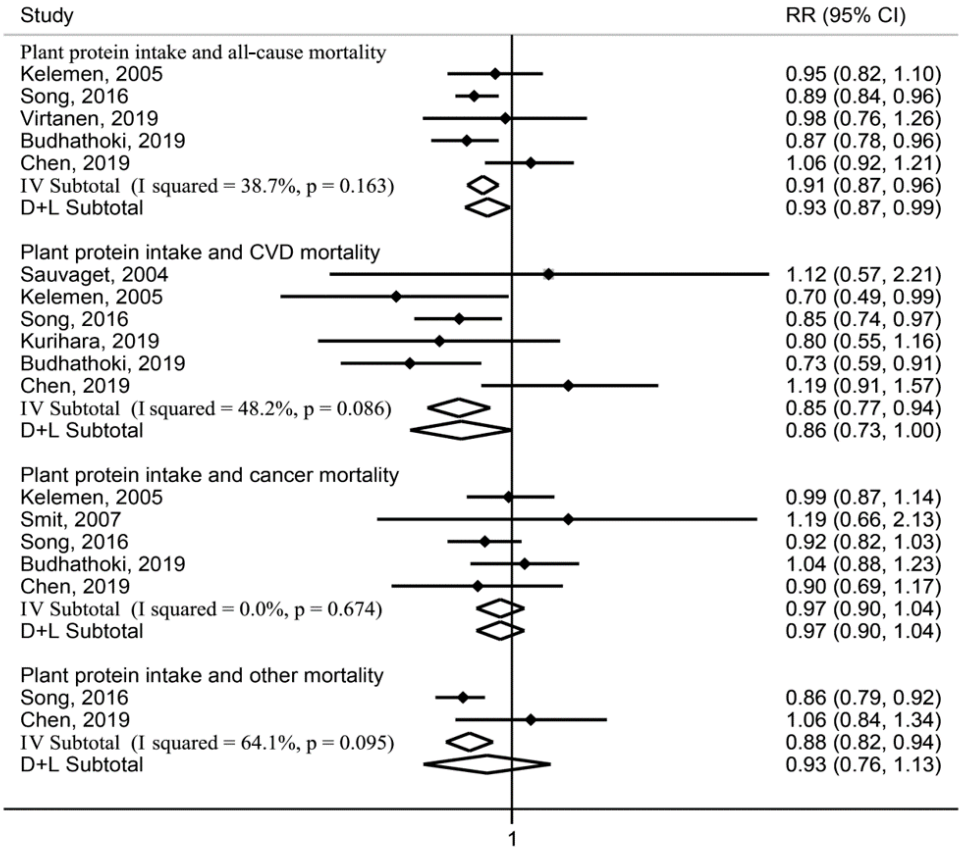


Figure 1 C. Plant protein and mortality

Solid dots denote individual HRs, horizontal lines demote individual 95% CIs, open diamonds correspond to the pooled RRs including the 95% CIs, p values denote $P_{heterogeneity}$ values, I-V Subtotal denotes fixed-effects analysis, and D+L Subtotal denotes random-effects analysis. Abbreviations: CVD mortality, cardiovascular mortality, RR, relative risk; CI, confidential interval

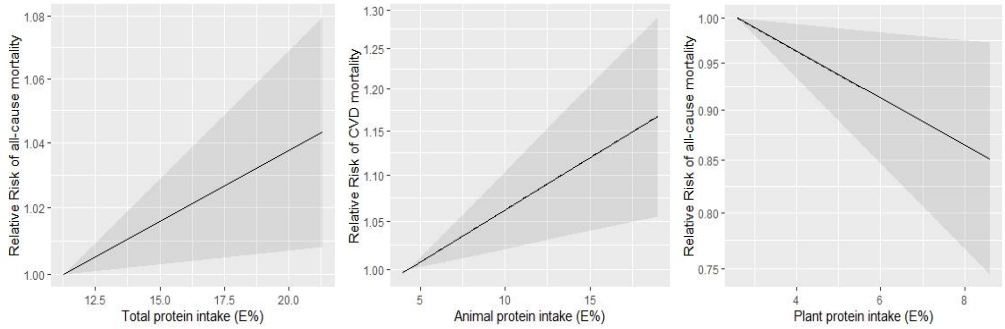


Figure 2. Combined dose–response associations between dietary protein intake with mortality (solid line) with 95% confidence intervals (shaded area).

The median total protein intake ranged from 11.3 through 25.0 E%; the median animal protein intake ranged from 4.3 through 20.0 E%, the median plant protein intake ranged from 2.6 through 8.4 E%.

DISCUSSION

Main findings

In the Rotterdam Study, we observed that higher total protein intake was associated with higher all-cause mortality, which was mainly driven by higher animal protein intake and CVD mortality. Plant protein intake was not associated with all-cause or cause-specific mortality. A meta-analysis of elven prospective cohort studies including the Rotterdam Study corroborated that higher total protein intake may increase risk of all-cause mortality, driven by a harmful association between animal protein and CVD mortality. Furthermore, our overall meta-analysis also indicated that higher plant protein may decrease all-cause mortality and CVD mortality. These overall meta-analysis results were modified by geographical study location. As we further observed that the harmful associations of total and animal protein were mainly among the North American and European populations, and the inverse associations of plant protein were mainly among the North American and Japanese populations.

Interpretations of our findings

In contrast to reported beneficial short-term effects of dietary protein intake on weight management, and cardiovascular risk factors,[2-5] we observed that a higher total protein intake was associated with higher all-cause mortality, which was mainly driven by a positive association between animal protein intake and CVD mortality.

The findings for higher animal protein intake and higher CVD mortality are supported by several biological mechanisms and pathways. First, animal protein is relatively high in dietary branched-chain and aromatic amino acids, which may result in insulin resistance[41, 42] and overweight,[42, 43] via mammalian target of rapamycin pathway.[44] These are strong risk factors for various cardiometabolic diseases, in turn increasing CVD mortality risk.[41] Second, higher protein intake, particularly from animal sources, may be harmful for kidney function, especially among individuals with impaired kidney function,[45] which presents another risk factor for CVD incidence and mortality.[46] Last, the association could be fueled or amplified by other components in animal-based foods, such as SFA and sodium from red and processed meat, which have both been linked to higher CVD risk.[12, 47] To investigate if the association of animal protein intake and CVD mortality would differ by more specific CVD causes, we further examined non-stroke CVD mortality and stroke mortality in the Rotterdam Study. We observed that the association of animal protein and CVD mortality was mainly driven by non-stroke CVD mortality, which is in line with previous studies which indicated a lack of association between animal protein intake with stroke.[40, 48] Moreover, we observed in subgroup analyses that these harmful associations were mainly observed in North American and European populations, not in Japanese populations. That could be partly explained by different levels and food sources of animal protein intake. In the North American[12] and European study populations,[14] the major animal protein sources were red and processed meat, whereas in the Japanese study populations, population levels of animal protein intake were lower, and the main animal protein source was fish.[40]

For plant protein, we observed inverse associations between plant protein intake and all-cause and CVD mortality in the overall meta-analysis. The difference of associations for animal protein and plant protein might be explained by their different amino acid composition. Unlike animal protein, plant protein is generally low in branched-chain acids and aromatic amino acids,[6, 49] thereby, resulting in decreased risks of CVD.[42] Furthermore, in subgroup analyses, we observed the inverse associations existed in North American and Japanese populations, but not in European populations. This may also be explained by different dietary plant protein sources among different populations. In the European populations, the main source was grains.[41] Among the North American populations in the study by Song et al, the main plant protein sources were legumes, whole grains and nuts,[12] and in the Japanese populations in the study by Budhathoki et al,[40] the main source was legumes.

Overall, the evidences provided herein indicates the importance of specific protein sources for overall health, especially CVD health, and support a replacement of animal protein intake with plant protein intake. For example, in our meta-analysis, we observed that those in the highest quantile of animal protein intake, may have an averagely 9% higher CVD mortality risk than those in the lowest quantile. Based on reports of the individual studies, we estimated that those in the highest quantile had a median animal protein intake of approximately 75 grams/day, and those in the lowest quantile around 38 grams/day. This suggests that a decrease in animal protein intake from 75 grams/day (e.g.

corresponding to around 220 grams red meat/day) to 42 grams/day (e.g., around 100 grams red meat/day), may attenuate risk of CVD mortality by around 9%, assuming other covariates remain stable. However, given that the populations in our meta-analysis were mainly general populations, and therefore, our results and public health implications cannot be generalized to patient groups who may have other protein requirements. For example, for severely ill patients or elderly, high dietary protein intake may be beneficial in recovery or to prevent sarcopenia.

Strengths and limitations

Our study has several strengths. First, the Rotterdam Study analysis was based on a prospective design and included comprehensive assessments of cause-specific deaths. Second, our meta-analysis is, to our knowledge, the first to summarize the associations of specific dietary protein intake with all-cause and cause-specific mortality, for which, we not only conducted highest versus lowest meta-analyses, but also dose-response meta-analyses. This can help to quantify the associations and test the shape of these possible associations. Third, the meta-analysis was based on several prospective cohort studies across various populations from different geographical locations. Moreover, the combined sample size was large, and the follow-up period was long, resulting in a substantial number of cases. Additionally, the cohort studies cohort studies in the meta-analysis were of medium to high quality, and their analyses included macronutrient substitution models as well as adjustments for other important confounding factors, such as total energy, physical activity, and BMI.

We also need to acknowledge several limitations. First, the Rotterdam Study and most studies in the meta-analysis measured dietary intake data based on self-reported FFQs, 24-hour dietary recalls, or food records, for which measurements errors are unavoidable. However, as these methods were expected to adequately rank subjects according to food and nutrient intake, we do not expect these measurement-errors to have largely affected associations. Second, in all studies except one, dietary intake data were measured only once at baseline, and changes in diet over time may affect associations. However, our results were generally consistent with results from the only study with repeated dietary measurements.[12] Third, in the Rotterdam Study analysis, a weak trend of an association between animal protein intake and other mortality might exist, but we could not further explore this due to limited numbers of cases for death from specific other causes. Fourth, we observed that the geographic study location modified the meta-analysis results. However, we could not further conduct subgroup analyses or meta-regression to explore other potential sources of the heterogeneity (e.g., age and sex). For example, we could not explore possible sex difference. Only two studies including the Rotterdam Study reported sex-stratified associations. The Rotterdam Study analysis observed that the association between animal protein intake and all-cause mortality, but not CVD mortality, differed by sex in the Rotterdam Study, with positive associations only in men. Only one other study examined sex differences for this association and observed null associations in both genders. Fifth, since all the

studies included in the meta-analysis were conducted in general populations, our results may not be generalizable to populations with other protein requirements. Last, as a meta-analysis of observational studies, the results could be subject to residual or unmeasured confounding. Thus, the associations we report should be interpreted with caution.

In conclusion, our study provides evidence that higher total protein intake is associated with higher all-cause mortality, primarily driven by a positive association between animal protein intake and CVD mortality. In contrast, higher plant protein intake is associated with lower all-cause and CVD mortality. Food source and level of protein intake may play a substantial role as we observed harmful associations of total and animal protein mainly in North American and European populations and beneficial associations of plant protein mainly in North American and Japanese populations. Further studies in other populations with different amounts and food sources of protein intakes or with different protein requirements are needed to improve global dietary recommendations and to define optimal ranges and sources of protein intake for different populations.

REFERENCES

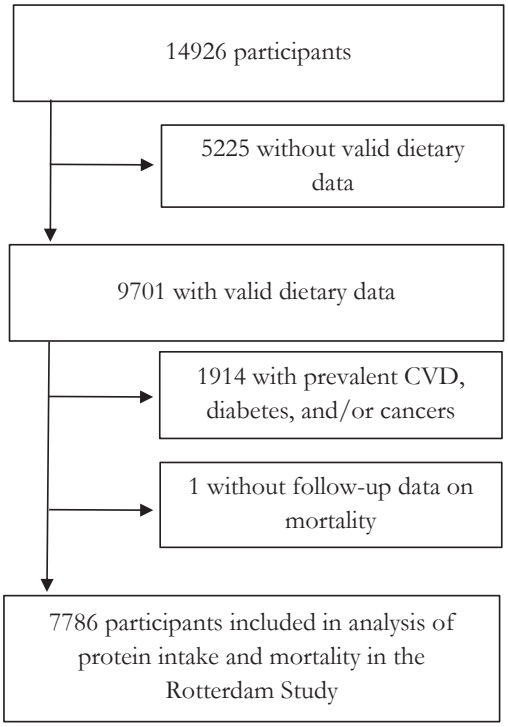
1. Bauer J, Biolo G, Cederholm T, Cesari M, Cruz-Jentoft AJ, Morley JE, et al. Evidence-Based Recommendations for Optimal Dietary Protein Intake in Older People: A Position Paper From the PROT-AGE Study Group. Journal of the American Medical Directors Association. 2013;14(8):542-59.
2. Appel LJ, Sacks FM, Carey VJ, Obarzanek E, Swain JF, Miller ER, et al. Effects of protein, monounsaturated fat, and carbohydrate intake on blood pressure and serum lipids: results of the OmniHeart randomized trial. Jama. 2005;294(19):2455-64.
3. Wycherley TP, Moran LJ, Clifton PM, Noakes M, Brinkworth GD. Effects of energy-restricted high-protein, low-fat compared with standard-protein, low-fat diets: a meta-analysis of randomized controlled trials. The American journal of clinical nutrition. 2012:ajcn. 044321.
4. Santesso N, Akl EA, Bianchi M, Mente A, Mustafa R, Heels-Ansdell D, et al. Effects of higher-versus lower-protein diets on health outcomes: a systematic review and meta-analysis. European journal of clinical nutrition. 2012;66(7):780-8.
5. Westerterp-Plantenga MS, Lemmens SG, Westerterp KR. Dietary protein - its role in satiety, energetics, weight loss and health. Br J Nutr. 2012;108 Suppl 2:S105-12.
6. Shang X, Scott D, Hodge AM, English DR, Giles GG, Ebeling PR, et al. Dietary protein intake and risk of type 2 diabetes: results from the Melbourne Collaborative Cohort Study and a meta-analysis of prospective studies. The American journal of clinical nutrition. 2016;104(5):1352-65.
7. Lagiou P, Sandin S, Lof M, Trichopoulos D, Adami H-O, Weiderpass E. Low carbohydrate-high protein diet and incidence of cardiovascular diseases in Swedish women: prospective cohort study. Bmj. 2012;344:e4026.
8. Sauvaget C, Nagano J, Hayashi M, Yamada M. Animal protein, animal fat, and cholesterol intakes and risk of cerebral infarction mortality in the adult health study. Stroke. 2004;35(7):1531-7.
9. Kelemen LE, Kushi LH, Jacobs Jr DR, Cerhan JR. Associations of dietary protein with disease and mortality in a prospective study of postmenopausal women. American journal of epidemiology. 2005;161(3):239-49.
10. Smit E, Garcia-Palmieri MR, Figueroa NR, McGee DL, Messina M, Freudenheim JL, et al. Protein and legume intake and prostate cancer mortality in Puerto Rican men. Hnuc. 2007;58(2):146-52.
11. Levine ME, Suarez JA, Brandhorst S, Balasubramanian P, Cheng C-W, Madia F, et al. Low protein intake is associated with a major reduction in IGF-1, cancer, and overall mortality in the 65 and younger but not older population. Cell metabolism. 2014;19(3):407-17.
12. Song M, Fung TT, Hu FB, Willett WC, Longo VD, Chan AT, et al. Association of animal and plant protein intake with all-cause and cause-specific mortality. JAMA internal medicine. 2016;176(10):1453-63.
13. Kurihara A, Okamura T, Sugiyama D, Higashiyama A, Watanabe M, Okuda N, et al. Vegetable Protein Intake was Inversely Associated with Cardiovascular Mortality in A 15-Year Follow-Up Study of A General Japanese Population. Journal of atherosclerosis and thrombosis. 2018:44172.

14. Virtanen HEK, Voutilainen S, Koskinen TT, Mursu J, Kokko P, Ylilauri M, et al. Dietary proteins and protein sources and risk of death: the Kuopio Ischaemic Heart Disease Risk Factor Study. The American journal of clinical nutrition. 2019;109(5):1462-71.
15. Tharrey M, Mariotti F, Mashchak A, Barbillon P, Delattre M, Fraser GE. Patterns of plant and animal protein intake are strongly associated with cardiovascular mortality: the Adventist Health Study-2 cohort. International journal of epidemiology. 2018;47(5):1603-12.
16. Halbesma N, Bakker SJL, Jansen DF, Stolk RP, De Zeeuw D, De Jong PE, et al. High protein intake associates with cardiovascular events but not with loss of renal function. Journal of the American Society of Nephrology. 2009;20(8):1797-804.
17. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol. 2017;32(9):807-50.
18. Chen Z, Schoufour JD, Rivadeneira F, Lamballais S, Ikram MA, Franco OH, et al. Plant-based diet and adiposity over time in a middle-aged and elderly population: the Rotterdam Study. Epidemiology. 2018.
19. Chen Z, Zuurmond MG, van der Schaft N, Nano J, Wijnhoven HAH, Ikram MA, et al. Plant versus animal based diets and insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. European journal of epidemiology. 2018;33(9):883-93.
20. Voortman T, Kiefte-de Jong JC, Ikram MA, Stricker BH, van Rooij FJA, Lahousse L, et al. Adherence to the 2015 Dutch dietary guidelines and risk of non-communicable diseases and mortality in the Rotterdam Study. European journal of epidemiology. 2017;32(11):993-1005.
21. Klipstein-Grobusch Kd, Den Breeijen JH, Goldbohm RA, Geleijnse JM, Hofman A, Grobbee DE, et al. Dietary assessment in the elderly: validation of a semiquantitative food frequency questionnaire. European journal of clinical nutrition. 1998;52(8):588.
22. van den Brandt PA, Goldbohm RA, Veer PVT, Volovics A, Hermus RJJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. Journal of clinical epidemiology. 1990;43(3):285-95.
23. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. The American journal of clinical nutrition. 1993;58(4):489-96.
24. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81(3):515-26.
25. Willett W. Implications of total energy intake for epidemiologic analyses. Nutritional epidemiology. 1998:273-301.
26. Resche-Rigon M, White IR. Multiple imputation by chained equations for systematically and sporadically missing multilevel data. Statistical methods in medical research. 2018;27(6):1634-49.
27. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine. 2009;151(4):264-9.
28. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Jama. 2000;283(15):2008-12.

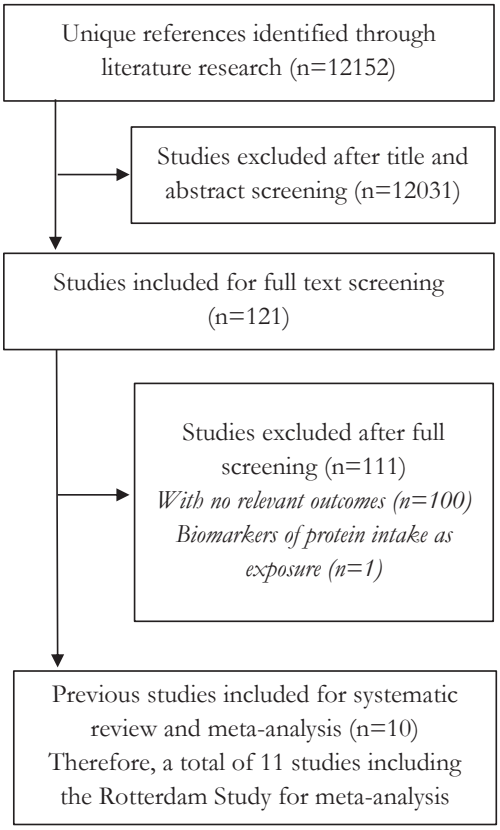
29. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. European journal of epidemiology. 2010;25(9):603-5.
30. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. Bmj. 2011;342:d549.
31. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj. 2003;327(7414):557-60.
32. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. American journal of epidemiology. 1992;135(11):1301-9.
33. Keum N, Greenwood DC, Lee DH, Kim R, Aune D, Ju W, et al. Adult weight gain and adiposity-related cancers: a dose-response meta-analysis of prospective observational studies. Journal of the National Cancer Institute. 2015;107(2):djv088.
34. Crippa A, Orsini N. Multivariate dose-response meta-analysis: the dosresmeta R package. J Stat Softw. 2016;72(1):1-15.
35. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics. 1994:1088-101.
36. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Bmj. 1997;315(7109):629-34.
37. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. Jama. 2006;295(6):676-80.
38. Organization WH. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WH. 2006.
39. Bates CJ, Mansoor MA, Pentieva KD, Hamer M, Mishra GD. Biochemical risk indices, including plasma homocysteine, that prospectively predict mortality in older British people: the National Diet and Nutrition Survey of People Aged 65 Years and Over. British journal of nutrition. 2010;104(6):893-9.
40. Budhathoki S, Sawada N, Iwasaki M, Yamaji T, Goto A, Kotemori A, et al. Association of Animal and Plant Protein Intake With All-Cause and Cause-Specific Mortality. JAMA internal medicine. 2019.
41. Chen Z, Franco OH, Lamballais S, Ikram MA, Schoufour JD, Muka T, et al. Associations of specific dietary protein with longitudinal insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. Clinical Nutrition. 2019.
42. Lynch CJ, Adams SH. Branched-chain amino acids in metabolic signalling and insulin resistance. Nature Reviews Endocrinology. 2014;10(12):723.
43. Lin Y, Bolca S, Vandevijvere S, De Vriese S, Mouratidou T, De Neve M, et al. Plant and animal protein intake and its association with overweight and obesity among the Belgian population. British journal of nutrition. 2011;105(7):1106-16.
44. Solon-Biet SM, Cogger VC, Pulpitel T, Wahl D, Clark X, Bagley EE, et al. Branched-chain amino acids impact health and lifespan indirectly via amino acid balance and appetite control. Nature Metabolism. 2019.
45. Kalantar-Zadeh K, Fouque D. Nutritional management of chronic kidney disease. New England Journal of Medicine. 2017;377(18):1765-76.

46. Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, Jafar TH, Heerspink HJL, Mann JF, et al. Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. The Lancet. 2013;382(9889):339-52.
47. Wang DD, Li Y, Chiuve SE, Stampfer MJ, Manson JE, Rimm EB, et al. Association of specific dietary fats with total and cause-specific mortality. JAMA internal medicine. 2016;176(8):1134-45.
48. Zhang Z, Xu G, Yang F, Zhu W, Liu X. Quantitative analysis of dietary protein intake and stroke risk. Neurology. 2014:10.1212/WNL. 0000000000000551.
49. Wu G. Amino acids: metabolism, functions, and nutrition. Amino acids. 2009;37(1):1-17.

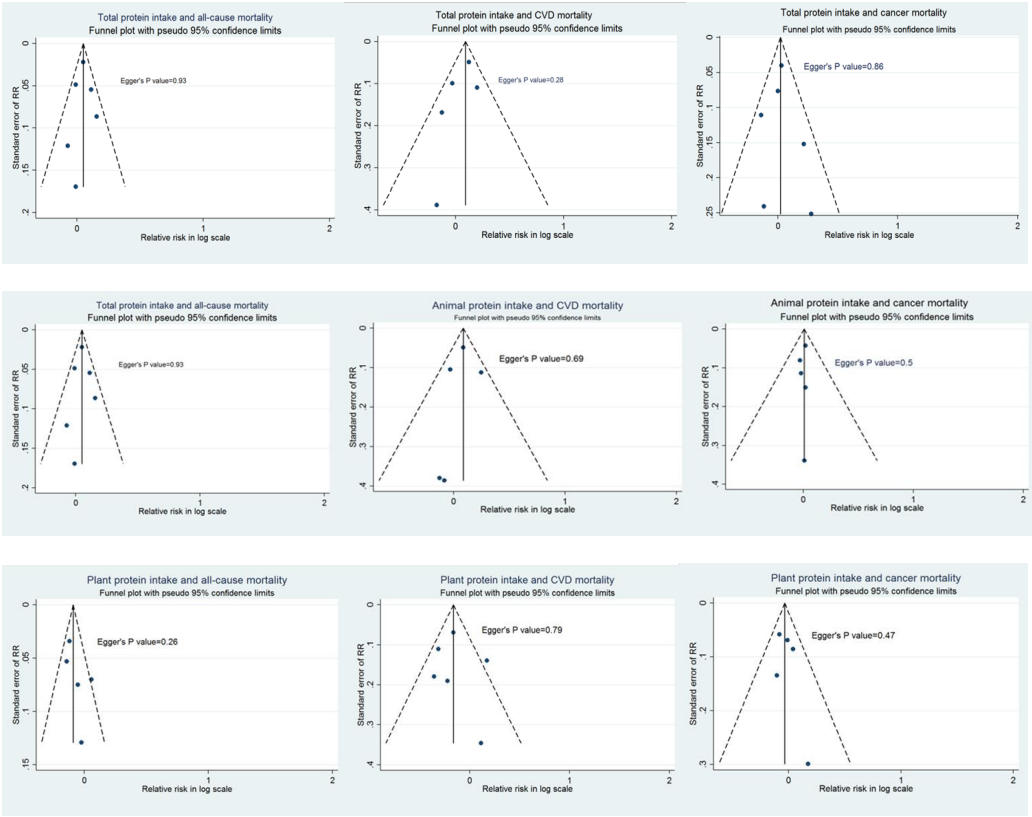
SUPPLEMENTAL MATERIAL



Supplemental Figure 1. Participants selection



Supplemental Figure 2. Selection of studies for meta-analysis



Supplemental Figure 3. Funnel plots and Egger's test results

Supplemental Table 1. Missing values in the Rotterdam study (n=7786)

	Physical activity	BMI	Education level	Smoking status
RS-I n=4309	NA=1515 (35.1%)	NA=29 (0.67%)	NA=23 (0.53%)	NA=28 (0.64%)
RS-II n=1249	NA=6 (0.48%)	NA=4 (0.32%)	NA=14 (1.1%)	NA=5 (0.40%)
RS-III n=2228	NA=195 (8.8%)	NA=62 (2.8%)	NA=7 (0.31%)	NA=6 (0.27%)
Total n=7786	NA=1716 (22.0%)	NA=95 (1.2%)	NA=44 (0.56%)	NA=39 (0.50%)

In our main analyses, only four variables: physical activity, BMI, education level, and smoking status were with missing values. Abbreviations: NA, not available, (numbers of participants with missing values); BMI, body mass index, RS, Rotterdam Study.

Supplemental Table 2. Detailed search terms and strategies

Database	Search term
Embase.com	('protein diet'/exp OR 'protein intake'/de OR 'plant protein'/de OR 'red meat'/exp OR 'dairy product'/exp OR 'nut'/exp OR 'soybean protein'/exp OR 'soybean milk'/exp OR 'protein restriction'/exp OR meat/exp OR (protein/de AND 'diet supplementation'/de) OR (((protein* OR nut OR nuts OR meat) NEAR/3 (intake OR diet* OR consum* OR nutrion OR food OR eating OR restrict* OR suppl* OR added OR rich OR enrich* OR meal*)) OR red-meat OR (milk NOT (breast-milk OR human-milk)) OR dairy OR cheese OR ((plant* OR animal OR soy) NEXT/1 protein*) OR yogurt OR yoghurt):ab,ti) AND ('cardiovascular disease'/de OR 'heart failure'/de OR 'congestive heart failure'/de OR 'heart disease'/de OR 'coronary artery disease'/de OR 'ischemic heart disease'/exp OR 'cerebrovascular accident'/de OR 'atherosclerotic cardiovascular disease'/de OR 'brain ischemia'/exp OR 'mortality'/exp OR 'diabetes mellitus'/de OR 'non insulin dependent diabetes mellitus'/de OR 'cardiovascular risk'/de OR (((cardiovascular OR coronar*) NEAR/3 (disease* OR event*)) OR cvd OR cvds OR ((ischemi* OR ischaemi* OR fail* OR insufficien*) NEAR/3 (heart OR cardia*)) OR (cerebrovascular* NEAR/3 accident*) OR cva OR stroke* OR ((brain OR cerebral) NEAR/3 (ischemi* OR ischaemi*)) OR mortalit* OR (diabet* NOT ((type-1 OR type-I OR DM-1 OR DM-I OR t1d OR gestation* OR iddm) NOT (type-2 OR type-ii OR type-2a OR type-iia OR type-2b OR type-iib OR DM-2 OR DM-ii OR t2d))) OR niddm OR t2d OR t2dm OR ((chd OR cvd OR cardiovascul*) NEAR/3 risk*)):ab,ti) NOT ([animals]/lim NOT [humans]/lim) NOT ([Conference Abstract]/lim OR [Letter]/lim OR [Note]/lim OR [Editorial]/lim) AND ('cohort analysis'/exp OR 'prospective study'/exp OR 'longitudinal study'/exp OR 'retrospective study'/exp OR 'follow up'/de OR 'case control study'/exp OR 'cross-sectional study'/exp OR 'clinical study'/exp OR 'meta analysis'/de OR 'clinical trial'/exp OR 'major clinical study'/de OR ((cross NEXT/1 section*) OR (case NEXT/1 control*) OR cohort* OR trial* OR ((clinical OR prospectiv* OR population* OR observation* OR retrospecti*):ab,ti)

Supplemental Table 2. Detailed search terms and strategies (Continued)

Database	Search term
Medline ovid	(Diet, Protein-Restricted/ OR exp Diet, High-Protein/ OR exp Dietary Proteins/ OR exp Protein Deficiency/ OR exp Plant Proteins/ OR exp Dairy Products/ OR exp nuts/ OR exp Soy Foods/ OR exp meat/ OR (proteins/ AND Dietary Supplements/) OR (((protein* OR nut OR nuts OR meat) ADJ3 (intake OR diet* OR consum* OR nutrion OR food OR eating OR restrict* OR suppl* OR added OR rich OR enrich* OR meal*)) OR red-meat OR (milk NOT (breast-milk OR human-milk)) OR dairy OR cheese OR ((plant* OR animal OR soy) ADJ protein*) OR yogurt OR yoghurt).ab,ti.) AND (Cardiovascular Diseases/ OR heart failure/ OR Heart Diseases/ OR Coronary Artery Disease/ OR exp Myocardial Ischemia/ OR stroke/ OR exp brain ischemia/ OR exp mortality/ OR exp Survival/ OR diabetes mellitus/ OR Diabetes Mellitus, Type 2/ OR (((cardiovascular OR coronar*) ADJ3 (disease* OR event*)) OR cvd OR cvds OR ((ischemi* OR ischaemi* OR fail* OR insufficien*) ADJ3 (heart OR cardia*)) OR (cerebrovascular* ADJ3 accident*) OR cva OR stroke* OR ((brain OR cerebral) ADJ3 (ischemi* OR ischaemi*)) OR mortalit* OR (diabet* NOT ((type-1 OR type-I OR DM-1 OR DM-I OR t1d OR gestation* OR iddm) NOT (type-2 OR type-ii OR type-2a OR type-iia OR type-2b OR type-iib OR DM-2 OR DM-ii OR t2d))) OR niddm OR t2d OR t2dm OR ((chd OR cvd OR cardiovascul*) ADJ3 risk*)).ab,ti.) NOT (exp animals/ NOT humans/) NOT (letter OR news OR comment OR editorial OR congresses OR abstracts).pt. AND (exp Cohort Studies/ OR Case-Control Studies/ OR cross-sectional study/ OR Meta-Analysis / OR exp clinical trial/ OR ((cross ADJ section*) OR (case ADJ control*) OR cohort* OR trial* OR ((clinical OR prospectiv* OR population* OR observation* OR retrospecti* OR intervention*) ADJ3 stud*) OR follow up OR (meta ADJ analy*) OR metaanaly* OR trial OR random*).ab,ti.)
Cochrane CENTRAL	((((protein* OR nut OR nuts OR meat) NEAR/3 (intake OR diet* OR consum* OR nutrion OR food OR eating OR restrict* OR suppl* OR added OR rich OR enrich* OR meal*)) OR red-meat OR (milk NOT (breast-milk OR human-milk)) OR dairy OR cheese OR ((plant* OR animal OR soy) NEXT/1 protein*) OR yogurt OR yoghurt):ab,ti) AND ((((cardiovascular OR coronar*) NEAR/3 (disease* OR event*)) OR cvd OR cvds OR ((ischemi* OR ischaemi* OR fail* OR insufficien*) NEAR/3 (heart OR cardia*)) OR (cerebrovascular* NEAR/3 accident*) OR cva OR stroke* OR ((brain OR cerebral) NEAR/3 (ischemi* OR ischaemi*)) OR mortalit* OR (diabet* NOT ((type-1 OR type-I OR DM-1 OR DM-I OR t1d OR gestation* OR iddm) NOT (type-2 OR type-ii OR type-2a OR type-iia OR type-2b OR type-iib OR DM-2 OR DM-ii OR t2d))) OR niddm OR t2d OR t2dm OR ((chd OR cvd OR cardiovascul*) NEAR/3 risk*)):ab,ti)

Supplemental Table 3. Characteristics of population of the Rotterdam Study across the quartiles of total, animal, and plant protein (n=7786)

	Total protein				Animal protein				Plant protein			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Protein intake (E%)	**≤14.4**	**14.4>, ≤16.2**	**16.2>, ≤18.1**	**>18.1**	**≤8.4**	**8.4>, ≤10.2**	**10.2>, ≤12.2**	**>12.2**	**≤5.2**	**5.2>, ≤5.9**	**5.9>, ≤6.7**	**>6.7**
Age (Year)	63.9 (9.4)	63.1 (8.8)	63.5 (8.3)	64.2 (8.2)	62.1 (9.1)	63.7 (8.8)	64.1 (8.5)	64.9 (8.2)*	66.2 (8.8)	65.1 (8.7)	62.9 (8.3)	60.6 (7.9)*
Sex (%)												
-Female	12.8	14.4	15.7	18.4*	12.9	13.9	16.0	17.9*	13.9	13.7	15.7	15.6*
-Male	12.7	10.6	9.3	6.6*	12.1	11.0	9.0	7.1*	11.1	9.4	9.3	9.4
BMI (kg/m^2)	25.8 (3.6)	26.3 (3.7)	26.6 (3.8)	27.5* (4.3)	25.8 (3.7)	26.3 (3.7)	26.7 (3.9)	27.4* (4.2)*	26.4 (3.7)	26.5 (3.8)	26.8 (4.0)	26.6 (4.1)
Smoking Status (%)												
-Never	7.2	8.3	8.5	9.3	7.8	8.1	8.5	8.9	6.8	8.6	8.9	9.1
-Ever	11.0	10.7	10.5	10.1	11.4	10.9	10.3	9.8	9.6	10.8	10.7	11.2
-Current	6.7	5.9	5.8	5.4*	5.7	5.9	6.0	6.2*	8.5	5.4	5.3	4.7*
Education level (%)												
-Primary	3.6	3.6	4.0	4.1	3.1	3.7	4.0	4.5	4.6	4.3	3.2	3.1
-Low	10.0	9.9	10.3	10.9	9.5	10.3	10.3	10.9	10.8	10.3	10.4	9.6
-Intermediate	7.2	6.9	6.6	6.6	7.1	6.8	7.0	6.4	6.6	6.9	7.2	6.5
-High	4.1	4.5	4.0	3.2	5.2	4.0	3.6	3.1*	2.9	3.3	4.0	5.7*
Physical activity (MET-hours/week)												
-RS-I and II	74.3 (47.8, 105.6)	79.0 (56.2, 112.1)	81.5 (55.9, 113.4)	83.9 (59.8, 117.5)*	76.7 (49.5, 108.7)	78.8 (56.2, 112.0)	80.7 (55.7, 113.6)	83.0 (58.9, 116.0)*	75.9 (47.6, 105.5)	80.1 (57.4, 113.8)	81.8 (56.4, 112.7)	84.8 (61.6, 121.0)*
-RS-III	41.0 (17.1, 84.3)	49.2 (21.0, 81.5)	42.2 (15.4, 82.9)	39.4 (15.0, 56.4)	42.8 (18.0, 84.1)	51.5 (21.0, 82.7)	38.5 (13.6, 83.3)	38.0 (15.0, 72.4)	38.0 (16.0, 72.4)	36.5 (15.3, 73.0)	45.1 (18.1, 86.6)	48.0 (18.7, 87.8)*

Supplemental Table 3. Characteristics of population of the Rotterdam Study across the quartiles of total, animal, and plant protein (n=7786) (Continued)

	Total protein				Animal protein				Plant protein			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Protein intake (E%)	**≤14.4**	**14.4>, ≤16.2**	**16.2, ≤18.1**	**>18.1**	**≤8.4**	**8.4>, ≤10.2**	**10.2>, ≤12.2**	**>12.2**	**≤5.2**	**5.2>, ≤5.9**	**5.9>, ≤6.7**	**>6.7**
Dietary intake												
Meat protein[1]	3.1	3.8	4.6	5.6	2.8	3.9	4.6	5.8	4.8	4.5	4.3	3.7
(E%)	(1.4)	(1.6)	(1.8)	(2.6)*	(1.3)	(1.5)	(1.8)	(2.6)*	(2.4)	(2.0)	(2.0)	(2.0)*
Dairy protein[2]	3.0	3.8	4.5	6.0	2.8	3.8	4.6	6.1	4.7	4.5	4.3	3.8
(E%)	(1.4)	(1.6)	(2.0)	(2.8)*	(1.3)	(1.5)	(1.9)	(2.8)*	(2.5)	(2.2)	(2.2	(2.1)*
Fish protein	0.39	0.58	0.61	0.77	0.44	0.53	0.59	0.74	0.50	0.53	0.57	0.63
(E%)	(0.0,	(0.13,	(0.14,	(0.18,	(0.05,	(0.11,	(0.14,	(0.16,	(0.04,	(0.10,	(0.14,	(0.17,
	0.81)	1.06)	1.26)	1.54)*	0.87)	1.05)	1.25)	1.49)*	1.08)	1.14)	1.17)	1.20)*
Eggs protein	0.27	0.31	0.36	0.39	0.26	0.32	0.36	0.40	0.35	0.34	0.33	0.28
(E%)	(0.16,	(0.18,	(0.20,	(0.21,	(0.15,	(0.18,	(0.20,	(0.22,	(0.20,	(0.19,	(0.19,	(0.16,
	0.42)	0.46)	0.51)	0.56)*	0.40)	0.46)	0.51)	0.56)*	0.51)	0.49)	0.49)	0.46)*
Total fat (E%)	35.3	35.8	35.7	34.5	34.3	35.4	36.1	35.6	38.1	36.0	34.7	32.3
	(7.3)	(6.4)	(6.1)	(6.2)	(7.3)	(6.1)	(6.2)	(6.5)*	(7.1)	(6.0)	(5.7)	(6.0)*
SFA (E%)	13.3	13.6	13.8	13.5	12.4	13.5	14.1	14.2	15.5	14.1	13.2	11.5
	(3.4)	(3.3)	(3.3)	(3.3)	(3.3)	(3.0)	(3.2)	(3.4)*	(3.5)	(3.0)	(2.7)	(2.6)*
MUFA (E%)	11.9.	11.9.	11.9	11.5	11.5	11.9	11.9	11.9	12.8	11.9	11.5	11.0
	(3.2)	(2.7)	(2.5)	(2.5)	(3.2)	(2.7)	(2.7)	(2.7)	(3.2)	(2.7)	(2.5)	(2.7)*
PUFA (E%)	7.2	7.2	7.0	6.5	7.4	7.2	7.0	6.5	6.8	7.0	7.0	7.2
	(2.7)	(2.5)	(2.3)	(2.5)	(2.7)	(2.5)	(2.5)	(2.5)*	(2.7)	(2.5)	(2.3)	(2.3)*
TSF (E%)	0.72	0.72	0.74	0.72	0.61	0.72	0.79	0.77	0.92	0.81	0.70	0.56
	(0.52,	(0.52,	(0.54,	(0.54,	(0.45,	(0.52,	(0.56,	(0.59,	(0.65,	(0.59,	(0.52,	(0.43,
	1.10)	1.10)	1.06)	0.97)	0.92)	1.10)	1.10)	1.06)	1.33)	1.13)	0.97)	0.74)*

Supplemental Table 3. Characteristics of population of the Rotterdam Study across the quartiles of total, animal, and plant protein (n=7786) (Continued)

	Total protein				Animal protein				Plant protein			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Protein intake (E%)	**≤14.4**	**14.4>, ≤16.2**	**16.2, ≤18.1**	**>18.1**	**≤8.4**	**8.4>, ≤10.2**	**10.2>, ≤12.2**	**>12.2**	**≤5.2**	**5.2>, ≤5.9**	**5.9>, ≤6.7**	**>6.7**
Diet quality score	6.3 (2.0)	6.7 (1.9)	6.9 (1.8)	7.2 (1.8)*	6.8 (2.0)	6.6 (1.9)	6.7 (1.8)	6.9 (1.8)	5.7 (1.7)	6.5 (1.7)	7.0 (1.8)	7.7 (1.7)*
Fiber (gram)	20.8 (15.4, 28.9)	20.6 (15.7, 28.6)	19.6 (15.2, 26.4)	17.7 (14.3, 22.6)*	24.6 (18.1, 33.9)	20.4 (15.8, 27.5)	18.5 (14.8, 239)	16.6 (13.6, 20.9)*	15.2 (12.2, 19.5)	18.1 (14.7, 23.3)	20.8 (16.6, 27.4)	26.2 (19.6, 35.2)*

Variables expressed as mean (SD), median (25th percentile–75th percentile), or percentage.

P-trend was assessed were tested with linear regression (continuous variables) or with chi-square test (categorical variables).

* indicates $P < 0.05$ for trend across quartiles.

[1] Red and processed meat comprised 80% of meat products.

[2] Milk and cheese comprised 75% of dairy products.

