University of Groningen

Analysis of lifestyle patterns and underlying factors for better type 2 diabetes prevention

Duan, Ming-Jie

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

[Link to publication in University of Groningen/UMCG research database](#)

Analysis of LIFESTYLE PATTERNS and UNDERLYING FACTORS for Better Type 2 Diabetes Prevention

Ming-Jie Frederick Duan 段鸣杰

Analysis of Lifestyle Patterns and Underlying Factors
for Better Type 2 Diabetes Prevention

Ming-Jie Frederick Duan 段鸣杰

university of groningen

Analysis of Lifestyle Patterns and Underlying Factors for Better Type 2 Diabetes Prevention

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Monday 28 August 2023 at 16:15 hours

by

Mingjie Duan

born on 12 May 1992
in Beijing, China

To my family and friends
致我的家人和朋友们

Paranymphs

İdil Esen

Vicente Artola Arita

Qiming Sun

Li Luo

CONTENTS

CHAPTER 1

Introduction

Type 2 Diabetes – A Major Global Health Crisis

Type 2 diabetes is a major global health problem that leads to considerable morbidity, mortality, and economic burden. The International Diabetes Federation (IDF) estimated that 783 million people will be living with diabetes globally by 2045, approximately 90% of which will be type 2 diabetes.[1] In The Netherlands, approximately 1.1 million diabetes patients were documented and registered at primary care practices in 2019, accounting for more than 6% of the total population. This escalating epidemic of diabetes becomes especially prominent in low-income and middle-income countries.[1]

For type 2 diabetes patients, the early asymptomatic phase may last for years, during which time unmanaged blood glucose levels may result in severe complications.[2] The IDF estimated that approximately 46% of all diabetes cases were undiagnosed in 2013.[3] Despite successful treatment and prevention of acute diabetes complications (such as ketoacidosis and hypoglycemia), diabetes patients continue to face a constant threat of chronic complications. Major chronic complications of diabetes include macrovascular complications (such as cardiovascular diseases) and microvascular complications (such as end-stage renal diseases, retinopathy, and neuropathy). Accumulating evidence also shows that conditions such as cancer, dementia, and infection may also be causally linked to diabetes.[4] Diabetes patients experience lower quality of life, higher risk of mortality, and lower life expectancy. The Global Burden of Diseases Study identified that of all communicable diseases, diabetes was the eighth major cause of reduced life expectancy from 2007 to 2017,[5] and became the fourth leading cause of disability-adjusted life years (DALYs) from 1990 to 2019.[6] Every year in The Netherlands, an average of 39,000 people over the age of 45 die with diabetes, accounting for more than a quarter of the total number of deaths.[7-9]

The Rationale and Evidence Base for Type 2 Diabetes Prevention
A Narrative Review

The causes of type 2 diabetes are multifactorial and interrelated. Major causes include genetic risk factors, obesity, unhealthy lifestyle, and the important interplay of them. Socioeconomic status is considered to be one of the most important underlying risk

factors (also see **Box 1** and **Fig. 1**) for type 2 diabetes. Targeting these risk factors – especially the modifiable ones – forms the evidence base and rationale for type 2 diabetes prevention.

Box 1/Fig. 1. A brief explanation of upstream, downstream, and underlying risk factors of type 2 diabetes