Abbreviations: MET, metabolic equivalent; E%, energy percent; SFA, saturated fat acids, MUFA, monounsaturated fat acids, PUFA, polyunsaturated fat acids, TSF, trans saturated fat acids

Supplemental Table 4. Associations of protein intake with all-cause and cause-specific mortality (n=7786, comparison isocaloric substitution for fat)

	Per 5 E% increment n= 7786	Quartile 1 n= 1906	Quartile 2 n=1906	Quartile 3 n=1905	Quartile 4 n=1906	P trend
			Total protein			
Median intake (E%)	16.2	13.3	15.3	17.0	19.7	
All-cause mortality						
Number of deaths	n=3,589	n=1,176	n=986	n=813	n=614	
Multivariate model	1.08 (1.01, 1.15)	1 (Reference)	1.04 (0.94, 1.15)	1.01 (0.91, 1.12)	1.10 (0.98, 1.22)	0.04
CVD mortality						
Number of deaths	n=877	n=169	n=216	n=219	n=273	
Multivariate model	1.14 (0.99, 1.30)	1 (Reference)	0.99 (0.81, 1.20)	0.98 (0.79, 1.20)	1.20 (0.97, 1.49)	0.07
Non-stroke CVD mortality						
Number of deaths	n=594	n=109	n=145	n=150	n=190	
Multivariate model	1.17 (1.004, 1.38)	1 (Reference)	0.95 (0.74, 1.22)	1.00 (0.78, 1.28)	1.20 (0.92, 1.55)	0.13
Stroke mortality						
Number of deaths	n=283	n=92	n=80	n=62	n=49	
Multivariate model	1.07 (0.84, 1.35)	1 (Reference)	1.05 (0.75, 1.49)	0.93 (0.64, 1.34)	1.22 (0.84, 1.79)	0.34
Cancer mortality						
Number of deaths	n=896	n=305	n=227	n=204	n=160	
Multivariate model	0.93 (0.81, 1.06)	1 (Reference)	0.96 (0.79, 1.16)	0.90 (0.73, 1.08)	0.87 (0.70, 1.08)	0.16
Other mortality						
Number of deaths	n=1,289	n=441	n=392	n=261	n=195	
Multivariate model	1.07 (0.96, 1.20)	1 (Reference)	1.11 (0.94, 1.31)	1.06 (0.89, 1.26)	1.16 (0.96, 1.39)	0.18
			Animal protein			
Median intake (E%)	10.2	7.2	9.3	11.1	13.9	
All-cause mortality						
Number of deaths	n=3,589	n=1,176	n=986	n=813	n=614	
Multivariate model	1.16 (1.02, 1.33)	1 (Reference)	1.05 (0.95, 1.17)	1.07 (0.97, 1.19)	1.17 (1.04, 1.30)	0.001

Supplemental Table 4. Associations of protein intake with all-cause and cause-specific mortality (n=7786, comparison isocaloric substitution for fat) (Continued)

	Per 5 E% increment n= 7786	Quartile 1 n= 1906	Quartile 2 n=1906	Quartile 3 n=1905	Quartile 4 n=1906	P trend
CVD mortality						
Number of deaths	n=877	n=169	n=216	n=219	n=273	
Multivariate model	1.13 (0.99, 1.30)	1 (Reference)	1.06 (0.86, 1.30)	0.99 (0.80, 1.22)	1.24 (0.99, 1.55)	0.05
Non-stroke CVD mortality						
Number of deaths	n=594	n=109	n=145	n=150	n=190	
Multivariate model	1.19 (1.002, 1.39)	1 (Reference)	1.11 (0.86, 1.43)	1.05 (0.81, 1.38)	1.34 (1.01, 1.75)	0.04
Stroke mortality						
Number of deaths	n=283	n=92	n=80	n=62	n=49	
Multivariate model	1.05 (0.83, 1.34)	1 (Reference)	1.00 (0.70, 1.42)	0.88 (0.61, 1.28)	1.09 (0.90, 2.03)	0.69
Cancer mortality						
Number of deaths	n=896	n=305	n=227	n=204	n=160	
Multivariate model	0.93 (0.81, 1.07)	1 (Reference)	1.11 (0.90, 1.34)	0.96 (0.77, 1.19)	0.98 (0.78, 1.23)	0.56
Other mortality						
Number of deaths	n=1,289	n=441	n=392	n=261	n=195	
Multivariate model	1.09 (0.98, 1.22)	1 (Reference)	1.04 (0.88, 1.24)	1.13 (0.95, 1.35)	1.24 (1.02, 1.49)	0.02
			Plant protein			
Median intake (E%)	5.9	4.7	5.5	6.2	7.3	
All-cause mortality						
Number of deaths	n=3,589	n=1,176	n=986	n=813	n=614	
Multivariate model	1.04 (0.85, 1.27)	1 (Reference)	0.93 (0.85, 1.02)	0.92 (0.83, 1.01)	1.00 (0.88, 1.13)	0.75
CVD mortality						
Number of deaths	n=877	n=169	n=216	n=219	n=273	
Multivariate model	1.22 (0.82, 1.81)	1 (Reference)	0.98 (0.82, 1.18)	1.08 (0.88, 1.32)	1.19 (0.93, 1.53)	0.14
Non-stroke CVD mortality						
Number of deaths	n=594	n=109	n=145	n=150	n=190	

Supplemental Table 4. Associations of protein intake with all-cause and cause-specific mortality (n=7786, comparison isocaloric substitution for fat) (Continued)

	Per 5 E% increment n= 7786	Quartile 1 n= 1906	Quartile 2 n=1906	Quartile 3 n=1905	Quartile 4 n=1906	P trend
Multivariate model	1.17 (0.79, 1.73)	1 (Reference)	0.92 (0.74, 1.15)	1.07 (0.85, 1.35)	1.12 (0.84, 1.48)	0.31
Stroke mortality						
Number of deaths	n=283	n=92	n=80	n=62	n=49	
Multivariate model	1.39 (0.80, 2.44)	1 (Reference)	1.06 (0.77, 1.45)	1.07 (0.76, 1.52)	1.36 (0.90, 2.03)	0.16
Cancer mortality						
Number of deaths	n=896	n=305	n=227	n=204	n=160	
Multivariate model	0.83 (0.59, 1.16)	1 (Reference)	0.82 (0.68, 0.98)	0.84 (0.70, 1.03)	0.89 (0.71, 1.11)	0.27
Other mortality						
Number of deaths	n=1,289	n=441	n=392	n=261	n=195	
Multivariate model	0.77 (0.58, 1.02)	1 (Reference)	1.00 (0.86, 1.15)	0.81 (0.69, 0.96)	0.92 (0.76, 1.12)	0.18

Effect estimates are hazard ratios (HRs) and 95%-confidence intervals (95%CIs) derived from Cox proportional hazards regression models with adjustment for carbohydrate (E%), total energy, alcohol (E%), fiber, age, sex, RS-cohorts (RS-I, -II, and -III), education level(primary, lower, intermediate, and high), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task-hours/week), diet quality score, BMI. Estimates are based on pooled results of imputed data.

Abbreviations: BMI, body mass index.

Supplemental Table 5. Associations of animal protein from various foods with mortality (n=7786, comparison isocaloric substitution for carbohydrate)

Animal protein from (Per 5E%)	All-cause mortality HR (95% CI)	CVD mortality HR (95% CI)	Non-stroke CVD mortality HR (95% CI)
- Meat	1.16 (1.07, 1.27)	1.36 (1.15, 1.61)	1.39 (1.14, 1.70)
- Fish	0.71 (0.58, 0.88)	0.92 (0.58, 1.45)	1.10 (0.64, 1.88)
- Dairy	1.19 (1.10, 1.28)	1.29 (1,11, 1.51)	1.36 (1.13, 1.64)
- Eggs	0.55 (0.29, 1.05)	1.49 (0.42, 5.32)	1.92 (0.42, 8.78)
Animal protein from (Per 5E%)	**Stroke mortality HR (95% CI)**	**Cancer mortality HR (95% CI)**	**Other mortality HR (95% CI)**
- Meat	1.30 (0.96, 1.76)	1.07 (0.90, 1.28)	1.26 (1.10, 1.45)
- Fish	0.58 (0.24, 1.40)	0.84 (0.54, 1.30)	0.68 (0.46, 1.01)
- Dairy	1.15 (0.87, 1.52)	1.16 (0.99, 1.36)	1.21 (1.07, 1.37)
- Eggs	0.86 (0.09, 8.58)	0.46 (0.13, 1.70)	0.63 (0.21, 1.86)

Effect estimates are hazard ratios (HRs) and 95%-confidence intervals (95%CIs) derived from Cox proportional hazards regression models with adjustment for SFA (E%), MUFA (E%), PUFA (E%), TSF (E%), total energy, alcohol (E%), fiber, age, sex, RS-cohorts, education level, smoking status, physical activity, diet quality score, and BMI. Protein from meat, fish, dairy, and eggs are mutually adjusted. Abbreviations: SFA, saturated fat acids, MUFA, monounsaturated fat acids, PUFA, polyunsaturated fat acids, TSF, trans fat acids; BMI, body mass index.

Supplemental Table 6. Associations of protein and mortality after excluding death cases within the first 2 years of follow-up in the Rotterdam study (n=7623, comparison isocaloric substitution for carbohydrate)

	Per 5 E% increment n= 7623	Quartile 1 n= 1906	Quartile 2 n=1906	Quartile 3 n=1905	Quartile 4 n=1906	P trend
			Total protein			
Median intake (E%)	16.2	13.3	15.3	17.0	19.7	
All-cause mortality						
Number of deaths	n=3,427	n=825	n=811	n=850	n=941	
Multivariate model	1.09 (1.02, 1.17)	1 (Reference)	1.06 (0.96, 1.19)	1.04 (0.93, 1.16)	1.15 (1.02, 1.28)	0.04
CVD mortality						
Number of deaths	n=806	n=200	n=175	n=186	n=245	
Multivariate model	1.16 (1.00, 1.35)	1 (Reference)	0.99 (0.79, 1.23)	1.02 (0.82, 1.28)	1.26 (1.00, 1.60)	0.05
Non-stroke CVD mortality						
Number of deaths	n=539	n=131	n=112	n=127	n=169	
Multivariate model	1.23 (1.04, 1.48)	1 (Reference)	0.93 (0.72, 1.22)	1.05 (0.80, 1.38)	1.34 (1.00, 1.77)	0.04
Stroke mortality						
Number of deaths	n=267	n=69	n=63	n=59	n=76	
Multivariate model	1.01 (0.76, 1.32)	1 (Reference)	1.12 (0.76, 1.63)	0.95 (0.63, 1.43)	1.09 (0.70, 1.70)	0.91
Cancer mortality						
Number of deaths	n=836	n=220	n=204	n=207	n=205	
Multivariate model	0.95 (0.82, 1.12)	1 (Reference)	0.92 (0.75, 1.14)	0.90 (0.73, 1.12)	0.90 (0.70, 1.15)	0.38
Other mortality						
Number of deaths	n=1,258	n=303	n=304	n=312	n=339	
Multivariate model	1.09 (0.97, 1.23)	1 (Reference)	1.14 (0.95, 1.35)	1.04 (0.87, 1.26)	1.17 (0.96, 1.43)	0.22

Supplemental Table 6. Associations of protein and mortality after excluding death cases within the first 2 years of follow-up in the Rotterdam study (n=7623, comparison isocaloric substitution for carbohydrate) (Continued)

	Per 5 E% increment n= 7623	Quartile 1 n= 1906	Quartile 2 n=1906	Quartile 3 n=1905	Quartile 4 n=1906	P trend
			Animal protein			
Median intake (E%)	10.2	7.2	9.3	11.1	13.9	
All-cause mortality						
Number of deaths	n=3,427	n=656	n=839	n=929	n=1003	
Multivariate model	1.09 (1.02, 1.17)	1 (Reference)	1.08 (0.97, 1.21)	1.11 (0.99, 1.23)	1.21 (1.07, 1.36)	0.01
CVD mortality						
Number of deaths	n=806	n=154	n=201	n=200	n=251	
Multivariate model	1.16 (1.00, 1.35)	1 (Reference)	1.09 (0.88, 1.38)	1.03 (0.82, 1.30)	1.32 (1.04, 1.70)	0.04
Non-stroke CVD mortality						
Number of deaths	n=539	n=97	n=134	n=133	n=175	
Multivariate model	1.23 (1.03, 1.48)	1 (Reference)	1.13 (0.86, 1.49)	1.12 (0.86, 1.43)	1.38 (1.04, 1.82)	0.04
Stroke mortality						
Number of deaths	n=267	n=57	n=67	n=67	n=76	
Multivariate model	1.00 (0.77, 1.32)	1 (Reference)	1.03 (0.69, 1.54)	0.98 (0.65, 1.48)	1.03 (0.66, 1.62)	0.17
Cancer mortality						
Number of deaths	n=836	n=168	n=226	n=216	n=226	
Multivariate model	0.95 (0.82, 1.12)	1 (Reference)	1.01 (0.80, 1.26)	0.93 (0.74, 1.17)	0.98 (0.76, 1.26)	0.41
Other mortality						
Number of deaths	n=1,258	n=233	n=294	n=356	n=375	
Multivariate model	1.09 (0.97, 1.23)	1 (Reference)	1.05 (0.88, 1.27)	1.11 (0.91, 1.34)	1.25 (1.00, 1.52)	0.06

Supplemental Table 6. Associations of protein and mortality after excluding death cases within the first 2 years of follow-up in the Rotterdam study (n=7623, comparison isocaloric substitution for carbohydrate) (Continued)

	Per 5 E% increment n= 7623	Quartile 1 n= 1906	Quartile 2 n=1906	Quartile 3 n=1905	Quartile 4 n=1906	P trend
			Plant protein			
Median intake (E%)	5.9	4.7	5.5	6.2	7.3	
All-cause mortality						
Number of deaths	n=3,427	n=1,110	n=943	n=787	n=587	
Multivariate model	1.08 (0.87, 1.36)	1 (Reference)	0.96 (0.87, 1.05)	0.95 (0.84, 1.06)	1.07 (0.93, 1.23)	0.68
CVD mortality						
Number of deaths	n=806	n=251	n=218	n=200	n=137	
Multivariate model	1.35 (0.85, 2.14)	1 (Reference)	1.04 (0.85, 1.28)	1.07 (0.84, 1.36)	1.26 (0.93, 1.70)	0.43
Non-stroke CVD mortality						
Number of deaths	n=539	n=165	n=143	n=141	n=90	
Multivariate model	1.31 (0.75, 2.29)	1 (Reference)	1.05 (0.82, 1.34)	1.09 (0.82, 1.45)	1.25 (0.87, 1.80)	0.56
Stroke mortality						
Number of deaths	n=267	n=86	n=75	n=59	n=47	
Multivariate model	1.45 (0.64, 3.25)	1 (Reference)	1.04 (0.72, 1.51)	1.04 (0.68, 1.60)	1.30 (0.75, 2.23)	0.35
Cancer mortality						
Number of deaths	n=836	n=285	n=214	n=190	n=147	
Multivariate model	0.79 (0.49, 1.30)	1 (Reference)	0.84 (0.68, 1.03)	0.85 (0.67, 1.07)	0.90 (0.66, 1.20)	0.14
Other mortality						
Number of deaths	n=1,258	n=433	n=384	n=250	n=191	
Multivariate model	1.09 (0.63, 1.40)	1 (Reference)	1.07 (0.91, 1.26)	0.87 (0.71, 1.05)	1.08 (0.84, 1.39)	0.89

Effect estimates are hazard ratios (HRs) and 95%-confidence intervals (95%CIs) derived from Cox proportional hazards regression models with adjustment for SFA (E%), MUFA (E%), PUFA (E%), TSF (E%), total energy, alcohol (E%), fiber, age, sex, RS-cohorts (RS-I, -II, and -III), education level(primary, lower, intermediate, and high), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task-hours/week), diet quality score, BMI. Estimates are based on pooled results of imputed data. Abbreviations: SFA, saturated fat acids, MUFA, monounsaturated fat acids, PUFA, polyunsaturated fat acids, TSF, trans fat acids; BMI, body mass index.

Supplemental Table 7. Characteristics of studies in the systematic review and the meta-analysis

Supplemental Table 7 can be obtained from the author.

Supplemental Table 8. Dose-response meta-analysis results

	Number of studies	Number of participants	Number of deaths	RR (95% CI)	I^2	$P_{heterogeneity}$
			Total protein			
All-cause mortality	5	218846	55863	1.02 (1.00, 1.04)	37.9%	0.17
CVD mortality	4	216205	13965	1.04 (0.997, 1.09)	37.4%	0.19
Cancer mortality	4	216205	19748	1.00 (0.95, 1.05)	11.6%	0.33
			Animal protein			
All-cause mortality	4	212465	53310	1.05 (0.99, 1.12)	70.1%	0.02
CVD mortality	3	209824	12753	1.05 (1.02, 1.09)	31.2%	0.23
Cancer mortality	3	209824	19110	1.01 (0.98, 1.04)	0.0%	0.76
			Plant protein			
All-cause mortality	4	212465	53310	0.87 (0.78. 0.98)	40.0%	0.17
CVD mortality	4	217568	13107	0.77 (0.52, 1.16)	73.2%	0.01
Cancer mortality	3	209824	19110	0.88 (0.71, 1.09)	58.1%	0.09

Risk ratios (RRs) and 95%-confidence intervals (95%CIs) reflect the difference in the risks per 5 E% increment of dietary protein intake based on a linear dose-response meta-analysis. When sufficient studies (n≥5) contributed to a dose-response meta-analysis, non-linearity of dose-response association was explored using restricted cubic splines with three knots (10%, 50%, and 90%) for the amount of dietary protein intake. No evidence for non-linear associations was observed (Wald test: p>0.05).

Supplemental Table 9. Subgroup meta-analysis by geographic study location

	Total protein				Animal protein				Plant protein			
	n[1]	**RR (95%CI)**	**I^2**	**P^2**	**n[1]**	**RR (95% CI)**	**I^2**	**P^2**	**n[1]**	**RR (95% CI)**	**I^2**	**P^2**
All-cause mortality												
US	3	1.05 (1.002, 1.09)	0.0	0.58	2	0.97 (0.81, 1.18)	46.0	0.17	3	0.90 (0.85, 0.96)	0.0	0.43
Europe	2	1.13 (1.04, 1.24)	0.0	0.67	2	1.17 (1.06, 1.28)	0.0	0.68	2	1.04 (0.92, 1.17)	0.0	0.59
Asia	1	0.99 (0.90,1.09)	-	-	1	0.98 (0.88, 1.09)	-	-	1	0.87 (0.78, 0.97)	-	-
CVD mortality												
US	3	1.06 (0.90, 1.25)	20.8	0.28	2	1.09 (0.99, 1.19)	0.0	0.58	3	0.77 (0.64, 0.92)	0.0	0.49
Europe	1	1.22 (0.99 1.51)	-	-	1	1.28 (1.03, 1.60)	-	-	1	1.19 (0.91, 1.56)	-	-
Asia	1	0.97 (0.80, 1.18)	-	-	2	0.97 (0.79, 1.18)	0.0	0.90	2	0.83 (0.73, 0.94)	0.0	0.31
Cancer mortality												
US	4	0.97 (0.80, 1.18)	0.0	0.44	3	1.02 (0.94, 1.10)	0.0	1.0	3	0.95 (0.87, 1.04)	0.0	0.54
Europe	1	1.00 (0.86, 1.16)	-	-	1	0.98 (0.78, 1.23)	-	-	1	1.04 (0.88, 1.23)	-	-
Asia	1	0.87 (0.70, 1.08)	-	-	1	0.97 (0.83, 1.14)	-	-	1	0.90 (0.69, 1.17)	-	-

RR (95% CI), I^2, and $P_{heterogeneity}$ values were obtained from a highest versus lowest meta-analysis with random effects

[1] Number of studies

[2] $P_{heterogeneity}$

Supplemental Table 10. Sensitivity analysis for meta-analysis by excluding each one study one at a time[1]

	RR (95% CI)	I^2	$P_{heterogeneity}$
Pooled association of total protein and all-cause mortality, after excluding			
- Kelemen et al, 2005	1.05 (1.00, 1.11)	26.2%	0.73
- Levine et al, 2014	1.06 (1.01, 1.10)	11.3%	0.31
- Song et al, 2016	1.05 (0.97, 1.14)	27.8%	0.97
- Virtanen et al, 2019	1.05 (1.01,1.08)	0%	0.20
- Budhathoki et al, 2019	1.06 (1.02, 1.10)	0%	0.19
- Chen et al, 2019	1.04 (1.00, 1.08)	0%	0.21
Pooled association of total protein and CVD mortality, after excluding			
- Kelemen et al, 2005	1.08 (0.96, 1.21)	34.1%	0.49
- Levine et al, 2014	1.11 (1.01, 1.21)	7.4%	0.18
- Song et al, 2016	1.03 (0.88, 1.21)	23.1%	0.29
- Budhathoki et al, 2019	1.12 (1.01,1.23)	7.3%	0.18
- Chen et al, 2019	1.05 (0.93, 1.18)	23.2%	0.29
Pooled association of total protein and cancer mortality, after excluding			
- Kelemen et al, 2005	1.01 (0.95, 1.08)	0%	0.19
- Smit et al, 2007	1.01 (0.95, 1.09)	3.5%	0.30
- Levine et al, 2014	1.02 (0.94, 1.11)	18.2%	0.57
- Song et al, 2016	1.01 (0.88, 1.15)	20.6%	0.67
- Budhathoki et al, 2019	1.02 (0.91,1.14)	22.1%	0.78
- Chen et al, 2019	1.04 (0.97, 1.11)	0%	0.13
Pooled associations of animal protein and all-cause mortality, after excluding			
- Kelemen et al, 2005	1.06 (0.98, 1.15)	58.6%	0.06
- Song et al, 2016	1.05 (0.93, 1.20)	65.8%	0.03
- Virtanen et al, 2019	1.04 (0.95,1.14)	64.7%	0.04
- Budhathoki et al, 2019	1.07 (0.97, 1.19)	61.3%	0.05
- Chen et al, 2019	1.02 (0.96, 1.08)	20.3%	0.28
Pooled associations of animal protein and CVD mortality, after excluding			
- Sauvaget et al, 2004	1.09 (0.99, 1.21)	16.8%	0.66
- Kelemen et al, 2005	1.09 (0.99, 1.21)	13.9%	0.57
- Song et al, 2016	1.08 (0.90, 1.30)	21.1%	0.97
- Budhathoki et al, 2019	1.11 (1.02, 1.21)	0%	0.23
- Chen et al, 2019	1.06 (0.98, 1.16)	0%	0.12

Supplemental Table 10. Sensitivity analysis for meta-analysis by excluding each one study one at a time[1] (Continued)

	RR (95% CI)	I^2	$P_{heterogeneity}$
Pooled associations of animal protein and cancer mortality, after excluding			
- Kelemen et al, 2005	1.01 (0.94, 1.08)	0%	0.93
- Smit et al, 2007	1.01 (0.94, 1.08)	0%	0.99
- Song et al, 2016	0.98 (0.87, 1.10)	0%	0.60
- Budhathoki et al, 2019	1.02 (0.94, 1.09)	0%	0.61
- Chen et al, 2019	1.01 (0.94, 1.08)	0%	0.80
Pooled associations of plant protein and all-cause mortality, after excluding			
- Kelemen et al. 2005	0.93 (0.85, 1.01)	51.9%	0.10
- Song et al, 2016	0.95 (0.86, 1.05)	42.4%	0.16
- Virtanen et al, 2019	0.93 (0.86,1.00)	51.9%	0.10
- Budhathoki et al, 2019	0.95 (0.87, 1.04)	44.5%	0.14
- Chen et al, 2019	0.90 (0.85, 0.94)	0%	0.69
Pooled associations of plant protein and CVD mortality, after excluding			
- Sauvaget et al, 2004	0.84 (0.71, 1.00)	55.5%	0.06
- Kelemen et al, 2005	0.88 (0.74, 1.05)	52.4%	0.08
- Song et al, 2016	0.86 (0.68, 1.09)	58.6%	0.05
- Kurihara et al, 2019	0.87 (0.72, 1.04)	58.1%	0.05
- Budhathoki et al, 2019	0.90 (0.75, 1.07)	45.4%	0.12
- Chen et al, 2019	0.81 (0.73, 0.90)	0%	0.57
Pooled associations of plant protein and cancer mortality, after excluding			
- Kelemen et al, 2005	0.95 (0.87, 1.04)	0%	0.66
- Smit et al, 2007	0.96 (0.89, 1.04)	0%	0.48
- Song et al, 2016	1.00 (0.91, 1.10)	0%	0.28
- Budhathoki et al, 2019	0.95 (0.87, 1.03)	0%	0.33
- Chen et al, 2019	0.97 (0.90, 1.05)	0%	0.59

Effect estimates are Risk ratios (RRs) and 95% confidence intervals (95%CIs) derived from random-effect highest versus lowest meta-analysis. [1]This sensitivity analysis was not conducted for other mortality, because there were only two studies this outcome.

Supplemental Table 11. Sensitivity analysis of dose-response meta-analysis

	RR (95% CI)	I^2	$P_{heterogeneity}$
Total protein and all-cause mortality			
The main dose-response meta-analysis results (Per 5 E%)	1.02 (1.004, 1.04)	37.9%	0.17
The result from the study by Bates et al (Per 5 E%)	0.86 (0.77, 0.97)	-	-
Pooled results (Per 5 E%)	0.94 (0.80, 1.12)	87.8%	0.004
Total protein and CVD mortality			
The main dose-response meta-analysis results (Per 5 E%)	1.04 (0.997, 1.09)	37.4%	0.19
The result from the study by Bates et al (Per 5 E%)	0.79 (0.67, 0.94)	-	-
Pooled results (Per 5 E%)	0.91 (0.70, 1.20)	89.4%	0.002
Animal protein and CVD mortality			
The main dose-response meta-analysis results (Per 5 E%)	1.05 (1.02, 1.09)	31.2%	0.23
The result from the study by Tharrey et al (Per 5 E%)	1.12 (1.05, 1.19)	-	-
Pooled results (Per 5 E%)	1.08 (1.01, 1.16)	68.6%	0.07
Plant protein and CVD mortality			
The main dose-response meta-analysis results (Per 5 E%)	0.77 (0.52, 1.16)	73.2%	0.01
The result from the study by Tharrey (Per 5 E%)	0.95 (0.89, 1.05)	-	-
Pooled results (Per 5 E%)	0.94 (0.87, 1.01)	2.3%	0.31

Effect estimates are Risk ratios (RRs) and 95%-confidence intervals (95%CIs) derived from random-effects meta-analysis.

Chapter 2.3

Plant-based diet and type 2 diabetes

Chen Z*, Zuurmond MG*, Van der Schaft N, Nano J, Wijnhoven HAH, Ikram MA, Franco OH, Voortman T. Plant versus animal based diets and insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. European journal of epidemiology. 2018;33(9):883-93.

**denotes equal contribution*

ABSTRACT

Background: Vegan or vegetarian diets have been suggested to reduce type 2 diabetes (T2D) risk. However, not much is known on whether variation in the degree of having a plant-based versus animal-based diet may be beneficial for prevention of T2D.

Objectives: We aimed to investigate whether level of adherence to a diet high in plant-based foods and low in animal-based foods is associated with insulin resistance, prediabetes, and T2D.

Methods: Our analysis included 6798 participants (62.7 ± 7.8 years) from the Rotterdam Study (RS), a prospective population-based cohort in the Netherlands. Dietary intake data were collected with food-frequency questionnaires at baseline of three sub-cohorts of RS (RS-I-1: 1989-93, RS-II-1: 2000-01, RS-III-1: 2006-08). We constructed a continuous plant-based dietary index (range 0-92) assessing adherence to a plant-based versus animal-based diet. Insulin resistance at baseline and follow-up was assessed using homeostasis model assessment of insulin resistance (HOMA-IR). Prediabetes and T2D were collected from general practitioners' records, pharmacies' databases, and follow-up examinations in our research center until 2012. We used multivariable linear mixed models to examine association of the index with longitudinal HOMA-IR, and multivariable Cox proportional-hazards regression models to examine associations of the index with risk of prediabetes and T2D.

Results: During median 5.7 years, and 7.3 years of follow-up, we documented 928 prediabetes cases and 642 T2D cases. After adjusting for sociodemographic and lifestyle factors, a higher score on the plant-based dietary index was associated with lower insulin resistance (per 10 units higher score: β= -0.09, 95% CI: -0.10, -0.08), lower prediabetes risk (HR=0.89, 95% CI: 0.81, 0.98), and lower T2D risk (HR=0.82 (0.73, 0.92)). After additional adjustment for BMI, associations attenuated and remained statistically significant for longitudinal insulin resistance (β= -0.05 (-0.06, -0.04)) and T2D risk (HR=0.87 (0.79, 0.99)), but no longer for prediabetes risk (HR=0.93 (0.85, 1.03)).

Conclusions: A more plant-based and less animal-based diet may lower risk of insulin resistance, prediabetes and T2D. These findings strengthen recent dietary recommendations to adopt a more plant-based diet.

INTRODUCTION

Diet is an important modifiable lifestyle determinant in the development of type 2 diabetes (T2D).[1] Among these dietary determinants, several plant-based foods such as root vegetables, green leafy vegetables, whole grains, nuts and peanut butter, have been associated with a lower risk of T2D.[2-5] By contrast, several animal-based foods, including red meat, processed meat, and daily consumption of eggs have been associated with an increased risk of T2D.[4, 6, 7]

Although multiple food groups seem to influence the risk of T2D, humans generally do not consume single food items or food groups, and the role of diet in health may be better described by overall dietary patterns.[8] Previous studies have observed that vegan or vegetarian diets are associated with improved glycemic control[9] and lower T2D risk.[10] However, these previous studies dichotomously classified participants, and only defined diets as vegetarian or vegan versus non-vegetarian diets. A dichotomous classification of vegans or vegetarians versus their non-vegetarian counterparts might not be an optimal approach in understanding the effect of a plant-based diet in Western countries, because it does not reflect dietary patterns of a large proportion of the population. For public health advice, it is interesting to know if a more plant-based and less animal-based diet may also influence insulin resistance and risk of prediabetes and T2D beyond strict adherence to a vegetarian or vegan diet. To our knowledge, only one previous study, a large prospective cohort study in the US, examined associations between variations in the degree of adherence to plant-based versus animal-based diets with T2D risk and observed that a more plant-based diet was associated with a lower T2D risk.[11] Studies on the associations of such plant-based dietary patterns with T2D risk in other populations are needed. In addition, the association of such plant-based dietary patterns with intermediate risk factors for T2D, such as insulin resistance and prediabetes remain unknown.

Therefore, we aimed to investigate whether adherence to a more plant-based, and less animal-based diet is associated with insulin resistance, and risk of prediabetes and T2D in a Dutch middle-aged and older general population.

METHODS

Study population

This study was carried out within three sub-cohorts of the Rotterdam Study (RS), a prospective cohort study of adult aged 45 years and older living in the well-defined district of Ommoord in Rotterdam, the Netherlands. A detailed description of the Rotterdam Study methodology is described elsewhere.[12] Briefly, recruitment of participants for the first sub-cohort (RS-I) started in the period of 1989-93 among inhabitants aged ≥ 55 years (n=7983). In 2000-01, the study was extended with a second sub-

cohort (RS-II) of new individuals (n=3011) who had become 55 years of age or moved into the study area after 1990. In 2006-08, a third sub-cohort (RS-III) was recruited with new individuals aged 45 years and older (n=3932). By the end of 2008, the overall study population contained 14926 participants. Upon entering the study, participants underwent home interviews and a series of examinations in our research center every 3-5 year.

The Rotterdam Study has been approved by the institutional review board (Medical Ethics Committee) of Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. The approval has been renewed every 5 years. All participants gave informed consent.

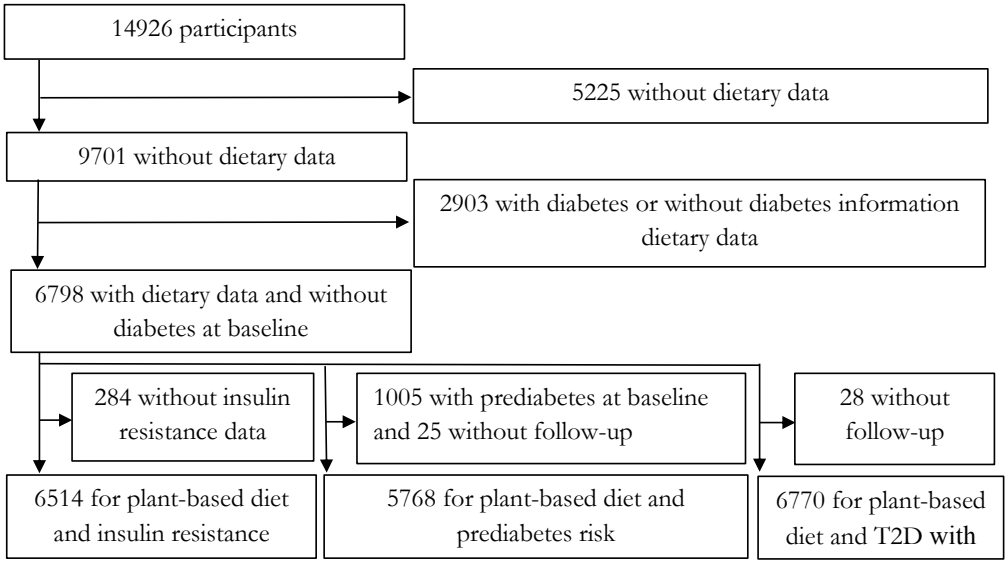


Figure 1. Participants selection

Population for current analyses

For the current study, we used data from all three sub-cohorts (Figure 1). Of the 14926 participants, we excluded those without valid dietary data (no dietary data (n=5141) or unreliable dietary intake according to a trained nutritionist or an estimated energy intake of <500 or >5000 kcal/day (n=84)[13]) at baseline (RS-I-1: 1989-93, RS-II-1: 2000-01, RS-III-1: 2006-08), and those without diabetes information or with prevalent T2D at baseline (n=2903), leaving 6798 participants included as main population for analysis.

From this group of 6798 participants, 6514 participants had data on HOMA-IR before onset of T2D and were included in the longitudinal HOMA-IR analyses. For the analyses on prediabetes risk, we excluded those with prevalent prediabetes at baseline (n=1005) or without follow-up of prediabetes (n=25), leaving 5768 participants. In the analyses assessing risk of T2D, we excluded participants without follow-up of T2D (n=28), leaving 6770 participants. The flow-diagram of the included participants is presented in Figure 1.

Dietary assessment

Dietary intake was assessed at baseline in all three sub-cohorts using semi-quantitative food-frequency questionnaires (FFQ) as described in more detail elsewhere.[13] We used an FFQ with 170 food items to assess dietary intake at baseline of RS-I (1989-93) and RS-II (2000-01);[14] and at baseline of RS-III (2006-08) we used an FFQ with 389 food items.[15] The 170-item FFQ was validated in a subsample of the Rotterdam Study (n=80) against fifteen 24-h food records and four 24h urinary urea excretion samples;[14] and the 389-item FFQ was previously validated in other Dutch population against measurement of biomarkers, against a 9-day dietary record, and against a 4 week dietary history.[16] In general, the validation studies demonstrated that the FFQs were able to adequately rank participants according to their intake.[13] Food intake data were converted to energy and nutrient intake based on Dutch Food Composition tables (NEVO).

Plant-based dietary index

We constructed an overall plant-based dietary index, which was a modified version of two previously created indices.[11, 17] More specifically, our index is similar to the "provegetarian food pattern" of Martínez-Gonzáles et al.[17] and to the "overall plant-based diet index" of Satija et al.,[11] but was adapted to include slightly different types and numbers of food categories.

First, the food items as measured by the FFQs were divided into 23 food categories (Supplemental Table 1), on the basis of the main food groups in the Dutch diet and the Dutch food-based dietary guidelines.[18, 19] Twelve of the categories were plant-based and eleven were animal-based. Food items that were not clearly animal-based or plant-based, such as pizza, as well dietary supplements, were not included in the food categories for the index.

Dietary intake for each of the 23 food categories (g/d) was calculated for each participant. Subsequently, for each category, the intake was divided into cohort-specific quintiles. Each quintile was assigned a value between 0 and 4. For the twelve plant-based food categories, consumption within the highest quintile was scored a 4, consumption within the second highest quintile was scored a 3, and so on, ending with consumption within the lowest quintile receiving a score of 0. The eleven animal-based food categories were scored reversely: consumption within the highest quintile was

scored a 0 consumption within the second highest quintile was scored a 1, ending with consumption within the lowest quintile receiving a score of 4. Furthermore, we ensured that all participants with null consumption were given the score belonging to the lowest quintile by re-scoring when necessary.

Finally, these category quintile-scores were added up for per participant to create their overall score on the plant-based dietary index. The resulting index yielded a score for each participant that measured adherence to a plant-based versus animal-based diet on a continuous scale, with a lowest possible score of 0 (low adherence to a plant-based diet) and a highest possible score of 92 (high adherence: high plant-based and low animal-based). Information on intake of each food category across quintiles of scores on the plant-based dietary index is shown in Supplemental Table 2.

Assessment of insulin resistance

Fasting blood samples were collected at RS-I (RS-I-3: 1997-99, RS-I-5: 2009-10), RS-II (RS-II-1: 2000-01, RS-II-3: 2010-11), and RS-III (RS-III-1: 2006-08, RS-III-2: 2011-12). Glucose levels were examined with the glucose hexokinase method. Serum insulin was measured by electro chemiluminescence immunoassay technology. Insulin resistance was calculated using the homeostasis model assessment of insulin resistance (HOMA-IR). The following formula was used: fasting insulin (mU/L) × fasting glucose (mmol/L) / 22.5.

Assessment of prediabetes and type 2 diabetes

Information on prediabetes and T2D was collected from general practitioners' records, pharmacies' databases, and follow-up examinations in our research center. Data of prediabetes and T2D in our analyses were collected until January 1, 2012. Prediabetes and T2D were identified according to WHO criteria: prediabetes was defined as a fasting blood glucose concentration of > 6.0 and < 7.0 mmol/L, or a non-fasting blood glucose concentration of > 7.7 mmol/L and < 11.1 mmol/L; T2D was defined as a fasting blood glucose concentration of ≥ 7.0 mmol/L, a non-fasting blood glucose concentration of ≥ 11.1 mmol/L (when fasting samples were unavailable), or the use of blood glucose-lowering drugs or dietary treatment and registration of the diagnosis diabetes. All possible cases of prediabetes and T2D were formally judged by two independently working study physicians or, in case of disagreement, by an endocrinologist.[20]

Assessment of covariates

Information on age, sex, smoking status, educational level, medication use, food supplement use, and family history of diabetes, was obtained from questionnaires at baseline. Information on physical activity was obtained using the adapted version of the Zutphen Physical Activity Questionnaire at RS-I-3 and RS-II-1 and using the LASA Physical Activity Questionnaire at RS-III-1. Physical activities were weighted according to intensity with Metabolic Equivalent of Task (MET), from the

Compendium of Physical Activities version 2011. To account for differences between the two questionnaires, questionnaire-specific z-scores of MET-hours per week were calculated. At our research center at baseline, body weight was measured using a digital scale and body height was measured using a stadiometer, while participants wore light clothing and no shoes, and BMI was calculated (kg/m^2). Information on hypertension, hypercholesterolemia, coronary heart disease (CHD), cancers, and stroke was obtained from general practitioners, pharmacies' databases, Nationwide Medical Register, or follow-up examinations in our research center.

Data analysis

To obtain a normal distribution for HOMA-IR, we applied a natural-log transformation. Non-linearity of associations of score on the plant-based dietary index with all outcomes were explored using natural cubic splines (degrees of freedom = 3). As no indications for non-linear associations for the main models were found, all primary analyses were performed using models assuming linearity. We examined the association between score on the plant-based dietary index with longitudinal HOMA-IR using linear mixed models, with a random-effects structure including a random intercept and slope (for time of repeated measurements of HOMA-IR). We examined the association between score on the plant-based dietary index and risk of prediabetes and risk of T2D using Cox proportional-hazards regressions. Hazard ratios (HRs) and regression coefficients (βs) were presented per 10 units higher score on the plant-based dietary index, along with the corresponding 95% confidence intervals (CIs). All analyses were performed in participants of the three sub-cohorts combined and in the three sub-cohorts separately.

All analyses were adjusted for energy intake, age, sex and RS sub-cohort in model 1, and for the analyses of longitudinal HOMA-IR we additionally adjusted for the time of repeated measurements of HOMA-IR. In model 2, we additionally adjusted for smoking status, educational level, physical activity, food supplement use, and family history of diabetes. Baseline BMI was added to model 3 to examine its potential mediating effect.

We examined effect modification by including interactions of the plant-based index with age, sex, or BMI for all outcomes in model 2.

Several sensitivity analyses were performed based on model 2. First, to check if the associations were driven by any specific components of the plant-based dietary index, we repeated our main analyses by excluding each one of the 23 components from the plant-based dietary index one by one at a time, and additionally adjusting for the excluded component. Second, to check if the associations were mainly driven by plant-based beverages combined, we examined the associations by excluding all plant-based beverages combined (category "coffee and tea", category "alcoholic beverages", and category "sugary beverages") from the plant-based dietary index at a time, and additionally adjusting

for them. Third, we examined the associations by excluding less healthy plant-based foods combined (category "sweets", category "sugary beverages", category "potatoes", and category "refined grains") from the plant-based dietary index at a time, and additionally adjusting for them. To further examine whether these less healthy plant foods contributed to the association of the plant-based dietary index; we created a less healthy plant foods score, for which, positive scores were given to these four types of less healthy plant-based food groups; and reverse scores were given to healthy plant food groups and animal food groups.[21] Fourth, to examine if potential associations of the plant-based dietary score with outcomes were independent of overall quality of the diet based on adherence to dietary guidelines, we examined the correlation between the plant-based dietary score and the dietary guidelines score; and we repeated analyses with additional adjustment for dietary guidelines score. Fifth, we additionally adjusted for hypertension and hypercholesterolemia. Sixth, we excluded the participants with chronic diseases at baseline, such as participants with CHD, cancers, or stroke, to exclude the possibility of a significant change of diet and lifestyle at follow-up. Last, we excluded the participants who developed prediabetes and T2D in the first 2 years of follow-up in the analyses for risk of prediabetes and T2D, respectively.

Missing values on covariates (ranging from 0.3% to 3.9%) were accounted for using multiple imputations (n=10 imputations). We used SPSS version 21 (IBM Corp., Armonk, NY, USA) and R version 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria) to perform these analyses.

RESULTS

Baseline characteristics

Baseline characteristics of the study population are shown in Table 1. In our population of 6798 participants, baseline scores on the plant-based dietary index (with a theoretical range from 0 to 92) ranged from 24 to 75, with a mean ± SD score of 49.3 ± 7.1. Mean age of the study population was 62.0 ± 7.8 years and 41.3% of the participants were male. Mean BMI was 26.6 ± 3.9 kg/m^2. Characteristics were similar before and after multiple imputation (Supplemental Table 3). Supplemental Table 4 shows baseline characteristics of the participants not included in our analyses.