There are many risk factors for type 2 diabetes, such as obesity,[10] unhealthy lifestyle,[11] and low socioeconomic status.[12] Obesity is largely attributed to unhealthy lifestyle.[11] Unhealthy lifestyle is thus considered the upstream risk factor (or determinant) of obesity in the development of type 2 diabetes; comparatively, obesity is the downstream risk factor for type 2 diabetes. Likewise, low socioeconomic status is a risk factor for both obesity and unhealthy lifestyle.[13] Hence, low socioeconomic status is considered the upstream risk factor – in the literature often also indicated as underlying factor – of these two risk factors in the development of type 2 diabetes. In the figure, such relationships between risk factors are indicated by lines with arrowheads. Generally, a cluster of downstream risk factors can be related to a common upstream/underlying risk factor, which renders the upstream/underlying risk factor a potential target for intervention and prevention. In this thesis, in order to avoid confusion, the term **underlying factor** is used, e.g., the underlying factors of lifestyle factors.
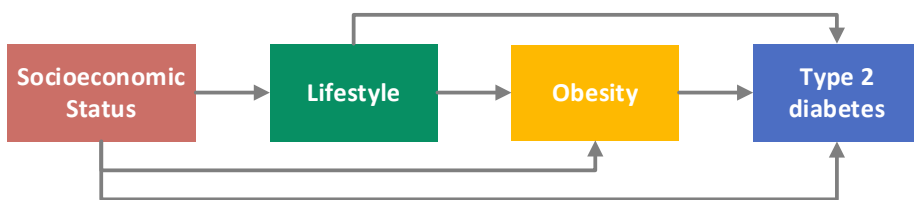
Genetic Predisposition

Until now, more than 36 genes and 80 single nucleotide polymorphisms (SNPs) have been found to be involved in the development of type 2 diabetes, which however in aggregate only explain 10% of the total heritable component of the disease. Most genes are found to be moderately associated with type 2 diabetes, each being a small risk contributor.[14] The TCF7L2 allele is the most significant genetic marker associated with 50% higher risk of type 2 diabetes.[15]

Obesity

Prospective studies have demonstrated obesity as a single independent predictor for type 2 diabetes risk, accounting for more than 50% of type 2 diabetes cases.[10] Chronic obesity has been linked to hyperglycemia, insulin resistance, and the subsequent development of type 2 diabetes. This is primarily due to an excessive accumulation of visceral adipose tissue, especially visceral and ectopic fat. Excess visceral adipose tissue induces increased production of inflammatory cytokines, dysregulation of hormone production, and hyperliposis and hypertriglyceridemia that leads to impairments in liver metabolism, all of which contribute to the pathogenesis of hyperglycemia and insulin resistance.[16] In common practice, direct measurements of visceral adipose tissue are not always available. A fast approximation is abdominal obesity, which is commonly measured by waist circumference, waist-hip ratio, or waist-height ratio.

Lifestyle

Unhealthy lifestyle substantially increases the risk of developing type 2 diabetes. These unhealthy lifestyle factors include unhealthy diets,[17] lack of physical activity,[18] smoking,[19] and excessive alcohol consumption.[20] Other than the aforementioned traditional ones, these unhealthy lifestyle factors also include high ultra-processed food intake (as a special dietary factor),[21,22] excessive TV watching (as an indicator for sedentary behavior),[23] and poor sleep quality.[24] These "non-traditional" lifestyle factors are often indicated as emerging lifestyle factors. Although these lifestyle factors mostly do not affect beta cell function directly, they are closely related to the pathogenesis of type 2 diabetes through numerous interrelated pathways.[11] Lifestyle interventions in

combination with weight control are fundamental in type 2 diabetes prevention and management.[11,25,26]

The effects of poor diets on the development of type 2 diabetes have been extensively studied. Dietary effects on the development of type 2 diabetes are not only demonstrated in the level of energy balance and single nutrients, but also manifested in the collective and synergistic effects among different dietary factors, such as dietary patterns. Evidence from randomized controlled trials has shown that low-carbohydrate diets (≤45% of energy from carbohydrates) and low-fat diets (≤30% of energy from fat) are both effective in reducing weight and waist circumference, as well as in improving metabolic risk factors.[27] Prospective studies have established that unhealthy dietary patterns may contribute to risk of type 2 diabetes, possibly through chronic inflammation,[28] abnormal blood lipid profiles,[29] and insulin resistance,[30] that are closely related to the pathogenesis of type 2 diabetes. Emerging evidence also shows that a higher ultra-processed food intake is associated with increased risk of type 2 diabetes.[21,22] Often considered as a single lifestyle factor, diet however consists of multiple aspects, affecting the development of type 2 diabetes through a wide spectrum of pathways. This distinct feature of diet thus makes it to be one of the most important causes of type 2 diabetes, as well as to be the primary intervention target for type 2 diabetes management and prevention.[31]