Table 1. Baseline characteristics of study participants (n=6798)

Characteristics	Mean (SD), median (IQR), or %
Age (years)	62.0 (7.8)
Sex (% male)	41.3 %
BMI (kg/m^2)	26.6 (3.9)
Smoking (%)	
- Never	32.2 %
- Ever	45.1 %
- Current	22.7 %
Physical activity[1] (MET-hours/week)	
- RS-I and RS-II (Zutphen Questionnaire, n=4393)	86.7 (44.7)
- RS-III (LASA Questionnaire, n=2194)	58.4 (55.8)
Hypertension (%)	42.3 %
Hypercholesterolemia (%)	45.4 %
Family history of diabetes (%)	10.8 %
Highest level of education (%)	
- Primary	11.8 %
- Lower	40.9 %
- Intermediate	29.0 %
- Higher	18.3 %
Current food supplement use (%)	16.5 %
Total energy intake (kcal/day)	2134 (615)
Plant-based food category intake (grams/day)	
- Fruit	212.2 (115.5, 332.3)
- Vegetables	209.1 (147.9, 286.87)
- Whole grains	105.7 (61.3, 152.5)
- Nuts	3.9 (0.0, 12.0)
- Legumes	4.1 (0.0, 19.4)
- Potatoes	99.7 (61.4, 148.2)
- Vegetable oils	19.7 (9.2, 30.0)
- Tea and coffee	758.9 (580.4, 1000)
- Sugary beverages	46.3 (0.0, 139.6)
- Refined grains	50.7 (23.9, 102.1)
- Sweets	63.8 (37.1, 97.4)
- Alcoholic beverages	56.4 (4.9, 159.8)
Animal-based food category intake (grams/day)	
- Low-fat milk	82.3 (0.0, 232.3)
- Full-fat milk	0.0 (0.0, 0.0)
- Low-fat yoghurt	56.1 (0.0, 164.6)
- Full-fat yoghurt	0.0 (0.0, 4.9)
- Cheese	30.8 (20, 47.1)

Table 1. Baseline characteristics of study participants (n=6798) (Continued)

Characteristics	Mean (SD), median (IQR), or %
- Unprocessed lean meat	10.7 (4.3, 18.1)
- Fish	15.9 (3.9, 30.7)
- Eggs	14.3 (7.1, 19.6)
- Animal fat	0.0 (0.0, 0.9)
- Desserts/ dairy with sugars	14.1 (0.0, 54.6)
- Processed meat/ red meat	86.8 (60.4, 118.9)
Plant-based dietary index (score)	49.3 (7.1)

Plant-based dietary index: a higher score indicates a higher adherence to a plant-based diet (theoretical range from 0 to 92). Values shown are based on pooled results of imputed data.
[1]Values shown for MET-hours are un-imputed; imputation was performed on z-scores of physical activity.
Abbreviations: MET, metabolic equivalent of task; SD, standard deviation.

Plant-based dietary index and insulin resistance

After adjustment for confounders in model 2, a higher score on the plant-based dietary index was associated with lower longitudinal HOMA-IR (per 10 units higher score on the index: β= -0.09, (95% CI: -0.10, -0.08)) (Table 2). Adding BMI to the model (Model 3), attenuated the association, but it remained statistically significant (β= -0.05 (-0.06, -0.04)).

Plant-based dietary index and incidence of prediabetes

During 43773 person-years of follow-up amongst 5768 participants (median follow-up 5.7 years), 928 participants developed prediabetes. After adjustment for confounders in model 2 (Table 2), a higher score on the plant-based dietary index was associated with a lower incidence of prediabetes (per 10 units higher score on the index: HR=0.89, (95%CI 0.81, 0.98)). After additional adjustment for BMI (Model 3) the association was attenuated, and no longer statistically significant (HR=0.93 (0.85, 1.03)).

Plant-based dietary index and incidence of type 2 diabetes

During 54024 person-years of follow-up amongst 6770 participants (median follow-up 7.3 years), 642 participants developed T2D. In model 2, a higher score on the plant-based dietary index was associated with a lower incidence of T2D (per 10 units higher score on the index: HR=0.82, (95%CI 0.73, 0.92)) (Table 2). Additional adjustment for BMI (Model 3) attenuated this association, but it was still statistically significant (HR=0.87 (0.79, 0.99)).

Table 2. Associations of the plant-based dietary index with longitudinal insulin resistance (HOMA-IR), risk of prediabetes, and risk of type 2 diabetes

	HOMA-IR n=6514 β (95% CI)	Prediabetes n=5768 HR (95% CI)	Type 2 diabetes n=6770 HR (95% CI)
Model 1	-0.09 (-0.10, -0.08)***	0.88 (0.80, 0.97)**	0.82 (0.73, 0.92)***
Model 2	-0.09 (-0.10, -0.08)***	0.89 (0.81, 0.98)*	0.82 (0.73, 0.92)**
Model 3	-0.05 (-0.06, -0.04)***	0.93(0.85, 1.03)	0.87 (0.79, 0.99)*

Effect estimates are regression coefficients (β) for ln HOMA-IR or hazard ratios (HRs) for incidence of prediabetes or type 2 diabetes with their 95%-confidence intervals (95%CIs), per 10 units higher score on the plant-based dietary index. Estimates are based on pooled results of imputed data.
Model 1 is adjusted for energy intake (kcal), sex (male or female), age (years) and RS sub-cohort (RS-I, -II, or -III); and only for the HOMA analyses additionally for the time measurements of longitudinal HOMA.
Model 2 is additionally adjusted for education (primary, lower, intermediate, or higher), smoking status (never, ever, current); family history of diabetes (yes, no, or unknown); physical activity (z-score of MET-hours/week); and food supplement use (yes or no).
Model 3 is additionally adjusted for BMI
*p<0.05; **p<0.01; ***p<0.001
Abbreviations: BMI, body mass index; CI, confidence interval; HR, hazard ratio; MET, metabolic equivalent of task; RS, Rotterdam-Study.

The associations between the plant-based dietary index with longitudinal insulin resistance, and risk of prediabetes and T2D were similar in three sub-cohorts (Supplemental Tables 5-7). Associations did not differ by age, sex or baseline BMI (p-values for all interaction terms were >0.05).

Sensitivity analyses

The exclusion of each one of 23 foods from the index one by one at a time did not substantially change the estimates (Supplemental Table 8). Excluding all plant-based beverages combined at a time (coffee and tea, alcoholic beverages and sugary beverages) did not substantially change the estimates (per 10 units higher score on the index, insulin resistance: β= -0.06 (-0.10, -0.03), prediabetes risk: HR=0.93 (0.84, 1.02), and T2D risk: HR=0.85 (0.80, 0.96)). The estimates also remained similar after excluding these less healthy plant-based foods combined at a time (sweets, sugary beverages, potatoes, and refined grains) (per 10 units higher score on the index, insulin resistance: β= -0.09 (-0.10, -0.07), prediabetes risk: HR=0.90 (0.84, 0.98), and T2D risk: HR=0.83 (0.74, 0.94)), but the less healthy plant foods score was not associated with insulin resistance or with risk of prediabetes or type 2 diabetes (insulin resistance: β= -0.002 (-0.01, 0.006), risk of prediabetes: HR=1.00 (-0.99, 1.01), and risk of type 2 diabetes: HR=0.99 (0.98, 1.00)). The Pearson's correlation coefficient between the plant-based

dietary score with the dietary guidelines score was 0.16 (P<0.05); and controlling for the dietary guidelines score did not substantially affect the estimates (per 10 units higher score on the index, insulin resistance: β= -0.09 (-0.10, -0.08), prediabetes risk: HR=0.91 (0.82, 1.00), and T2D risk: HR=0.81 (0.71, 0.91)).

Additional adjustment for hypertension and hypercholesterolemia did not change effect estimates (per 10 units higher score on the index, insulin resistance: β= -0.08 (-0.10, -0.07), risk of prediabetes: HR=0.90 (0.82, 0.99), and risk of T2D: HR=0.84 (0.75, 0.94)), and estimates remained similar after excluding participants with chronic diseases at baseline (per 10 units higher score on the index, insulin resistance: β= -0.09 (-0.11, -0.07), prediabetes risk: HR=0.88 (0.79, 0.97), and T2D risk: HR=0.81 (0.72, 0.92)). Finally, excluding participants who developed T2D or prediabetes in the first 2 years of follow-up modestly attenuated the associations for prediabetes (per 10 units higher score on the index, HR=0.91 (0.83, 1.01)), and T2D (HR=0.82 (0.73, 0.92)).

DISCUSSION

In this large population-based cohort, we observed that a diet higher in plant-based foods and lower in animal-based foods was associated with lower insulin resistance, and a lower risk of prediabetes and T2D, suggesting a protective role of a more plant-based opposed to a more animal-based diet in the development to T2D, beyond strict adherence to a vegetarian or vegan diet.

The inverse association between plant-based diets and T2D risk is in agreement with previous research showing lower T2D risk for vegans or vegetarians, compared to non-vegetarians.[10] Moreover, our observed associations confirmed the observations of Satija and colleagues in a US sample,[11] the only other prospective study examining adherence to plant-based diets in a continuous graduation with risk of T2D. Compared to this previous study in the US population, we have extended this evidence by also showing associations between plant-based diets in a continuous graduation with earlier stages of the development of T2D: insulin resistance, and prediabetes in a European population.

Our results imply a beneficial effect of adherence to a diet higher in plant-based foods and lower in animal-based foods on the development of T2D, irrespective of general healthfulness of the specific plant-based and animal-based foods. With these results, we provide a different view on what a healthy diet may entail. However, we acknowledge that our plant-based diet included positive scoring for some components that are not necessarily healthy choices for prevention of T2D, or a healthy diet in general. Sugary beverages, for example, have been associated with adverse effects for T2D in other studies.[22, 23]

To further clarify whether these less healthy plant foods contributed to the observed associations, we examined the associations between less healthy plant-based diet score with insulin resistance, and risk of prediabetes and T2D in our sensitivity analyses, and observed null associations; suggesting beneficial associations were mainly driven by higher intake of healthy plant-based food groups and lower intake of animal-based food groups. This emphasizes that it is important to also consider the quality of plant-based foods consumed, which has important public health implications. Furthermore, the estimates for the plant-based dietary index remained similar after excluding these plant-based beverages combined, or after excluding the less healthy plant-based foods combined, which indicated that our results were stable in diverse versions of plant-based diets, thus increased our confidence in the validity of the findings. We also observed that excluding each one of 23 components one by one at a time resulted in similar associations as observed for the total plant-based index, indicating that the associations were not mainly explained by any one specific food group, which supports the importance of recognizing overall plant-based diet. Finally, we extended our analyses to examine if adherence to a plant-based diet was independent of adherence to current Dutch dietary guidelines. In line with results from the large prospective cohort study in the US which examined if adherence to a plant-based diet was independent of general healthy dietary patterns that have been linked to prevention of T2D, such as the Mediterranean diet, the alternative Healthy Eating Index (aHEI), and the Dietary approaches to stop hypertension (DASH) diet.[24-26] We observed that associations of the plant-based dietary index with outcomes remained similar after additional adjustment for adherence to current Dutch dietary guidelines. This lends support to novelty of the plant-based dietary index.

Taken together, a more plant-based, less animal-based diet may help prevent the development of T2D. Still more important, a more plant-based diet, does not require a radical change in diet or a total elimination of meat or animal products but instead can be achieved in various ways, increasing the potential for population-wide health recommendations. For example, if a participant in our cohort would increase fruits intake from 95 grams per day to 200 grams per day, increase vegetables intake from 100 grams to 260 grams, and at the same time decrease red meat intake from 129 grams per day to 55 grams per day, this would improve the plant-based dietary index by 10 units, which may decrease risk of T2D by 13%, assuming other covariates remain stable.

Potential biological mechanisms

Several mechanisms behind the inverse associations could involve the intermediate conditions of T2D, such as obesity and inflammation, can offer explanations for the observed protection and T2D. On the one hand, a plant-based diet usually has more fiber, chlorogenic acids, certain amino acids, unsaturated fatty acids, and antioxidants. For example, vegetables and fruits are the main sources of fiber, anti-oxidants, and chlorogenic acids; nuts are rich in poly-unsaturated fatty acids; soy and beans are main sources of plant protein; whole grains are rich in fiber and plant protein; and coffee and tea

are rich in anti-oxidants and phenol chlorogenic acid. These beneficial components may influence the development of T2D through impact on the potential intermediate conditions, such as obesity and inflammation. Fiber is known to lower gastric emptying and thereby glycemic responsiveness,[27] and might improve inflammation,[28, 29] and obesity.[30] Chlorogenic acids can improve inflammation, glucose tolerance and glucose levels, and improve increasing insulin secretion.[31] Soy protein contains high amounts of the amino acids arginine and glycine, which have been associated with a decrease in cholesterol levels.[32] High intake of unsaturated fatty acids has also been associated to lower inflammation and less obesity.[28, 33] Phenol chlorogenic acid was reported to reduce insulin resistance.[34] On the other hand, a plant-based diet, usually has less animal protein, saturated fatty acids, and heme iron. Animal protein is rich in branched-chain amino acids and aromatic amino acids and may impair glucose metabolisms and increase T2D risk;[35-38] animal protein is also rich in heme iron, which has been suggested to increase risk of cardio-metabolic diseases.[39-41] Higher saturated fatty acids have been suggested to be associated higher inflammation,[33] higher risk of obesity[33] and T2D.[42, 43] Besides, other nutrients from processed red meat, such as sodium and nitrites, may increase risk of cardio-metabolic diseases.[41] More research is needed to explore whether the mechanisms also involve an effect of plant foods on gut microbiome. Finally, these different mechanisms may influence each other because of inter-relations between different food components. This also highlights the relevance of examining overall diets in additional to isolated food items, as this enables capturing of the combined effects of the potential pathways.

Strengths and limitations

This study has several strengths. First, to our knowledge, we are the first to investigate the associations between plant-based diets with longitudinal insulin resistance and prediabetes, for which we had longitudinal data from long follow-up available. Studying these early risk stages help minimize reverse causation, understand how plant-based diet influences the development of T2D. Second, we observed that the potential beneficial effect of a more plant-based diet was independent of less healthy plant foods, such as sweets, sugary beverages and refined grains, emphasizing the importance of considering the quality of plant-based foods consumed. We also observed associations of the plant-based dietary score independent of overall adherence to dietary guidelines, indicating that the plant-based diet score may reflect more than only a healthful dietary pattern as reflected by current dietary guidelines. Other strengths also included the population-based nature of the study, the detailed and thorough data collected on the outcomes and the assessment of the extent to which diets were plant-based and animal based, based upon overall dietary intake patterns of the general population.

Nevertheless, there are several limitations we should consider. First, the assessment of a plant-based diet with this index has its limitations as several sometimes-arbitrary decisions had to be made. A decision was, for example, to add up food items within categories based on the intake in grams per

day. As a result, products that were high in water-content will have contributed less energy or nutrients compared to products containing less water in the same category. However, using grams per day reflects intake of foods as they are consumed and recommended.[19] Also, decisions had to be made for the categorization of foods and the number of categories. We chose categories reflecting those used in the Dutch dietary guidelines, which are based on similarities of the food items in (botanical) origin, nutrient composition, and nutrient density;[18] thereby reducing nutritional differences between food items within one category. Furthermore, in our main analyses, we treated all plant-based foods equally by giving all plant-based foods positive scores, and all animal-based foods equally by giving all animal-based foods reverse scores, irrespective of their nutrient-density or previous evidence for a role in T2D prevention and general health. For example, less healthy plant-based foods, such as sugary beverages and refined grains, were included as positive scores, although sugary beverages,[23] and refined grains[44] have been linked to higher T2D risk; by contrast, healthy animal-based foods, such as dairy and fish, were included as reverse scores, although dairy[45] and fish[46] have been linked to lower T2D risk or mortality risk. That is because our study aimed to emphasize an overall plant-based diet including various increased plant-based foods consumption and decreased animal-based foods consumption, which would increase the potential for population-wide recommendation. However, in our sensitivity analyses, excluding any one of alcoholic beverages, sugary beverages, sweets, potatoes, refined grains, fish, and dairy did not substantially change our estimates.

In addition to the choices we had to make in the construction of the index, this study has some other limitations. First, dietary data were derived from self-reported diet measured with FFQs, making measurement-errors likely. However, because we used relative scores (quintiles) of intake and the FFQs were shown in several validation studies to adequately rank subjects according to intake,[13-16] we do not expect these measurement-errors to have largely affected our results. Second, we did not have dietary data for many of the participants of the original cohort, which might have resulted in selection bias if associations of plant-based diets with T2D risk differed in those included and those not included in our current analyses. Third, we assumed stable diets over time. However, the estimates were similar after excluding the participants who were likely to change their diet during follow-up, such as participants with CHD, stroke, and cancers at baseline. Last, our results may be generalizable only to people of similar age and race.

Conclusions

In this large population-based cohort, higher adherence to an overall plant-based diet is associated with lower longitudinal insulin resistance, and lower risk of prediabetes and T2D, indicating a protective role of diets high in plant-based foods and low in animal-based foods in the development to T2D beyond strict adherence to a vegetarian or vegan diet. These promising findings call for further exploration of overall plant-based dietary recommendations aimed at T2D prevention.

REFERENCES

1. Ley SH, Hamdy O, Mohan V, Hu FB. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. The Lancet. 2014;383(9933):1999-2007.
2. Cooper AJ, Forouhi NG, Ye Z, Buijsse B, Arriola L, Balkau B, et al. Fruit and vegetable intake and type 2 diabetes: EPIC-InterAct prospective study and meta-analysis. Eur J Clin Nutr. 2012;66(10):1082-92.
3. Aune D, Norat T, Romundstad P, Vatten LJ. Whole grain and refined grain consumption and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis of cohort studies. Eur J Epidemiol. 2013;28(11):845-58.
4. Schwingshackl L, Hoffmann G, Lampousi AM, Knuppel S, Iqbal K, Schwedhelm C, et al. Food groups and risk of type 2 diabetes mellitus: a systematic review and meta-analysis of prospective studies. Eur J Epidemiol. 2017;32(5):363-75.
5. Jiang R, Manson JE, Stampfer MJ, Liu S, Willett WC, Hu FB. Nut and peanut butter consumption and risk of type 2 diabetes in women. Jama. 2002;288(20):2554-60.
6. van Woudenbergh GJ, Kuijsten A, Tigcheler B, Sijbrands EJ, van Rooij FJ, Hofman A, et al. Meat consumption and its association with C-reactive protein and incident type 2 diabetes: the Rotterdam Study. Diabetes Care. 2012;35(7):1499-505.
7. Djoussé L, Gaziano JM, Buring JE, Lee IM. Egg consumption and risk of type 2 diabetes in men and women. Diabetes Care. 2009;32(2):295-300.
8. Salas-Salvado J, Martinez-Gonzalez MA, Bullo M, Ros E. The role of diet in the prevention of type 2 diabetes. Nutr Metab Cardiovasc Dis. 2011;21 Suppl 2:B32-48.
9. Yokoyama Y, Barnard ND, Levin SM, Watanabe M. Vegetarian diets and glycemic control in diabetes: a systematic review and meta-analysis. Cardiovascular diagnosis and therapy. 2014;4(5):373.
10. Tonstad S, Stewart K, Oda K, Batech M, Herring RP, Fraser GE. Vegetarian diets and incidence of diabetes in the Adventist Health Study-2. Nutrition, Metabolism and Cardiovascular Diseases. 2013;23(4):292-9.
11. 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 Med. 2016;13(6):e1002039.
12. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. European Journal of Epidemiology. 2017:1-44.
13. Voortman T, Kiefte-de Jong JC, Ikram MA, Stricker BH, van Rooij FJA, Lahousse L, et al. Adherence to the 2015 Dutch dietary guidelines and risk of non-communicable diseases and mortality in the Rotterdam Study. Eur J Epidemiol. 2017;32(11):993-1005.
14. Klipstein-Grobusch K, den Breeijen JH, Goldbohm RA, Geleijnse JM, Hofman A, Grobbee DE, et al. Dietary assessment in the elderly: validation of a semiquantitative food frequency questionnaire. Eur J Clin Nutr. 1998;52(8):588-96.

15. Goldbohm RA, van den Brandt PA, Brants HA, van't Veer P, Al M, Sturmans F, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur J Clin Nutr. 1994;48(4):253-65.
16. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. The American journal of clinical nutrition. 1993;58(4):489-96.
17. Martinez-Gonzalez MA, Sanchez-Tainta A, Corella D, Salas-Salvado J, Ros E, Aros F, et al. A provegetarian food pattern and reduction in total mortality in the Prevencion con Dieta Mediterranea (PREDIMED) study. Am J Clin Nutr. 2014;100:320S-8S.
18. Dutch Nutrition Center. Guidelines Wheel of Five. The Hague: Dutch Nutrition Center, 2016.
19. Health Council of the Netherlands. Guidelines healthy nutrition 2015. The Hague: Health Council of the Netherlands, 2015 2015/24.
20. Ligthart S, van Herpt TT, Leening MJ, Kavousi M, Hofman A, Stricker BH, et al. Lifetime risk of developing impaired glucose metabolism and eventual progression from prediabetes to type 2 diabetes: a prospective cohort study. The Lancet Diabetes & Endocrinology. 2016;4(1):44-51.
21. Satija A, Bhupathiraju SN, Spiegelman D, Chiuve SE, Manson JE, Willett W, et al. Healthful and Unhealthful Plant-Based Diets and the Risk of Coronary Heart Disease in U.S. Adults. J Am Coll Cardiol. 2017;70(4):411-22.
22. Ma J, Jacques PF, Meigs JB, Fox CS, Rogers GT, Smith CE, et al. Sugar-Sweetened Beverage but Not Diet Soda Consumption Is Positively Associated with Progression of Insulin Resistance and Prediabetes. The Journal of nutrition. 2016;146(12):2544-50.
23. Imamura F, O'Connor L, Ye Z, Mursu J, Hayashino Y, Bhupathiraju SN, et al. Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. Bmj. 2015;351:h3576.
24. Alhazmi A, Stojanovski E, McEvoy M, Garg ML. The association between dietary patterns and type 2 diabetes: a systematic review and meta-analysis of cohort studies. Journal of Human Nutrition and Dietetics. 2014;27(3):251-60.
25. Jannasch F, Kroger J, Schulze MB. Dietary Patterns and Type 2 Diabetes: A Systematic Literature Review and Meta-Analysis of Prospective Studies. J Nutr. 2017;147(6):1174-82.
26. Schwingshackl L, Bogensberger B, Hoffmann G. Diet Quality as Assessed by the Healthy Eating Index, Alternate Healthy Eating Index, Dietary Approaches to Stop Hypertension Score, and Health Outcomes: An Updated Systematic Review and Meta-Analysis of Cohort Studies. J Acad Nutr Diet. 2018;118(1):74-100 e11.
27. Livesey G, Tagami H. Interventions to lower the glycemic response to carbohydrate foods with a low-viscosity fiber (resistant maltodextrin): meta-analysis of randomized controlled trials. The American journal of clinical nutrition. 2009;89(1):114-25.
28. Eichelmann F, Schwingshackl L, Fedirko V, Aleksandrova K. Effect of plant-based diets on obesity-related inflammatory profiles: a systematic review and meta-analysis of intervention trials. Obes Rev. 2016;17(11):1067-79.

29. Wannamethee SG, Whincup PH, Thomas MC, Sattar N. Associations between dietary fiber and inflammation, hepatic function, and risk of type 2 diabetes in older men: potential mechanisms for the benefits of fiber on diabetes risk. Diabetes care. 2009;32(10):1823-5.
30. Papathanasopoulos A, Camilleri M. Dietary fiber supplements: effects in obesity and metabolic syndrome and relationship to gastrointestinal functions. Gastroenterology. 2010;138(1):65-72. e2.
31. Santos RM, Lima DR. Coffee consumption, obesity and type 2 diabetes: a mini-review. Eur J Nutr. 2016;55(4):1345-58.
32. Sanchez A, Hubbard RW. Plasma amino acids and the insulin/glucagon ratio as an explanation for the dietary protein modulation of atherosclerosis. Medical hypotheses. 1991;36(1):27-32.
33. Bray GA, Lovejoy JC, Smith SR, DeLany JP, Lefevre M, Hwang D, et al. The influence of different fats and fatty acids on obesity, insulin resistance and inflammation. The Journal of nutrition. 2002;132(9):2488-91.
34. Shearer J, Farah A, de Paulis T, Bracy DP, Pencek RR, Graham TE, et al. Quinides of roasted coffee enhance insulin action in conscious rats. The Journal of nutrition. 2003;133(11):3529-32.
35. Wittenbecher C, Muhlenbruch K, Kroger J, Jacobs S, Kuxhaus O, Floegel A, et al. Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes. Am J Clin Nutr. 2015;101(6):1241-50.
36. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes. 2013;62(2):639-48.
37. Guasch-Ferre M, Hruby A, Toledo E, Clish CB, Martinez-Gonzalez MA, Salas-Salvado J, et al. Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. Diabetes Care. 2016;39(5):833-46.
38. Batch BC, Shah SH, Newgard CB, Turer CB, Haynes C, Bain JR, et al. Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness. Metabolism. 2013;62(7):961-9.
39. Ascherio A, Willett WC, Rimm EB, Giovannucci EL, Stampfer MJ. Dietary iron intake and risk of coronary disease among men. Circulation. 1994;89(3):969-74.
40. de Oliveira Otto MC, Alonso A, Lee D-H, Delclos GL, Bertoni AG, Jiang R, et al. Dietary Intakes of Zinc and Heme Iron from Red Meat, but Not from Other Sources, Are Associated with Greater Risk of Metabolic Syndrome and Cardiovascular Disease–3. The Journal of nutrition. 2012;142(3):526-33.
41. Micha R, Wallace SK, Mozaffarian D. Red and processed meat consumption and risk of incident coronary heart disease, stroke, and diabetes mellitus: a systematic review and meta-analysis. Circulation. 2010;121(21):2271-83.
42. Van Dam RM, Willett WC, Rimm EB, Stampfer MJ, Hu FB. Dietary fat and meat intake in relation to risk of type 2 diabetes in men. Diabetes care. 2002;25(3):417-24.
43. de Souza RJ, Mente A, Maroleanu A, Cozma AI, Ha V, Kishibe T, et al. Intake of saturated and trans unsaturated fatty acids and risk of all cause mortality, cardiovascular disease, and type 2 diabetes: systematic review and meta-analysis of observational studies. Bmj. 2015;351:h3978.
44. Hu EA, Pan A, Malik V, Sun Q. white rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. Bmj. 2012;344:e1454.

45. Chen M, Sun Q, Giovannucci E, Mozaffarian D, Manson JE, Willett WC, et al. Dairy consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. BMC medicine. 2014;12(1):215.
46. Zhao LG, Sun JW, Yang Y, Ma X, Wang YY, Xiang YB. Fish consumption and all-cause mortality: a meta-analysis of cohort studies. European journal of clinical nutrition. 2016;70(2):155.

SUPPLEMENTAL MATERIAL

Supplemental Table 1. Food categories used for the plant-based diet index and examples of food items included in each of the food categories

Plant-based food categories	
Fruits	Apple, banana, pear, orange, strawberry, grapes, other fruits
Vegetables	Cauliflower, broccoli, spinach, carrots, onion, lettuce, tomato, cabbage, cooked vegetables
Whole grains	Whole grain bread, dark bread, rye bread, whole grain breakfast oats, whole grain pasta, brown rice
Nuts	Peanuts, walnuts, other nuts, peanut butter
Legumes	Legumes, tofu, soybeans, other soy products
Potatoes	Potatoes, fries
Vegetable oils	Olive oil, vegetable oils used for cooking, and all margarines
Tea and coffee	Black tea, green tea, herbal tea, coffee
Sugary beverages	Carbonated beverages with sugar, non-carbonated beverages with sugar, orange juice, fruit juice
Refined grains	Cornflakes, white bread, croissants, raisin bread, white pasta, white rice
Sweets	Sugar, cookies, cake, chocolate, candy-bars, honey, sweets, chocolate toppings, other sweet toppings
Alcoholic beverages	Red wine, white wine, beer, liquor, Dutch-eggnog
Animal-based food categories	
Low-fat Yoghurt	Skimmed yoghurt, semi-skimmed yoghurt, skimmed quark, buttermilk
Full-fat Yoghurt	Full-fat yoghurt, semi-skimmed quark, full quark
Low-fat milk	Skimmed milk, semi-skimmed milk, skimmed coffee creamer, semi-skimmed coffee creamer
Full-fat milk	Full-fat milk, cream, coffee-cream
Cheese	Full fat cheese, low fat cheese, cheese fondue, other cheese
Fish	Salmon, tuna, trout, herring, mussels, other fish
Eggs	Boiled eggs, fried eggs
Animal fat	Butter on bread, butter used for cooking, lard
Desserts and sugary dairy	Custard, cream, ice cream, mousse, cream, chocolate milk, fruit yoghurt, yoghurt drinks
Unprocessed lean meat	Chicken
Processed meat and red meat	Beef, pork, meatballs, sate, bacon, liver, processed meats

Supplemental Table 2. Baseline intake of 23 food categories of participants in quintiles of plant-based dietary index

Plant-based dietary index	Score≤43 n=1417	43<score≤47 n=1311	47<score ≤51 n=1559	51<score≤55 n=1226	score>55 n=1285
Food intake (grams/day)					
- Fruits	168.0 (83.4, 274.5)	197.4 (104.0, 320)	215.7 (115.2, 340.3)	226.7 (127.3, 351.9)	258.5 (161.1, 395.1)
- Vegetables	181.6 (128.0, 252.9)	199.4 (143.9, 277.1)	205.2 (146.4, 283.3)	216.9 (156.4, 297.7)	241.3 (180.4, 331.4)
- Whole grains	88.3 (46.6, 125.0)	99.5 (50.0, 140.6)	108.3 (63.0, 151.1)	114.7 (67.6, 160.0)	135.0 (80.0, 188.0)
- Legumes	0.0 (0.0, 8.9)	0.0 (0.0, 16.9)	4.1 (0.0, 18.0)	7.8 (0.0, 24.0)	13.5 (0.0, 35.6)
- Nuts	13.5 (0, 6.0)	2.1 (0.0, 8.8)	3.6 (0.0, 11.8)	5.6 (0.4, 14.1)	9.0 (2.7, 19.2)
- Vegetable oils	12.0 (3.3, 21.4)	16.6 (7.2, 26.0)	20.6 (10.4, 30.0)	24.0 (13.3, 32.6)	27.7 (18.1, 38.5)
- Tea and coffee	705.4 (500.0, 875.0)	750.0 (525.0, 937.5)	767.9 (597.1, 1000.0)	812.5 (625.0,1044.6)	900.0 (705.4,1125.0)
- Refined grains	37.7 (17.1, 76.8)	50.0 (22.7, 97.6)	50.6 (23.5, 101.3)	60.0 (30.4, 115.6)	61.2 (30.9, 122.2)
- Potatoes	83.6 (45.9, 122.0)	88.2 (57.0, 131.0)	97.9 (61.7, 142.5)	108.3 (71.2, 163.1)	126.0 (85.5, 178.1)
- Sweets	50.3 (26.6, 81.7)	57.2 (32.6, 87.5)	64.2 (38.2, 95.6)	71.3 (43.5, 105.2)	71.3 (43.5, 105.2)
- Sugary beverages	15.0 (0.0, 89.6)	40.0 (0.0, 139.3)	42.9 (0.0, 139.6)	42.9 (0.0, 139.6)	59.8 (1.2, 152.6)
- Alcoholic beverages	31.8 (2.5, 124.7)	47.7 (3.6, 155.3)	58.8 (4.9, 160.3)	65.4 (8.4, 167.9)	81.9 (14.2, 189.3)
- Low-fat yoghurt	82.3 (5.4, 192.9)	64.1 (0.0, 166.1)	60.0 (0.0, 164.5)	53.6 (0.0, 162.0)	32.1 (0.0, 149.6)
- Full-fat yoghurt	0.0 (0.0, 34.8)	0.0 (0.0, 13.4)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
- Low-fat milk	111.0 (1.9, 278.6)	100.8 (0.88, 263.6)	91.0 (0.0, 224.4)	59.0 (0, 224.4)	48.0 (0.0, 196.5)
- Full-fat milk	0.0 (0.0, 7.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
- Cheese	32.9 (21.3, 47.1)	32.6 (20.3, 50.0)	30.3 (20.0, 46.6)	28.4 (18.2, 44.6)	29.9 (17.8, 47.0)
- Fish	21.4 (7.1, 33.8)	18.9 (5.9, 33.0)	14.6 (4.2, 30.2)	14.4 (2.4, 28.6)	11.0 (0.0, 25.9)

Supplemental Table 2. Baseline intake of 23 food categories of participants in quintiles of plant-based dietary index (Continued)

Plant-based dietary index	Score≤43 n=1417	43<score≤47 n=1311	47<score ≤51 n=1559	51<score≤55 n=1226	score>55 n=1285
- Eggs	14.3 (8.9, 21.4)	14.3 (7.1, 21.4)	14.3 (7.1, 17.9)	14.3 (7.1, 17.1)	10.7 (7.1, 17.1)
- Animal fat	0.7 (0.0, 12.0)	0.0 (0.0, 2.3)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
- Desserts/dairy with sugars	21.4 (1.5, 63.9)	18.4 (0.4, 60.5)	14.9 (0.0, 59.6)	10.2 (0.0, 48.1)	6.4 (0.0, 35.8)
- Unprocessed lean meat	14.3 (6.9, 21.4)	14.3 (7.1, 21.4)	11.4 (4.3, 18.6)	10.7 (4.3, 17.8)	7.6 (0.0, 14.9)
- Processed/red meat	93.2 (65.4, 127.5)	89.3 (63.4, 127.5)	86.9 (60.0, 118.0)	85.5 (60.4, 117.9)	80.0 (52.5, 112.3)

Variables expressed as median (IQR) because of their skewed distributions.

Supplemental Table 3. Baseline characteristics of participants in original and multiple imputed dataset

Characteristics	Original data Mean (SD) or valid %	After imputation Mean (SD) or %
Age (years)	62.0 (7.8)	NI
missing (%)	–	–
Gender (% male)	41.3 %	NI
missing (%)	–	–
BMI (kg/m^2)	26.6 (3.9)	26.6 (3.9)
missing (%)	1.3 %	–
Smoking (%)		
- never	32.2 %	32.2 %
- ever	45.1 %	45.1 %
- current	22.7 %	22.7 %
missing (%)	0.5 %	–
Physical activity[1] (MET-hours/week)		
- RS-III (assessed with LASA	58.4 (55.8)	58.4 (55.8)
Questionnaire, n=2194)		
- RS-I and RS-II (assessed with Zutphen	86.7 (44.7)	86.7 (44.7)
Questionnaire, n=4393)		
missing (%)	3.9 %	–
Hypertension (%)	42.3 %	42.3 %
missing (%)	0.9 %	–
Hypercholesterolemia (%)	45.6 %	45.4 %
missing (%)	1.6 %	–
Family history of type 2 diabetes (%)	10.8 %	NI
missing (%)	–	–
Education level (%)		
- primary	11.8 %	11.8 %
- lower	40.9 %	40.9 %
- intermediate	29.0 %	29.0 %
- higher	18.3 %	18.3 %
missing (%)	0.6 %	–
Current food supplement use (%)	16.5 %	16.5 %
missing (%)	0.3 %	–
Total energy intake (kcal/day)	2134 (615)	NI
missing (%)	–	–
Food category intake[2] (grams/day)		
- Fruits	212.2 (115.5, 332.3)	NI
- Vegetables	209.1 (147.9, 286.87	NI
- Whole grains	105.7 (61.3, 152.5)	NI

Supplemental Table 3. Baseline characteristics of participants in original and multiple imputed dataset (Continued)

Characteristics	Original data Mean (SD) or valid %	After imputation Mean (SD) or %
- Nuts	3.9 (0.0, 12.0)	NI
- Legumes	4.1 (0.0, 19.4)	NI
- Potatoes	99.7 (61.4, 148.2)	NI
- Vegetable oils	19.7 (9.2, 30.0)	NI
- Tea and coffee	758.9 (580.4, 1000)	NI
- Sugary beverages	46.3 (0.0, 139.6)	NI
- Refined grains	50.7 (23.9, 102.1)	NI
- Sweets	63.8 (37.1, 97.4)	NI
- Alcoholic beverages	56.4 (4.9, 159.8)	NI
- Low-fat milk	82.3 (0.0, 232.3)	NI
- Full-fat milk	0.0 (0.0, 0.0)	NI
- Low-fat yoghurt	56.1 (0.0, 164.6)	NI
- Full-fat yoghurt	0.0 (0.0, 4.9)	NI
- Cheese	30.8 (20, 47.1)	NI
- Unprocessed lean	10.7 (4.3, 18.1)	NI
- Fish	15.9 (3.9, 30.7)	NI
- Eggs	14.3 (7.1, 19.6)	NI
- Animal fat	0.0 (0.0, 0.9)	NI
- Desserts/dairy with sugars	14.1 (0.0, 54.6)	NI
- Processed meat/ red meat	86.8 (60.4, 118.9)	NI
Plant-based dietary index (score)	49.3 (7.1)	NI

Plant-based dietary index: a higher score indicates a higher adherence to a plant-based diet (theoretical range from 0 to 92).

[1]Values shown are un-imputed; imputation was performed on z-scores of physical activity.

[2]Variables expressed as median (IQR) because of their skewed distributions.

Abbreviations: MET, metabolic equivalent of task; NI, not imputed; SD, standard deviation.

Supplemental Table 4. Non-response analyses

Covariates	Participants without valid dietary data n=5225 Mean (SD) or %	Participants with valid dietary data n=9701 Mean (SD) or %	P value T-test or X^2 test
Age (years)	64.9 (12.7)	62.0 (7.8)	P<0.05
Sex (%)			
- Female	59.0%	41.8%	P<0.05
- Male	38.8%	58.0%	
BMI (kg/m^2)	27.0 (4.4)	26.6 (3.9)	P<0.05
Physical activity (MET-hours/week)			P<0.05
- RS-I and RS-II (Zutphen Questionnaire)	72.4 (42.5)	83.5 (44.6)	P<0.05
- RS-III (LASA Questionnaire)	65.3(43.5)	58.2(59.3)	
Education level (%)			p>0.05
- Lower	25.0%	11.8%	
- Primary	37.2%	40.9%	
- Intermediate	24.4%	29.0%	
- Higher	13.3%	18.4%	
Smoking status (%)			
- Never	35%	32.2%	p>0.05
- Ever	39%	45.06%	
- Current	25.6%	22.7%	
Current food supplement use (%)			
- Yes	16.9%	16.5%	p>0.05
- No	83.1%	83.2%	
Family history of diabetes (%)			
- Yes	9.0%	10.8%	p>0.05
- No	39.8%	45.8%	
- Unknown	51.3%	43.4%	

Supplemental Table 4. Non-response analyses (Continued)

Covariates	Participants not included in analyses n=8128 Mean (SD) or %	Included participants in analyses n=6798 Mean (SD) or %	P value T-test or X^2 test
Age (years)	69.3 (11.4)	62.0 (7.8)	P<0.05
Sex (%)			
- Female	59.5%	57%	p>0.05
- Male	40.1%	41.3 %	
BMI (kg/m^2)	27.1 (4.3)	26.6 (3.9)	P<0.05
Physical activity (MET-hours/week)			
- RS-I and RS-II (Zutphen Questionnaire)	72.1 (42.5)	86.7 (44.7)	P<0.05
- RS-III (LASA Questionnaire)	61.6 (79.9)	58.4 (55.8)	
Education level (%)			
- Primary	23.6%	11.8 %	p>0.05
- Lower	37.0%	40.9 %	
- Intermediate	23.6%	29.0 %	
- Higher	11.1%	18.3 %	
Smoking status (%)			
- Never	32.5%	32.2 %	p>0.05
- Ever	38.4%	45.1 %	
- Current	24.3%	22.7 %	
Current food supplement use (%)			
- Yes	14.6%	16.5%	P<0.05
- No	84.6%	83.5%	
Family history of diabetes (%)			
- Yes	13.9%	45.8%	p>0.05
- No	49.1%	10.8%	
- Unknown	36.9%	43.4%	

t-test was performed for continuous variables, and X^2 was performed for categorical variable

Supplemental Table 5. Associations of the plant-based dietary index with longitudinal insulin resistance (HOMA-IR) for the three sub-cohorts separately

	β for HOMA-IR (95% CI)		
	RS-I (n=2892)	**RS-II (n=1389)**	**RS-III (n=2233)**
Model 1	-0.09 (-0.10, -0.08)***	-0.07 (-0.11, -0.03)***	-0.11 (-0.14, -0.07)***
Model 2	-0.09 (-0.10, -0.08)***	-0.06 (-0.10, -0.02)**	-0.10 (-0.13, -0.07)***
Model 3	-0.05 (-0.07, -0.03)*	-0.01 (-0.05, 0.02)	-0.06 (-0.09, -0.03)***

Effect estimates are βs for ln-transformed HOMA-IR per 10 units higher score on the plant-based dietary index and are based on pooled results of the imputed dataset. Model 1 is adjusted for energy intake (kcal), sex (male or female), age (years), and time (years) of repeated measurements of longitudinal insulin resistance. Model 2 is additionally adjusted for education (primary education, lower/intermediate general education or lower vocational education, intermediate vocational education or higher general education, higher vocational education or university), smoking status (never, ever, current), family history of diabetes (yes, no, or unknown); physical activity (z-score of MET-hours/week); and food supplement use (yes or no). Model 3 is additionally adjusted for BMI (kg/m2). *p<0.05; **p<0.01; ***p<0.001. Abbreviations: CI, confidence interval; HOMA-IR, homeostasis model assessment for insulin resistance; MET, metabolic equivalent of task; RS, Rotterdam-Study.