Similar to diet, other lifestyle factors also influence the development of type 2 diabetes through multiple pathways. For instance, physical activities, including aerobic and resistance exercise, can optimize immune function, increase mitochondrial density, enhance skeletal muscle oxidative capacity, and increase insulin sensitivity, even without weight loss.[25] Accumulating evidence also supports the pathophysiological links between emerging lifestyle factors and the development of type 2 diabetes. For example, disrupted circadian body clocks due to poor sleep quality result in disturbances of circadian gene expressions, such as the synthesis and secretion of melatonin and its receptors, which are suggested to be important mediators of circadian regulation of insulin sensitivity. These potential mechanisms related to poor sleep quality are also found to have interactive effects with other lifestyle factors in the development of type

2 diabetes.[32]

Socioeconomic Status

Socioeconomic status has been increasingly recognized as one of the most important underlying factors of type 2 diabetes and ill health. Conventionally, it is considered that socioeconomic status indirectly influences the risk of type 2 diabetes, because socioeconomic status is the upstream determinant of several downstream risk factors such as lifestyle, obesity, and access to medical resources.[12,33] Increasing evidence from recent years indicates, however, that the interplay between socioeconomic status and other contextual and downstream risk factors, and the associations of these risk factors with eventual health risks are not as simple as that. In fact, the associations between socioeconomic status and type 2 diabetes are only partly explained by the abovementioned downstream risk factors, even when a wide spectrum of risk factors is investigated.[34,35] Accordingly, the mechanisms of how socioeconomic status affects health remains incompletely understood. Some evidence suggests that adverse socioeconomic status is related to epigenetic status, which may influence gene expression that contributes to the development of type 2 diabetes.[36] Importantly, adverse socioeconomic status may influence type 2 diabetes risk through the interplay with other social-health factors, such as neighborhood socioeconomic status,[37] social insecurity, and lack of social and human capital.[38]

Real-life Practice
Achievements of Lifestyle Interventions in Type 2 Diabetes Prevention

Evidence from both biological and epidemiological studies has demonstrated the essential role of lifestyle factors in the development of type 2 diabetes. These findings provide the main evidence base and the rationale that guides the design and practice of lifestyle intervention programs. Amongst all causes of type 2 diabetes, while genetic causes and aging are generally unmodifiable, lifestyle and obesity naturally become the targets in public health prevention programs, of which obesity intervention is often achieved through lifestyle modifications. In the past decades, several milestone clinical trials in type 2 diabetes prevention have consecutively demonstrated that intensive

lifestyle interventions are effective in lowering the disease incidence, delaying the disease progression, and preventing major complications of type 2 diabetes.

In the Chinese Da Qing study, lifestyle interventions were conducted in people with impaired glucose tolerance. Lifestyle modification goals focused on increasing consumption of vegetables, reducing consumption of alcohol and simple sugar, and 20 minutes of moderate physical activity per day. After 6 years of follow-up, intervention groups showed 31-46% risk reduction in type 2 diabetes, compared with the untreated control group.[39] Remarkably, the 30-year follow-up results (even longer than many cohort studies) showed that compared with the control groups, intervention groups had a median delay in type 2 diabetes onset of 4 years, and had 26% and 35% lower risks of cardiovascular disease events and microvascular complications, respectively.[40] Results from the Da Qing study suggest that the benefits of lifestyle interventions in the prevention and progression of type 2 diabetes are sustained and may last for decades.[40]

Other major clinical trials have achieved similar remarkable outcomes. For instance, results from the Finnish Diabetes Prevention Study demonstrated significant clinical impacts of intensive lifestyle modifications in overweight participants. The intervention group maintained a 32% relative risk reduction in type 2 diabetes incidence 9 years after the randomization and 6 years after the active intervention.[41] The largest lifestyle intervention trial for the prevention of type 2 diabetes so far – the US Diabetes Prevention Program – has found that intensive lifestyle interventions were effective in reducing weight and delaying the onset of type 2 diabetes. Another important finding is that lifestyle interventions are more cost-effective compared with pharmacological interventions with metformin.[42,43]

Lifestyle Patterns and Underlying Factors of Lifestyle Factors
Major Evidence Gaps

As extensively discussed, there is potent evidence demonstrating the essential role of lifestyle factors in the prevention of type 2 diabetes. Nevertheless, we may only have solved part of the problems. The flip side of the great achievements of the abovementioned lifestyle intervention trials is that they mostly focused on high-

risk populations and are commonly small-scale, e.g., people with impaired glucose tolerance or obesity. There is a substantial lack of progress and knowledge of improving lifestyle and preventing type 2 diabetes in the general population.

Population-wide lifestyle intervention programs may require different approaches focusing on multiple dimensions of lifestyle and its interrelated factors, that are beyond intensive lifestyle interventions in the high-risk population emphasizing individual lifestyle behavioral changes. However, the evidence for designing such population-level lifestyle interventions is considerably lacking. More specifically, current population-level lifestyle intervention programs mainly put focus on single lifestyle factors and often adopt a generic one-size-fits-all approach.[44] There is an apparent lack of attention to other co-occurring lifestyle factors and how they are distributed and cluster in the target population. The relevance of the clustering of lifestyle factors in the development of type 2 diabetes also remains under-explored.[45] In addition, socioeconomic status, as an important underlying factor of lifestyle factors and health, has rarely been successfully translated and integrated into the design and practice of lifestyle intervention programs.[26] Having a clear understanding of these issues may provide important evidence that guides the design of more effective and better-targeted lifestyle intervention programs at the population level. Details of these major evidence gaps are discussed below.

The Relevance of Lifestyle Patterns in the Development of Type 2 Diabetes

The development of type 2 diabetes is not due to a single lifestyle factor. Lifestyle factors often co-occur with each other, commonly in the form of lifestyle patterns.[45-47] However, in much of the research and public health practice focusing on lifestyle factors and type 2 diabetes risk, a single lifestyle factor approach has been widely applied.[44] The relevance of lifestyle patterns in the development and prevention of type 2 diabetes remains largely unknown.

Epidemiological evidence suggests that there are interdependent effects among different lifestyle factors.[48-50] Numerous studies have demonstrated that having multiple healthy lifestyle behaviors (such as non-smoking, sufficient physical activity, and healthy

diet) may substantially lower the risk of developing type 2 diabetes (e.g., 75% lower hazard reported in a meta-analysis).[44] However, these studies commonly investigated an unweighted additive score of lifestyle factors. Residual confounding from other co-occurring lifestyle factors may result in biased estimates. This is well illustrated in dietary patterns (as a form of lifestyle patterns).[31] For example, while refined grains and high-fat dairy are often identified as major components in a dietary pattern that is associated with higher risk of type 2 diabetes,[17] they have not been consistently linked to type 2 diabetes risk when they are studied separately.[51,52] Nevertheless, different combinations of lifestyle factors, especially concerning lifestyle patterns, and their relevance in the development of type 2 diabetes have rarely been studied.