Supplemental Table 6. Associations of the plant-based dietary index with incidence of prediabetes for the three sub-cohorts separately

	HR (95% CI) for prediabetes		
	RS-I (n=2492)	**RS-II (n=1151)**	**RS-III (n=2125)**
Model 1	0.93 (0.82, 1.05)	0.94 (0.78, 1.14)	0.65 (0.51, 0.84)***
Model 2	0.94 (0.83, 1.06)	0.94 (0.78, 1.14)	0.66 (0.52, 0.85)**
Model 3	0.96 (0.85, 1.09)	1.00 (0.83, 1.21)	0.70 (0.54, 0.90)**

Effect estimates are HRs (95% CIs) for incidence of prediabetes per 10 units higher score on the plant-based dietary index and are based on pooled results of the imputed dataset. Model 1 is adjusted for energy intake (kcal), sex (male or female), and age (years). Model 2 is additionally adjusted for education (primary education, lower/intermediate general education or lower vocational education, intermediate vocational education or higher general education, higher vocational education or university), smoking status (never, ever, current), family history of diabetes (yes, no, or unknown); physical activity (z-score of MET-hours/week); and food supplement use (yes or no). Model 3 is additionally

adjusted for BMI (kg/m2). *p<0.05; **p<0.01; ***p<0.001. Abbreviations: BMI, body mass index; CI, confidence interval; HR, hazard ratio; MET, metabolic equivalent of task; RS, Rotterdam-Study.

Supplemental Table 7. Associations of the plant-based dietary index with incidence of type 2 diabetes for the three sub-cohorts separately

	HR (95% CI) for type 2 diabetes		
	RS-I (n=2975)	**RS-II (n=1411)**	**RS-III (n=2384)**
Model 1	0.85 (0.73, 0.98)*	0.82 (0.65, 1.02)	0.74 (0.54, 1.02)
Model 2	0.86 (0.74, 0.997)*	0.86 (0.69, 1.07)	0.75 (0.54, 1.04)
Model 3	0.91 (0.78, 1.05)	0.93 (0.74, 1.16)	0.80 (0.58, 1.12)

Effect estimates are HRs (95% CIs) for incidence of type 2 diabetes per 10 units higher score on the plant-based dietary index and are based on pooled results of the imputed dataset. Model 1 is adjusted for energy intake (kcal), sex (male or female), and age (years). Model 2 is additionally adjusted for education (primary education, lower/intermediate general education or lower vocational education, intermediate vocational education or higher general education, higher vocational education or university), smoking status (never, ever, current), family history of diabetes (yes, no, or unknown); physical activity (z-score of MET-hours/week); and food supplement use (yes or no). Model 3 is additionally adjusted for BMI (kg/m2). *p<0.05; **p<0.01; ***p<0.001. Abbreviations: BMI, body mass index; CI, confidence interval; HR, hazard ratio; MET, metabolic equivalent of task; RS, Rotterdam-Study.

Supplemental Table 8. Associations of the plant-based dietary index with longitudinal insulin resistance (HOMA-IR), risk of prediabetes and type 2 diabetes excluding each of 23 components at a time

Plant-based dietary index with 22 components instead of 23 components	β (95% CI) for HOMA-IR n=6514	HR (95% CI) for Prediabetes risk n=5768	HR (95% CI) for T2D risk n=6770
Excluding fruits	-0.08 (-0.10, -0.07) ***	0.89 (0.81, 0.98) *	0.82 (0.73, 0.92) **
Excluding vegetables	-0.09 (-0.10, -0.09) ***	0.89 (0.81, 0.98) *	0.81 (0.72, 0.92) **
Excluding whole grains	-0.09 (-0.10, -0.09) ***	0.89 (0.81, 0.98) *	0.81 (0.73, 0.92) **
Excluding nuts	-0.07 (-0.09, -0.06) ***	0.91 (0.81, 1.00)	0.84 (0.76, 0.95) **
Excluding legumes	-0.08 (-0.10, -0.07) ***	0.90 (0.82, 0.99) *	0.83 (0.74, 0.92) **
Excluding vegetable oils	-0.08 (-0.10, -0.07) ***	0.90 (0.82, 0.99) *	0.82 (0.73, 0.92) **
Excluding tea and coffee	-0.07 (-0.09, -0.06) ***	0.91 (0.83, 0.99) *	0.84 (0.75, 0.95) **
Excluding potatoes	-0.09 (-0.10, -0.09) ***	0.89 (0.81, 0.98) *	0.82 (0.73, 0.92) **
Excluding sugary beverages	-0.09 (-0.10, -0.08) ***	0.89 (0.81, 0.98) *	0.82 (0.72, 0.92) **
Excluding refined grains	-0.09 (-0.10, -0.08) ***	0.89 (0.81, 0.98) *	0.82 (0.73, 0.92) **
Excluding sweets	-0.08 (-0.10, -0.08) ***	0.90 (0.82, 0.99) *	0.81 (0.73, 0.92) **
Excluding alcoholic beverages	-0.08 (-0.10, -0.06) ***	0.89 (0.82, 0.98) *	0.83 (0.71, 0.95) **
Excluding red and processed meat	-0.07 (-0.08, -0.07) ***	0.93 (0.84, 0.99) *	0.84 (0.76, 0.95) **
Excluding unprocessed lean meat	-0.07(-0.08, -0.07) ***	0.90 (0.82, 0.99) *	0.84 (0.76, 0.95) **
Excluding fish	-0.08 (-0.10, -0.07) ***	0.90 (0.81, 0.99) *	0.84 (0.74, 0.94) **
Excluding eggs	-0.09 (-0.10, -0.08) ***	0.89 (0.80, 0.98) *	0.82 (0.73, 0.92) **
Excluding animal fat	-0.08 (-0.10, -0.08) ***	0.89 (0.79, 0.99) *	0.83 (0.70, 0.95) **
Excluding cheese	-0.08 (-0.10, -0.07) ***	0.91 (0.82, 0.99) *	0.84 (0.75, 0.94) **
Excluding low-fat milk	-0.08 (-0.10, -0.06) ***	0.86 (0.79, 0.95) *	0.81 (0.72, 0.92) **
Excluding full-fat milk	-0.08 (-0.10, -0.07) ***	0.90 (0.82, 0.99) *	0.83 (0.72, 0.93) **
Excluding low-fat yoghurt	-0.08 (-0.10, -0.07) ***	0.89 (0.81, 0.98) *	0.82 (0.74, 0.92) **
Excluding full-fat yoghurt	-0.09 (-0.10, -0.09) ***	0.86 (0.78, 0.94) *	0.80 (0.70, 0.90) **
Excluding desserts/dairy with sugars	-0.08 (-0.10, -0.08) ***	0.90 (0.81, 0.99) *	0.83 (0.71, 0.94) **

Effect estimates are regression coefficients (β) for ln HOMA-IR or hazard ratios (HRs) for incidence of prediabetes or type 2 diabetes with their 95%-confidence intervals (95%CIs), per 10 units higher score on the plant-based dietary index by excluding one of 23 foods at a time

and additionally adjusting for the excluded food group. Estimates are adjusted for total energy, age, sex, Rotterdam Study sub-cohort, education, smoking status, family history diabetes, physical activity, and food supplement use (only for the HOMA-IR analyses, additionally adjusted for the time measurements of longitudinal HOMA-IR), based on pooled results of imputed data. *p<0.05; **p<0.01; ***p<0.001.

Chapter 2.4

Plant-based diet and obesity

Chen Z, Schoufour JD, Rivadeneira F, Lamballais S, Ikram MA, Franco OH, Voortman T. Plant-based diet and adiposity over time in a middle-aged and elderly population: the Rotterdam Study. Epidemiology. 2019;30(2):303-10.

ABSTRACT

Background/Aims: We explored whether the degree of adherence to a plant-based diet was associated with body mass index (BMI kg/m^2), waist circumference (cm), fat mass index (kg/m^2), and body fat percentage over time in a middle-aged and elderly population.

Methods: We included 9633 participants from the Rotterdam Study, a prospective cohort in the Netherlands. Dietary data were collected using food-frequency questionnaires at baseline of three sub-cohorts of the Rotterdam Study (1989-93, 2000-01, 2006-08). We created a plant-based diet index by giving plant-based foods positive scores and animal-based foods reverse scores. A higher score on the index reflected an overall more plant-based and less animal-based diet. Data on anthropometrics and body composition (using dual energy X-ray absorptiometry) were collected every 3-5 years from 1989-2016. We used multivariable linear mixed models to analyze the associations.

Results: In the 9633 participants, baseline plant-based diet score ranged from 21.0 to 73.0 with a mean ± SD of 49.0 ± 7.0. In multivariable-adjusted analyses, higher adherence to a plant-based diet was associated with lower BMI, waist circumference, fat mass index, and body fat percentage across a median follow-up period of 7.1 years (per 10 points higher score, BMI: β= -0.70 kg/m^2, (95% CI -0.81, -0.59); waist circumference: -2.0 cm (-2.3, -1.7); fat mass index: -0.66 kg/m^2 (-0.80, -0.52); body fat percentage: -1.1 (-1.3, -0.84)).

Conclusions: Higher adherence to plant-based diets beyond vegan or vegetarian diets may prevent obesity, irrespective of general healthfulness of the specific plant- and animal-based foods.

INTRODUCTION

Diet is an important modifiable lifestyle determinant of body adiposity. Several studies have indicated that plant-based diets may lower body mass index (BMI).[1-5] Potential mechanisms behind the link between plant-based diets with BMI may involve numerous biological pathways, such as changes in satiety,[6] inflammation,[7-10] and gut microbiome composition.[11] However, most of these studies classified participants dichotomously, and only defined plant-based diets as vegetarian or vegan versus non-vegetarian diets;[4] and few studies roughly classified plant-based diets as semi-vegetarian, lactovegetarian, and vegan diets, but they still did not address the gradual variation of plant-based diets.[1, 2, 5] Since the majority of the general populations do not follow strict vegan and vegetarian diets, and are more likely to adopt plant-based diets rich in plant-based foods and low in animal-based foods, from a clinical and public health point of view, it is interesting to question if and how the degree of adherence to an overall more plant-based and less animal-based diet influences body adiposity. Furthermore, the degree of adherence to an overall more plant-based and less animal-based diet can be assessed using a continuous plant-based diet score.[12, 13] Recent evidence has indicated that a plant-based diet score may represent a novel assessment of quality of dietary patterns, and reflect a complementary approach of what a healthful diet entails, different from the other diet quality scores, such as the Mediterranean Diet score, the Dietary Approaches to Stop Hypertension diet score, and diet quality scores on the basis of dietary guidelines.[13, 14] For example, Satija et al observed low or moderate correlations between a plant-based diet score with the Mediterranean Diet score and the Dietary Approaches to Stop Hypertension diet score.[13] We previously reported low correlation between a plant-based diet score with a diet quality score reflecting adherence to Dutch dietary guidelines.[14] These low or moderate correlations may be explained by the fact that some, but not all components are between the different scores and different scoring criteria. For example, the Mediterranean Diet score usually includes positive scores not only for healthy plant-based foods that appear beneficial for general health, such as whole grains, fruits, vegetables, legumes, nuts, and olive oil, but also for healthy animal-based foods, such as fish. A plant-based diet score, however, is allowed to include positive scores for plant-based foods and negative scores for animal-based foods irrespective of their known healthfulness.

Therefore, we aimed to examine the associations between the degree of adherence to plant-based diet assessed by a plant-based diet score with changes in measures of adiposity including BMI, waist circumference, fat mass index, and body fat percentage in a large Dutch middle-aged and elderly population with a median follow-up of 7.1 years (range 0-25 years).

METHODS

Study population

This study was embedded in three sub-cohorts of the Rotterdam Study (RS), a prospective cohort of adults living in the district of Ommoord in Rotterdam, the Netherlands. A detailed description of the Rotterdam Study methodology is described elsewhere.[15] Briefly, the first sub-cohort (RS-I) started in 1990 with participants aged ≥55 years (n=7983). The study was extended with a second sub-cohort (RS-II) in 2000 with new participants aged ≥55 years (n=3011), and a third sub-cohort (RS-III) in 2006 (n=3932), in which new participants aged ≥45 years were included. In each sub-cohort, follow-up examinations were performed in a research center every 3-5 years. The Rotterdam Study has been approved by the Medical Ethics Committee of Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. Informed consent was obtained from all participants.

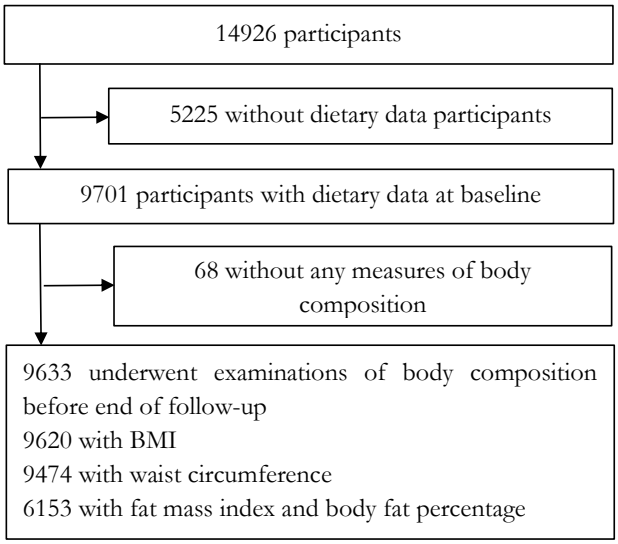


Figure 1. Participants selection

Population for current analyses

For our current analyses, of the 14926 participants from the three sub-cohorts combined, we excluded 5225 participants without valid dietary data (5141 without dietary data or unreliable dietary intake according to a trained nutritionist and 84 with an estimated energy intake of 500 or >5000 kcal/day), leaving 9701 participants with valid dietary data at baseline.[14] Of the 9701 participants, 9633 participants had at least one time measurement of body composition: resulting in 9620 for longitudinal BMI analyses, 9474 for longitudinal waist circumference analyses, and 6153 for longitudinal fat mass index and body fat percentage analyses. Figure 1 shows details of the participant selection.

Dietary assessment and plant-based diet index

Dietary intake was assessed at baseline of all sub-cohorts using a semi-quantitative food-frequency questionnaires (FFQ), as described in detail elsewhere.[16] Briefly, for RS-I (RS-I-1: 1989-93) and RS-II (RS-II-1: 2000-01) an FFQ with 170 food items was used;[17] and for RS-III (RS-III-1: 2006-08) an FFQ with 389 food items was used.[18] The validity of the questionnaires has been described previously.[16-18] Based on the dietary data, we constructed a plant-based diet index to assess variation in degree of adherence to a plant-based diet, which was a modified version of two previously created indices.[12, 13] First, the food items measured by FFQs were divided into 23 food groups (Supplemental Table 1) based on the Dutch food-based dietary guidelines, which were on the basis of similarities of the food items in (botanical) origin, nutrient composition. Of the 23 food groups, twelve food groups were plant-based (fruits, vegetables, whole grains, nuts, legumes, potatoes, vegetable oils, tea and coffee, sugary beverages, refined grains, sweets, alcoholic beverages), and eleven food groups were animal-based (low-fat milk, low-fat yoghurt, full-fat milk, full-fat yoghurt, cheese, fish, eggs, animal fat, unprocessed lean meat, processed and red meat, dessert and sugary dairy). For each food group, we divided the intake (gram) into cohort-specific quintiles. Each quintile was scored between 0 to 4. We gave plant-based foods positive scores. Consumption of plant-based foods within the highest quintile was scored a 4, consumption of plant-based foods within the second highest quintile was scored a 3, ending with consumption of plant-based food within the lowest quintile was scored a 0. By contrast, we gave animal-based foods reverse scores. Consumption of animal-based foods within the highest quintile was scored a 0, consumption of animal-based food within the second highest quintile was scored a 1, ending with consumption within the lowest quintile was scored a 4. Additionally, all participants with null consumption were given the score belonging to the lowest quintile by re-scoring when necessary. Finally, these category quintile-scores were added up for each participant to create a plant-based diet index, which measured degree of adherence to a plant-based diet on a continuous scale, with a lowest possible score of 0 (low adherence to a diet high in plant-based foods and low in animal-based foods) and a highest possible score of 92 (high adherence: high plant-based and low

animal-based). Further details of this index were described elsewhere.[14] Information on intake of each food group across quintiles of this plant-based diet score is shown in Supplemental Table 2.

Assessment of anthropometrics and body composition

Anthropometrics and body composition were repeatedly measured in our research center (Supplemental Table 3). Body weight was measured using a digital scale and body height was measured using a stadiometer, while participants wore light clothing and no shoes. BMI (kg/m^2) was calculated: Body weight (kg) / (Height (m) × Height (m)). We measured height and weight at six time points in RS-I (1989-2015); at four time points in RS-II (2000-16); and at two time points in RS-III (2006-14). Waist circumference (cm) was measured at the level midway between the lower rib margin and the iliac crest with the participants in a standing position. We measured waist circumference at five time points in RS-I (1989-2015), at four time points in RS-II (2000-16), and at two time points in RS-III (2006-14). Body fat and fat-free mass were measured with dual energy X-ray absorptiometry (DXA) Prodigy and iDXA devices (starting in 2002). Data for these outcomes were therefore available for the three time points in RS-I (2002-15) and RS-II (2004-16), respectively; and for two time points for RS-III (2006-14). From the DXA data we calculated adiposity outcomes: fat mass index (Fat mass (kg) / (Height (m) × Height (m)), and body fat percentage (fat mass (kg) / weight (kg)*100)). We also calculated fat-free mass index (Fat-free mass (kg) / (Height (m) × Height (m)).

Assessment of covariates

Information on smoking status and educational level was obtained during home interviews at baseline. Physical activity was assessed with an adapted version of the Zutphen Physical Activity Questionnaire at RS-I-3 and RS-II-1, and with the LASA Physical Activity Questionnaire at RS-III-1.[19] To account for differences between the two questionnaires, questionnaire-specific z-scores of metabolic equivalent of task-hours per week were calculated. Obesity was defined as BMI ≥ 30 kg/m^2.[20] Information of diabetes, coronary heart disease, and cancers were obtained from general practitioners, pharmacies' databases, Nationwide Medical Register, and follow-up examinations in our research center.[21-23]

Data analyses

We specified linear mixed models to analyze associations of the score on the plant-based diet index with adiposity outcomes over time. Likelihood ratio test, an objective model selection tool,[24] was used to determine random-effect structure and fixed-effect structure. We constructed 2 models with a fixed-effect structure that included the plant-based diet score and possible confounders and a random-effect structure including a random intercept and slope (for time of repeated measurements of adiposity outcomes). Non-linearity of associations of the score with outcomes using cubic splines

(degree of freedom = 3) were explored, as no indications for non-linear associations for the main models were found, all primary analyses were performed using models assuming linearity. The plant-based diet score was entered in models per 10 points higher score as 1 unit. Model 1 included plant-based diet score, baseline age, sex, total energy intake (kcal/day), RS sub-cohort, time of repeated measurements of BMI, waist circumference, fat mass index, or body fat percentage. Model 2 additionally included smoking status, education levels, physical activity, and food supplement use. The effect estimate for the plant-based diet score in the models indicates associations of the plant-based diet score with adiposity outcomes averaged across the median follow-up of 7.1 years. To explore whether an annual change in adiposity related to the plant-based diet score existed, i.e., whether the association between the plant-based diet score with adiposity differed across the follow-up time, a plant-based diet score × time interaction term was added to model 2 in a subsequent step.

We also conducted several additional analyses. First, we examined whether the associations differed by baseline age or sex by including interaction with baseline age or sex in model 2. Second, we repeated our main analyses by examining the index categorized into quintiles with the lowest quartile as reference. Last, we analyzed the associations with fat-free mass index based on model 2.

We performed sensitivity analyses based on model 2. First, we analyzed the associations with adiposity by excluding 'alcoholic beverages' from plant-based diet index. Second, to examine whether the associations of the plant-based diet with adiposity were independent of diet quality on the basis of dietary guidelines, we additionally adjusted for a diet quality score reflecting adherence to current Dutch dietary guidelines. Third, to examine whether our main results were robust after incorporating potential effect of dietary intake at follow-up, we further adjusted for plant-based diet score measured at RS-I-5 and RS-II-3 (20 years after RS-I baseline and 10 years after RS-II baseline) among participants with these data available. In this sensitivity analysis, we also adjusted for physical activity at RS-I-5 and RS-II-3. Fourth, to examine the individual contributions of healthy plant-based foods combined (fruits, vegetables, whole grains, legumes, nuts, vegetable oils, coffee and tea) and less healthy plant-based foods combined (sweets, sugary beverages, refined grains, potatoes) in the potential associations, we repeated our analyses by excluding these less healthy plant-based foods combined at a time, or these healthy plant-based foods combined at a time from the plant-based diet index and additionally adjusting for the excluded food groups. Fifth, we examined the association between a plant-based diet that is also high in healthy animal-based foods including fish, eggs, low-fat milk and low-fat yoghurt with adiposity. Sixth, we additionally adjusted for baseline health conditions including baseline diabetes, coronary heart disease, obesity, and cancers. Seventh, we excluded the participants with diabetes, coronary heart disease, obesity, or cancers at baseline, and further censored body composition data measured after onset of diabetes, coronary heart disease, and cancers during follow-up, and examined the associations. Last, we repeated our main analyses in three sub-cohorts, respectively.

All results were examined based on the combined data from RS-I, RS-II, and RS-III. All variables included in analyses were used to predict missingness patterns. Missing values on covariates (Supplemental Table 4) were assumed to be missing at random and accounted for using multiple imputations (m=10 imputations). We used SPSS version 21 (IBM Corp., Armonk, NY, USA) and R version 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria) to perform these analyses.

RESULTS

Baseline characteristics

Baseline characteristics of the study population are shown in Table 1. In 9633 participants, the baseline plant-based diet score (with a theoretical range from 0.0 to 92.0) ranged from 21.0 to 73.0, with a mean ± SD score of 49.0±7.0. Mean age of the study population at baseline was 64.2±8.7 years. Mean baseline BMI, waist circumference, fat mass index, and body fat percentage was 26.8±4.0 kg/m^2, 91.6±11.8 cm, 9.2±3.4 kg/m^2, and 34.2±8.4, respectively. Compared with the participants in the lowest quintile of the score, the participants in the highest quintile were older, more active, more highly educated, and less likely to smoke.

Repeated measurements of adiposity were performed during a median follow-up of 7.1 years (range 0-25 years) (Supplemental Table 3). Of the 9620 participants with BMI measurements, 8215 underwent at least two examinations of BMI; of the 9474 participants with waist circumference measurements, 6196 underwent at least two examinations of waist circumference; and of 6153 participants with fat mass index and body fat percentage measurements, 3806 underwent at the least two examinations of fat mass index and body fat percentage.

Degree of adherence to a plant-based diet and adiposity

After multivariable adjustment, more adherence to a plant-based diet was associated with lower BMI, waist circumference, fat mass index, and body fat percentage averaged across the median follow-up of 7.1 years (per 10 points higher score, BMI: β= -0.70, 95% CI: -0.81, -0.59; waist circumference: β= -2.0, 95% CI: -2.3, -1.7; fat mass index: β= -0.66, 95% CI: -0.80, -0.52; body fat percentage: β= -1.1, 95% CI: -1.3, -0.84) (Table 2). Interactions of the plant-based diet score with time were not statistically significant for any of the outcomes (Table 2), indicating that no change in the strength of associations with BMI, fat mass index, waist circumference, and body fat percentage across the follow-up period. Therefore, our models estimated that for participants having a 10 points higher score on the plant-based diet index, their mean BMI, waist circumference, fat mass index, and body fat percentage were 0.70 kg/m^2 lower, 2.0 cm lower, 0.66 kg/m^2 lower, and 1.1 lower across the median follow-up of 7.1 years.

Table 1. Baseline characteristics of population

Characteristics	(n=9633)	Quintile 1[b] (n=2212)	Quintile 5[b] (n=1635)
Age (years)	64.2 (8.7)	65.7 (9.0)	62.8 (8.0)
Sex, male (%)	42.1	32.1	56.3
Physical activity (MET-hours/week)			
- Zutphen Physical ActivityQuestionnaire	76.6 (51.5, 108.7)	72.7 (47.6, 106.0)	76.2 (52.6, 106.0)
- LASA Physical Activity Questionnaire	42.3 (17.7, 82.0)	34.6 (14.0, 74.1)	48.0 (20.5, 91.5)
Education level (%)			
- Higher	15.4	18.0	12.4
- Intermediate	27.8	42.3	39.1
- Lower	40.8	26.5	27.5
- Primary	15.5	12.8	19.8
Smoking (%)			
- Never	31.5	35.8	28.7
- Ever	44.0	38.6	48.9
- Current	24.3	25.2	21.5
BMI (kg/m^2)	26.8 (4.0)	27.5 (4.4)	26.0 (3.5)
Waist circumference (cm)	91.6 (11.8)	92.6 (12.2)	90.7 (11.0)
Fat mass index(kg/m^2)[a]	9.2 (3.4)	10.6 (3.8)	8.2 (2.9)
Body fat percentage (%)[a]	34.2 (8.4)	36.5 (8.3)	31.1 (8.4)
Plant-based diet score	49.0 (7.0)	40.0 (37.0, 42.0)	59.0 (57.0, 61.0)
Foods groups intake (grams/day)			
- Fruits	203.2 (145.5, 275.1)	178.7 (127.4, 244.2)	234.4 (178.1, 313)
- Vegetables	109.6 (68.2, 154.0)	93.4 (56.0, 129.9)	138.8 (80.9, 188.7)
- Whole grains	109.6 (68.2, 154.0)	93.4 (56.0, 129.9)	138.8 (80.9, 188.7)
- Legumes	0.0 (0.0, 17.1)	0.0 (0.0, 7.8)	10.7 (0.0, 31.0)

Table 1. Baseline characteristics of population (Continued)

Characteristics	(n=9633)	Quintile 1[b] (n=2212)	Quintile 5[b] (n=1635)
- Nuts	2.7 (0.0, 10.8)	0.0 (0.0, 4.4)	8.5 (2.2, 18.1)
- Tea and coffee	745 (594, 1000)	750 (500, 875)	933 (750, 1125)
- Vegetable oils	20.5 (9.7, 31.0)	12.0 (3.0, 22.0)	28.7 (19.6, 39.7)
- Refined grains	45.0 (20.5, 92.9)	33.5 (14.1, 70.0)	58.4 (30.0, 115.1)
- Potatoes	106.9 (68.4, 151.0)	85.5 (56.6, 128.2)	128.2 (87.3, 178.1)
- Sweets	63.4 (36.3, 98.5)	48.3 (26.1, 81.3)	82.9 (53.6, 118.3)
- Sugary beverages	39.9 (0.0, 139.6)	9.2 (0.0, 79.8)	87.3 (15.0, 174.5)
- Alcoholic beverages	43.2 (2.7, 143.1)	21.4 (0.3, 106.3)	71.4 (10.8, 176.4)
- Low-fat yoghurt	53.8 (0.0, 164.6)	74.8 (0.0, 192.9)	21.4 (0.0, 149.6)
- Full-fat yoghurt	0.0 (0.0, 4.4)	0.0 (0.0, 37.4)	0.0 (0.0, 0.0)
- Low-fat milk	82.3 (0.0, 240.4)	107.1 (1.6, 278.6)	(0.0, 196.5)
- Full-fat milk	0 (0.0, 0.0)	0.0 (0.0, 23.5)	0.0 (0.0, 0.0)
- Cheese	31.3 (20.0, 47.0)	33.4 (21.4, 48.1)	29.3 (18.6, 46.2)
- Fish	14.3 (2.1, 29.2)	18.7 (5.9, 32.1)	8.9 (0.0, 23.9)
- Eggs	14.3 (7.1, 17.9)	14.3 (8.6, 21.4)	10.1 (7.1, 16.6)
- Animal fat	0.0 (0.0, 1.2)	0.7 (0.0, 14.7)	0.0 (0.0, 0.0)
- Desserts/dairy with sugars	13.0 (0.0, 54.9)	20.5 (0.0, 63.9)	5.4 (0.0, 41.0)
- Unprocessed lea meat	10.7 (4.1, 17.9)	14.3 (5.4, 21.4)	7.1 (0.0, 14.9)
- Processed meat/red meat	87.6 (61.7, 120.0)	92.5 (65.5, 126.5)	81.6 (55.0, 114.8)

Values are percentages for categorical variables, mean (SD) for continuous variables with a normal distribution, and median (25th percentile–75th percentile) for continuous variables with a skew distribution; on the basis of unimputed data.

[a]Data on baseline fat mass index, and body fat percentage are from participants in RS-III

[b]Quintile 1: Score ≤43; Quintile 5: score >55

Abbreviations: MET, metabolic equivalent of task; BMI, body mass index; RS-III, the third sub-cohort; SD, standard deviation; NA, not available.

Table 2. Associations between plant-based diet index with longitudinal BMI, waist circumference, fat mass index, and body fat percentage

	BMI (kg/m^2) n=9620 β (95% CI)	Waist circumference (cm) n=9474 β (95% CI)	Fat mass index (kg/m^2) n=6153 β (95% CI)	Body fat percentage n=6153 β (95% CI)
Model 1				
Plant-based diet score[a]	-0.68 (-0.79, -0.57)	-2.0 (-2.3, -1.7)	-0.64 (-0.76, -0.51)	-1.1 (-1.3, -0.84)
Plant-based diet score × time[b]	-0.003 (-0.009, 0.003)	-0.02 (-0.04, 0.01)	0.01 (-0.002, 0.02)	-0.01 (-0.03, 0.01)
Model 2				
Plant-based diet score[a]	-0.70 (-0.81, -0.59)	-2.0 (-2.3, -1.7)	-0.66 (-0.80, -0.52)	-1.1 (-1.3, -0.84)
Plant-based diet score × time[b]	-0.004 (-0.01, 0.003)	-0.02 (-0.04, 0.01)	0.01 (-0.003, 0.02)	-0.01 (-0.03, 0.01)

Values are regression coefficients and 95%CIs based on linear mixed models, and reflect differences in BMI, waist circumference, fat mass index, and body fat percentage averaged across a median follow-up of 7.6 years per 10 points higher score.

Model 1 is adjusted for baseline age (years), sex, total energy (kcal), RS sub-cohort (RS-I, -II, or -III), time of repeated measurements of longitudinal BMI, waist circumference, fat mass index, or body fat percentage.

Model 2 is additionally adjusted for education level (primary, lower, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task--hours/week), food supplement use (yes or no).

[a]Estimates are from the models without interactions of plant-based diet and time.

[b]All interactions of plant-based diet score with time are not statistically significant, indicating no change in the strength of associations with BMI, waist circumference, fat mass index, or body fat percentage averaged across the median follow-up of 7.1 years; furthermore, addition of the interaction term of plant-based score with time to the models does not improve model fit.

Abbreviations: BMI, body mass index; RS, Rotterdam Study.

Table 3. Associations of quintiles of plant-based diet index with BMI, waist circumference, fat mass index, and body fat percentage

Multivariate adjustment[a]	BMI (kg/m^2) n=9620 β (95% CI)	Waist circumference (cm) n=9474 β (95% CI)	Fat mass index (kg/m^2) n=6153 β (95% CI)	Body fat percentage n=6153 β (95% CI)
Quintile 1 (Score ≤43, median =40)	Reference	Reference	Reference	Reference
Quintile 2 (43< score ≤47, median=46)	-0.51 (-0.77, -0.26)	-1.5 (-2.1, -0.83)	-0.59 (-0.89, -0.29)	-0.94 (-1.6, -0.28)
Quintile 3 (47< score ≤50, median=49)	-0.57 (-0.84, -0.30)	-2.0 (-2.7, -1.3)	-0.83 (-1.1, -0.52)	-1.7 (-2.3, -0.98)
Quintile 4 (50< score ≤55, median =53)	-0.81 (-1.1, -0.55)	-2.7 (-3.4, -2.1)	-1.0 (-1.3, -0.76)	-2.0 (-2.6, -1.3)
Quintile 5 (score >55, median =59)	-1.3 (-1.6, -0.99)	-4.1 (-4.8, -3.3)	-1.5 (-1.8, -1.2)	-2.9 (-3.6, -0.25)
P for trend[b]	P<0.0001	P<0.0001	P<0.0001	P<0.0001

Values are β coefficients and 95%CIs, and reflect the differences in BMI, waist circumference, fat mass index, and body fat percentage for quintiles of the plant-based dietary index compared to the lowest quartile, indicating that higher score on plant-based diet index is associated with lower BMI, waist circumference, fat mass index, and body fat percentage averaged across the median follow-up of 7.1 years.

[a]Models are adjusted for baseline age (years), sex, total energy (kcal), RS sub-cohort (RS-I, -II, or -III), time of repeated measurements of longitudinal BMI, waist circumference, fat mass index, and body fat percentage, education level (primary, lower, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task-hours/week), and food supplement use (yes or no).

[b]Trend tests across quintiles of plant-based diet score are performed by assigning median values for these quintiles and treating the variables as continuous terms in the models.

Abbreviations: BMI, body mass index; RS, Rotterdam Study.

Additional results

The interaction of the index with baseline age or sex was not statistically significant (Supplemental Table 5). However, we observed that compared to the participants in the lowest quintile, those in the highest quintile had a lower BMI, waist circumference, fat mass index, and body fat percentage (Table 3). Furthermore, we observed that more adherence to a plant-based diet was associated with slightly lower fat-free mass index based on model 2 (per 10 points higher score, fat-free mass index: -0.16 (-0.21, -0.11) averaged across the median follow-up of 7.1 years).

Sensitivity analyses results

Exclusion of alcoholic beverages from plant-based diet index did not substantially change our estimates (Supplemental Table 6). The estimates were similar after additional adjustment for dietary intake and physical activity at RS-I-5 and RS-II-3 or for baseline diet quality reflecting adherence to dietary guidelines (Supplemental Table 7). Exclusion of the less healthy plant-based foods combined from the plant-based diet index did not substantially change the estimates; while exclusion of the healthy plant-based foods combined from the plant-based diet index moderately attenuated the inverse associations (Supplemental Table 8). The associations were also moderately attenuated by giving fish, eggs, low-fat milk, and low-yoghurt positive scores (Supplemental Table 9). Adjustment for baseline health conditions or exclusion of participants with obesity, diabetes, coronary heart disease, or cancers at baseline and censoring body composition data collected after onset of these diseases did not substantially affect our findings (Supplemental Table 10). The estimates were similar in the three sub-cohorts (Supplemental Table 11).

DISCUSSION

In the present study, we observed that higher adherence to a plant-based diet was associated with lower adiposity status averaged across the median follow-up of 7.1 years, and the inverse associations with adiposity remained stable over time. Our results were in line with the results from previous studies reporting reverse associations of vegetarian or vegan diets with BMI.[1-5] More importantly, we extended this evidence by showing associations of adherence to a plant-based diet beyond vegetarian or vegan diets irrespective of general healthfulness of the specific plant- and animal-based foods, and by presenting that this is not only associated with BMI, but with detailed measures of adiposity over time.

However, we acknowledge that our plant-based diet included less healthy plant-based foods (sweets, potatoes, refined grains, sugary beverages). That is because we took into account the fact that most of populations are not likely to completely avoid less healthy plant-based foods intake in real life. To further clarify the individual contributions of these healthy plant-based foods combined and less

healthy plants-based foods combined to the inverse associations with adiposity, we examined the associations of the plant-based diet score by excluding less healthy plant-based foods combined or healthy plant-based foods combined from the index in sensitivity analyses. We observed that a higher plant-based diet score remained strongly associated with less adiposity irrespective of exclusion of healthy plant-based foods or less healthy plant-based foods, although exclusion of the healthy plant-based foods moderately attenuated the associations. This indicates that the beneficial associations of the plant-based diet score were contributed to by both of substitution of the healthy plant-based foods and the less plant-based foods for animal-based foods, although substitution of the healthy plant-based foods for animal-based foods appeared to contribute more. Our findings suggest a beneficial effect of an overall plant-based diet on adiposity, irrespective of general healthfulness of the specific plant- and animal-based foods, which increase potentials of recommendation for population. Our findings also suggest that healthy plant-based foods may contribute more to the beneficial effect, which emphasizes that it is important to also consider the quality of plant-based foods consumed.

Potential mechanisms underlying the inverse association with adiposity

The inverse associations of a plant-based diet with adiposity could be partly explained by more intake of certain components of plant-based foods.[25] A diet high in plant-based foods usually contains more fiber, chlorogenic acids, antioxidants, plant protein and plant unsaturated fatty acids. For example, vegetables and fruits are the main sources of fiber, antioxidants, and chlorogenic acids; nuts are rich in poly-unsaturated fatty acids; soy and pulses are main sources of plant protein; and coffee and tea are rich in antioxidants and phenol chlorogenic acid. These components have been suggested to reduce adiposity through different pathways and intermediate conditions, such as satiety,[6] inflammation,[7-10] oxidative stress,[10] and gut microbiome composition.[11] Lower intake of certain components of animal-based foods also may explain our findings. A diet low in animal-based foods contains less animal protein and saturated fatty acids. Lower intake of these components has been suggested to be beneficial for prevention of obesity.[9, 26]

Important implication

Our findings have important public health implications. In our study, based on the comparison of food components in lowest quintile of the score as reference and highest quintile of the score that was associated with lower adiposity status, we observed that a beneficial plant-based diet for improvement of adiposity does not require a total elimination of meat or animal products, but instead can be achieved by a moderate decrease in animal-based foods intake, and a moderate increase in plant-based foods intake, increasing the potential for population-wide health recommendations. For example, we observed that the participants in the highest quintile of the score might have an averagely 4.1 cm lower waist circumference and 1.3 kg/m^2 lower BMI across the median follow-up of 7.1 years, compared

with those in the lowest quintile of the score, yet the participants in the highest quintile had a median red meat consumption of 81.6 g per day and a median vegetables consumption of 234.4 g per day, relative to a median red meat consumption of 92.5 g per day and a median vegetable consumption of 178.7 g per day of the participants in the lowest quintile .

Study strengths and limitations

Our study has several strengths. First, to our knowledge, this is the first study to examine associations of degree of adherence to a plant-based diet with adiposity over time, for which we had longitudinal detailed data on adiposity including BMI, waist circumference, fat mass index, and body fat percentage. Second, to assess the gradations of adherence to a plant-based diet, we used a novel plant-based diet score. Previous studies have indicated low and moderate correlations between the novel plant-based diet score with other known diet scores, such as the Mediterranean Diet score. Furthermore, our results showed beneficial associations of an overall plant-based diet score with adiposity, independent of adherence to current dietary guidelines, which indicated that our plant-based diet score might reflect another distinguishing aspect of a healthful diet more than solely assessment of diet quality according to current guidelines. Last, our study highlights that higher adherence to plant-based diets beyond vegan or vegetarian diets may help prevent obesity, irrespective of general healthfulness of the specific plant- and animal-based foods, which increase the potential for population-wide health recommendations.

However, we also acknowledge some limitations. First, dietary information was derived from self-report, measurement-errors was possible. However, because the FFQs used in our cohort were shown in several validation studies to adequately rank subjects according to food and nutrient intake,[17, 18] and because we used relative quintiles of foods intake (gram) to create the score, we do not expect these measurement-errors to have largely affected the index. Second, we only used baseline measurement of dietary intake in main analyses, whereas diet could change over time, and repeated measurements of diet over time would be preferable. We also only adjusted for baseline covariates, instead of time-varying covariates in main analyses, whereas, these covariates were not necessarily constant through the follow-up. However, we explored the potential effect of dietary intake and physical activity at follow-up on the associations in a subgroup of participants with these data available and observed similar results. Furthermore, after excluding the participants who were likely to change their diet and lifestyle at follow-up, such as participants with diabetes and cancers at baseline, the estimates were still similar. Combined, these results indicate that our findings were robust. Third, many of the participants of the original cohort were excluded due to report of invalid dietary information, which might have led to selection bias if associations of plant-based diet with adiposity differed in those included and those not included in our current analyses. Fourth, we used two different FFQs to measure dietary intake and two different physical activity questionnaires to measure physical activity level at different

sub-cohorts. However, we do not expect the use of different questionnaires to considerably influence our findings, since the associations were similar in the three RS sub-cohorts. Fifth, we noticed that more adherence to a plant-based diet was associated with slightly lower fat-free mass index in additional analyses, which indicated that more adherence to a plant-based diet might not be beneficial for prevention of fat-free mass loss. However, the inverse association of a plant-based diet was much stronger for fat mass than for fat-free mass, suggesting overall beneficial effect on adiposity, which was also reflected by the lower body fat percentage. Finally, our results may not be generalizable to people of other race and age, therefore replication in other populations is warranted.

Conclusions

A diet higher in plant-based foods and lower in animal-based foods beyond strict vegan or vegetarian diet, was associated with lower adiposity status over time, irrespective of healthfulness of specific plant- and animal- based foods.

REFERENCES

1. Tonstad S, Butler T, Yan R, Fraser GE. Type of vegetarian diet, body weight, and prevalence of type 2 diabetes. Diabetes care. 2009;32(5):791-6.
2. Newby P, Tucker KL, Wolk A. Risk of overweight and obesity among semivegetarian, lactovegetarian, and vegan women. The American journal of clinical nutrition. 2005;81(6):1267-74.
3. Turner-McGrievy G, Mandes T, Crimarco A. A plant-based diet for overweight and obesity prevention and treatment. Journal of Geriatric Cardiology : JGC. 2017;14(5):369-74.
4. Huang R-Y, Huang C-C, Hu FB, Chavarro JE. Vegetarian diets and weight reduction: A meta-analysis of randomized controlled trials. Journal of general internal medicine. 2016;31(1):109-16.
5. Rosell M, Appleby P, Spencer E, Key T. Weight gain over 5 years in 21 966 meat-eating, fish-eating, vegetarian, and vegan men and women in EPIC-Oxford. International journal of obesity. 2006;30(9):1389.
6. Westerterp-Plantenga MS, Nieuwenhuizen A, Tome D, Soenen S, Westerterp KR. Dietary protein, weight loss, and weight maintenance. Annual review of nutrition. 2009;29:21-41.
7. Watzl B, Kulling SE, Möseneder J, Barth SW, Bub A. A 4-wk intervention with high intake of carotenoid-rich vegetables and fruit reduces plasma C-reactive protein in healthy, nonsmoking men. The American journal of clinical nutrition. 2005;82(5):1052-8.
8. Butcher JL, Beckstrand RL. Fiber's impact on high-sensitivity C-reactive protein levels in cardiovascular disease. Journal of the American Association of Nurse Practitioners. 2010;22(11):566-72.
9. Bray GA, Lovejoy JC, Smith SR, DeLany JP, Lefevre M, Hwang D, et al. The influence of different fats and fatty acids on obesity, insulin resistance and inflammation. The Journal of nutrition. 2002;132(9):2488-91.
10. Liang N, Kitts DD. Role of Chlorogenic Acids in Controlling Oxidative and Inflammatory Stress Conditions. Nutrients. 2016;8(1):16.
11. Holscher HD, Caporaso JG, Hooda S, Brulc JM, Fahey GC, Swanson KS. Fiber supplementation influences phylogenetic structure and functional capacity of the human intestinal microbiome: follow-up of a randomized controlled trial. The American journal of clinical nutrition. 2015;101(1):55-64.
12. 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(Supplement 1):320S-8S.
13. 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 Med. 2016;13(6):e1002039.
14. Chen Z, Zuurmond MG, van der Schaft N, Nano J, Wijnhoven HAH, Ikram MA, et al. Plant versus animal based diets and insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. Eur J Epidemiol. 2018.

15. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. European Journal of Epidemiology. 2017;32(9):807-50.
16. Voortman T, Kiefte-de Jong JC, Ikram MA, Stricker BH, van Rooij FJA, Lahousse L, et al. Adherence to the 2015 Dutch dietary guidelines and risk of non-communicable diseases and mortality in the Rotterdam Study. European Journal of Epidemiology. 2017:1-13.
17. Klipstein-Grobusch K, den Breeijen JH, Goldbohm RA, Geleijnse JM, Hofman A, Grobbee DE, et al. Dietary assessment in the elderly: validation of a semiquantitative food frequency questionnaire. Eur J Clin Nutr. 1998;52(8):588-96.
18. Feunekes GI, Van Staveren WA, De Vries J, Burema J, Hautvast J. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. The American journal of clinical nutrition. 1993;58(4):489-96.
19. Koolhaas CM, Dhana K, Schoufour JD, Ikram MA, Kavousi M, Franco OH. Impact of physical activity on the association of overweight and obesity with cardiovascular disease: The Rotterdam Study. European journal of preventive cardiology. 2017;24(9):934-41.
20. Ravussin E, Swinburn BA. Pathophysiology of obesity. The Lancet (USA). 1992.
21. Leening MJG, Kavousi M, Heeringa J, van Rooij FJA, Verkroost-van Heemst J, Deckers JW, et al. Methods of data collection and definitions of cardiac outcomes in the Rotterdam Study. European journal of epidemiology. 2012;27(3):173-85.
22. Khan SR, Chaker L, Ruiter R, Aerts JGJV, Hofman A, Dehghan A, et al. Thyroid Function and Cancer Risk: The Rotterdam Study. The Journal of Clinical Endocrinology & Metabolism. 2016;101(12):5030-6.
23. Ligthart S, van Herpt TT, Leening MJ, Kavousi M, Hofman A, Stricker BH, et al. Lifetime risk of developing impaired glucose metabolism and eventual progression from prediabetes to type 2 diabetes: a prospective cohort study. The Lancet Diabetes & Endocrinology. 2016;4(1):44-51.
24. Akaike H. Information theory and an extension of the maximum likelihood principle. Selected papers of hirotugu akaike: Springer; 1998. p. 199-213.
25. Satija A, Hu FB. Plant-based diets and cardiovascular health. Trends in Cardiovascular Medicine. 2018.
26. Van Dam RM, Willett WC, Rimm EB, Stampfer MJ, Hu FB. Dietary fat and meat intake in relation to risk of type 2 diabetes in men. Diabetes care. 2002;25(3):417-24.

SUPPLEMENTAL MATERIAL

Supplemental Table 1. 23 Food groups used for the plant-based diet index and examples of food items included in each of the food groups

12 Plant-based food groups	
Fruits	Apple, banana, pear, orange, strawberry, grapes, other fruits
Vegetables	Cauliflower, broccoli, spinach, carrots, onion, lettuce, tomato, cabbage, cooked vegetables
Whole grains	Whole grain bread, dark bread, rye bread, whole grain breakfast oats, whole grain pasta, brown rice
Nuts	Peanuts, walnuts, other nuts, peanut butter
Legumes	Legumes, tofu, soybeans, other soy products
Vegetable oils	Olive oil, vegetable oils used for cooking, and all margarines
Tea and coffee	Black tea, green tea, herbal tea, coffee
Potatoes	Potatoes, fries
Refined grains	Cornflakes, white bread, croissants, raisin bread, white pasta, white rice
Sweets	Sugar, cookies, cake, chocolate, candy-bars, honey, sweets, chocolate toppings, other sweet toppings
Sugary beverages	Carbonated beverages with sugar, non-carbonated beverages with sugar, orange juice, fruit juice
Alcoholic beverages	Red wine, white wine, beer, liquor, Dutch-eggnog
11 Animal-based food groups	
Low-fat milk	Skimmed milk, semi-skimmed milk, skimmed coffee creamer, semi-skimmed coffee creamer
Low-fat yoghurt	Skimmed yoghurt, semi-skimmed yoghurt, skimmed quark, buttermilk
Full-fat milk	Full-fat milk, cream, coffee-cream
Full-fat yoghurt	Full-fat yoghurt, semi-skimmed quark, full quark
Cheese	Full fat cheese, low fat cheese, cheese fondue, other cheese
Desserts and sugary dairy	Custard, cream, ice cream, mousse, cream, chocolate milk, fruit yoghurt, yoghurt drinks
Unprocessed lean meat	Chicken
Fish	Salmon, tuna, trout, herring, mussels, other fish
Eggs	Boiled eggs, fried eggs
Animal fat	Butter on bread, butter used for cooking, lard
Processed and red meat	Beef, pork, meatballs, sate, bacon, liver, processed meats

Supplemental Table 2. Intake of 23 food groups across quintiles of the plant-based diet index

Plant-based diet index	Quintile 1 Score≤43 Median=40 n=2212	Quintile 2 43<score≤47 Median=46 n=1976	Quintile 3 47<score ≤50 Median=49 n=1639	Quintile 4 50<score≤55 Median=53 n=2171	Quintile 5 score>55 Median=59 n=1635
Food intake (grams/day)[a]					
Fruits	170.6 (91.2, 266.2)	195.0 (109.6, 299.5)	213.7 (122.4, 322.7)	220.4 (124.5, 329.2)	249.6 (150.6, 375.6)
Vegetables	178.7 (127.4, 244.2)	196.4 (141.4, 268.3)	201.8 (146.7, 272.3)	210.7 (151.5, 283.9)	234.4 (178.1, 313)
Whole grains	93.4 (56.0, 129.9)	102.4 (57.0, 142.3)	110.0 (70.0, 151.6)	116.1 (70.1, 162.5)	138.8 (80.9, 188.7)
Legumes	0.0 (0.0, 7.8)	0.0 (0.0, 12.9)	0.0 (0.0, 15.8)	4.5 (0.0, 19.4)	10.7 (0.0, 31.0)
Nuts	0.0 (0.0, 4.4)	1.1 (0.0, 8.1)	2.4 (0.0, 9.6)	4.3 (0.0, 12.8)	8.5 (2.2, 18.1)
Vegetable oils	12.0 (3.0, 22.0)	18.0 (8.0, 27.4)	21.0 (11.5, 30.7)	24.5 (14.6, 34.1)	28.7 (19.6, 39.7)
Tea and coffee	750.0 (500.0, 875.0)	750.0 (550.0,1000.0)	754.5 (593.8, 1000.0)	825.0 (625.0,1025.0)	933.1 (750.0,1125.0)
Refined grains	33.5 (14.1, 70.0)	42.3 (20.0, 86.2)	42.7 (20.4, 88.3)	52.2 (26.2, 104.9)	58.4 (30.0, 115.1)
Potatoes	85.5 (56.6, 128.2)	96.4 (64.1, 142.5)	106.9 (71.2, 149.6)	114.3 (73.3, 170.0)	128.2 (87.3, 178.1)
Sweets	48.3 (26.1, 81.3)	57.1 (31.8, 89.1)	63.9 (39.0, 96.7)	71.2 (42.3, 105.8)	82.9 (53.6, 118.3)
Sugary beverages	9.2 (0.0, 79.8)	26.8 (0.0, 139.3)	50.0 (0.0, 139.6)	53.8 (0.0, 142.5)	87.3 (15.0, 174.5)
Alcoholic beverages	21.4 (0.3, 106.3)	37.5 (1.6, 134.9)	42.7 (3.3, 147.0)	56.9 (4.3, 149.6)	71.4 (10.8, 176.4)
Low-fat yoghurt	74.8 (0.0, 192.9)	60.2 (0.0, 166.1)	64.1 (0.0, 164.6)	53.6 (0.0, 158.6)	21.4 (0.0, 149.6)
Full-fat yoghurt	0.0 (0.0, 37.4)	0.0 (0.0, 9.9)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
Low-fat milk	107.1 (1.6, 278.6)	100.2 (2.2, 264.4)	94.6 (0.0, 231.7)	79.0 (0.0, 224.4)	42.9 (0.0, 196.5)
Full-fat milk	0.0 (0.0, 23.5)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
Cheese	33.4 (21.4, 48.1)	32.4 (20.4, 48.4)	30.8 (20.0; 46.6)	29.6 (19.9, 45.5)	29.3 (18.6, 46.2)
Fish	18.7 (5.9, 32.1)	15.9 (3.3, 31.5)	13.7 (2.1, 28.5)	12.8 (1.0, 28.5)	8.9 (0.0, 23.9)
Eggs	14.3 (8.6, 21.4)	14.3 (7.1, 21.4)	14.3 (7.1, 17.8)	14.3 (7.1, 17.1)	10.1 (7.1, 16.6)

Supplemental Table 2. Intake of 23 food groups across quintiles of the plant-based diet index (Continued)

Plant-based diet index	Quintile 1 Score≤43 Median=40 n=2212	Quintile 2 43<score≤47 Median=46 n=1976	Quintile 3 47<score ≤50 Median=49 n=1639	Quintile 4 50<score≤55 Median=53 n=2171	Quintile 5 score>55 Median=59 n=1635
Animal fat	0.7 (0.0, 14.7)	0.0 (0.0, 3.0)	0.0 (0.0, 0.3)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
Desserts/dairy with sugars	20.5 (0.0, 63.9)	17.1 (0.0, 61.6)	13.8 (0.0, 52.3)	10.3 (0.0, 51.7)	5.4 (0.0, 41.0)
Unprocessed lean meat	14.3 (5.4, 21.4)	12.4 (5.4, 21.4)	10.7 (3.9, 17.9)	10.0 (3.7, 16.4)	7.1 (0.0, 14.9)
Red/processed meat	92.5 (65.5, 126.5)	88.7 (64.6, 118.3)	88.3 (62.0, 120.4)	85.7 (60.1, 116.6)	81.6 (55.0, 114.8)

[a]Values are median (25th percentile–75th percentile) for continuous variables with a skew distribution.

Supplemental Table 3. Repeated measurements of BMI, waist circumference, fat mass index, and body fat percentage

		Time[a] (years)	BMI	Waist circumference	Time[b] (years)	Fat mass index	Body fat percentage
RS-I	RS-I-1 (1989-93)	0.0 (0.0)	5397	5083	NA	NA	NA
n=5416	RS-I-2 (1993-95)	2.0 (0.6)	4483	NA	NA	NA	NA
	RS-I-3 (1997-99)	6.5 (0.4)	3512	3517	NA	NA	NA
	RS-I-4 (2002-04)	11.0 (0.5)	2540	2652	11.6 (0.5)	2329	2329
	RS-I-5 (2009-11)	17.6 (0.5)	1477	1473	17.2 (0.5)	1297	1297
	RS-I-6 (2014-15)	22.5 (0.7)	741	741	22.1 (0.3)	698	698
RS-II	RS-II-1 (2000-01)	0.0 (0.0)	1620	1623	NA	NA	NA
n=1624	RS-II-2 (2004-05)	4.0 (0.5)	1319	1339	4.0 (0.4)	659	659
	RS-II-3 (2011-12)	10.6 (0.4)	1033	1035	10.8 (0.3)	993	993
	RS-II-4 (2015-16)	15.0 (0.2)	797	820	15.0 (0.2)	797	797
RS-III	RS-III-1 (2006-08)	0.0 (0.0)	2561	2481	0.0 (0.0)	2079	2079
n=2581	RS-III-2 (2012-14)	5.6 (0.4)	2142	2139	5.8 (0.4)	2059	2059

At baseline and follow-up in RS-I, RS-II, and RS-III, 9620 participants had at least one time measurement of BMI; 9474 participants had at least one time measurement of waist circumference; 6153 participants had at least one time measurement of fat mass index, and body fat percentage.

[a]Time variable (Mean (SD)) is calculated as time scale for each measurement of BMI and waist circumference by subtracting baseline date from date of each measurement of BMI and waist circumference at follow-up period.

[b]Time variable (Mean (SD)) is calculated as time scale for each measurement of fat mass index and body fat percentage by subtracting baseline date from date of each measurement of fat mass index and body fat percentage at follow-up period.

Abbreviations: BMI, body mass index; RS, Rotterdam Study; NA, not available

Supplemental Table 4. Missing values[a]

	Physical activity	Education level	Smoking status
RS-I n=5426	NA=1551 (28.6%)	NA=26 (0.48%)	NA=32 (0.59%)
RS-II n=1624	NA=11 (0.68%)	NA=23 (1.4%)	NA=8 (0.49%)
RS-III n=2583	NA=238 (9.2%)	NA=7 (0.27%)	NA=6 (0.23%)
Total n=9633	NA=1800 (18.7%)	NA=56 (0.58%)	NA=46 (0.47%)

[a]In our main analyses, only three variables: physical activity, education level, and smoking status were with missing values.
Abbreviations: NA, not available, (numbers of participants with missing values); RS, Rotterdam Study

Supplemental Table 5. Interactions between plant-based diet score and baseline age, or sex for longitudinal BMI, waist circumference, fat mass index, and body fat percentage

	BMI (kg/m^2) n=9620	Waist circumference (cm) n=9474	Fat mass index (kg/m^2) n=6153	Body fat percentage n=6153
	β (95% CI), p value	β (95% CI), p value	β (95% CI), p value	β (95% CI), p value
Plant-based diet score × baseline age	0.01 (-0.01, 0.02), p=0.32	0.03 (-0.01, 0.07), p=0.11	0.02 (-0.01, 0.05), p=0.22	0.03 (-0.03, 0.10) p=0.31
Plant-based diet score × sex	-0.13 (-0.46, 0.20), P=0.45	-0.32 (-0.92, 0.27), p=0.29	-0.21 (-0.57, 0.15), p=0.35	-0.09 (-0.51, 0.34) p=0.69

Values are regression coefficients and 95%CIs for interactions between plant-based diet score and baseline age or sex for body adiposity outcomes.
Models include plant-based diet score, age (years), sex, total energy (kcal/day), RS sub-cohort (RS-I, -II, or -III), time of repeated measurements of BMI, waist circumference, fat mass index, and body fat percentage education level (primary, lower, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task -hours/week), and food supplement use (yes or no), and interaction between plant-based diet score and baseline age or sex.
Abbreviations: BMI, body mass index; RS, Rotterdam Study

Supplemental Table 6. Associations of plant-based diet score with BMI, waist circumference, fat mass index, and body fat percentage by excluding alcoholic beverages

	BMI (kg/m^2) n=9620 β (95% CI)	Waist circumference (cm) n=9474 β (95% CI)	Fat mass index (kg/m^2) n=6153 β (95% CI)	Body fat percentage n=6153 β (95% CI)
Excluding alcoholic beverages	-0.63 (-0.74, -0.51)	-1.8 (-2.1, -1.5)	-0.53 (-0.64, -0.42)	-0.94 (-1.2, -0.72)

Values are regression coefficients and 95%CIs based on linear mixed models, and reflect differences in BMI, waist circumference, fat mass index, and body fat percentage averaged across the median follow-up of 7.1 years per 10 points higher score on the plant-based diet score without alcoholic beverages.

Models are adjusted for baseline age (years), sex, total energy (kcal/day), RS sub-cohort (RS-I, -II, or -III), time of repeated measurements of BMI, waist circumference, fat mass index, and body fat percentage education level (primary, lower, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task -hours/week), and food supplement use (yes or no).

Abbreviations: BMI, body mass index; RS, Rotterdam Study

Supplemental Table 7. Associations of plant-based diet score with BMI, waist circumference, fat mass index, and body fat percentage by additionally adjusting for baseline diet quality reflecting dietary guidelines or for dietary intake at RS-I-5 and RS-II-3

	BMI (kg/m^2) β (95% CI)	Waist circumference (cm) β (95% CI)	Fat mass index (kg/m^2) β (95% CI)	Body fat percentage β (95% CI)
Adjusting for diet quality reflecting adherence to dietary guidelines[a]	n=9620 -0.67 (-0.78, -0.55)	n=9474 -1.8 (-2.2, -1.5)	n=6153 -0.59 (-0.70, -0.48)	n=6153 -1.0 (-1.2, -0.79)
Adjusting for dietary intake at RS-I-5 and RS-II-3[b]	n=2178 -0.57 (-0.79, -0.34)	n=2178 -1.5 (-2.2, -0.92)	n=2133 -0.42 (-0.61, -0.24)	n=2133 -0.75 (-1.1, -0.40)

Values are regression coefficients and 95%CIs based on linear mixed models, and reflect differences in BMI, waist circumference, fat mass index, and body fat percentage averaged across the median follow-up of 7.1 years per 10 points higher score on the plant-based diet score without alcoholic beverages or with additional adjustment for diet quality score reflecting adherence to dietary guidelines.

[a]Models are adjusted for baseline age (years), sex, total energy (kcal/day), RS sub-cohort (RS-I, -II, or -III), time of repeated measurements of BMI, waist circumference, fat mass index, and body fat percentage education level (primary, lower, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task-hours/week), food supplement use (yes or no) and diet quality reflecting adherence to dietary guidelines.

[b]Analyses are performed with adjustment for baseline age (years), sex, total energy (kcal/day), RS sub-cohort (RS-I, or -II), time of repeated measurements of longitudinal BMI, waist circumference, fat

mass index, and body fat percentage, education level (primary, lower/intermediate, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task -hours/week) at baseline and at RS-I-5 and RS-II-3, food supplement use (yes or no), and plant-based diet score at RS-I-5 and RS-II-3.

Abbreviations: BMI, body mass index; RS, Rotterdam Study

Supplemental Table 8. Associations of plant-based diet score with BMI, waist circumference, fat mass index, and body fat percentage by excluding less healthy plant-based foods or healthy plant-based foods

	BMI (kg/m^2) n=9620 β (95% CI)	Waist circumference (cm) n=9474 β (95% CI)	Fat mass index (kg/m^2) n=6153 β (95% CI)	Body fat percentage n=6153 β (95% CI)
Excluding less healthy plant-based foods	-0.68 (-0.80, -0.56)	-1.9 (-2.2, -1.6)	-0.64 (-0.74, -0.53)	-1.4 (-2.3, -0.60)
Excluding healthy plant-based foods	-0.42 (-0.56, -0.28)	-1.1 (-1.5, -0.71)	-0.40 (-0.54, -0.26)	-0.64 (-0.92, -0.36)

Values are regression coefficients and 95%CIs based on linear mixed models, and reflect differences in BMI, waist circumference, fat mass index, and body fat percentage averaged across the median follow-up of 7.1 years per 10 points higher score on the plant-based diet index without less healthy plant-based foods or healthy plant-based foods. Models are adjusted for baseline age (years), sex, total energy (kcal/day), RS sub-cohort (RS-I, -II, or -III), time of repeated measurements of BMI, waist circumference, fat mass index, and body fat percentage education level (primary, lower, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task -hours/week), and food supplement use (yes or no).

Abbreviations: BMI, body mass index; RS, Rotterdam Study

Supplemental Table 9. Associations of plant-based diet score with BMI, waist circumference, fat mass index, and body fat percentage after adding positive scores from fish, eggs, low-fat milk, and low-fat yoghurt

	BMI (kg/m^2) n=9620 β (95% CI)	Waist circumference (cm) n=9474 β (95% CI)	Fat mass index (kg/m^2) n=6153 β (95% CI)	Body fat percentage n=6153 β (95% CI)
Plant-based diet score	-0.54 (-0.76, -0.32)	-1.4 (-1.9, -0.93)	-0.46 (-0.68, -0.18)	-0.71 (-0.90, -0.52)

Values are regression coefficients and 95%CIs based on linear mixed models, and reflect differences in BMI, waist circumference, fat mass index, and body fat percentage averaged across the median follow-up of 16.1 years in participants from RS-I and RS-II with available data per 10 points higher score.

Analyses are performed for associations of modification of plant-based diet by scoring fish, egg, low-fat milk, and low-fat yoghurt positive scores with adjustment for baseline age (years), sex, total energy (kcal/day), RS sub-cohort (RS-I, -II or -III,), time of repeated measurements of longitudinal BMI, waist circumference, fat mass index, and body fat percentage education level (primary, lower, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task -hours/week), and food supplement use (yes or no).

Abbreviations: BMI, body mass index; RS, Rotterdam Study

Supplemental Table 10. Associations of plant-based diet score with BMI, waist circumference, fat mass index, and body fat percentage by additionally adjusting for baseline health condition or excluding participants with obesity, diabetes, coronary heart disease, cancers at baseline and censoring body composition data collected after onset of diabetes, coronary heart disease, and cancers at follow-up

	BMI (kg/m^2) β (95% CI)	Waist circumference (cm) β (95% CI)	Fat mass index (kg/m^2) β (95% CI)	Body fat percentage β (95% CI)
Additionally adjusting for baseline	n=9620	n=9474	n=6153	n=6153
health condition[a]	-0.67 (-0.82, -0.51)	-1.9 (-2.1, -1.7)	-0.63 (-0.75, -0.50)	-1.0 (-1.2, -0.87)
Excluding participants with baseline	n=6364	n=6080	n=2741	n=2741
chronic diseases and censoring data[b]	-0.63 (-0.76, -0.49)	-1.5 (-1.9, -1.1)	-0.48 (-0.62, -0.34)	-0.72 (-0.96, -0.46)

Values are regression coefficients and 95%CIs based on linear mixed models, and reflect differences in BMI, waist circumference, fat mass index, and body fat percentage averaged across the follow-up period per 10 points higher score on the plant-based diet index.

[a]Analyses are performed with adjustment for baseline age (years), sex, total energy (kcal/day), RS sub-cohort (RS-I, -II, or -III), time of repeated measurements of longitudinal BMI, waist circumference, fat mass index, or body fat percentage, education level (primary, lower, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of MET-hours/week), food supplement use (yes or no), baseline diabetes (yes or no), baseline obesity (yes or no), baseline coronary heart disease (yes or no), and baseline cancers (yes or no).

[b]Analyses are performed based on body composition data collected before onset of diabetes, coronary heart disease, and cancers at follow-up in participants without obesity, diabetes, coronary heart disease, or cancers at baseline and adjusted for baseline age (years), sex, total energy (kcal), RS sub-cohort (RS-I, -II, or -III), time of repeated measurements of BMI, waist circumference, fat mass index, or body fat percentage, education level (primary, lower, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task-hours/week), and food supplement use (yes or no).

Abbreviations: BMI, body mass index; RS, Rotterdam Study

Supplemental Table 11. Associations of plant-based diet score with BMI, waist circumference, fat mass index, and body fat percentage in the three sub-cohorts

	BMI (kg/m^2) β (95% CI)	Waist circumference (cm) β (95% CI)	Fat mass index (kg/m^2) β (95% CI)	Body fat percentage β (95% CI)
RS-I[a]	-0.62 (-0.78, -0.46)	-1.6 (-2.1, -1.2)	-0.44 (-0.62, -0.26)	-0.85 (-1.2, -0.50)
RS-II[b]	-0.63 (-0.97, -0.29)	-2.2 (-3.0, -1.5)	-0.76 (-1.0, -0.51)	-1.3 (-1.8, -0.87)
RS-III[c]	-0.88 (-1.2, -0.61)	-2.6 (-3.3, -1.9)	-0.68 (-0.88, -0.48)	-1.2 (-1.6, -0.81)

Values are regression coefficients and 95%CIs based on linear mixed models, and reflect differences in BMI, waist circumference, fat mass index, and body fat percentage averaged across the follow-up period per 10 points higher score on the plant-based diet index in three sub-cohorts, respectively. Analyses were performed with adjustment for baseline age (years), sex, total energy (kcal/day), time of repeated measurements of BMI, waist circumference, fat mass index, or body fat percentage, education level (primary, lower, intermediate, or higher), smoking status (never, ever, current), physical activity (z-score of metabolic equivalent of task -hours/week), and food supplement use (yes or no).

[a]In RS-I, 5416 participants with BMI measurements; 5281 participants with waist circumference measurements; 2477 participants with fat mass index, and body fat percentage measurements.

[b]In RS-II, 1624 participants with BMI measurements; 1624 participants with waist circumference measurements; 1189 participants with waist circumference measurements, and body fat percentage measurements.

[c]In RS-III, 2580 participants with BMI measurements; 2569 participants with waist circumference measurements; 2487 participants with fat mass index, and body fat percentage measurements.

Abbreviations: BMI, body mass index; RS, Rotterdam Study

Chapter 3

Gut Microbiome and Type 2 Diabetes

Chapter 3.1

Gut microbiome, insulin resistance, and type 2 diabetes

Chen Z*, Radjabzadeh D*, Chen L*, Klurilshikov A, Ikram MA, Ahmadizar F, Kavousi M, Uitterlinden A, Zhernakova A, Fu J, Kraaij R, Voortman T. Gut microbiome, insulin resistance and type 2 diabetes: results from two large population-based studies.

**denotes equal contribution*

ABSTRACT

Background: Compositional alterations in gut microbiome among type 2 diabetes (T2D) patients have been reported, yet specific microbial taxa associated with T2D have not been confirmed.

Objectives: To examine associations of gut microbiome composition with insulin resistance and T2D in a population-based setting.

Methods: Cross-sectional analyses embedded within two Dutch prospective cohort studies: the Rotterdam Study (n=1418); and the Lifelines-Deep Study (n=748). In both studies, we collected stools in 2012-2013, and used the 16S rRNA method to measure gut microbiome composition. We measured insulin resistance (Homeostatic Model Assessment for Insulin Resistance, HOMA-IR), and identified prevalent T2D cases in 2012-2013. We used linear or logistic regression to examine associations between alpha diversity (Shannon, richness, and Inverse Simpson indices) and between taxa (from phylum to genus level) with insulin resistance and T2D. We used adonis permutation p value calculation to examine whether Bray-Curtis distance of beta diversity differed by insulin resistance and T2D. Analyses were adjusted for multiple covariates, including e.g. age, sex, energy intake, physical activity, education, smoking, BMI, and medication use.

Results: Our study included 159 T2D cases among 1418 participants from the Rotterdam Study and 17 T2D cases among 748 participants from the Lifelines-Deep Study. Lower Shannon index and richness were associated with higher HOMA-IR ($p<0.05$). Compared with the non-diabetic participants, T2D patients had lower richness ($P<0.05$). Bray-Curtis distance was also associated to insulin resistance ($p<0.05$). Furthermore, lower insulin resistance was associated with higher abundance of family Christensenellaceae, and the genera *Marvinbryantia*, *RuminococcaceaeUCG005*, *RuminococcaceaeUCG008*, *RuminococcaceaeUCG010*, and *RuminococcaceaeNK4A214group* ($p<0.0005$). Compared with the non-diabetic participants, T2D patients had lower abundance of family Clostridiaceae, Peptostreptococcaceae, and genera *Clostridiumsensustricto*, *Intestinibacter*, and *Romboutsia* ($p<0.0005$).

Conclusions: Our findings from two large population-based cohorts suggest that higher alpha diversity, along with more butyrate-producing gut bacteria, is associated with less T2D prevalence and with lower insulin resistance among non-diabetics. We identified 11 groups of butyrate-producing bacteria linked to insulin resistance or T2D, replicating previously reported associations for family Clostridiaceae and genus *Clostridiumsensustricto*, and discovering novel associations for 9 additional butyrate-producing bacteria. These findings may help provide insight into the etiology and pathogenesis of T2D.

INTRODUCTION

Type 2 diabetes (T2D) is a common complex metabolic disorder. Currently, much research focuses on how diet and lifestyle, metabolic, and genetic factors influence the development of T2D. Recently, some human studies have indicated a role of gut microbiome in T2D.[1-4] Differences in gut microbiome composition with T2D status may provide pathways on how dietary and other risk factors affect T2D risk. Overall, most studies have indicated that, compared to healthy participants, T2D patients have a significant reduction in overall alpha diversity of gut microbiome composition, along with a decrease of some butyrate-producing bacteria, such as class Clostridia and genus *Faecalibacterium*.[1-4] Pathways behind potential links of gut microbiome composition with T2D remain unclear, but may involve metabolism of butyrate, the molecule produced by several obligate symbiotic bacteria. Butyrate may induce beneficial metabolic effects through enhancement of mitochondrial function, improvement of energy metabolism, activation of intestinal gluconeogenesis, and prevention of metabolic endotoxemia and inflammation.[5, 6] Furthermore, some non-butyrate-bacteria such as species *Haemophilus parainfluenzae*, have also been observed to be associated with T2D in previous studies.[7] However, these previous studies had some limitations. They were limited by small sample sizes, ranging from 202 to 784 participants,[1] which may limit statistical power to pick up true associations. Therefore, it is necessary to further increase sample size. Moreover, in most of these previous studies, key confounders, such as lifestyle and socioeconomic status, were not controlled for, and findings could thus be confounded by associations of these factors with gut microbiome and with T2D, which may obscure observed associations. Besides, as almost all these previous studies were conducted in patient populations or case-control settings, it is unclear whether these findings are applicable to real-world settings and whether associations can already be observed for subclinical parameters such as insulin resistance among non-diabetic groups. As gut microbiome composition can vary across different geographical locations, gender, ages, ethnicity, associations with T2D may differ in different populations, which calls for more research among different populations. Therefore, large population-based studies based on different populations, considering potential effect of key confounders are needed to replicate the analyses to extend existing evidence.

Therefore, we aimed to examine associations between gut microbiome composition and insulin resistance, and T2D in two large population-based cohorts (n=2166).

METHODS

Study design

The current cross-sectional analyses were embedded within two ongoing prospective cohorts: the Rotterdam Study and the Lifelines-Deep Study. We first conducted independent analyses in two cohorts separately, and then summarized the results from the two cohorts using fixed-effects meta-analysis.

The Rotterdam Study

The Rotterdam Study is a prospective cohort study of participants aged ≥ 45 years at baseline living in the Ommoord District of Rotterdam, the Netherlands. Details on the design of Rotterdam Study are described elsewhere.[8] Briefly, the first sub-cohort of the Rotterdam Study (RS-I) started in 1989-1993 and was since extended with a second sub-cohort (RS-II) in 2000-2001, and with a third sub-cohort (RS-III)) in 2006-2008. Measurements are performed at baseline and subsequently every 3 to 5 years. The current analyses in the Rotterdam Study population were performed within the second follow-up visit of the third sub-cohort (RS-III-2: 2012-2014), in which 3132 subjects participated. The Rotterdam Study has been approved by the institutional review board (Medical Ethics Committee) of Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. All participants gave informed consent.

The Lifelines-Deep study

The Lifelines-Deep Study is a prospective population-based cohort study of 1539 participants ≥ 18 years living in the three provinces in the northern part of the Netherlands: Groningen, Friesland and Drenthe. More details are described in elsewhere.[9] The current analyses in the Lifelines-Deep Study population were embedded within the baseline visit (LLD: 2013). The Lifelines-Deep Study was approved by the ethics committee of the University Medical Centre Groningen. All participants signed an informed consent prior to enrolment.

Participants in the current analyses

In the two cohorts, of the 4671 participants (n=3132 from RS-III-2, n=1539 from LLD), 2607 had data on gut microbiome composition (n=1427 from RS-111-2, n=1180 from LLD). From these 2607 participants, we excluded the participants without information on T2D, resulting in 2166 participants in analyses for gut microbiome composition and T2D (n=1418 from RS and n=748 from LLD). Furthermore, from these 2166 participants, we additionally excluded participants with prevalent T2D or without data on insulin resistance for analyses on insulin resistance, resulting in 1984 participants (n=1253 for RS and n=731 for LLD) (Supplemental Figure 1).

Collection of gut microbiome data

Details on microbiome data collection in the Rotterdam Study and the Lifelines-Deep Study are described elsewhere.[9-11] For the Rotterdam Study, participants were requested to collect a stool sample at their home in sterile tubes in the Rotterdam Study in 2012-2013. These stool samples were sent by mail to the laboratories of Erasmus MC (the Rotterdam Study). The time each sample was in the mail between sample production and receipt was recorded (time in mail). Hereafter, an automated stool DNA isolation kit (Diasorin, Saluggia, Italy) was used to isolate bacterial DNA. A confounding effect driven by DNA isolation batches was observed and recorded. The V3 and V4 hypervariable regions of the bacterial 16S rRNA gene were amplified and sequenced on the IIIumina Miseq platform. For the Lifelines-Deep Study, the stool samples were picked up by students of University Medical Center Groningen in 2013. DNA was isolated with the AllPrep DNA/RNA Mini Kit (Qiagen; cat. #80204). The V4 hypervariable region of the bacterial 16S rRNA gene were amplified and sequenced on the IIIumina MiSeq platform. In both cohort studies, we implemented the 16S data processing pipeline, which comprised a naive Bayesian classifier from the Ribosomal Database Project, and the recent, SILVA database release 128: we only analyzed taxonomical results using genus and higher taxonomic levels. All additional steps have been standardized across both cohorts, including down-sampling to 10000 reads with fixed seed to allow for replicability, procedures of transformations, and the thresholds set for bacterial taxa to be included in the analysis (any taxon should be present in more than 10% of the cohort's samples). This filtering effectively reduces the total number of tests and also makes cross-validation and meta-analysis among the two cohorts possible.[9] As a result, in the Rotterdam Study, the microbiome data contained 8 phyla, 15 classes, 18 orders, 33 families, and 126 genera. In the Lifelines-Deep Study, the microbiome data contained 12 phyla, 21 classes, 27 orders, 48 families, and 184 genera. Between the two cohorts, data on the same taxa were available for: 7 phyla, 14 classes, 16 orders, 30 families, 103 genera. We also calculated alpha diversity (Shannon index, richness, and Inverse Simpson index), and beta diversity (Bray-Curtis distance) at genus level using the R package 'vegan', in both cohorts separately.

Assessment of insulin resistance

In both cohorts, fasting blood samples of the participants were collected in 2012-2013. Glucose levels were examined with the glucose hexokinase method. Serum insulin was measured by electro chemiluminescence immunoassay technology on a Roche Modular Analytics E170 analyzer (Roche Diagnostics GmbH, Mannheim, Germany). Insulin resistance was calculated using the homeostasis model assessment of insulin resistance (HOMA-IR): fasting insulin (mU/L) × fasting glucose (mmol/L) / 22.5.[12]

Assessment of T2D

In the Rotterdam Study, information on T2D was collected from general practitioners' records, pharmacies' databases, hospital discharge letters, and follow-up examinations. In the Lifelines Deep Study, information on T2D was collected through self-reported questionnaires. Data on prevalent T2D were identified in the Rotterdam Study in 2013, and in the Lifelines-Deep cohort study in 2013, respectively. Cases of T2D were identified according to WHO criteria[13]: T2D was defined as a fasting blood glucose concentration of ≥ 7.0 mmol/L, a non-fasting blood glucose concentration of ≥ 11.1 mmol/L (when fasting samples were unavailable), or the use of blood glucose-lowering drugs or dietary treatment and registration of the diagnosis diabetes.

Assessment of covariates

Except for data on education level in the Rotterdam Study, which were collected at baseline in 2006-2008, data on all covariates were collected in both cohorts in 2012-2013 and included: smoking status, physical activity, total energy intake, alcohol intake, body mass index (BMI), use of lipid lowering medications, and use of proton pump inhibitors. Specifically, information on education, smoking status, dietary intake and physical activity were assessed through interviews and questionnaires. In the Rotterdam Study, physical activity was assessed with the LASA Physical Activity Questionnaire, activities were weighted by their intensity with Metabolic Equivalent of Task (MET)[14] and expressed in MET-hours per week. In the Lifelines-Deep Study, physical activity was assessed with the Short Questionnaire to assess Health-enhancing physical activity, and a physical activity composition score per week was calculated.[15, 16] To measure dietary intake, a 389-item food frequency questionnaire was used in the Rotterdam Study,[17] a 125-item food questionnaire was used in the lifelines-Deep cohort, from which total energy intake and alcohol intake were calculated.[16] Height and weight were measured in each study's research centers and body mass index ((kg/m^2)) was calculated. In both cohorts, information on medication use, CVD cases, or cancers cases was obtained from general practitioners, pharmacies' databases, Nationwide Medical Registry, or follow-up examinations.[16, 18]

Data analyses

Gut microbial alpha diversity and insulin resistance and T2D

We natural-log transformed HOMA-IR for all analyses. The associations of alpha diversity indices (Shannon index, richness, and Inverse Simpson index) with HOMA-IR were examined using linear regression. The associations of alpha diversities with T2D prevalence were examined using logistic regression.

Gut microbial beta diversity and insulin resistance and T2D

We explored whether variation of beta diversity of gut microbial composition was explained by insulin resistance and T2D using permutation analysis of variance (PERMANOVA, 1000 permutations) based on the function "adonis" from the R package "vegan".[20]

Gut microbial taxa and insulin resistance and T2D

We first added 1 to all taxa counts to prevent missingness derived from log zero. To reduce the skewness of the distribution of microbial taxa counts, we performed natural log transformation. The associations between microbial taxa and insulin resistance and T2D were assessed by linear regression and logistic regression, respectively.

For the analyses, we adjusted for a series of confounders. In model 1, we adjusted for age, sex, technical covariates including time in mail (only in RS), and DNA batch effect (only in RS). In model 2, we additionally adjusted for alcohol intake, energy intake, smoking, education level (RS), and physical activity. In model 3, we additionally adjusted for BMI. Last, in model 4, we additionally adjusted for use of lipid lowering drugs and protein-pump inhibitors.

Additional analyses

We examined interaction effects of alpha diversity, or taxa with age, sex, or BMI by additionally including these interaction terms (such as, Shannon index × age, and Shannon index × sex) one at a time into model 4. In case of statically significant interaction terms, stratified analyses by these factors would be conducted.

We first conducted all analyses in the two studies separately, and then combined the associations for alpha diversity and for taxa available in both cohorts using fixed-effects meta-analysis. Associations of beta diversity cannot be pooled and are presented for each cohort separately. P values <0.05 were considered statistically significant for analyses of alpha diversity and beta diversity. P values <0.0005 were considered statistically significant for taxa analyses based on a method of Li and Ji to calculate a number of independent tests.[19] There were 103 independent tests among microbial taxa. Therefore, experimental-wide significance threshold needed in order to keep type I error rate at 5% was set at (0.05/103.1082): 0.0005. All analyses were conducted in R version 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Baseline characteristics

Characteristics of the study population are shown in Table 1. For the 1418 participants in the Rotterdam Study, the mean age was 62.4 ± 5.9 years, 57.5% were female, and 176 had T2D. In the 748 participants of the Lifelines-Deep Study, the mean age was 44.7 ± 13.4 years, 57.6% were female, and 17 had T2D.

Table 1. Characteristics of participants

	The Rotterdam Study n=1418	**The Lifelines Deep Study n=748**
Age (years)	62.4 (5.9)	44.7 (13.4)
Sex (%)		
- Female	57.5%	57.6%
Smoking (%)		
- Current	13.6%	20.7%[d]
- Ever	49.8%	79.3%[e]
- Never	36.6%	
Education (%)[a]		
- Primary	7.6	\
- Lower	33.4	\
- Intermediate	28.0	\
- Higher	30.7	\
Physical activity[b]	42·9 (17·7-82·8)	55.5 (25.8-57.8)
BMI (kg/m^2)	27.5 (4.5)	25.2 (4.1)
Alcohol (g/day)	8.1 (1.4-19.7)	6.0 (1.6-11.8)
Total energy (kcal/day)	2243.2 (1869.4 – 2733.3)	1862.0 (1526.1-2282.8)
Lipid lowering medication[c]		
- Yes	27.7%	4.3%
Proton-pump Inhibitors		
- Yes	18.1%	8.42%

Variables expressed as mean (SD), median (25th percentile–75th percentile), or percentage.

[a]The Lifelines Deep Study does not collect data on education level.

[b]Physical activity level is expressed as metabolic equivalent of task-hours per week in the Rotterdam Study, and is calculated as a continuous score in the Lifelines Deep Study

[c]Lipid lowering medication is defined as statin intake in the Lifelines Deep Study

[d]Indicates smoker in the Lifelines Deep Study.

[e]Indicates non-smoker in the Lifelines Deep Study.

Association between gut microbial alpha diversity with insulin resistance and T2D

After adjustment for age, sex, time in mail (RS), DNA batch (RS), total energy, alcohol intake, education, smoking, BMI, proton-pump inhibitors (PPIs) and lipid-lowering medications, lower Shannon index, and richness were associated with higher HOMA-IR (e.g. meta-analysis results, Shannon index, -0.06, 95%CI (-0.10, -0.02), p=0.02); while Inverse Simpson was not associated with HOMA-IR (Table 2, Model 4). For T2D, after adjustment for the multiple covariates in model 4, higher richness was associated with a lower prevalence of T2D (OR= 0.93, 95%CI (0.88, 0.99), p=0.04)). Higher Shannon index was suggestively associated with a lower prevalence of T2D (Table 3, Model 4). Inverse Simpson index was not associated with T2D.