In addition, in terms of lifestyle intervention goals, besides traditional intervention targets such as vegetable and fruit consumption, alcohol consumption, smoking, and physical activity, there are also large needs for improvements in other important emerging lifestyle factors, such as TV watching, ultra-processed food intake, and sleep quality. Evidence on emerging lifestyle factors for the prevention of type 2 diabetes is scarce. For example, ultra-processed food forms a highly heterogeneous food category, especially in terms of its nutritional composition, product types, and contribution to a habitual diet.[22] It is unclear whether previous research treating the total intake of ultra-processed food as one single risk factor for type 2 diabetes applies to all underlying consumption patterns that fall under this "umbrella term". Whether a conventional healthy diet may compensate for the detrimental effects of ultra-processed food also remains unclear.

"Personalized" Lifestyle Interventions for Improved Effectiveness at the Population Level

For type 2 diabetes prevention, current evidence supports the relevance of targeting multiple lifestyle risk factors simultaneously through a personalized approach,[45-48,53-55] which can be best implemented in small-scale interventions. However, full personalization at the population level is generally laborious, costly, and time-consuming. Public health programs on the other hand generally apply a generic one-

size-fits-all approach, which compromises their effectiveness. Previous studies using the Lifelines cohort data have identified several dietary patterns, robust after adjustment for confounders.[56] This provides an excellent starting point to introduce "personalized" lifestyle intervention strategies at the population level, as tailored interventions can be designed and manifested at an intermediate level of the population, i.e., targeting lifestyle pattern groups who share similar lifestyle characteristics. Nevertheless, for the "whole package" of lifestyle factors, how do they cluster in the general population remains largely unknown. Having a clear understanding of this issue may help clarify the overall lifestyle risk profiles of the target populations and therefore facilitates the design of better-targeted lifestyle intervention programs.

Lack of Integration of Underlying Factors

While prevalent lifestyle intervention trials emphasize the importance of individual behavioral changes, less attention has been given to the underlying factors of lifestyle factors, particularly socioeconomic status and contextual factors. Studies on determinants of effective lifestyle interventions underline the importance of considering preferences, habits, and contextual factors as success factors for sustained efficacy.[54,57,58] However, these underlying factors have rarely been successfully translated and integrated into the design of lifestyle intervention schemes. A comprehensive narrative review found that such relevant evidence bases are surprisingly scarce.[26] While describing lifestyle patterns provides the starting point to map the lifestyle needs and habits of individuals, investigations into the interplay among socioeconomic status, contextual factors, and lifestyle factors take a step further to characterize facilitators and barriers to changes towards healthy lifestyle.

There is convincing evidence showing that socioeconomic status is one of the most important underlying factors of lifestyle factors and the development of type 2 diabetes.[12,34-37,46,47,59,60] However, conventional studies often study these two topics separately, overlooking the interrelationships among socioeconomic status, lifestyle factors, and type 2 diabetes risk.[26] This has led to the fragmentation of evidence for the design of prevention programs. In addition, definitions of low socioeconomic status

1

remain inconsistent in both research and public health practice. Different dimensions of socioeconomic status such as education and income may impact on health through different pathways.[12,33,34]

Contextual factors, especially neighborhood-level factors, add another layer to the complex interplay between socioeconomic status and lifestyle factors.[37,56,58,61-63] As exposed to different surrounding environments, individuals may engage in different lifestyle behaviors dependent on their own socioeconomic status.[37,58,61,63] This additional layer of contextual factors on lifestyle also requires further investigations, which will guide the design of neighborhood-level interventions to halt the rise in type 2 diabetes prevalence.

Most importantly, socioeconomic inequalities in health persist in society. Efforts have been made to tackle health inequalities over the past decades mainly through lifestyle interventions focusing on socioeconomically-disadvantaged people, the effects of which, however, are often limited.[64] Emerging evidence has suggested that socioeconomic inequalities exist in the relationships between lifestyle and health. For socioeconomically-disadvantaged people, they are found to have disproportionately higher risks of type 2 diabetes and mortality compared with their less disadvantaged counterparts, even if both of them have the same healthy lifestyle.[59,60] Improving lifestyle alone is therefore not sufficient to relieve the persistent socioeconomic inequalities in health. Great challenges remain in tackling health inequalities in addition to lifestyle interventions.

Outline of This Thesis

General Aims

The aim of this thesis is to provide empirical evidence for better type 2 diabetes prevention. As outlined above, evidence gaps remain in characterization of lifestyle patterns and in their relevance in the development of type 2 diabetes, as well as in the interplay between lifestyle factors and their underlying factors, and the relevance of this interplay in the development of type 2 diabetes. The main approach of this thesis is epidemiological.

More specifically, the aims are two-fold. **Part 1** of this thesis aims to describe the clustering of lifestyle factors – lifestyle patterns in the general population, with a special focus on dietary and lifestyle patterns. The relevance of dietary and lifestyle patterns in the development of type 2 diabetes is extensively analyzed and discussed. **Part 2** of this thesis aims to investigate the relationships of lifestyle factors with their underlying factors, with a special focus on individual socioeconomic status and neighborhood socioeconomic status. How these factors collectively affect type 2 diabetes risk is subsequently investigated. A graphical outline of the thesis chapters is presented in **Fig. 2**.

General Design – The Lifelines Cohort Study

Studies included in this thesis were performed using the data from the population-based Lifelines Cohort Study. The Lifelines Cohort Study is a multidisciplinary prospective population-based cohort study that applies a unique three-generation design to study the health and health-related behaviors of 167,729 persons living in the north of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, and physical factors, which contribute to health and disease of the general population. Before study entry, a signed informed consent form was obtained from each participant. The Lifelines study is conducted according to the principles of the Declaration of Helsinki and approved by the Medical Ethics Committee of the University Medical Center Groningen, The Netherlands (approval number 2007/152).

So far, three rounds of follow-up have taken place. After the baseline assessment (T1, years 2007 to 2013), all participants were invited for new rounds of assessments approximately every 5 years. In between assessments, follow-up questionnaires were completed approximately once every 1.5-2.5 years. The studies included in this thesis used data from the baseline assessment T1 and the second assessment T4, as well as the two follow-ups (T2 and T3) in between. Currently, the third round of assessment is on-going. Comprehensive physical examinations, biobanking, and questionnaires were conducted at T1 and T4. Follow-up questionnaires for the status of type 2 diabetes were issued to participants at T2, T3, and T4. The overall design and rationale of the study have been described in detail elsewhere.[65,66]

Part 1 – Lifestyle Patterns and Type 2 Diabetes Risk

Chapter 2 describes specific dietary patterns and their relationships with risk of type 2 diabetes. The dietary patterns identified specifically reflect the variation in blood lipid profiles of the study population, which are important intermediate risk factors for the development of type 2 diabetes.

Chapter 3 describes the associations between ultra-processed food intake, as an important emerging lifestyle factor, and risk of type 2 diabetes. A special focus has been put on the underlying consumption patterns of ultra-processed food, as a special form of dietary pattern.

Chapter 4 describes the clustering of lifestyle factors in the general population, i.e., lifestyle patterns. This study further investigates how lifestyle patterns are associated with risk of type 2 diabetes.

Part 2 – Underlying Factors of Lifestyle Factors

Chapter 5 examines a conceptual model, that describes multiple modifiable risk factors (socioeconomic status, lifestyle factors, obesity status, and clinical biomarkers) and their interrelationships with type 2 diabetes.

Chapter 6 investigates the associations of education and income with incident type 2 diabetes and cardiovascular diseases. In addition, the extent to which modifiable risk factors, including lifestyle factors, obesity status, and clinical biomarkers, explain the associations of education and income with these two health outcomes is studied.

Chapter 7 investigates the separate and combined effects of individual and neighborhood socioeconomic status on lifestyle factors. This study also investigates whether neighborhood socioeconomic status modifies the associations between individual socioeconomic status and lifestyle factors.