Table 2. Association of alpha diversity and insulin resistance

HOMA-IR	Rotterdam Study n=1253	Lifelines-Deep Study n=731	Meta-analysis n=1984
		Shannon index	
Model 1	-0.12 (-0.18, -0.07)	-0.09 (-0.16, -0.02)	-0.11 (-0.15, -0.07), p=0.01
Model 2	-0.08 (-0.14, -0.03)	-0.06 (-0.12, 0.0004)	-0.07 (-0.11, -0.03), p=0.01
Model 3	-0.08 (-0.14, -0.03)	-0.06 (-0.12, -0.03)	-0.07 (-0.11, -0.03), p=0.01
Model 4	-0.07 (-0.12, -0.01)	-0.05 (-0.11, 0.004)	-0.06 (-0.10, -0.02), p=0.02
		Richness	
Model 1	-0.12 (-0.18, -0.07)	-0.12 (-0.19, -0.05)	-0.12 (-0.16, -0.08), p=0.01
Model 2	-0.08 (-0.14, -0.02)	-0.06 (-0.12, -0.01)	-0.07 (-0.11, -0.03), p=0.03
Model 3	-0.08 (-0.14, -0.02)	-0.06 (-0.12, -0.01)	-0.07 (-0.11, -0.03), p=0.03
Model 4	-0.07 (-0.12, -0.02)	-0.06 (-0.12, -0.001)	-0.07 (-0.11, -0.03), p=0.03
		Inverse Simpson index	
Model 1	-0.22 (-0.40, -0.06)	-0.30 (-0.78, 0.18)	-0.23 (-0.40, -0.06), p=0.03
Model 2	-0.06 (-0.11, -0.004)	-0.03 (-0.09, 0.03)	-0.05 (-0.09, -0.01), p=0.04
Model 3	-0.06 (-0.11, -0.004)	-0.03 (-0.09, 0.03)	-0.05 (-0.09, -0.01), p=0.04
Model 4	-0.04 (-0.10, 0.01)	-0.03 (-0.09, 0.03)	-0.04 (-0.08, 0.002), p=0.05

Values are β coefficients and 95% confidence intervals (CI) from linear regression models for associations between Shannon index, richness and Inverse Simpson index with HOMA-IR.
Model 1: Age, sex, time in mail (only in the Rotterdam Study), and Batch (only in the Rotterdam Study)
Model 2: Model 1+smoking+education level (only in the Rotterdam Study) + physical activity + alcohol intake + total energy intake
Model 3: Model 2 + BMI
Model 4: Model 3 + lipid lowering medication + proton pump inhibitors

Table 3. Association of alpha diversity and type 2 diabetes risk

T2D	Rotterdam Study n=1418	Lifelines-Deep Study n=748	Meta-analysis n=2166
		Shannon index	
Model 1	0.80 (0.67, 0.96)	0.74 (0.46, 1.20)	0.79 (0.67, 0.94), p=0.03
Model 2	0.71 (0.57, 0.89)	0.85 (0.53, 1.38)	0.73 (0.60, 0.90), p=0.03
Model 3	0.76 (0.60, 0.96)	0.84 (0.51, 1.38)	0.78 (0.63, 0.96), p=0.03
Model 4	0.80 (0.63, 1.02)	0.94 (0.55, 1.62)	0.83 (0.66, 1.03), p=0.06
		Richness	
Model 1	0.74 (0.64, 0.91)	0.78 (0.48, 1.28)	0.76 (0.64, 0.90), p=0.04
Model 2	0.73 (0.58, 0.90)	0.83 (0.50, 1,36)	0.74 (0.61, 0.90), p=0.04
Model 3	0.78 (0.62, 0.98)	0.86 (0.52, 1.45)	0.79 (0.64, 0.98), p=0.04
Model 4	0.80 (0.63, 1.02)	0.95 (0.87, 1.00)	0.93 (0.88, 0.99), p=0.04
		Inverse Simpson index	
Model 1	0.88 (0.73, 1.05)	0.91 (0.57, 1.45)	0.88 (0.74, 1.04), p=0.10
Model 2	0.79 (0.64, 0.99)	1.05 (0.66, 1.66)	0.84 (0.68, 1.02), p=0.08
Model 3	0.84 (0.66, 1.06)	1.00 (0.63, 1.62)	0.87 (0.70, 1.07), p=0.23
Model 4	0.88 (0.69, 1.13)	1.08 (0.64, 1.82)	0.91 (0.73, 1.14), p=0.25

Values are ORs and 95% confidence intervals (CI) from logistical regression models for associations between Shannon index, richness and Inverse Simpson index with T2D.
Model 1: Age, sex, time in mail (only in the Rotterdam Study), and Batch (only in the Rotterdam Study)
Model 2: Model 1+smoking+education level (only in the Rotterdam Study) + physical activity + alcohol intake + total energy intake
Model 3: Model 2 + BMI
Model 4: Model 3 + lipid lowering medication + proton pump inhibitors

Associations between gut microbial beta diversity with insulin resistance and T2D

We observed that beta diversity (Bray-Curtis distance) was associated with insulin resistance (R^2= 0.004, p=0.001 in the Rotterdam Study and R^2=0.005, p=0.002 in the Lifelines-Deep Study). Furthermore, Bray-Curtis distance was associated with T2D prevalence in the Rotterdam Study, although not in the Lifelines-Deep Study (R^2= 0.003, p=0.001 in the Rotterdam Study, and R^2=0.001, p=0.65 in the Lifelines-Deep Study).

Associations between gut microbial communities with insulin resistance and T2D

Between the two cohorts, data on the same taxa were available for: 7 phyla, 14 classes, 16 orders, 30 families, 103 genera. After adjustment for age, sex, time in mail (in RS), DNA batch (in RS), smoking, education (in RS), alcohol intake, total energy, BMI, PPI, and lipid-lowering medications, we observed 6 taxa associated with HOMA-IR and 5 taxa with T2D in the meta-analysis. Specifically, a higher

abundance of family Christensenellaceae, genus *Marvinbryantia*, *RuminococcaceaeUCG005*, *RuminococcaceaeUCG008*, *RuminococcaceaeUCG010*, and *RuminococcaceaeNK4A214group* was associated with lower HOMA-IR (family Christensenellaceae, β=-0.08, 95%CI (-0.12, -0.03), P=0.0003; genus: *Marvinbryanti*, -0.07 (-0.11, -0.03), p=0.0002, *RuminococcaceaeUCG005*, -0.09 (-0.13, -0.05), p<0.0001, *RuminococcaceaeUCG008*, -0.07 (-0.11, -0.03), p=0.00045, *RuminococcaceaeUCG010*, -0.08 (-0.12, -0.04), p<0.0001, *RuminococcaceaeNK4A214group*, -0.09 (-0.13, -0.05), p<0.0001) (Table 4, Model 4). A higher abundance of family Clostridiaceae1, Peptostreptococcaceae, and genus *Clostridiumsensustricto*, *Intestinibacter*, and *Romboutsia* was associated with less T2D (family: Clostridiaceae1, OR=0.51, 95%CI (0.41, 0.65), p<0.0001, and Peptostreptococcaceae, 0.56 (0.45, 0.70), p<0.0001; genus: *Clostridiumsensustricto1*, 0.51 (0.40, 0.64), p<0.0001, *Intestinibacter*, 0.60 (0.48, 0.76), p<0.0001, *Romboutsia*, 0.55 (0.41, 0.70), p<0.0001) (Table 5, Model 4). Supplemental Tables 1-2 show all results in models 1-4 for all taxa in separated analyses and meta-analysis of the two studies.

Additional analyses

Associations for alpha diversity or taxa with HOMA-IR or T2D did not differ by age, sex, or BMI (All P values >0.05 for the interactions such as Shannon index × age in alpha diversity analysis, all p values >0.0005 in taxa analysis).

DISCUSSION

In this large population-based sample, we observed associations of gut microbiome composition with T2D prevalence and with insulin resistance about non-diabetics, independent of several sociodemographic and lifestyle factors. We observed that higher alpha diversity was associated with lower insulin resistance and lower prevalence of T2D, that variations of gut microbial beta diversity are linked to insulin resistance. Furthermore, we identified 11 groups of butyrate-producing bacteria linked to insulin resistance and T2D. Higher abundance of the 11 groups of butyrate-producing bacteria: family Christensenellaceae, Clostridiaceae1, and Peptostreptococcaceae, and the genus *Marvinbryantia*, *RuminococcaceaeUCG005*, *RuminococcaceaeUCG008*, *RuminococcaceaeUCG010*, *RuminococcaceaeNK4A214group*, *Clostridiumsensustricto1*, *Intestinibacter*, and *Romboutsia* may benefit insulin resistance and T2D.

Comparison with previous studies

Several previous studies have examined these associations, but were limited by smaller sample size,[1, 2, 4] and by the lack of adjustment for important confounders, such as diet, physical activity, and social economic factors.[1, 2, 4] Our study is the first to comprehensively investigate the associations between

gut microbiome composition with insulin resistance and T2D in a large population-based sample (n=2166), for which we adjusted for a series of important confounders.

Similar with previous studies,[6] we found that higher alpha diversity was associated with lower insulin resistance and T2D prevalence. Furthermore, we extended our analysis by examining the association between gut microbial beta diversity and insulin resistance and T2D. We found Bray-Curtis distance linked to insulin resistance, which further confirmed that variation of gut microbiome composition is closely related to the development of T2D.

We further explored which specific bacteria were associated with insulin resistance and T2D. Overall, we identified 11 butyrate-producing bacteria linked to insulin resistance or T2D. In line with previous studies,[6, 7] we observed that a higher abundance of the two butyrate-producing bacteria: family Clostridiaceae, and genus *Clostridumsensustricto* were associated with lower insulin resistance or lower prevalence of T2D. However, our study has also yielded nine novel associations. We observed that the nine butyrate-producing bacteria were inversely associated with insulin resistance or T2D: family Christensenellaceae, and Peptostreptococcaceae, and genus *Marvinbryantia*, *RuminococcaceaeUCG005*, *RuminococcaceaeUCG008*, *RuminococcaceaeUCG010*, *RuminococcaceaeNK4A214group*, *intestinibacter*, and *Romboutsia*. These novel findings further extend the evidence that butyrate-producing bacteria may influence the development of T2D. Furthermore, some of the novel findings were supported by previous evidence on obesity. For example, a previous study by Goodrich et al.[20] reported that higher abundance of family Christensenellaceae was link to lower BMI that is associated with lower HOMA-IR and lower T2D risk. We also noticed that, among the 11 associations, all the six associations with HOMA-IR (e.g. the associations of family Christensenellaceae and Marvinbryantia) and one association with T2D (genus *Romboutsia*) were consistent across the Rotterdam Study and the Lifelines-Deep Study, indicating the seven associations were successfully replicated between the two cohorts. Yet, the associations between T2D and four butyrate-producing bacteria (e.g. family Clostridiaceae1 and Peptostreptococcaceae) were heterogeneous across the two cohorts. The pooled associations of the four bacteria were mainly driven by the Rotterdam Study, and null associations for the four bacteria and T2D were observed in the Lifelines-Deep Study. This may be explained by the small number of T2D cases (n=17) in the Lifelines-Deep Study, limiting statistical power.

As for the possible explanations for the associations of the identified groups of bacteria, this may involve potential beneficial effects of butyrate produced by these bacteria.[6] Butyrate is a short-chain fatty acid fermented by butyrate-producing bacteria from dietary fiber.[21] And butyrate has been suggested to induce beneficial metabolic effects through enhancement of mitochondrial, improvement of energy metabolism, activation of intestinal gluconeogenesis, and prevention of metabolic endotoxemia and inflammation via different routes of gene expression and hormone regulation.[6] For example, butyrate has been reported to increase mitochondrial activity by increasing expression of

peroxisome proliferator-activated receptor gamma coactivator 1-alpha, which may contribute to more adenosine triphosphate production via the tricarboxylic acid cycle, benefiting energy metabolism.[5, 22, 23] Moreover, butyrate may improve glucose metabolism by activating intestinal gluconeogenesis gene expression through a cyclic adenosine monophosphate-dependent mechanism.[24] Additionally, butyrate may helpful in preventing inflammation by maintaining intestinal integrity via G protein-coupled receptors signaling pathways, which is associated lower insulin resistance and T2D risk.[24, 25]

Future research should aim to validate the hypothesis of butyrate-producing bacteria as role players and their products as signaling molecules in human glucose and lipid metabolism, for which double-blinded randomized controlled trials using either short-chain fatty acids supplementation (given either orally or rectally) or donor fecal microbiota transplantation are needed.

Strengths and limitations

First, our current study is the largest population-based study (n=2166) to date to investigate associations of gut microbial composition with insulin resistance and T2D. Such a large population-based study provided more statistical power to pick up associations. For example, our study identified nine novel butyrate-producing bacteria (e.g. family Peptostreptococcaceae, genus *Marvinbryantia*, genus *intestinibacter*, and genus *Romboutsia*) in relation to insulin resistance or T2D. Furthermore, among the 11 associations identified in our study, the 7 associations (e.g. family Christensenellaceae and Marvinbryantia, and genus *Romboutsia*) were successfully replicated between the two cohorts of our current study. Additionally, the data from the two cohorts had been standardized to pool the results, thus helping minimize information bias. Second, we adjusted for a broad range of confounders in our analyses, such as diet, physical activity, BMI, education level, and smoking status. Most of previous studies did not adjust for these important confounders. Last, to minimize reverse bias, we not only examined associations between gut microbiome with T2D status, but also with insulin resistance among non-diabetic participants, and observed similar results, which suggests that microbiome composition may play a role already in earlier phases of development of T2D. However, we have several limitations to consider. First, our study is a cross-sectional observational study, and thus our ability to assess causality or temporality is limited. Second, we determined gut microbiome composition from stool samples. As gut microbiome composition varies throughout the gut with respect to the anatomic location along the gut and at the given site, a more complete picture of the gut microbiome might be obtained by getting samples from different locations along the intestines in the future. Third, due to the use of 16S rRNA data, we could not explore associations at species and lower levels, or associations of functional profiles of gut microbiome composition. Metagenomics approaches could overcome these limitations. Fourth, we cannot exclude the possibility of residual confounding. Last, our study may be generalized to populations with similar age and race, but more studies in diverse populations are needed.

Conclusions

Our findings from a large population-based sample and adjusting for a broad range of confounders, indicate that gut microbiome composition may influence the development of T2D. An increased gut microbial diversity, along with more butyrate-producing bacteria may benefit insulin resistance and risk of T2D. Our findings may help provide new insight into etiology, mechanisms and thereby into the prevention, and therapy of T2D.

Table 4. Statistically significant pooled associations between taxa and insulin resistance[a]

	Rotterdam Study n=1253	Lifelines-Deep Study n=731	Pooled results[a] n=1984
Family Christensenellaceae.id.1866	-0.09 (-0.14, -0.03)	-0.06 (-0.12, -0.002)	-0.08 (-0.12, -0.03) p=0.0003
Genus *Marvinbryantia.id.2005*	-0.07 (-0.13, -0.02)	-0.08 (-0.13, -0.02)	-0.07 (-0.11, -0.03) p=0.0002
Genus *RuminococcaceaeUCG.005.id.11363*	-0.11 (-0.16, -0.05)	-0.07 (-0.12, -0.01)	-0.09 (-0.13, -0.05) p<0.0001
Genus *RuminococcaceaeUCG.008.id.11365*	-0.10 (-0.14, -0.04)	-0.04 (-0.09, 0.02)	-0.07 (-0.11, -0.03) p=0.00045
Genus *RuminococcaceaeUCG.010.id.11367*	-0.10 (-0.16, -0.05)	-0.06 (-0.12, 0.0001)	-0.08 (-0.12, -0.04) p<0.0001
Genus *RuminococcaceaeNK4A214group.id.11358*	-0.09 (-0.15, -0.04)	-0.08 (-0.14, -0.03)	-0.09 (-0.13, -0.05) p<0.0001

Effect estimates are β coefficients for a natural-transformed HOMA-IR with their 95%-confidence intervals (95%CIs) based on linear regression model among non-diabetic participants. Pooled estimates are calculated based on a mixed-effect meta-analysis.

Multiple model (corresponding to Model 4) is adjusted for time in mail (in the Rotterdam Study), DNA batch (in the Rotterdam Study), age, gender, alcohol intake, total energy intake, smoking status, physical activity, BMI, PPI, lipid-lowering medication, and education level (in the Rotterdam Study).

[a]The current meta-analysis combined associations for 7 phyla, 14 classes, 16 orders, 30 families, and 103 genera. This table only shows significant pooled associations. The results for all taxa and insulin resistance in separated analyses and meta-analysis of the two studies are shown in the Supplemental Table 1. Level of significance: $p < 0.0005$

Table 5. Statistically significant pooled associations between taxa and type 2 diabetes[a]

	Rotterdam Study	Lifelines-Deep Study	Pooled results[a]
Number of participants	**n=1418**	**n=748**	**n=2166**
T2D cases	**n=159**	**n=17**	**n=176**
Family Clostridiaceae1.id.1869	0.42 (0.70, 0.88)	1.07 (0.83, 1.39)	0.51 (0.41, 0.65) p<0.0001
Family Peptostreptococcaceae.id.2042	0.52 (0.46, 0.59)	0.89 (0.67, 1.20)	0.56 (0.45, 0.70) p<0.0001
Genus *Clostridiumsensustricto1.id.1873*	0.42 (0.36, 0.48)	1.08 (0.83, 1.41)	0.51 (0.40, 0.64) p<0.0001
Genus *Intestinibacter.id.11345*	0.50 (0.43, 0.57)	1.06 (0.84, 1.33)	0.60 (0.48, 0.76) p<0.0001
Genus *Romboutsia.id.11347*	0.56 (0.49, 0.63)	0.44 (0.24, 0.81)	0,55 (0.44, 0.70) p<0.0001

Effect estimates are odd ratios for T2D with their 95%-confidence intervals (95%CIs) based on cox regression model. Pooled estimates are calculated based on a mixed-effect meta-analysis.

Multiple model (corresponding Model 4) is adjusted for time in mail (in the Rotterdam Study), DNA batch (in the Rotterdam Study), age, gender, total energy, smoking status, physical activity, BMI, PPI, lipid-lowering medication, and education level (in Rotterdam Study).

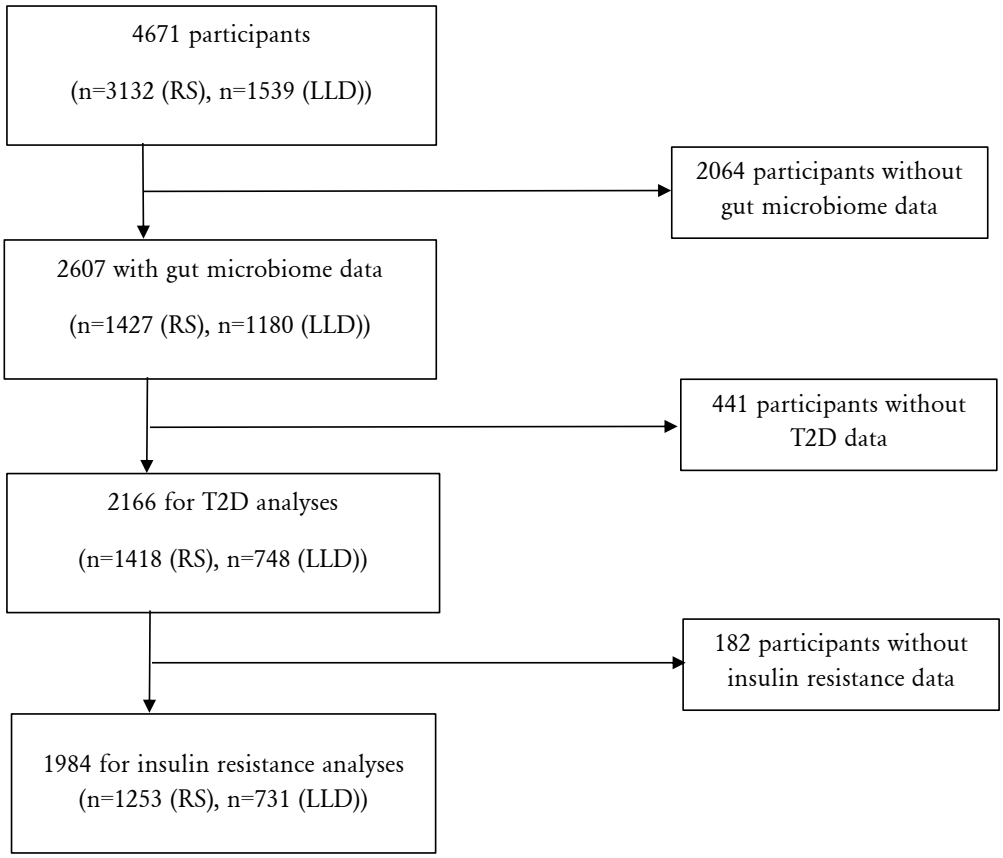
[a]The current meta-analysis combined associations for 7 phyla, 14 classes, 16 orders, 30 families, and 103 genera. This table only shows significant pooled associations. The results for all taxa and T2D in separated analyses and meta-analysis of the two studies are shown in the Supplemental Table 2. Level of significance: $p < 0.0005$

REFERENCES

1. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature. 2015;528(7581):262.
2. Larsen N, Vogensen FK, Van Den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PloS one. 2010;5(2):e9085.
3. Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 2013;498(7452):99.
4. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55.
5. Davie JR. Inhibition of histone deacetylase activity by butyrate. The Journal of nutrition. 2003;133(7):2485S-93S.
6. Hartstra AV, Bouter KEC, Bäckhed F, Nieuwdorp M. Insights into the role of the microbiome in obesity and type 2 diabetes. Diabetes care. 2015;38(1):159-65.
7. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490:55.
8. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. European Journal of Epidemiology. 2017:1-44.
9. Tigchelaar EF, Zhernakova A, Dekens JAM, Hermes G, Baranska A, Mujagic Z, et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. BMJ open. 2015;5(8):e006772.
10. Radjabzadeh D, Uitterlinden AG, Kraaij R. Microbiome measurement: Possibilities and pitfalls. Best Practice & Research Clinical Gastroenterology. 2017.
11. Kurilshikov A, van den Munckhof ICL, Chen L, Bonder MJ, Schraa K, Rutten JHW, et al. Gut Microbial Associations to Plasma Metabolites Linked to Cardiovascular Phenotypes and Risk: A Cross-Sectional Study. Circulation research. 2019;124(12):1808-20.
12. Matthews D, Hosker J, Rudenski A, Naylor B, Treacher D, Turner R. Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28(7):412-9.
13. World Health O. Global report on diabetes: World Health Organization; 2016.
14. Chen Z, Schoufour JD, Rivadeneira F, Lamballais S, Ikram MA, Franco OH, et al. Plant-based diet and adiposity over time in a middle-aged and elderly population: the Rotterdam Study. Epidemiology. 2018.
15. Wendel-Vos GCW, Schuit AJ, Saris WHM, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. Journal of clinical epidemiology. 2003;56(12):1163-9.

16. Scholtens S, Smidt N, Swertz MA, Bakker SJL, Dotinga A, Vonk JM, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. International journal of epidemiology. 2014;44(4):1172-80.
17. Chen Z, Zuurmond MG, van der Schaft N, Nano J, Wijnhoven HAH, Ikram MA, et al. Plant versus animal based diets and insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. European journal of epidemiology. 2018;33(9):883-93.
18. Bos D, Portegies ML, van der Lugt A, Bos MJ, Koudstaal PJ, Hofman A, et al. Intracranial carotid artery atherosclerosis and the risk of stroke in whites: the Rotterdam Study. JAMA Neurol. 2014;71(4):405-11.
19. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity. 2005;95(3):221.
20. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. 2014;159(4):789-99.
21. Scheppach W. Effects of short chain fatty acids on gut morphology and function. Gut. 1994;35(1 Suppl):S35-S8.
22. Gao Z, Yin J, Zhang J, Ward RE, Martin RJ, Lefevre M, et al. Butyrate improves insulin sensitivity and increases energy expenditure in mice. Diabetes. 2009;58(7):1509-17.
23. Donohoe DR, Garge N, Zhang X, Sun W, O'Connell TM, Bunger MK, et al. The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. Cell metabolism. 2011;13(5):517-26.
24. Delaere F, Duchampt A, Mounien L, Seyer P, Duraffourd C, Zitoun C, et al. The role of sodium-coupled glucose co-transporter 3 in the satiety effect of portal glucose sensing. Molecular metabolism. 2013;2(1):47-53.
25. Samuel BS, Shaito A, Motoike T, Rey FE, Backhed F, Manchester JK, et al. Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. Proceedings of the National Academy of Sciences. 2008;105(43):16767-72.

SUPPLEMENTAL MATERIAL



Supplemental Figure 1. Participants selection

Supplemental Table 1. Associations between all taxa and insulin resistance

Supplemental Table 2. Associations between all taxa and type 2 diabetes

Supplemental Tables 1-2 can be obtained from the author.

Chapter 4

Nutrition and Gut Microbiome

Chapter 4.1

Diet quality and gut microbiome

Chen Z, Radjabzadeh D, Ikram MA, Uitterlinden A, Kraaij R, Voortman T. Diet quality and gut microbiome: a large population-based study.

ABSTRACT

Aims To examine associations between diet quality and gut microbiome composition in a large population-based cohort.

Methods We analyzed data of 1130 participants (median age 57 years) from the Rotterdam Study, a prospective cohort study in the Netherlands. We measured dietary intake in 2006-2008 using a semi-quantitative 389-item food-frequency questionnaire and assessed adherence to dietary guidelines for 14 food groups separately and combined into an overall diet quality score (0-14). We collected stool samples in 2012-2013 and assessed gut microbiome composition using 16S rRNA gene sequencing. We used linear regression to examine associations of diet quality score and each one of 14 food groups with alpha diversity (Shannon index and richness). We also used adonis permutation P value calculation to examine whether Bray-Curtis distance of beta diversity differed by diet quality and specific food groups. Furthermore, we used Multivariate Association with Linear Models to examine associations of diet quality and specific food items with taxa (from phylum to genus level). Models were adjusted for technical covariates, total energy intake, age, sex, physical activity, education level, smoking status and BMI.

Results After multivariate adjustment, higher overall diet quality was suggestively associated with higher Shannon index ($p=0.07$) and richness ($p=0.06$). Overall diet quality significantly explained the variations in beta diversity ($p=0.001$). Furthermore, higher diet quality was associated with higher relative abundance of 17 taxa, and lower relative abundance of 12 taxa ($q<0.1$). Specifically, higher diet quality was associated with higher relative abundance of family Ruminococcaceae, Christensenellaceae, Pasteurellaceae, ClostridualesvadinBB60 group, and genus *RuminococcaceaeUCG010*, *RuminococcaceaeUCG005*, *Xylanophilum group*, *RuminococcaceaeUCG002*, *RuminococcaceaeNK4A214group*, *Eligens group*, *Lachnospira, ChristensenellaceaeR7 group*, *RuminococcaceaeUCG003*, *Phascolarctobacterium, Ruminococcus1*, *Haemophilus*, and *RikenellaceaeRC9gut group*. And higher diet quality was associated with lower relative abundance of class Erysipelotrichia, order Erysipelotrichales, family Erysipelotrichaceae, Lachnospiraceae, and genus *Torques group*, *Blautia, Coprococcus3*, *Senegalimassilia, Hallii group*, *Ventriosum group*, *RuminococcaceaeUCG004*, and *ErysipelotrichaceaeUCG003*. Results were not explained by any single food group: higher intake of fruits and vegetables, but also of whole grains, nuts, and tea, and lower intake of red meat and alcohol were all related to microbiome composition.

Conclusions High diet quality may improve overall gut microbial diversities. Diet quality may also influence abundance of certain gut microbial communities, several of which have previously been linked to lower risk of metabolic and inflammatory diseases, such as obesity, diabetes, and Crohn's disease.

INTRODUCTION

An increasing number of studies have uncovered potential associations between gut microbiome composition with chronic diseases, such as inflammatory and metabolic diseases,[1-3] and cancers.[4] For example, a higher diversity of gut microbiome composition has been associated with lower risk of metabolic and inflammatory diseases.[2, 3, 5] Firmicutes and Enterobacteriaceae have also been linked to Crohn's disease.[1] Given this, altering the composition of the gut microbiome has been suggested as a novel and promising approach for the prevention and treatment of these diseases. Many factors, either exogenous or endogenous, may influence the gut microbiome composition.[6] Of the factors studied to date, diet seems to be easiest to modify and presents the simplest route to improve the gut microbiome composition.[7] Recently, a few human studies have observed associations of intake of certain nutrients, such as fat and protein, and certain food items, such as fruits and vegetables, with gut microbiome composition.[8-12] However, with the exception of one study among 1135 participants, most of these previous studies were limited by small sample sizes (n<200). Besides, for other commonly consumed foods that have been shown to contribute to human health, such as whole grains, nuts, fish, and tea, it is largely unknown if and how they relate to gut microbiome composition.

Furthermore, individuals do not consume nutrients or foods in isolation. A single food or nutrient approach may be inadequate in taking into account correlations and interactions in intake of foods and nutrients and their effects on health. A whole diet analysis represents a broader picture of food and nutrient consumption and may be more predictive of the composition of gut microbiome than intake of individual foods or nutrients. To our knowledge, so far only three small observational studies in adults investigated the link between overall diet quality and gut microbial composition.[13-15] Claesson et al. observed in a study among 178 adults, that a higher healthy food diversity index was associated with higher alpha diversity of microbial composition.[13] Mitsou et al. reported that higher adherence to the Mediterranean diet was associated with lower Escherichia coli counts, and higher Bifidobacterium:E. coli ratio in a study among 120 participants.[14] And De Filippis et al. investigated that high-level adherence to a Mediterranean diet was associated with more beneficial gut microbiota in 153 individuals.[15] However, these three studies were limited by small sample sizes, lack of adjustment for important potential confounders, such as energy intake, BMI, physical activity, education level or smoking status.

Therefore, we aimed to examine the associations of overall diet quality and the components (14 food groups) of diet quality with gut microbial composition, including microbial diversity and relative abundance of taxa (from phylum to genus) in a large population-based sample of Dutch middle-aged and elderly.

METHODS

Study population

The current study was embedded within the third sub-cohort of the Rotterdam Study (RS-III), a population-based cohort including 3932 people aged ≥45 years living in the Ommoord District of Rotterdam, the Netherlands. Details on the design of Rotterdam Study are described elsewhere.[16] The baseline examination of the third sub-cohort (RS-III-1) was performed in 2006-2008. Follow-up examinations (RS-III-2) were performed in 2012-2013. The Rotterdam Study has been approved by the institutional review board (Medical Ethics Committee) of Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. The approval has been renewed every 5 years. All participants gave informed consent.[16]

Population for current analyses

Of all 3932 subjects participating in RS-III, 2644 participants had valid data on dietary intake at baseline (RS-III-1). Of this group, we excluded participants who used antibiotics or did not have gut microbiome data available at the second visit (RS-III-2), leaving 1130 participants included in our current analyses (Supplemental Figure 1).

Dietary intake assessment

Dietary intake at RS-III-1 (2006-2008) was assessed with a self-administered semi-quantitative 389-item food-frequency questionnaire (FFQ).[17] Food intake data was converted to energy and nutrient intake using the Dutch Composition table 2011. This FFQ was previously validated in two other Dutch populations using a 9-day dietary record[18] and a 4-week dietary history,[19] which showed Pearson's correlations for intakes of different nutrients varying from 0.40 to 0.86.

Quantification of overall diet quality

We assessed diet quality of the participants as adherence to the Dutch Dietary Guidelines.[20] As described in detail elsewhere,[17] we created an overall diet quality score based on the participants' intake of 14 food groups as specified in these dietary guidelines to measure overall diet quality.[17] The 14 food groups included fruits, vegetables, whole grains, legumes, nuts, tea, percentage of whole grains in total grains, percentage of unsaturated fat in total fat, dairy, fish, red meat, sugary-sweetened beverages, alcohol, and salt. We scored every participant as adhering to the recommendation for each of the food groups ('yes' scored as 1) or not adhering to the item ('no' scored as 0). Supplemental Table 1 shows cut-off values of recommendation for each food group and adherence to the recommendation of the participants. The sum of the number of items adhered to, with a theoretical range from 0 (no adherence) to 14 (full adherence) was used as measure of overall diet quality.[17]

Collection of gut microbiome data

Gut microbiome composition was analyzed at RS-III-2 (2012-2013), as described in detail elsewhere.[21] Briefly, participants were requested to collect a stool sample at their home in sterile tubes and to send the sample by mail to the laboratory of Erasmus MC, Rotterdam, the Netherlands. The time each sample was in the mail between sample production and receipt at Erasmus MC was recorded (time in mail). Subsequently, an automated stool DNA isolation kit (Diasorin, Saluggia, Italy) was used to isolate bacterial DNA. A batch effect during DNA isolation was observed and recorded (run batch). The V3 and V4 hypervariable regions of the bacterial 16S rRNA gene were amplified and sequenced on the Illumina MiSeq platform. After quality filtering and rarefaction analysis, the reads were subsampled at 10000 reads per sample. Reads were clustered into Operational Taxonomic Units (OTUs) using UPARSE (USEARCH version 8). As it has been demonstrated that quality-filtering 16S amplicon sequence reads can greatly improve accuracy of microbial community analysis, OTUs with <0.005% of total sequence reads were filtered out to account for sequencing errors. Taxonomy was assigned to the obtained OTUs using RDP classifier (version 2.12). Therefore, the resulting OTU table contained microbial information at different taxonomic levels, and their abundances for each OTU per sample based on the number of reads in that OTU. Our microbiome data consisted of 11 phyla, 19 classes, 25 orders, 44 families, and 184 genera. From these data, we calculated Shannon index and richness at the OTU-level to quantify alpha diversity, which reflects diversity within communities; and we calculated Bray-Curtis distance at the genus level of gut microbial communities to quantify beta-diversity, which reflects diversity between communities.

Assessment of covariates

Information on smoking status (never, ever, current), education level (primary, lower, intermediate, higher), physical activity, was assessed at baseline (RS-III-1). Information on smoking, and education, was obtained during home interviews. Physical activity was assessed with the LASA Physical Activity Questionnaire and physical activities were weighted by their intensity with Metabolic Equivalent of Task (MET), obtained from the 2011 version of the Compendium of Physical Activities. Height and weight were measured at the research center at RS-III-2 and body mass index (kg/m^2) was calculated. Cases of diabetes, cardiovascular diseases (CVD) and cancers were collected at RS-III-1 and RS-III-2. Information on diabetes, CVD, and cancers was obtained from general practitioners, pharmacies' databases, Nationwide Medical Registry, and follow-up examinations in our research center.[22, 23]

Data analyses

First, we examined associations of the overall diet quality score. Subsequently, we further examined the associations for each one of 14 food groups included the overall diet quality score to explore which

food groups contributed to the associations of overall diet quality with gut microbiome composition. Continuous variables of the 14 different food groups were analyzed and mutually adjusted in analyses.

Association analyses of diet and gut microbial diversity

We used linear regression models to examine associations of overall diet quality score or each of the 14 different food groups with the Shannon index and richness. Furthermore, we assessed how much variations of Bray-Curtis distance could be explained by overall diet quality or the 14 individual food groups to using the special "adonis" function from the Vegan package.

Association analyses between diet and gut microbial communities

We used Multivariate Association with Linear Models (MaAsLin) to examine the associations between overall diet quality score and the 14 individual food groups with relative abundance of gut microbial communities at multiple taxonomic levels: phylum, class, order, family, and genus. These analyses were carried out in R package "MaAsLin".[1] Before entering the models, the percentage of relative abundance of each phylum, class, order, family, and genus was transformed by taking the arcsine of the square root of the proportional value of relative abundance of each genus, family, order, class, and phylum within framework of package of "MaAsLin".

Adjusting for covariates

In the alpha diversity and relative abundance analyses, we adjusted for potential confounding factors including total energy intake, age, sex, time in mail of stool samples, and run batch, physical activity, education level, and smoking status and BMI.

For association analyses of gut microbial diversity, a p value of 0.05 was used to determine statistical significance. For association analyses of relative abundance of gut microbial communities at taxonomic levels, q value of 0.1 (The false discovery rate corrected p value using the Benjamini and Hochberg method[1]) was used to determine statistical significance.

Additional analyses

Effect modification was examined by including interactions of the diet quality score and the 14 food groups with age, sex, and BMI in analyses of alpha diversity of gut microbial composition. In case of significant interactions ($p<0.05$ in alpha diversity analyses), the analyses would be stratified.

We performed sensitivity analyses by excluding participants with diabetes, cardiovascular diseases (CVD), or cancers at RS-III-1 and RS-III-2 to examine whether our results from diversity analyses and relative abundance analyses were robust, since these diseases could change diet and lifestyle habits overtime.

To adjust for potential bias associated with missing data (Supplemental Table 2), missing values on covariates were imputed. For the diversity analyses, a multiple imputation procedure (n=10) was performed using R package "mice". In the analyses of associations of relative abundance of individual species, all missing data of covariates were imputed to the median value using R package "MaAsLin". Statistical procedures were performed with the use of SPSS version 21.0 (IBM Corp, Armonk, NY) and with R version 3.3.1 (The R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Baseline characteristics

Median age of the 1130 participants at baseline was 57 (interquartile-range 53-60) years. Participants had a mean energy intake of 2328 (±693) kcal/d, and a mean diet quality score of 7.1 (±1.9). None of the participants fully adhered to all the guidelines. Table 1 shows intake of the 14 individual food groups included in the diet quality score.

The 1130 participants had high inter-variability of diversity on Shannon index and richness (Figure 1). Table 2 shows relative abundances of the two most abundant taxa in our populations. For example, Firmicutes and Bacteroidetes were most abundant at phylum level, and *Blautia,* and *Bacteroides* were most abundant at genus level.

Associations between diet and gut microbial diversity

For alpha diversity, in multivariable analyses (Table 3, Model 3), a higher diet quality score was suggestively associated with higher Shannon index and richness (Shannon index: β = 0.02, 95%CI (-0.002, 0.03), p=0.07 per 1 unit increase of diet score; richness: β = 4.82, 95%CI (-0.22, 9.85), p=0.06, per 1 unit increase of diet score) (Table 3). For beta diversity, the overall diet quality score explained the variation of Bray-Cutis distance (Table 4). Furthermore, overall diet quality explained more of the variation of Bray-Curtis distance, compared to other factors such as age, sex, physical activity, BMI, education, and smoking (Supplemental Table 3).

When we examined the 14 individual food groups in the diet score, we observed that higher vegetables intake was associated higher Shannon index (β = 0.0002, 95%CI(0.00001, 0.0004), p=0.003), and suggestively associated with higher richness (β = 0.05, 95%CI(-0.006, 0.10), p=0.08); while higher legumes intake was associated lower Shannon index (β = -0.002, 95%CI(-0.004, -0.0009), p=0.001) and richness (β = -0.45, 95%CI (-0.89, -0.01), p=0.04), and that intake of the other 12 individual foods was not associated with Shannon index or richness (Table 5). Among 14 individual food groups, we observed that five food groups, explained to the variations of beta diversity (Bray-Curtis distance):

fruits, vegetables, whole grains, red and processed meat, and tea; and that, two other food groups: sugar-containing beverages and salt, suggestively explained variations of Bray-Curtis distance (Table 4).

Table 1 - Baseline characteristics of study participants (n=1130)

Characteristics	Mean (SD), median (IQR), or %
Age (years)	57.0 (53.0-60.0)
Sex (% female)	58.8%
Physical activity (MET-hours/week)	45.0 (19.0-83.0)
BMI (kg/m^2)	27.3 (4.5)
Education (%)	
- Primary	7.3%
- Lower	33.6%
- Intermediate	27.2%
- Higher	31.9%
Smoking (%)	
- No	33.0%
- Ever	45.7%
- Current	21.3%
Total energy intake (kcal/day)	2328 (693)
Diet quality score	7.1 (1.9)
Food intake (grams/day, unless specified otherwise)	
-Vegetables	219.9(130.9-332.2)
- Fruit	264.9 (125.1-480.8)
- Whole grains	121.3 (79.8-171.2)
- Legumes	8.3 (0-19.4)
- Nuts	8.6 (3.1-17.6)
- Dairy	297.5 (175.6-446.5)
- Fish	16.3 (8.6-28.0)
- Tea	133.9 (13.4-241.1)
- Red and processed meat	74.6 (51.2-106.0)
- Alcohol	8.3 (1.7-20.0)
- Sugar-containing beverages	64.1 (13.4-139.7)
- Salt	5.9 (4.7-7.2)
- Whole grains percentage	71% (53.9%-83.6%)
- Healthy fats percentage	63.9% (50.0%-76.7%)
Shannon index	4.01 (0.50)
Richness	470.5 (152.94)

Overall diet quality score: a higher score indicates a higher adherence to dietary guidelines (theoretical range from 0 to 14).

Abbreviations: MET, metabolic equivalent of task; SD, standard deviation.

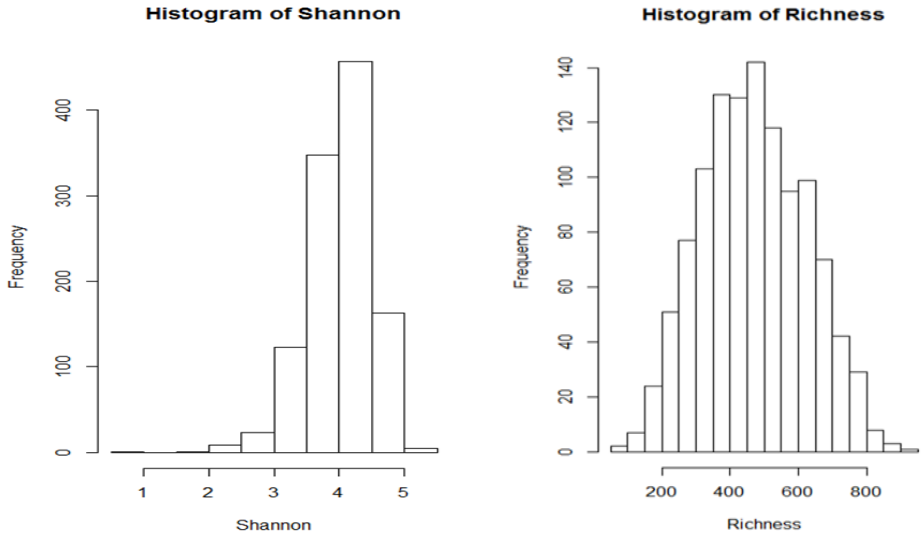


Figure 1. Shannon and richness

Table 2. The two most abundant microbial taxa identified at multiple taxonomic levels

Microbial taxa	Relative abundance (%)
Phylum	
Firmicutes	81.4 (70.0-88.4)
Bacteroidetes	8.4 (3.1-17.4)
Class	
Clostridia	74.9 (63.3-82.6)
Bacteroidia	8.4 (3.1-17.4)
Order	
Clostridiales	74.9 (63.3-82.6)
Bacteroidales	8.4 (3.1-17.4)
Family	
Lachnospiraceae	38.7 (28.8-50.0)
Ruminococcaceae	25.2 (18.7-32.0)
Genus	
Blautia	9.1 (5.1-15.3)
Bacteroides	3.9 (1.4-8.9)

Values are expressed as median (25th percentile-75th percentile IQR) because of their skewed distributions.

Table 3. Associations between overall diet quality score with Shannon index and richness

	Shannon index (n=1130)		Richness (n=1130)	
	β (95% CI)	**P value**	**β (95% CI)**	**P value**
Model 1	0.03 (0.009, 0.04)	0.002	8.09 (3.19, 13.00)	0.001
Model 2	0.02 (0.0006, 0.03)	0.04	5.28 (0.26, 10.31)	0.04
Model 3	0.02 (-0.002, 0.03)	0.07	4.82 (-0.22, 9.85)	0.06

Effect estimates are linear regression coefficients (β) for Shannon index or richness with their 95% confidence intervals (95%CIs) per 1 unit higher overall diet quality score.

Model 1 is adjusted for batch, time-in-mail, total energy intake (kcal), sex (male or female), and age (years).

Model 2 is additionally adjusted for education level (primary, lower/intermediate, intermediate, or higher), smoking status (never, ever, current); physical activity (MET-hours/week).

Model 3 is additionally adjusted for BMI (kg/m^2).

Abbreviations: BMI, body mass index; CI, confidence interval; MET, metabolic equivalent of task.

Results with p values < 0.05 are regarded as statistically significant associations.

Table 4. Associations of overall diet quality score and intake of the 14 food groups with Bray-Curtis distance

	Bray-Curtis distance (n=1130)
	R^2 (p value)
Overall diet quality	0.006 (0.001)
Food groups	
- Fruits	0.005 (0.001)
- Vegetables	0.002 (0.008)
- Whole grains	0.002 (0.002)
- Nuts	0.001 (0.31)
- Legumes	0.001 (0.10)
- Whole grains percentage	0.002 (0.008)
- Dairy	0.0006 (0.81)
- Fish	0.001 (0.21)
- Healthy fat percentage	0.0008 (0.59)
- Red meat	0.002 (0.008)
- Tea	0.001 (0.04)
- SCBs	0.001 (0.05)
- Alcohol	0.001 (0.14)
- Salt	0.001 (0.09)

The associations with Bray-Curtis distance were examined using function adonis from R package vegan. The summary statistics of the adonis analysis are summarized in the table, including partial R-squares (R^2), and p values out of 1000 permutations.
Results with p values < 0.05 are regarded as statistically significant associations
Abbreviations: SCBs, sugar-containing beverages

Table 5. Associations of 14 foods groups with Shannon index and richness

Food groups	Shannon index (n=1130)		Richness (n=1130)	
	β (95% CI)	P value	β (95% CI)	P value
Fruits	0.00007 (-0.00004. 0.0001)	0.21	-0.004 (-0.03, 0.03)	0.83
Vegetables	0.0002 (0.00001, 0.0004)	0.03	0.05 (-0.006, 0.10)	0.08
Whole grains	0.0001 (-0.0004, 0.0008)	0.50	0.07 (-0.12, 0.25)	0.48
Nuts	0.001 (-0.0009, 0.003)	0.29	0.33 (-0.26, 0.93)	0.27
Legumes	-0.002 (-0.004, -0.0009)	0.001	-0.45 (-0.89, -0.01)	0.04
Whole grain percentage	0.0002 (-0.002, 0.002)	0.81	0.25 (-0.28, 0.79)	0.35
Dairy	-0.00001 (-0.0001, 0.0001)	0.89	-0.01 (-0.05, 0.03)	0.53
Fish	0.0006 (-0.0008, 0.002)	0.37	0.20 (-0.25, 0.65)	0.38
Healthy fat percentage	-0.001 (-0.002, 0.0003)	0.12	-0.20 ()-0.59, 0.20)	0.33
Red &processed meat	0.0001 (-0.0007, 0.0008)	0.87	0.03 (-0.21, 0.27)	0.81
Tea	-0.00004 (-0.0002, 0.0001)	0.55	-0.002 (-0.04, 0.04)	0.93
SCBs	-0.00002 (-0.0003, 0.0002)	0.55	-0.04 (-0.12, 0.04)	0.33
Alcohol	0.0004 (-0.002, 0.003)	0.68	0.11 (-0.58, 0.80)	0.34
Salt	0.000005 (-0.00003, 0.00004)	0.73	-0.005 (-0.01, 0.004)	0.34

Effect estimates are linear regression coefficients (β) for Shannon index or richness with their 95% confidence intervals (95% CI). The multivariate linear regression model included 14 food groups, and the covariates: batch, time-in-mail, total energy intake (kcal), sex (male or female), age (years), education level (primary, lower/intermediate, intermediate, or higher), smoking status (never, ever, current); physical activity (MET-hours/week), and BMI (kg/m^2) (equivalent to model 3 in table 2).
Results with p values < 0.05 are regarded as significant associations
Abbreviations: BMI, body mass index; CI, confidence interval; MET, metabolic equivalent of task; SCBs, sugar-containing beverages

Associations between diet and relative abundance of gut microbial taxa

After multivariable adjustment, higher diet quality was associated with higher relative abundance of 17 taxa, and lower relative abundance of 12 taxa (q<0.10). Specifically, higher diet quality was associated with higher relative abundance of family Ruminococcaceae, Christensenellaceae, Pasteurellaceae, ClostridualesvadinBB60 group, and genus *RuminococcaceaeUCG010*,

RuminococcaceaeUCG005, *Xylanophilum group*, *RuminococcaceaeUCG002*, *RuminococcaceaeNK4A214group*, *Eligens group*, *Lachnospira*, *ChristensenellaceaeR7 group*, *RuminococcaceaeUCG003*, *Phascolarctobacterium*, *Ruminococcus1*, *Haemophilus*, and *RikenellaceaeRC9gut group*. And higher diet quality was associated with lower relative abundance of class Erysipelotrichia, order Erysipelotrichales, family Erysipelotrichaceae, Lachnospiraceae, and genus *Torques group*, *Blautia*, *Coprococcus3*, *Senegalimassilia*, *Hallii group*, *Ventriosum group*, *RuminococcaceaeUCG004*, and *ErysipelotrichaceaeUCG003* (Table 6). Supplemental Table 4 shows the results for diet quality score and all taxa: 11 phyla, 19 classes, 25 orders, 44 families, 184 genera. For the 14 individual food groups, we observed that fruits, vegetables, whole grains, nuts, tea, red and processed meat, and alcohol were associated with 19 taxa (Table 7). For example, higher intake of fruits was associated with higher relative abundance of the family Ruminococcaceae, and the genus *Lachnospira*, and *xylanophilum group*. Higher intake of tea was associated with higher relative abundance of genus *LachnospiraceaeNK4A136 group*. Supplemental Table 5 shows the results for the 14 individual food groups and all taxa.

Additional analyses

The estimates of overall diet quality score or individual food groups with gut microbial diversity and relative abundance of taxa were similar in analyses restricted to 1013 participants without prevalent diabetes, CVD or cancers at RS-III-1 and RS-III-2 (Supplemental Tables 6-9). The estimates of overall diet quality or individual food groups with gut microbial alpha diversity did not differ by age, sex, or BMI (all interactions p values > 0.05).

Table 6. Statistically significant associations of overall diet quality score with relative abundance of gut microbial taxa

Gut microbial taxa	β coefficient	P value	Q value
Class Erysipelotrichia	-0.004	0.00005	0.002
Order Erysipelotrichales	-0.004	0.00005	0.002
Family Erysipelotrichaceae	-0.004	0.00005	0.002
Family Ruminococcaceae	0.008	0.0002	0.005
Family Lachnospiraceae	-0.01	0.0003	0.007
Family Christensenellaceae	0.005	0.001	0.01
Family Pasteurellaceae	0.0005	0.005	0.07
Family ClostridialesvadinBB60 group	0.0005	0.007	0.09
Genus *Torques group*	-0.004	0.00006	0.00008
Genus *RuminococcaceaeUCG010*	0.003	0.00006	0.00008
Genus *RuminococcaceaeUCG005*	0.003	0.00002	0.0002
Genus *Xylanophilum group*	0.002	0.00002	0.001
Genus *RuminococcaceaeUCG002*	0.005	0.0004	0.002
Genus *Blautia*	-0.008	0.00006	0.003
Genus *RuminococcaceaeNK4A214 group*	0.003	0.0001	0.005
Genus *Eligens group*	0.003	0.0003	0.01
Genus *Coprococcus3*	-0.003	0.0004	0.01
Genus *Senegalimassilia*	-0.001	0.0004	0.01
Genus *Lachnospira*	0.002	0.0007	0.02
Genus *Hallii group*	-0.004	0.001	0.02
Genus *ChristensenellaceaeR7 group*	0.005	0.001	0.03
Genus *Ventriosum group*	-0.001	0.001	0.03
Genus *RuminococcaceaeUCG003*	0.001	0.002	0.03
Genus *Phascolarctobacterium*	0.003	0.003	0.06
Genus *RuminococcaceaeUCG004*	-0.001	0.004	0.06
Genus *Ruminococcus1*	0.003	0.005	0.07
Genus *Haemophilus*	0.0005	0.005	0.07
Genus *ErysipelotrichaceaeUCG003*	-0.002	0.005	0.07
Genus *RikenellaceaeRC9gut group*	0.001	0.007	0.09

Multivariate analysis was performed using MaAsLin to examine the associations between overall diet quality score and relative abundance of gut microbial communities at multiple taxonomical levels. Only findings with $q < 0.10$ are presented in the table. In total we analyzed associations of overall diet quality score with 11 phyla, 19 classes, 25 orders, 44 families, and 184 genera as outcomes. Batch, time-in-mail, age (years), sex (male or female), total energy intake (kcal), education level (primary, lower/intermediate, intermediate, or higher), smoking status (never, ever, current); physical activity (MET-hours/week), and BMI (kg/m^2) were included as covariates (model 3). Results with q values < 0.10 are regarded as significant associations.

Abbreviations: MaAsLin, Multivariate Association with Linear Models; BMI, body mass index; MET, metabolic equivalent of task.

Table 7. Statistically significant associations of 14 food groups with relative abundance of gut microbial taxa

Food groups	β coefficient	P value	Q value
Fruits			
Famlily Ruminococcaceae	0.00006	0.0001	0.02
Family Lachnospiraceae	-0.00006	0.002	0.08
Genus *Lachnospira*	0.00002	0.00007	0.009
Genus *xylanophilumgroup*	0.00001	0.0001	0.02
Genus *eligensgroup*	0.00002	0.0008	0.05
Genus *LachnospiraceaeUCG001*	0.00008	0.001	0.07
Vegetables			
Class Verrucomicrobiae	0.00002	0.002	0.08
Family Verrucomicrobiaceae	0.00002	0.002	0.09
Whole grain percentage			
Genus *RuminococcaceaeNK4A214group*	0.0002	0.0007	0.05
Tea			
Genus *LachnospiraceaeNK4A136group*	0.00003	0.0004	0.04
Genus *LachnospiraceaeUCG004*	0.00001	0.001	0.06
Nuts			
Genus *Anaerofilum*	-0.0001	0.0006	0.05
Alcohol			
Class Actinobacteria	-0.0008	0.0008	0.04
Order Bifidobacteriales	-0.0008	0.0005	0.04
Family Bifidobacteriaceae	-0.0008	0.0005	0.0006
Genus *Bifidobacterium*	-0.0008	0.001	0.06
Genus *Butyricicoccus*	0.0002	0.002	0.096
Fish			
Genus *PrevotellaceaeNK3B31group*	0.0003	0.002	0.08
Red and processed meat			
Genus *Slackia*	0.00004	0.001	0.08

Multivariate analysis was performed using MaAsLin to examine the associations between 14 food groups and relative abundance of microbial communities at multiple taxonomical levels. Only findings with $q < 0.10$ are presented in the table. In total we analyzed associations with 11 phyla, 19 classes, 25 orders, 44 families, and 184 genera as outcomes. Batch, time-in-mail, age (years), sex (male or female), total energy intake (kcal), education level (primary, lower/intermediate, intermediate, or higher), smoking status (never, ever, current); physical activity (MET-hours/week), and BMI (kg/m^2) were included as covariates (model 3).

Results with q values < 0.10 are regarded as statistically significant associations.

Abbreviations: MaAsLin, Multivariate Association with Linear Models; BMI, body mass index; MET, metabolic equivalent of task.

DISCUSSION

In this large population-based cohort, we observed that higher overall diet quality was suggestively associated with more alpha diversity, independent of a wide range of other lifestyle factors. Diet quality also explained variations of beta diversity. Furthermore, we found that higher overall diet quality was associated with higher relative abundance of 17 taxa and lower relative abundance of 12 taxa. Of the 29 taxa, five (Erysipelotrichaceae, Ruminococcaceae, Lachnospiraceae, Christensenellaceae, *Torques group*) have previously been linked to obesity, diabetes, and Crohn's disease.[24-27] Additionally, we also observed that the associations of overall diet quality with gut microbiome were not fully explained by any single food group, although main contributors were fruits and vegetables. Besides fruits and vegetables, whole grains, tea, red meat, fish, and alcohol were also associated with gut microbiome composition. This suggests the importance of overall diet quality in shaping gut microbiome composition.

Overall, evidence for associations between overall diet quality and gut microbiome composition is limited. To date, only three previous studies investigated associations between overall diet quality and gut microbiome composition in adults.[13-15] They reported that a healthy food diversity index was positively associated with alpha diversity: Shannon index and richness;[13] that greater adherence to a Mediterranean diet was associated with more counts of Bacteroides spp., and lower counts of Escherichia coli.;[14] and with increased Prevotella and some fiber-degrading Firmicutes.[15] However, these previous studies were limited by small sample sizes (n<200),[13-15] lack of adjustment for important confounders,[13-15] and lack of investigation on beta diversity.[14, 15] Our study is the first to examine the associations between overall diet quality with gut microbial diversity including alpha diversity and beta diversity and relative abundance of gut microbial taxa in large population (n=1130), taking into account a wide range of potential confounders. Our study not only confirmed that higher diet quality was associated with more alpha diversity of gut microbiome, which is in line with the study conducted by Claesson et al.,[13] but also added novel evidence that diet quality also explained beta diversity. Furthermore, from analyses of beta diversity, we found that diet explained more variation than physical activity, BMI, or smoking, highlighting the importance of diet in improving gut microbiome composition. Furthermore, we observed 29 associations between overall diet quality and taxa, which not only confirmed association of diet quality with the family Lachnospiraceae and the genus *Lachnospira*, but also reported novel associations of overall diet quality with more specific gut microbial communities, such as the families: Ruminococcaceae, Christensenellaceae, and Erysipelotrichaceae,

and the genera: *RuminococcaceaeUCG002*, *RuminococcaceaeUCG003*, *RuminococcaceaeUCG005, RuminococcaceaeUCG010*, *xylanophilum group*, *Blautia, RuminococcaceaeNK4A214 group*, *Eligens group*, and *Coprococcus3*. Some of these specific taxa have been linked to chronic diseases. For example, higher relative abundance of the class Erysipelotrichia, family Ruminococcaceae and Christensenellaceae might protect against obesity[28] and inflammatory bowel disease, like Crohn's disease.[24] However, the role of some taxa that were also observed in our analyses, such as *Torques group* and *Eligens group* in human health warrants further research.

To clarify which food groups were mainly responsible for the associations between overall diet quality with gut microbiome composition, we further examined the associations between the 14 individual food groups of overall diet quality with gut microbial diversity, and relative abundance of gut microbial taxa. We found that beneficial associations of overall diet quality were not fully driven by any single food group, but main contributors were higher intake of fruits and vegetables, and to a lesser extent whole grains, nuts, and tea, and lower intake of red meat and alcohol. Few previous studies observed associations or effect of individual foods on gut microbial diversity and relative abundance of gut microbial communities at multiple taxonomical levels.[3, 4] In line with our findings, Zhernakova et al.[8] observed that higher intakes of fruit and vegetables were correlated with more alpha diversity of gut microbiome.[5] However, the previous study by Zhernakova et al. was limited by lack of adjustment for important confounders, such as physical activity, smoking status, and education level. Another important different point was that Zhernakova et al. investigated the associations of individual food groups with relative abundance of gut microbial taxa at species level, and our study investigated the associations at genus, family, order, class, and phylum levels, which could be complementary approaches to understand how diet may affect gut microbial taxa.

Finally, we also found that diet quality may be differentially associated with members with similar features from the same family, the same order, the same class, or the same phylum. For example, we found that overall diet quality was positively associated with genus *Lachnospira*, a member of the family Lachnospiraceae, but overall diet quality was inversely associated with genus *Coprococcus3*, another member of family Lachnospiraceae. These findings support the importance of recognizing that individual members even with certain similar features and from the same family, order, class, or phylum might be influenced heterogeneously by diet.

Implications

Our findings shed light on gut microbiome composition as potential link behind the associations between diet and several chronic diseases. We found positive associations of diet quality with gut microbial diversity and previous studies have indicated that more gut microbial diversity may reduce risk of T2D,[2] obesity and Crohn's disease.[5] Furthermore, we observed that better diet quality was

associated with more abundance of the families Ruminococcaceae and Christensenellaceae and lower abundance of the families Erysipelotrichaceae and Lachnospiraceae. Previous studies have shown that higher relative abundance of family Ruminococcaceae might protect against Crohn's disease and other inflammatory bowel disease;[24] that family Christensenellaceae was enriched in participants with lower BMI;[25] that family Erysipelotrichaceae was enriched in participants with Crohn's diseases;[26] and that family Lachnospiraceae increased susceptibility to Crohn's disease risk.[27] This suggests that diet may influence risk of Crohn's diseases, obesity, diabetes and other chronic diseases through potential effects on gut microbial diversity and abundance of certain gut microbial communities, such as family Ruminococcaceae, Christensenellaceae, Erysipelotrichaceae, and Lachnospiraceae.

Our findings also have important clinical and public health implications. We found that no single individual food groups could fully explain the beneficial associations of diet quality with gut microbial diversity and gut microbial taxa, highlighting the importance of an overall healthy dietary pattern. Although intake of fruits and vegetables were, as expected, important for the associations, also other food groups such as nuts and tea seemed to contribute to gut microbial composition. Furthermore, the gut microbial taxa related to nuts and tea were not the same as those related to fruits and vegetables, further highlighting the importance of overall diet rich in various healthy foods on gut microbiome composition. Our findings support that high diet quality characterized by higher intake of fruits, vegetables, whole grains, and tea; and lower intake of red meat and alcohol may be a strategy to improve gut microbiome composition and thereby improve health.

Study strengths and limitations

The strengths of our study include the large sample size (n=1130); the use of not only overall diet quality but also individual foods; the comprehensive analysis of alpha diversity, beta diversity and gut microbial communities at multiple taxonomic levels; and the adjustment for several important confounders, such as energy intake, physical activity, education level, and smoking status.

Our study has limitations we should consider. First, we only measured both dietary intake and gut microbiome composition only once, whereas both may change over time. However, the estimates were similar after exclusion participants with diabetes, CVD, or cancers who were likely to change dietary habits and whose microbiome may be affected by disease. Second, misclassification of dietary intake could have occurred. However, given the prospective study design, this measurement error was likely to be non-differential. Non-differential measurement error in dietary intake would have attenuated observed associations. Third, although we adjusted for many potential confounders, the possibility of residual confounding cannot be dismissed. Finally, our results may be generalizable only to people of similar age and race and replication in other populations is needed.

Conclusions

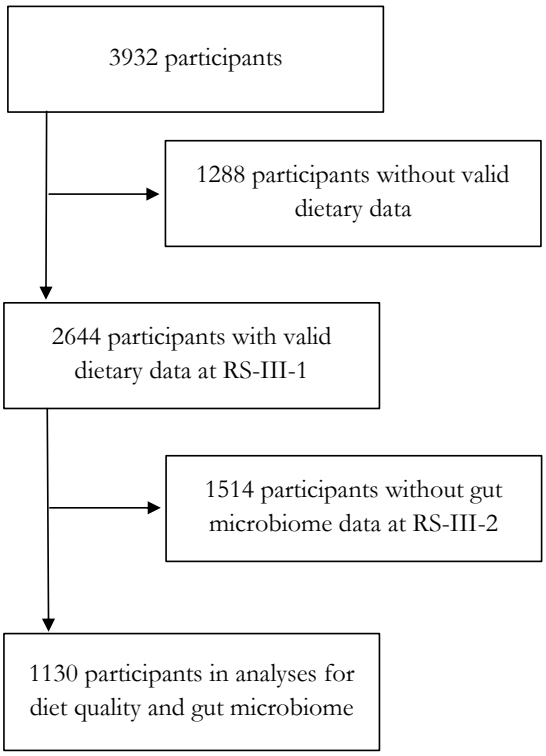
High overall diet quality may improve overall gut microbial diversity and enhance relative abundance of certain microbial communities that have previously been linked to lower risk of metabolic and inflammatory diseases, thereby likely facilitating a more beneficial microbiome composition for prevention of chronic diseases. Furthermore, our results suggest that the potential benefit of better diet quality for gut microbiome composition is not driven by any single food group but is explained by various foods including high intake of fruits, vegetables, whole grains, nuts, and tea and low intake of red meat.

REFERENCES

1. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13(9):R79.
2. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55-60.
3. Hartstra AV, Bouter KEC, Bäckhed F, Nieuwdorp M. Insights into the role of the microbiome in obesity and type 2 diabetes. Diabetes care. 2015;38(1):159-65.
4. Schwabe RF, Jobin C. The microbiome and cancer. Nature Reviews Cancer. 2013;13(11):800-12.
5. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature. 2012;489(7415):220-30.
6. Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, et al. Metagenomic analysis of the human distal gut microbiome. science. 2006;312(5778):1355-9.
7. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334(6052):105-8.
8. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016;352(6285):565-9.
9. Bamberger C, Rossmeier A, Lechner K, Wu L, Waldmann E, Altenhofer J, et al. Walnut-Enriched Diet Beneficially Alters Gut Microbiome in Healthy Caucasian Subjects. Am Heart Assoc; 2017.
10. Henning SM, Yang J, Shao P, Lee R-P, Huang J, Ly A, et al. Health benefit of vegetable/fruit juice-based diet: Role of microbiome. Scientific Reports. 2017;7.
11. Lisko DJ, Johnston GP, Johnston CG. Effects of Dietary Yogurt on the Healthy Human Gastrointestinal (GI) Microbiome. Microorganisms. 2017;5(1):6.
12. Sheflin AM, Melby CL, Carbonero F, Weir TL. Linking dietary patterns with gut microbial composition and function. Gut microbes. 2017;8(2):113-29.
13. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, et al. Gut microbiota composition correlates with diet and health in the elderly. Nature. 2012;488(7410):178-84.
14. 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(12):1645-55.
15. De Filippis F, Pellegrini N, Vannini L, Jeffery IB, La Storia A, Laghi L, et al. High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. Gut. 2016;65(11):1812-21.
16. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. European Journal of Epidemiology. 2017:1-44.
17. Voortman T, Kiefte-de Jong JC, Ikram MA, Stricker BH, van Rooij FJA, Lahousse L, et al. Adherence to the 2015 Dutch dietary guidelines and risk of non-communicable diseases and mortality in the Rotterdam Study. Eur J Epidemiol. 2017.

18. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. The American journal of clinical nutrition. 1993;58(4):489-96.
19. Goldbohm RA, Van den Brandt PA, Brants HAM, Van't Veer P, Al M, Sturmans F, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. European journal of clinical nutrition. 1994;48(4):253-65.
20. Kromhout D, Spaaij CJ, de Goede J, Weggemans RM. The 2015 Dutch food-based dietary guidelines. Eur J Clin Nutr. 2016;70(8):869-78.
21. Radjabzadeh D, Uitterlinden AG, Kraaij R. Microbiome measurement: Possibilities and pitfalls. Best Practice & Research Clinical Gastroenterology. 2017.
22. Bos D, Portegies ML, van der Lugt A, Bos MJ, Koudstaal PJ, Hofman A, et al. Intracranial carotid artery atherosclerosis and the risk of stroke in whites: the Rotterdam Study. JAMA Neurol. 2014;71(4):405-11.
23. Zhangling Chen OHF, Sander Lamballais, M. Arfan Ikram, Josje D. Schoufour, Taulant Muka, Trudy Voortman. Associations of protein intake with longitudinal insulin resistance and prediabetes and type 2 diabetes and the role of obesity over time. Submitted. 2017.
24. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology. 2014;146(6):1489-99.
25. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. 2014;159(4):789-99.
26. Kaakoush NO. Insights into the Role of Erysipelotrichaceae in the Human Host. Frontiers in cellular and infection microbiology. 2015;5.
27. Martinez-Medina M, Denizot J, Dreux N, Robin F, Billard E, Bonnet R, et al. Western diet induces dysbiosis with increased E coli in CEABAC10 mice, alters host barrier function favouring AIEC colonisation. Gut. 2013:gutjnl-2012-304119.
28. Pindjakova J, Sartini C, Lo Re O, Rappa F, Coupe B, Lelouvier B, et al. Gut dysbiosis and adaptive immune response in diet-induced obesity vs. systemic inflammation. Frontiers in microbiology. 2017;8:1157.

SUPPLEMENTAL MATERIAL



Supplemental Figure 1. Participants selection

Supplemental Table 1. Adherence to the 14 food groups of the diet quality score of 1130 participants

Food intake (grams/day)	Percentages (%)
Vegetables ≥200 g/day	56.3
Fruit ≥200	59.3
Whole grains ≥ 90	69.4
Legumes ≥ 135 g/week	30.4
Nuts ≥15 g/day	29.8
Dairy ≥ 350g/day	38.8
Fish ≥ 100 g/week	56.1
Tea ≥ 450 mL/day	9.8
Red and processed meat ≤ 300 g/week	18.2
Alcohol ≤ 10 g/day	53.7
Sugar-sweetened beverages (SCBs) ≤ 150 mL/day	77.6
Salt ≤ 6 g/day	52.7
Whole grains ≥ 50% of total grains	79.4
Unsaturated fats and oils ≥ 50% of total fat	75.2

Categorical variable expressed as percentages

Supplemental Table 2. Number and percentages of missing values of covariates in 1130 participants before multiple imputation

Covariates	Number (percentages)
Education level	3 (0.27%)
Smoking status	2 (0.18 %)
Physical activity	81 (7.1%)

For all covariates, there were only missing values on education level, smoking status and physical activity before imputation.

Supplemental Table 3. Variations of Bray-Curtis explained by age, sex, BMI, physical activity, education level, and smoking status

	Bray-Curtis distance (n=1130) R^2 (p value)
Total energy	0.002 (0.04)
Age	0.001 (0.06)
Sex	0.002 (0.05)
BMI	0.0008 (0.51)
Physical activity	0.0007 (0.62)
Education level	0.003 (0.14)
Smoking status	0.0008 (0.51)

The associations of total energy, age, sex, BMI, physical activity, education level, and smoking status with Bray-Curtis distance are examined using function adonis from R package "Vegan". The summary statistics of the adonis analysis are summarized in the table, including partial R-squares (R^2), and p values out of 1000 permutations.
Results with p values < 0.05 are regarded as significant associations.

Supplemental Table 4. Associations between overall diet quality score and all gut microbial taxa: 11 phyla, 19 classes, 25 orders, 44 families, and 184 genera

Supplemental Table 5. Associations between 14 food groups and all gut microbial taxa: 11 phyla, 19 classes, 25 orders, 44 families, and 184 genera

Supplemental Tables 4-5 can be obtained from the author.

Supplemental Table 6. Associations of overall diet quality score and 14 food groups with Shannon index and richness in sensitivity analyses restricting to the participants without prevalent diabetes, cardiovascular diseases, and cancers (n=1013)

	Shannon index (n=1013) β (95% CI) and p value	**Richness (n=1013) β (95% CI) and p value**
Overall diet quality	0.01 (-0.003, 0.03)	4.31 (-0.98, 9.60)
Fruits	0.00005 (-0.00007, 0.0002)	-0.01 (-0.05, 0.03)
Vegetables	0.0002 (0.00003, 0.0004)	0.05 (-0.01, 0.11)
Whole grains	0.0002 (-0.0004, 0.0008)	0.08 (-0.12, 0.28)
Nuts	0.001 (-0.0009, 0.003)	0.43 (-0.20, 1.05)
Legumes	-0.002 (-0.004, -0.0009)	-0.38 (-0.84, 0.08)
Whole grain percentage	0.0002 (-0.002, 0.002)	0.26 (-0.32, 0.84)
Dairy	-0.00006 (-0.0002, 0.00007)	-0.02 (-0.07, 0.02)
Fish	0.0009 (-0.0007, 0.003)	0.27 (-0.22, 0.76)
Healthy fat percentage	-0.001 (-0.003, 0.0001)	-0.27 (-0.69, 0.14)
Red and processed meat	0.0001 (-0.0007, 0.001)	0.08 (-0.18, 0.33)
Tea	-0.00003 (-0.0002, 0.0001)	-0.0007 (-0.04, 0.04)
SCBs	-0.00009 (-0.0004, 0.0002)	-0.04 (-0.12, 0.04)
Alcohol	-0.0002 (-0.003, 0.002)	-0.08 (-0.82, 0.65)
Salt	0.000006 (-0.00003, 0.00004)	-0.005 (-0.02, 0.005)

Effect estimates are linear regression coefficients (β) for Shannon index or richness with their 95% confidence intervals, from multivariate linear regression models including overall diet quality score or 14 food groups, and adjusting for batch, time-in-mail, total energy intake (kcal), sex (male or female), age (years), education (primary, lower/intermediate, intermediate, or higher), smoking status (never, ever, current); physical activity (MET-hours/week), and BMI (kg/m^2).

Abbreviations: SCBs, sugar-containing beverages; BMI, body mass index; CI, confidence interval; MET, metabolic equivalent of task.

Results with p values < 0.05 are regarded as significant associations.

Supplemental Table 7. Associations of overall diet quality score and 14 food groups with Bray-Curtis distance in sensitivity analyses restricting to the participants without prevalent diabetes, cardiovascular diseases, and cancers

	Bray-Curtis distance (n=1013) R^2 (p value)
Overall diet quality	0.06 (0.001)
Fruits	0.004 (0.001)
Vegetables	0.002 (0.03)
Whole grains	0.002 (0.04)
Nuts	0.001 (0.30)
Legumes	0.001 (0.17)
Whole grains percentage	0.002 (0.007)
Dairy	0.0008 (0.63)
Fish	0.001 (0.38)
Healthy fat percentage	0.0009 (0.57)
Red meat	0.002 (0.02)
Tea	0.001 (0.09)
SCBs	0.002 (0.03)
Alcohol	0.001 (0.18)
Salt	0.002 (0.08)

The associations of overall diet quality score and the 14 food groups with Bray-Curtis distance were examined using the function adonis from R package vegan. The statistics of the Adonis analysis are summarized in the table, including partial R-squares (R^2), and p values out of 1000 permutations.
Abbreviations: SCBs, sugar-containing beverages.
Results with p values < 0.05 are regarded as significant associations.

Supplemental Table 8. Associations of overall diet quality score with all gut microbial taxa: 11 phyla, 19 classes, 25 orders, 44 families, and 184 genera, in sensitivity analyses restricting to participants without prevalent diabetes, cardiovascular diseases, or cancers (n=1013)

Supplemental Table 9. Associations of 14 food groups with all gut microbial taxa: 11 phyla, 19 classes, 25 orders, 44 families, and 184 genera, in sensitivity analyses restricting to participants without prevalent diabetes, cardiovascular diseases, or cancers (n=1013)

Supplemental Tables 8-9 can be obtained from the author.

Chapter 5

General Discussion

GENERAL DISCUSSION

AIMS

The main aim of this thesis was to investigate the role of nutrition and gut microbiome in type 2 diabetes (T2D) risk. Regarding nutrition, I was interested in dietary protein intake and plant-based diet, because evidence for dietary protein intake was inconsistent, and because evidence for plant-based diets, defined as a low frequency of animal-based foods, and T2D risk was very limited. To better understand the role of nutrition in T2D, including its early stages and its consequences, we also investigated associations with obesity, insulin resistance, prediabetes, and mortality. Another interest in my research was gut microbiome composition as an important potential determinant of T2D risk, which in turn may be modified by diet. In this part, I was interested in not only microbial alpha and beta diversities, but also gut microbial taxa at phylum, order, class, family, and genus levels. We studied associations between this detailed composition of the microbiome and T2D risk, and we also investigated associations of overall diet quality and food groups intake with gut microbiome composition.

MAIN FINDINGS

Chapter 2: The role of nutrition in T2D risk

In Chapter 2, we studied associations of dietary protein intake and plant-based diet with T2D risk, and additionally with obesity, insulin resistance, prediabetes risk, and risk of all-cause and cause-specific mortality. We observed that in middle-aged and elderly Dutch participants, higher intake of animal protein was associated with higher insulin resistance, and risk of prediabetes and T2D, which did not differ by protein from meat, fish or dairy. In contrast, plant protein intake was not associated with insulin resistance, and risk of prediabetes and T2D, which also did not differ by more specific protein sources, including protein from legumes and nuts, from potatoes, from grains, or from vegetables and fruits. We also observed that higher total or animal protein intake was associated with higher all-cause mortality and cardiovascular diseases (CVD) mortality, although not with cancer mortality and other mortality. And plant protein intake was not associated with all-cause and cause-specific mortality in middle-aged and elderly Dutch participants. These findings for animal protein intake and mortality were supported in a meta-analysis pooling results from the Rotterdam Study and other cohorts. However, this meta-analysis indicated that higher plant protein intake was associated with lower all-cause and CVD mortality. In line with our findings on animal and plant protein, in another separate analysis, we observed a beneficial association between an overall more plant-based and less animal-based diet with adiposity, insulin resistance, and risk of prediabetes and T2D.

Chapter 3: The role of gut microbiome in T2D risk

In chapter 3, we investigated associations between gut microbiome with insulin resistance and T2D risk. In the Rotterdam Study we observed that higher alpha diversity (higher Shannon index, richness, or Inverse Simpson index) was associated with lower insulin resistance or T2D risk. Insulin resistance and T2D also may explain beta diversity (Bray-Curtis distance). Furthermore, we also observed that more abundance of family Christensenellaceae, genus *Marvinbryantia,* genus *RuminococcaceaeUCG005,* genus *RuminococcaceaeUCG008,* genus *RuminococcaceaeUCG010,* and genus *RuminococcaceaeNK4A214group* was associated with lower insulin resistance; and that more abundance of family Clostridiaceae, family Peptostreptococcaceae, genus *Clostridiumsensustricto,* genus *Intestinibacter,* or genus *Romboutsia* was associated with lower prevalence of T2D. Moreover, we also observed similar results for alpha diversity, beta diversity, and gut microbial taxa in the Lifelines-Deep Study. Furthermore, a meta-analysis of results from the two cohort studies corroborated the results of the Rotterdam Study.

Chapter 4: The link between nutrition and gut microbiome

In chapter 4, we studied the associations between overall diet quality as adherence to Dutch dietary guidelines and the 14 food groups included in the diet quality score with gut microbiome composition. In the Rotterdam Study, we observed that higher overall diet quality was suggestively associated with higher alpha diversity (higher Shannon index, and richness); that diet quality explained the variation of the beta diversity (Bray-Curtis distance); and that overall diet quality was associated with relative abundance of 29 gut microbial taxa. Some of the taxa, such as the family Erysipelotrichia, and Ruminococcaceae, have previously suggested to be related to inflammatory and metabolic diseases. Furthermore, we also observed that most of the individual food groups included in the diet quality were associated with gut microbiome composition. For example, higher intake of fruit and vegetable was associated with higher alpha diversity. Fruits, vegetables, legumes, whole grains, fish, and meat explained beta diversity. Fruits, vegetables, legumes, nuts, tea, whole grains, and meat were all associated with relative abundance of certain gut microbial taxa.

Taken together, our studies have indicated that nutrition and gut microbiome composition may influence the development of T2D, and that gut microbiome composition may be modified by nutrition. On basis of these results, I think that gut microbiome might be a mechanism or mediator for the associations between nutrition and the development of T2D.

METHODOLOGICAL CONSIDERATIONS

In this section, I present and discuss some of the methodological issues that I faced in identifying the associations of nutrition and gut microbiome with T2D. In particularly, I focus on methodological

issues of the overall study design, and the measurement of nutrition and gut microbiome composition. Furthermore, I also highlight emerging methodological trends in the assessment of nutrition and gut microbiome, in the context of the aims of my research.

Study design and study population

For the studies in this thesis, most data were from the Rotterdam Study, and replication analyses were performed in the Lifelines-Deep Study, which are both population-based prospective cohorts. The Rotterdam Study consists of middle-aged and elderly participants in Ommoord district of the Rotterdam city.[1] The Linelines-Deep Study contains participants aged 18 years or older from the northern regions of the Netherlands.[2]

Therefore, all the data were of observational nature. When interpreting results of our studies, both internal validity and external validity should be considered. Regarding internal validity, three different types of bias should be taken into account, i.e. selection bias, information bias, and confounding. Here I first discuss selection bias and confounding; information bias is discussed in the next paragraph on dietary assessment.

In this thesis, the analyses population from the Rotterdam Study tended toward a selection of a healthier population with a higher social-economic status, compared to the populations from the Rotterdam Study that could not be included into our analyses due to various reasons, such as lack of measures of dietary intake data at baseline. However, previous studies have indicated that selective non-participation at baseline is not likely to be related to future risk of diseases and therefore do not strongly influence associations, making bias due to selection less likely.[3] However, the selection of participants still may affect the external validity of our findings, which should be considered when extending the application of our findings into other populations. For example, the Rotterdam Study and the Lifelines-Deep Study included general populations living in a Rotterdam suburb and the northern parts of the Netherlands, respectively. Therefore, our findings from these studies may not be completely generalizable to populations in other regions or countries where populations for example may have different dietary patterns and social economic status, such as Asian and African populations. A second type of bias that can threaten the internal validity in our observational studies, is confounding. Although the rich data in the Rotterdam Study and the Lifelines-Deep Study allowed us to adjust for various possible confounders in different associations studied in this thesis, the possibility of residual confounding cannot be completely ruled out. For example, in chapter 2.2 level of physical activity was measured at the third visit of the Rotterdam Study, while dietary intake questionnaires were completed in the first visit of the first sub-cohort of the Rotterdam Study. In chapters 3 and 4, I could not adjust for the status of the stool samples. Therefore, we cannot fully exclude residual confounding by the levels of physical activity, and the status of stools. The residual

confounding can lead to either overestimation or underestimation of the observed effect estimate. Given this, a simplistic and favorite response to concern about residual confounding and causality is to conduct a randomized controlled trial. Conducting a randomized controlled trial for research on nutrition and gut microbiome, or for gut microbiome and glucose metabolism could be possible, because evidence has shown that gut microbiome composition could be changed in 24 hours by diet,[4] and that insulin resistance of individuals with metabolic syndrome could be improved by fecal microbiota transplantation in a six-weeks trial.[5] However, doing randomized trails is often infeasible in research on nutrition and chronic diseases, because decades of follow-up are needed for clinically relevant outcomes, such as T2D and CVD, to develop. When the potential for interventional research is limited, several other approaches such as Mendelian randomization analysis[6] and Directed Acyclic Graphs,[7] are considered to help to infer causality. However, these approaches face other challenges. A main challenge is that these approaches are based on a few underlying assumptions that are hard to verify in practice. For Mendelian randomization for example, strong claims of causality cannot be justified when the assumptions required for the instrumental variable analysis such as reproducible in multiple independent samples, and functionally related to the exposures, would be violated.[8] Another helpful way of inferring causality could be to consider different types of exposure (i.e., dietary patterns, foods, nutrients, and biomarkers) and different types of data, such as longitudinal data[9] within frameworks of well-conducted prospective cohort studies. We would consider whether findings from well-conducted prospective cohort studies keep in line with findings from other types of studies, such as animal studies, mechanic studies in human, randomized trials of intermediate trials of intermediate outcomes, and the Mendelian randomization analysis, if possible. If these findings are taken together to arrive a consensus, which can strengthen the inference of causality. For example, adherence to a plant-based diet has been suggested to have a beneficial effect on cardiometabolic intermediate risk factors in observational studies[10] and RCTs.[11] Staples of a plant-based diet such as fruits, vegetables, and whole grains, have been individually linked to lower risk of cardiometabolic risk factors.[12, 13] Reviews of major nutrients abundant in these foods, such as fibers, unsaturated fats and polyphenols, have confirmed this finding as well.[10] Furthermore, adherence to a plant-based diet has also been associated with lower risk of cardiovascular hard endpoints, such as adiposity,[14] T2D,[15] and cardiovascular mortality.[16-18] Besides, these findings are supported by some biological mechanisms and pathways.[10] Such a convergence among these studies provides convincing support for adoption of a plant-based diet in prevention of CVD. Overall, corroborating data from multiple study types and populations can enhance the weight of evidence and help to infer causality. Furthermore, valid conclusions and policy decisions for dietary recommendation also need to evaluate and quantify sources of biases and to use the totality of the best available evidence, which is an iterative process.[19]

Nutrition assessments

For our studies involving dietary protein intake and plant-based diet, we used semi-food questionnaires (FFQs) to collect dietary intake data. The FFQ method is most widely used for dietary assessment in epidemiological studies for two reasons.[19, 20] One reason is that researchers are generally interested in average relative long-term dietary intake, rather than on one or several specific days, and an FFQ measures this habitual dietary intake. The other reason is that FFQs are relatively easy to complete for study participants and relatively easy to process in large quantities without high cost compared to measurements of dietary biomarkers; these practical aspects make an FFQ as the method of choice for large studies. The main limitation of an FFQ is that a self-reported retrospective dietary assessment method, making measurement-errors likely. This measurement error is an important source of information bias in studies on dietary intake in relation to health or disease. This measurement error, or misclassification of exposure, is assumed to be mainly non-differential, indicating that it is random and not related to the outcomes under study. Non-differential measurement error of the exposure may result in attenuation of the observed associations and in wider confidence intervals. Hence, it leads to underestimation of associations and reduces statistical power to detect associations. However, errors in dietary intake assessment may also be differential, i.e., related to outcome. For example, evidence has indicated that obese people are more likely to underreport their habitual food intake than people with a normal weight.[21] When one examines the association of diet with obesity or outcomes closely related to obesity, this underreporting could therefore lead to differential misclassification of exposure, which could be associated an overestimation or underestimation of the associations. We expect that the dietary measurement errors in our studies are unlikely to be strongly related to outcomes. However, differential measurement error cannot be ruled out.

In our studies, we took some statistical-related methods to attempt to account for these potential measurement errors of dietary intake. For example, in analyses for dietary protein intake or a plant-based diet score and outcomes, we adjusted for total energy intake. Energy adjustment addresses not only confounding by energy, but also the measurement error that is related to energy intake.[22] Furthermore, in analyses for dietary protein intake, we also used nutrients-density models.[23] Additionally, in analyses for a plant-based diet score, we created this plant-based diet score by using relative scores (quintiles) of intake of individual food items rather than absolute intakes, and the FFQs were shown in several validation studies to adequately rank subjects according to intake.[24-27]

Another aspect of dietary assessment is repeated measurements of diet using FFQs. I acknowledge that repeated measurements of diet are also particularly useful in representing long-term dietary habits. Unfortunately, we did not measure dietary intake repeatedly among all participants of the Rotterdam Study, only in a subgroup of participants, with many years apart and using a different updated FFQ.

Given this, we only conducted sensitivity analyses using these repeated dietary intake data available, and similar results were observed after adjustment for dietary intake at six to eight years of follow-up. Moreover, we also conducted several sets of sensitivity analyses by excluding participants who were expected to have changed their dietary intake over time and observed similar results. Additionally, our findings for dietary protein intake, plant-based diets and T2D risk were also similar with those observed in Nurse Health Study, Nurse Health Study II, and Health Professionals Follow-Up Study that had repeated measurements of dietary intake over time among a large US population during a long-time follow-up.

Although FFQs have been widely used to measure dietary data in nutritional epidemiology research, and well-designed FFQs have been shown to adequately rank participant according to their usual diet, novel methods may improve the accuracy of dietary assessment, thereby benefit research of nutritional epidemiology. To date, objective biomarkers of nutrient intake or nutrient status (such as urine nitrogen levels for protein intake, blood levels of fat acids for certain fat intake, and carotenoids for certain plant-based foods) have been considered as powerful complementary tools to further improve the accuracy of dietary assessment.[19] Especially, new omics technologies such as metabolomics might hold potentials as biomarkers of dietary intake or overall diet patterns, because metabolomics can measure the full profile of small-molecule metabolites in biofluids, thereby probably providing a comprehensive picture of an individual's overall dietary intake.[28] Overall, much effort is being paid to improve assessment of biomarkers for individual nutrients or foods and overall dietary patterns. However, improvement of dietary assessment is challenging, for example, some dietary biomarkers, such as fat biomarkers, appear not to accurately reflect specific dietary fat intake, which may be caused by some factors, such as genetic variability, lifestyle, and physiological factors.[29] Furthermore, while many studies have indicated the association between dietary patterns and metabolomics profiles, only limited studies have shown the ability to classify or assign people into certain dietary patterns based on the metabolomics profiles as biomarkers.[28] And these concentration biomarkers will not reflect only intake but also metabolism of the individual.[19] Besides, these dietary biomarkers measures are generally more expensive and invasive, such as need of blood samples, which limits the wide use in large-scale epidemiological studies.

Given this, I would consider that these objective measures, as complementary tools, rather than a replacement of self-reported FFQ to further improve assessment of dietary intake in nutritional epidemiologic studies.[11] Further research on dietary biomarkers should be directed at: 1) refining existing dietary biomarkers by accounting for confounders, such as genetic variability, and lifestyles; 2) discovering new valid biomarkers of individual nutrients, foods, and overall dietary patterns; and 3) developing new measure technologies that would be cost-effective, non-invasive, and rapid for use among large populations.

Gut microbiome composition

For our studies involving gut microbiome, we measured gut microbiome composition using 16S rRNA method in the Rotterdam Study and in the Lifelines-Deep Cohort Study.

Briefly, we collected the stool samples from the participants. Once all samples were collected, we used the 16S rRNA method to detect gut microbiome composition. So far, 16S rRNA profiling is the most direct and cost-effective approach to obtain phylogenetic profiles. Nevertheless, 16S rRNA profiling has several typical limitations, including bias introduced by hyper-variable region selection and the profiling pipeline, the inability to detect novel, unknown operational taxonomic units (OTUs), overestimation of alpha diversity, and difficulty to compare samples with varying numbers of reads.[30] The effect of some of these limitations can be covered in the bioinformatic pipeline as well as by large integrated studies.[30] Therefore, to minimize these limitations, in the process of bioinformatics pipeline of our Rotterdam Study, we only included OTUs clustering on basis of homology of the reads, and OTUs with <0.005% of total sequence reads were filtered out to account for sequencing errors. We also excluded OTUs presented in less than 10% of samples. Furthermore, when we analyzed the role of gut microbiome in the development of T2D risk by analyzing data from the Rotterdam Study and the Lifelines-Deep Study, to provide a platform for robust and reliable results, we further standardized all the procedures and protocols for the Rotterdam Study and the Lifelines-Deep Study, for which we implemented the 16S data processing pipeline, which comprised a naive Bayesian classifier from the Ribosomal Database Project, and the recent, SILVA database release 128: we only analyzed taxonomical results using genus and higher taxonomic levels.[31] This OTU-independent approach was utilized to decrease domain-dependent bias. However, there are also some main limitations of the 16S rRNA method that could not be covered in the bioinformatics pipeline and using larger integrated studies. For example, in this method specific genes are not directly sequenced, but rather predicted based on the OUT, therefore, the 16S rRNA method often reports less precise gut microbiome data at the species level. An alternative approach to the 16S rRNA method is whole metagenome shotgun sequencing in which random fragments of genome are sequenced.[32, 33] Compared with 16S rRNA method, whole metagenome shotgun sequencing can capture sequences from all the organisms, including accurate taxa at the species and lower levels, viruses and fungi. Furthermore, whole metagenome shotgun sequencing can be used to identify rare or novel organisms in the community, which 16S rRNA method cannot do. Additionally, it is less susceptible to the biases that are inherent in targeted gene amplification.[33] Perhaps most interestingly, whole metagenome shotgun sequencing method can also provide direct information about the presence or absence of specific functional pathways in samples, also known as the 'hologenome'. This can provide potentially important information about the capabilities and functions of the organisms in the community.[33] However, whole metagenome shotgun method is more expensive and requires more extensive data analysis. Recently, another new approach to measure gut microbiome composition, the metatranscriptome, has

been developed.[34] Compared with the 16S rRNA and the whole metagenome shotgun methods, the metatranscriptome method estimates which microorganisms in a community are actively transcribing, and inherently discriminates between active live organisms versus dormant or dead microorganisms and extracelluar DNA. Therefore, it can capture dynamic intra-individual variation, and directly evaluates microbial activity, including responses to intervention and event exposure. However, this latter method is more expensive. To summarize, 16S rRNA method highlights high-level community profiling, whole metagenome shotgun sequencing highlights functional profiling, and metatranscriptome sequencing highlights real-time functional profiling. Therefore, after conducting 16S rRNA method to gain a low-resolution understanding of the gut microbiome composition, researchers could move on to metsgenome sequencing and metatranscriptome sequencing to further capture function profile of gut microbiome composition.[34]

Additionally, like data on dietary intake in our main analysis, gut microbiome data were also measured once. However, gut microbiome is a complex, and dynamic ecosystem, which can easily change over time,[34] thereby, repeated measurements of gut microbiome over time in longitudinal cohort study are particularly useful to further understand gut microbiome.

PUBLIC HEALTH AND CLINICAL IMPLICATIONS & DIRECTIONS FOR FUTURE RESEARCH

In the studies presented in this thesis, I have sought to respond to a series of research questions related to the role of nutrition and gut microbiome in T2D risk. Additionally, I also investigated the associations of nutritional factors with obesity, insulin resistance, prediabetes and mortality. In this section, I conclude by briefly foregrounding some of the studies' implication for public health and clinical practice, and some of the directions for future research that stem from these studies and expand to this whole field.

Public health and clinical implications

I conducted the studies, with special attention to the public health and clinical practice whereby my studies made the results knowledgeable for researchers, medical professionals, policy makers, and even public readers. Accordingly, the first major public health and the clinical practical contribution derives from our findings on dietary protein intake, and a plant-based diet and T2D risk. Our findings point out that high total protein intake, especially high animal protein intake may increase T2D risk; instead, adherence a plant-based diet may reduce T2D risk. Overall, these studies have indicated the importance of foods sources, supporting more plant-based foods intake and less animal-based foods intake. However, I have also felt that more effective strategies and actions are needed to effectively

translate these existing nutritional knowledges or dietary guidelines into public health practice, because in our research I saw that diets of individuals in the Rotterdam Study and several other studies remained far from optimal. In these populations, the total amount of protein intake was usually higher than the amount recommended by WHO, and animal-based foods were usually the main source.[35] Furthermore, in our Rotterdam Study population, the individuals had only 7 or less points out of 14 on a scale of adherence to the most recent Dutch dietary guidelines.[24] In this sense, I believe that our research is especially timely, which calls for the communities to further improve nutrition practice, such as lower intake of animal-based foods, and also call for more effective strategies by the scientists, physicians, policy-makers, nutritionists, medias, and the communities to better transfer these existing nutritional knowledges into public health practice. For example, nutritional education in schools and communities should be greatly encouraged.

A second important implication of our research derives from our findings on associations between gut microbiome with insulin resistance and T2D. These findings may provide insights into the etiology of T2D, potential targets for the therapies, and safety and effectiveness of the treatment. For example, it is possible, that increasing gut microbial diversity and abundance of certain bacteria, such as butyrate-producing bacteria, (e.g. family Clostridum) might be a promising approach to prevent and treat T2D. Furthermore, as some drugs, (e.g. antibiotics) could have adverse effect on the gut microbiome composition, which might further fuel unbalanced gut microbiome composition of T2D patients, I would advise caution in use of these drugs among T2D patients. Additionally, previous evidence has indicated that gut microbiome composition is related to response to chemotherapy and immunotherapy, thereby, the response to chemotherapy and immunotherapy might differ among for example cancers patients with and without T2D.[36]

Finally, our findings suggest that high overall diet quality may improve gut microbial diversity, along with a beneficial change of abundance of certain bacteria, which seems to be explained by various food items, not by any single foods item. These findings have indicated the importance of nutritional factors, especially overall diet quality for gut microbiome composition. Therefore, it is very likely that in a next future, a targeted modulation of the gut microbiome through ad hoc dietary interventions, used along or combined with the administration of mixtures of gut microbial species, may improve gut microbiome composition, which would benefit prevention and treatment of T2D and other diseases. Gut microbiome composition, in turn, might also be used to personalize diet, which together may thereby hold potential for enhancing public health.[36]

Directions for future research

Our studies have answered some research questions about nutrition, gut microbiome, and T2D; but also raised a number of additional questions for future research. More research will in fact be needed to refine, elaborate and extend most of our novel findings.

First, in line with previous studies, our studies in this thesis have indicated that lower intake of animal protein intake, and a more plant-based diet are associated with lower insulin resistance, and lower risk of prediabetes and T2D and other health events. However, our studies and most previous studies were embedded in European or North American populations. In these populations, a western dietary pattern is more likely and nutrition excess is of concern. For example, the total amount of protein intake is usually higher than the amount recommended by WHO, and animal products are usually the main source. Therefore, further studies in other populations who are more likely to have different dietary patterns, such as Asian and African populations, are needed. Additionally, further research is needed not only in general populations but also in more specific populations with health conditions, where nutrition requirements may differ. These efforts will help make targeted dietary recommendations and define optimal nutrients ranges and overall dietary patterns for different populations in different geographic locations and health stages.[19] Moreover, further research on mechanisms through which nutritional factors influence health is needed. New molecular fields of nutritional epidemiological research have developed by remarkable advances in omics technologies, including genomics, metabolomics, and proteomics, and by the study of the human gut microbiome. Research on these new fields will provide molecular insights on mechanisms pathways, which will help to discover novel biomarkers of nutritional factors, understand individual variability in dietary responses, and identify high-risk T2D populations to target for intervention. Additional aspects of nutrition for T2D risk deserve to be investigated further, such as effects of contaminants, food processing, and cooking methods. Our food supply and personal choices are constantly changing over time, so that new issues are continuously emerging, such as effects of highly manufactured meat alternatives and gluten-free diets.[37] Last, as I have addressed above, further research regarding how to effectively translate existing nutritional knowledge or dietary guidelines into public health practice is also needed.

Second, we have observed the associations of gut microbiome composition with insulin resistance and T2D. However, our study was based on a cross-sectional study, which failed to distinguish whether alterations of the gut microbiome were a cause or consequence of changes in difference of insulin resistance and T2D risk. Therefore, future research should further attempt to explore the temporal direction and causality in the framework of longitudinal repeated measures of gut microbiome and clinical interventional studies. In this process, we could also in turn explore how T2D influences gut microbiome composition. Furthermore, future research could further explore the mechanisms behind

the role of gut microbiome composition in T2D risk. For this aim, much work is needed. For example, we could extend our research from investigation of effect of gut composition profile into that of gut microbiome function using metagenome sequencing and metatranscriptome sequencing data. We also could explore the effect of metabolomics of gut microbiome on T2D risk. Besides, more replication analyses for gut microbiome and T2D risk among various populations are needed, as gut microbiome composition varied to some extent by different populations. Finally, based on the existing knowledge, we should further develop more effective strategies to apply these existing knowledges to early prevent progression or even the overt manifestation of T2D in public health and clinical practice settings (e.g. dietary interventions including prebiotics, probiotics, and FMT).

Third, we have explored the associations between nutrition and gut microbiome in Chapter 4. Similar to our study on gut microbiome and T2D risk, the study on overall diet quality and gut microbiome could also be extended: 1) to replicate the findings in various populations; 2) to infer causality of the findings; and 3) to further explore mechanisms behind the associations of diet and gut microbiome. Furthermore, given that diet habit could change over time, we could further elaborate if and how change of nutrition including overall diet quality and specific foods items over time influences gut microbiome composition over time. Moreover, it would be necessary to extend the existing evidence by exploring the associations between prebiotic foods and organic foods and gut microbiome composition. For example, we could explore if and how prebiotic foods, such as garlic and onions, influence gut microbiome composition; how effects of natural prebiotic foods compare to probiotic supplements; and if and how organic foods influence gut microbiome composition. Besides, further research could take a perspective of clinical practice and ask how to improve gut microbiome through dietary intervention in various specific patients, such as cancer patients. Additionally, we could also investigate whether gut microbiome can influence food choices and appetite, which could lead to positive feedback loops when these dietary changes in turn alter the gut microbiome. Overall, to date, insufficient public health and clinical evidence exists to draw clear conclusions or firm recommendations based on gut microbiome composition. Further research is needed to infer the causality for the known associations, and to further explore the potential effect for probiotic foods, organic foods, and food additives. The potential research will help to update dietary guidelines and develop precision nutrition approach to benefit public health and clinical practice.

Finally, on basis of all the studies presented in the thesis, I think that gut microbiome could be a mechanism and mediator behind the associations of nutrition and T2D. However, the current work in the thesis has not shown more specific evidence of how the gut microbiome mediates the associations between nutrition and T2D, therefore, more work is needed to examine how gut microbiome mediates the associations in detail, which will help to develop precision nutrition strategies for preventing and treating T2D in clinical and public health settings.

Precision nutrition for preventing and treating T2D is an emerging new research direction. It aims to tailor personalized dietary interventions or recommendations by integrating traditional nutritional factors research and new molecular mechanisms research (e.g. gut microbiome research, genetics research, and metabolomics research).[38] Currently, precision nutrition for T2D and other diseases is still in its infancy and much research is needed before it can be widely used in clinical and public health settings. There are many challenges to be faced in the field of precision nutrition, such as a lack of robust and reproducible results, the high cost of omics technologies, and methodological issues in study design as well as high-dimensional data analyses and interpretation.[38] Further research is needed to address these issues. Furthermore, as precision nutrition research is moving towards prevention and treatment of T2D, parallel efforts, such as precision medicine, are also needed to make the precision approaches more completed. Overall, personalized precision nutrition approach by integrating findings from traditional nutritional factors research and new molecular mechanisms research, such as gut microbiome research, together with other parallel efforts, might have the potential to reduce the burden of illness and disability due to T2D and its related disorders, which thereby points to a new direction for research of prevention and treatment of T2D and its related diseases.

CONCLUSIONS

The findings of the studies in this thesis provide new recommendations and implications for prevention of the development of T2D. Specifically, lower animal protein intake, and higher degree of adherence to plant-based diet may reduce T2D risk. More gut microbial diversity and beneficial change of certain gut microbial communities (e.g. butyrate-producing bacteria) may benefit T2D risk, which might be achieved by improving overall diet quality and higher intake of specific plant-based foods, such as vegetables, fruits, nuts, whole grains, and lower intake of certain animal-based foods, such as red and processed meat. Overall, our findings give novel insights regarding pathophysiology of T2D and indicates potential mechanisms related to gut microbiome underlying associations between nutrition and T2D. Awaiting further research, these findings carry potential to contribute to improvement of T2D and its related cardio-metabolic events, treatment, and prognosis.

REFERENCES

1. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. European journal of epidemiology. 2017;32(9):807-50.
2. Tigchelaar EF, Zhernakova A, Dekens JAM, Hermes G, Baranska A, Mujagic Z, et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. BMJ open. 2015;5(8):e006772.
3. Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? Epidemiology. 2006:413-8.
4. 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(7484):559.
5. Vrieze A, Van Nood E, Holleman F, Salojärvi J, Kootte RS, Bartelsman JFWM, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. Gastroenterology. 2012;143(4):913-6. e7.
6. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. PLoS medicine. 2008;5(8):e177.
7. Textor J, Hardt J, Knüppel S. DAGitty: a graphical tool for analyzing causal diagrams. Epidemiology. 2011;22(5):745.
8. Qi L. Mendelian randomization in nutritional epidemiology. Nutrition reviews. 2009;67(8):439-50.
9. Satija A, Yu E, Willett WC, Hu FB. Understanding nutritional epidemiology and its role in policy. Advances in nutrition. 2015;6(1):5-18.
10. Satija A, Hu FB. Plant-based diets and cardiovascular health. Trends in Cardiovascular Medicine. 2018.
11. Wang F, Zheng J, Yang B, Jiang J, Fu Y, Li D. Effects of vegetarian diets on blood lipids: a systematic review and meta-analysis of randomized controlled trials. Journal of the American Heart Association. 2015;4(10):e002408.
12. Wang X, Ouyang Y, Liu J, Zhu M, Zhao G, Bao W, et al. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. Bmj. 2014;349:g4490.
13. 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. bmj. 2016;353:i2716.
14. Chen Z, Schoufour JD, Rivadeneira F, Lamballais S, Ikram MA, Franco OH, et al. Plant-based diet and adiposity over time in a middle-aged and elderly population: the Rotterdam Study. Epidemiology. 2018.
15. Chen Z, Zuurmond MG, van der Schaft N, Nano J, Wijnhoven HAH, Ikram MA, et al. Plant versus animal based diets and insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. European journal of epidemiology. 2018;33(9):883-93.

16. Satija A, Bhupathiraju SN, Spiegelman D, Chiuve SE, Manson JE, Willett W, et al. Healthful and unhealthful plant-based diets and the risk of coronary heart disease in US adults. Journal of the American College of Cardiology. 2017;70(4):411-22.
17. Baden MY, Liu G, Satija A, Li Y, Sun Q, Fung TT, et al. Abstract MP38: Association of Changes in Plant-Based Diet Indices with Total and Cause-Specific Mortality. Circulation. 2019;139(Suppl_1):AMP38-AMP.
18. Kim H, Caulfield LE, Rebholz CM. Healthy plant-based diets are associated with lower risk of all-cause mortality in US adults. The Journal of nutrition. 2018;148(4):624-31.
19. Hu FB, Willett WC. Current and Future Landscape of Nutritional Epidemiologic Research. Jama. 2018.
20. Willett W. Nutritional epidemiology: Oxford University Press; 2012.
21. Goris AHC, Westerterp-Plantenga MS, Westerterp KR. Undereating and underrecording of habitual food intake in obese men: selective underreporting of fat intake. The American journal of clinical nutrition. 2000;71(1):130-4.
22. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. The American journal of clinical nutrition. 1997;65(4):1220S-8S.
23. Hu FB, Stampfer MJ, Rimm E, Ascherio A, Rosner BA, Spiegelman D, et al. Dietary fat and coronary heart disease: a comparison of approaches for adjusting for total energy intake and modeling repeated dietary measurements. American journal of epidemiology. 1999;149(6):531-40.
24. Voortman T, Kiefte-de Jong JC, Ikram MA, Stricker BH, van Rooij FJA, Lahousse L, et al. Adherence to the 2015 Dutch dietary guidelines and risk of non-communicable diseases and mortality in the Rotterdam Study. European journal of epidemiology. 2017;32(11):993-1005.
25. Klipstein-Grobusch Kd, Den Breeijen JH, Goldbohm RA, Geleijnse JM, Hofman A, Grobbee DE, et al. Dietary assessment in the elderly: validation of a semiquantitative food frequency questionnaire. European journal of clinical nutrition. 1998;52(8):588.
26. Goldbohm RA, van den Brandt PA, Brants HAM, van't Veer P, Al MDM, Sturmans F, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. European journal of clinical nutrition. 1994;48(4):253-65.
27. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. The American journal of clinical nutrition. 1993;58(4):489-96.
28. Bhupathiraju SN, Hu FB. One (small) step towards precision nutrition by use of metabolomics. The Lancet Diabetes & Endocrinology. 2017;5(3):154-5.
29. Arab L. Biomarkers of fat and fatty acid intake. The Journal of nutrition. 2003;133(3):925S-32S.
30. Radjabzadeh D, Uitterlinden AG, Kraaij R. Microbiome measurement: Possibilities and pitfalls. Best Practice & Research Clinical Gastroenterology. 2017.
31. Wang J, Kurilshikov A, Radjabzadeh D, Turpin W, Croitoru K, Bonder MJ, et al. Meta-analysis of human genome-microbiome association studies: the MiBioGen consortium initiative. BioMed Central; 2018.
32. Radjabzadeh D, Uitterlinden AG, Kraaij R. Microbiome measurement: Possibilities and pitfalls. Best Practice & Research Clinical Gastroenterology. 2017;31(6):619-23.

33. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochemical and biophysical research communications. 2016;469(4):967-77.
34. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nature Reviews Microbiology. 2018:1.
35. Chen Z, Franco OH, Lamballais S, Ikram MA, Schoufour JD, Muka T, et al. Associations of specific dietary protein with longitudinal insulin resistance, prediabetes and type 2 diabetes: The Rotterdam Study. Clinical nutrition. 2019.
36. Valdes AM, Walter J, Segal E, Spector TD. Role of the gut microbiota in nutrition and health. Bmj. 2018;361:k2179.
37. Willett WC. Diet and health—finding a path to Veritas. European journal of epidemiology. 2018;33(2):127-35.
38. Wang DD, Hu FB. Precision nutrition for prevention and management of type 2 diabetes. The lancet Diabetes & endocrinology. 2018;6(5):416-26.

Chapter 6

Summaries

Summary

Chapter 1 provides a general introduction on nutrition, gut microbiome, and type 2 diabetes (T2D), and a description of studies on which this thesis is based.

In Chapter 2, we examined associations between nutrition and T2D. Chapter 2.1 provides associations between dietary protein intake and insulin resistance, and the risk of prediabetes and T2D. We observed that higher total or animal protein intake was associated with higher insulin resistance, and risk of prediabetes and T2D, while plant protein intake was not associated with these outcomes. Chapter 2.2 shows associations between dietary protein intake and all-cause and cause-specific mortality. We observed that higher total protein intake was associated with higher all-cause mortality and cardiovascular mortality, which was mainly driven by higher animal protein intake. Higher plant protein intake was associated with lower all-cause and cardiovascular mortality. We further extended our research by investigating associations between a plant-based diet and the development of T2D and adiposity in Chapter 2.3, and Chapter 2.4. In Chapter 2.3, we observed that higher adherence to a plant-based diet was associated with lower insulin resistance, and lower risk of prediabetes and T2D. In Chapter 2.4, we observed that higher adherence to a plant-based diet was associated with less adiposity over time.

Chapter 3 focuses on associations of gut microbiome and insulin resistance and T2D. In this chapter, we observed that higher Shannon index or richness was associated with lower insulin resistance and prevalence of T2D. Insulin resistance and T2D may explain beta diversity of gut microbiome. Furthermore, we also observed that more abundance of family Christensenellaceae, genus *Marvinbryantia*, genus *RuminococcaceaeUCG005*, genus *RuminococcaceaeUCG008*, genus *RuminococcaceaeUCG010*, and *genus RuminococcaceaeNK4A214group* was associated with lower insulin resistance; and that more abundance of family Clostridiaceae1, family Peptostreptococcaceae, genus *Clostridiumsensustricto1*, genus *Intestinibacter*, or genus *Romboutsia* was associated with lower prevalence of T2D.

In Chapter 4, we observed associations between nutrition and gut microbiome. We observed that better overall diet quality was associated with a higher Shannon index. Overall diet quality also may explain beta diversity of gut microbiome. Furthermore, overall diet quality was associated with 21 microbial genera, 6 microbial families, 1 microbial order, and 1 microbial class. Additionally, we observed that the associations between overall diet quality and gut microbiome composition could not fully be explained by any single component of the diet quality. Most of the components, such as fruits, vegetables, whole grains, nuts, and meat were all associated with gut microbiome composition.

In Chapter 5, the interpretation of the main findings is discussed as well as the methodological considerations, implications, and recommendations for future studies.

Samenvatting

Hoofdstuk 1 geeft een algemene inleiding over voeding, darmmicrobioom en type 2 diabetes, en een beschrijving van de studies waarop dit proefschrift is gebaseerd.

In hoofdstuk 2 hebben we de associaties tussen voeding en diabetes type 2 onderzocht. Hoofdstuk 2.1 gaat over de associaties tussen eiwitinname en insulineresistentie en het risico op prediabetes en type 2 diabetes. We zagen dat een hogere inname van totaal eiwit of van eiwit uit dierlijke bronnen geassocieerd was met hogere insulineresistentie en een hoger risico op prediabetes en type 2 diabetes; terwijl eiwit uit plantaardige bronnen niet was geassocieerd met deze uitkomsten. Hoofdstuk 2.2 laat de associaties zien tussen eiwitinname en sterfte. We hebben gevonden dat hogere totale eiwitinname geassocieerd was met hogere mortaliteit en met cardiovasculaire mortaliteit, maar niet met mortaliteit door kanker of andere oorzaken. Dit verband van eiwit werd voornamelijk veroorzaakt door een hogere inname van dierlijke eiwitten. Een toename van plantaardige eiwitinname was geassocieerd met een lagere mortaliteit en met lagere cardiovasculaire mortaliteit. We hebben ons onderzoek verder uitgebreid door associaties tussen een meer plantaardig voedingspatroon en de ontwikkeling van type 2 diabetes te onderzoeken. We hebben gevonden dat een meer plantaardig en minder dierlijk voedingspatroon was geassocieerd met minder adipositas (Hoofdstuk 3.4) en met minder insulineresistentie en een lager risico op prediabetes en type 2 diabetes (Hoofdstuk 2.3).

Hoofdstuk 3 richt zich op de associaties van het darmmicrobioom, insulineresistentie en type 2 diabetes. In dit hoofdstuk hebben we vastgesteld dat een hogere Shannon index of richness geassocieerd was met lagere insulineresistentie en prevalentie van type 2 diabetes. Bovendien kunnen insulineresistentie en type 2 diabetes de bèta diversiteit van het darmmicrobioom tot op zekere hoogte verklaren. Nog belangrijker is dat we observeerde dat een overvloed van de familie Christensenellaceae, genus *Marvinbryantia*, genus *RuminococcaceaeUCG005*, genus *RuminococcaceaeUCG008*, genus *RuminococcaceaeUCG010*, en genus *RuminococcaceaeNK4A214group* geassocieerd was met een lagere insuline resistentie; terwijl een overvloed van de familie Clostridiaceae1, familie Peptostreptococcaceae, genus *Clostridiumsensustricto1*, genus *Intestinibacter*, en genus *Romboutsia* geassocieerd was met een lagere prevalentie van type 2 diabetes resistentie.

In hoofdstuk 4 hebben we associaties gevonden tussen voeding en darmmicrobioom. We constateerden dat een betere algehele kwaliteit van het voedingspatroon geassocieerd was met een hogere Shannon-index en met 21 bacteriën op genus-niveau, 6 bacteriën op familie-niveau, 1 bacterie op order-niveau en 1 bacterie op klasse-niveau. De algehele kwaliteit van het voedingspatroon leek de bètadiversiteit ook enigszins te verklaren. Bovendien hebben we onderzocht welke componenten van de voeding deze associaties kunnen verklaren. We hebben vastgesteld dat de verschillende componenten, zoals fruit, groenten, volle granen, noten en vlees allemaal geassocieerd zijn met het darmmicrobioom.

In hoofdstuk 5 wordt de interpretatie van de belangrijkste bevindingen besproken, evenals de methodologische overwegingen, implicaties en aanbevelingen voor toekomstige studies.

Authors' affiliations

Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands
Zhangling Chen, *Trudy Voortman*, *M. Arfan Ikram*, *Oscar H. Franco*, *Josje D. Schoufour*, *Maria G. Zuurmond*, *Taulant Muka*, *Marija Glisic*, *Sander Lamballais*, *Hamid Aliahmad*, *Xiaofang Zhang, Alice Moumdjian, Valentina Gonzalez-Jaramillo, Niels van der Schaft*

Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands
Djawad Radjabzadeh, Robert Kraaij, André Uitterlinden, Fernando Rivadeneira

Medical Library, Erasmus MC, Rotterdam, the Netherlands
Wichor M. Bramer

Department of Health Sciences, Faculty of Earth and Life SciencesVrije Universiteit, Amsterdam, the Netherlands
Hanneke Wijnhoven

Department of Genetics and Department of Pediatrics, University of Groningen, University Medical Center Groningen, Groningen, RB, Groningen, the Netherlands
Jingyuan Fu, Lianmin Chen, Alexander Klurilshikov, Alexander Zhernakova

Department of Epidemiology and Department of Nutrition, Harvard TH Chan School of Public Health, Harvard University, USA
Mingyang Song

List of manuscripts

1. Chen Z, Franco OH, Lamballais S, Ikram MA, Schoufour JD, Muka T, Voortman T. Associations of specific dietary protein with longitudinal insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. Clinical Nutrition. 2019. DOI: 10.1016/j.clnu.2019.01.021

2. Chen Z, Schoufour JD, Rivadeneira F, Lamballais S, Ikram MA, Franco OH, Voortman, T. Plant-based diet and adiposity over time in a middle-aged and elderly population: the Rotterdam Study. Epidemiology. 2019;30(2):303-10.

3. Chen Z*, Zuurmond MG*, Van der Schaft N, Nano J, Wijnhoven HAH, Ikram MA, Franco OH, Voortman T. Plant versus animal based diets and insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. European Journal of Epidemiology. 2018;33(9):883-93.

4. Chen Z, Glisic M, Song M, Aliahmad H, Zhang X, Moumdjian A, Gonzalez-Jaramillo V, Van der Schaft N, Bramer WM, Ikram MA, Voortman T. Dietary protein intake and all-cause and cause-specific mortality: results from the Rotterdam Study and a meta-analysis of prospective studies. (Under review)

5. Chen Z*, Radjabzadeh D*, Chen L*, Klurilshikov A, Ikram MA, Uitterlinden A, Zhernakova A, Fu J, Kraaij R, Voortman T. Gut microbiome, insulin resistance, and type 2 diabetes: results from large population-based studies. (Manuscript)

6. Chen Z, Radjabzadeh D, Ikram MA, Uitterlinden A, Kraaij R, Voortman T. Diet quality and gut microbiome: results from large population-based studies. (Manuscript)

7. Braun KVE, Chen Z, Girschik C, Kavousi M, Franco OH, Voortman T. Associations between macronutrient intake and coronary heart disease: The Rotterdam Study. (Submitted)

**denotes equal contribution*

PhD portfolio

Name PhD student: Zhangling Chen

Erasmus MC department: Epidemiology

Research school: NIHES

PhD period: February 2016 – December 2019

Promotor: Prof.dr. M. Arfan Ikram

Co- promotor: Dr.ir. Trudy Voortman

Training	Year	ECTS
Courses		
Study Design	2016	4.3
Biostatistical Methods I: Basic Principles	2016	5.7
Biostatistical Methods II: Classical Regression Models	2016	4.3
English Language	2016	1.4
Health Economics	2016	0.7
The Practice of Epidemiologic Analysis	2016	0.7
Introduction to Medical Writing	2017	2.0
Clinical Epidemiology	2017	3.7
Principles of Research in Medicine and Epidemiology	2017	0.7
Methods of Public Health Research	2017	0.7
Clinical Trials	2017	0.7
Fundamentals of Medical Decision Making	2017	0.7
Clinical Translation to Epidemiology	2017	2.0
Methodological Topics in Epidemiologic Research	2017	1.4
Markers and Prognostic Research	2017	0.7
Erasmus Summer Lectures	2017	0.4
Advanced Topics in Clinical Trials	2017	1.9
Courses for Quantitative Researcher	2017	1.4
Cardiovascular Epidemiology	2017	0.9
Causal Medication Analysis	2018	0.7
Repeated Measurements in Clinical Studies	2018	1.4
Missing Values in Clinical Research	2018	1.4

Attended Seminars

Seminars of the Department of Epidemiology	2016-2019	2.0
2020 Meetings	2016-2019	2.0
Nutrition-lifestyle Group Meetings	2016-2019	2.0
Cardiovascular Group Meetings	2016-2019	2.0

Inter(national) conferences

Oral Presentation at Annual Dutch Nutritional Science Days, Heeze	2016	1.0
Oral Presentation at the 53rd annual meeting of the European Association for the Study of Diabetes, Lisbon	2017	1.0
Oral Presentation at Annual Dutch Nutritional Science Days, Heeze	2017	1.0
Oral Presentation at Annual meeting of the Dutch Association for the study of Diabetes, Utrecht	2017	1.0
Oral Presentation at the 25th European Congress on Obesity, Vienna	2018	1.0
Oral Presentation at the 54th annual meeting of the European Association for the study of Diabetes, Berlin	2018	1.0

Grants

Travel grant, granted by the 53rd annual meeting of the European Association for the Study of Diabetes, Lisbon	2016
Travel grant, granted by the 54th annual meeting of the European Association for the Study of Diabetes, Berlin	2017
Albert Renold Travel Fellowship Programme, granted by European Foundation for the Study of Diabetes	2018
International Fellowship, granted by Nutricia Research Foundation	2019
2018 best scientific paper award, granted by alpro Foundation	2019

Supervision

Maria G Zuurmond, MSc thesis	2017	2.0

Others

Peer review of articles for scientific journals	2017-2019	1.0
Research visit at School of Public Health, Harvard University	2019	1.0

About the author

Zhangling Chen was born on September 23, 1983 in Guang'an, Sichuan Province, China. She grew up and attended primary, middle, and high schools in her hometown. In 2003, she moved to Changsha to start a bachelor's program in Medicine at Central South University. In 2008, she earned her bachelor's degree in Medicine. Subsequently she started a master's program in Medicine at Central South University and obtained her master's degree in Medicine in 2011. In the next few years she worked as a Medical Doctor in the Second Xiangya Hospital of Central South University. In 2016, she started her PhD study at the Department of Epidemiology of the Erasmus Medical Center Rotterdam in the Netherlands. She worked in the Nutrition & Lifestyle group under supervision of Prof. M. Arfan Ikram and Dr. Trudy Voortman. Her PhD research focused on the role of nutrition related to gut microbiome in the development of type 2 diabetes. In 2018, she also received a master's degree in Clinical Epidemiology at the Netherlands Institute of Health Sciences. In 2019, she worked as a visiting researcher at the Department of Nutrition, Harvard School of Public Health Boston, under supervision of Prof. Frank Hu. After finalizing her PhD study, Zhangling will return to China where she will combine her clinical work with further research on nutrition, gut microbiome, and cardiometabolic health.

Acknowledgements

During this enjoyable journey, I have been privileged to be supported by my families, colleagues, friends and others, who have contributed to this book directly or indirectly. All my success and knowledge, I owe to you.

First, I would like to express my sincere gratitude to my promotor Prof. Arfan Ikram, co-promotor Dr. Trudy Voortman, and former promotor Prof. Oscar Franco.

Dear Prof. Arfan Ikram, thank you for your suggestions and supervisions during my PhD study. I am so impressed by your inspiring comments on my work. I also want to thank you for all your support during my PhD study and for your extraordinary ability to guide us. I am very lucky to have you as my promotor.

Dear Dr. Trudy Voortman, I would like to thank you for the enormous support and encouragement during my entire PhD study. You are a brilliant, open-minded, and hard-working scientist. Your enthusiasm in science really motivates me. I fully appreciate the way you trained me to conduct research. You always patiently supervised me in many aspects and gave me useful suggestions on how to improve my scientific writing. I would like to especially thank you for offering me various opportunities, believing in me, and giving me the necessary pep-talks whenever I started doubting myself. Thank you very much! I am very lucky to have you as my co-promotor.

Dear Prof. Oscar Franco, thank you so much for giving me the opportunity to perform my PhD research with the Rotterdam Study. I am very grateful that you were my promotor. Thank you so much for all your encouragement, guidance and support.

During my PhD study, I was deeply inspired by Prof. Albert Hofman and Prof. Frank Hu.

Dear Prof. Albert Hofman, your vision to establish the Rotterdam Study gave us the opportunity to study and work in such a great platform. Thank you so much! And I would also like to thank you for your wonderful lectures and for being an inspirational example.

Dear Prof. Frank Hu, thank you for granting me the opportunity to expand my insights and skills through the research visit to your department. I know the insights and skills I've gained as a result will help catapult my success, and for that I am truly grateful. I also want to thank you for being an inspirational example.

I would also like to thank the committee members: Prof. André Uitterlinden, Prof. Eric Boersma, Dr. Sabita Soedamah-Muthu, Prof. Jingyuan Fu, and Prof. Jaap Deckers, thank you for your time and consideration to read my thesis and hear my defense.

Many thanks also to my colleagues and friends in the Nutrition & Lifestyle research group, the Cardiovascular research group, and the Gut Microbiome research group, and amazing colleagues from the Epidemiology and Internal Medicine departments.

Dear Kim, Vincent, Mirjam, Hamid, Anh Nhi, Marija, Carolina, Eralda, Yuchan, Xiaofang, Zuolin, Kan, Irma, Xuan, Juna, Josje, Jessica, Adela, Laura, Taulant, Jana, Maca, Chantal, Niels, Moshen, Silvana, Valentina, Alice, Frank, Oscar Leonel, Michelle, Oscar Rueda, Lyda, Marlou, Amy, Arash, Banafsheh, Elif, Magda, Jaap, Sven, Gabriëlle, Solange, Andreas, Yolanda, Andrea, Nano, Sadaf, Fariba, Nicole, Dimitris, Amber, Tim, Marten, and Tom. I thank you for listening, offering me advice, and supporting me through this entire process. Thank you for all your help. I wish you all success in the future.

Dear Sander, I am so thankful for your support and care. You are a very thoughtful and helpful person. Thank you for helping me solve R and other problems. I cannot express my gratitude enough. I am very lucky and proud of having you in my PhD life.

Dear Djawad, thank you for sharing your knowledge on the gut microbiome and the analytical methods with me. Even though we have had many differences and don't always agree, I want to thank you for being an open-minded person. I appreciate your time, support, and cooperation.

Dear Runyu, thank you for sharing your knowledge and the interesting life stories with me. Thank you and your wife, Jie, for all your help. I wish you and your wife, Jie, all the best.

Dear Hoyan, thank you for taking care of me in numerous aspects.

A heartfelt thanks go out to Prof. André Uitterlinden, Dr. Robert Kraaij, Prof. Jingyuan Fu, Lianmin, Prof. Alexander Zhernakova, and Dr. Alexander Kurilshikov. I am very happy to collaborate with you on gut microbiome projects. Thank you so much for your help.

I sincerely thank Maria for working with me on the project: "Plant-based diet and type 2 diabetes". Thank you, dear Maria. I wish you all the best. Also thank Dr. Wichor Bramer, Dr. Fernando Rivadeneira, Dr. Maryam Kavousi, and Dr. Hanneke Wijnhoven for your help in my research projects.

A special thanks to my dear Chinese friends scattered around Erasmus MC, TU Delft, and the Netherlands: Guannan, Mannan, Lu, Junwen, Jinluan, Ying, Dream, Wenhao, Kai, Shihao, Tong, Fang, Xingxing, Laobing, Jiapeng, Yuan, Jun, and Jiao. Thank you for your thoughts, well-wishes, phone calls, emails, texts, visits, editing advice, and being there whenever I needed a friend.

I had a good time in Boston this summer. Here I would also like to express my sincere thanks to my colleagues and friends in Boston. Dear Jean-Philippe, Yixin, Qiaoli, Wei, Yuan, Shilpa, Qi, Edward,

Zhilei, Megu, Shuang, Cuilin, Mingyang, Yanping, Gang, Siyu, Jessie, Patrice, Walter, Zhenfei, Liyi, Fan, Liang, and Uncle Zheng. Thank you so much!

To Frank Qian (Qianqian), thank you for your great help in many aspects. I look forward to seeing you in Changsha next summer. I would like to enjoy delicious foods with you in Changsha.

I want to express my sincere gratitude for the scholarships, research funding, travel grants, and scientific awards from the China Scholarship Council, the Nutricia Research Foundation, the European Foundation for the Study of Diabetes, and the Alpro Foundation. To Dr. Wang, Dr. Jacques Bindels, the EASD committee members, and Dr. Stephanie De Vriese, thank you so much for helping me to obtain these scholarships, funding, grants, and awards.

Special thanks are given to my beloved: The Rotterdam Study, the Department of Epidemiology, NIHES, and Erasmus Medical Center. I would like to express my sincere gratitude for providing me excellent material, knowledge sources, logistical and human conditions, helping me in these first steps of scientific demand, and consequently, in the development and conclusions of my research work. I also thank the Lifelines-Deep Study for collaborating with me on my gut microbiome research.

I would also like to thank all former colleagues from the Second Xiangya Hospital of Central South University for your great help.

My acknowledgement would be incomplete without thanking the biggest source of my strength, my family.

Dear mom, father- and mother-in-law, uncles, aunts, sister-in-law and cousins, thank you so much for your help and support. I love you.

Dear Dad, I miss you so much. I wish you would have been here to see me graduate. You would have been so proud. Thank you for everything, my dear dad.

Finally, to my husband, Xu, thank you so much. We started our PhD study journey together in the Netherlands. We have been each other's inspiration, motivation, support, and listener through all the ups and downs. I cherish my every moment with you. Without you, I wouldn't be the person I am today. I wouldn't be where I am now. I love you so much and feel so blessed to have you in my life.

Zhangling Chen

October 11, 2019