Fig. 2. Graphical outline of the thesis chapters

Chapter 6: effects of education and income on type 2 diabetes risk

Chapter 7: separate and combined effects of individual and neighborhood socioeconomic status on lifestyle factors

Education

Income

Individual socioeconomic status

+

Neighborhood socioeconomic status

Underlying Factors

Chapter 2: blood lipids-related dietary patterns

Chapter 3: ultra-processed food consumption patterns

Chapter 4: lifestyle patterns and risk of type 2 diabetes

Physical activity

TV watching

Ultra-processed food

Smoking

Risk drinking

Diet

Lifestyle Patterns

Chapter 5: analyzing a conceptual model that includes multiple modifiable risk factors (such as lifestyle factors and socioeconomic status) and their interrelationships with type 2 diabetes

Type 2 Diabetes

References

1.   Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022; 183: 109119.

2.   Clark NG, Fox KM, Grandy S, Group ftSS. Symptoms of Diabetes and Their Association With the Risk and Presence of Diabetes: Findings from the Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD). Diabetes Care 2007; 30(11): 2868-73.

3.   Beagley J, Guariguata L, Weil C, Motala AA. Global estimates of undiagnosed diabetes in adults. Diabetes Res Clin Pract 2014; 103(2): 150-60.

4.   Harding JL, Pavkov ME, Magliano DJ, Shaw JE, Gregg EW. Global trends in diabetes complications: a review of current evidence. Diabetologia 2019; 62(1): 3-16.

5.   Roth GA, Abate D, Abate KH, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2018; 392(10159): 1736-88.

6.   Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet 2020; 396(10258): 1204-22.

7.   Ligthart S, van Herpt TT, Leening MJ, et al. Lifetime risk of developing impaired glucose metabolism and eventual progression from prediabetes to type 2 diabetes: a prospective cohort study. Lancet Diabetes Endocrinol 2016; 4(1): 44-51.

8.   Nielen M, Poos R, Korevaar J. Diabetes mellitus in Nederland. Prevalentie en incidentie: heden, verleden en toekomst. Utrecht, The Netherlands: Nivel, 2020.

9.   Poos R, Nielen M, Hilderink H. Analyseren van het effect van het hebben van diabetes op de sterftekans en levensverwachting. Bilthoven, The Netherlands: RIVM, 2021.

10.   Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018; 14(2): 88-98.

11.   Kolb H, Martin S. Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. BMC Med 2017; 15(1): 131.

12.   Agardh E, Allebeck P, Hallqvist J, Moradi T, Sidorchuk A. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. Int J Epidemiol 2011; 40(3): 804-18.

13.   Lakerveld J, Mackenbach J. The Upstream Determinants of Adult Obesity. Obes Facts 2017; 10(3): 216-22.

14.   Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. Nature 2016; 536(7614): 41-7.

15.   Cauchi S, El Achhab Y, Choquet H, et al. TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. J Mol Med (Berl) 2007; 85(7): 777-82.

16.   Neeland IJ, Ross R, Despres JP, et al. Visceral and ectopic fat, atherosclerosis, and cardiometabolic disease: a position statement. Lancet Diabetes Endocrinol 2019; 7(9): 715-25.

17.	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.

18.	Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ. Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. Eur J Epidemiol 2015; 30(7): 529-42.

19.	Pan A, Wang Y, Talaei M, Hu FB, Wu T. Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis. Lancet Diabetes Endocrinol 2015; 3(12): 958-67.

20.	Knott C, Bell S, Britton A. Alcohol Consumption and the Risk of Type 2 Diabetes: A Systematic Review and Dose-Response Meta-analysis of More Than 1.9 Million Individuals From 38 Observational Studies. Diabetes Care 2015; 38(9): 1804-12.

21.	Almarshad MI, Algonaiman R, Alharbi HF, Almujaydil MS, Barakat H. Relationship between Ultra-Processed Food Consumption and Risk of Diabetes Mellitus: A Mini-Review. Nutrients 2022; 14(12).

22.	Duan MJ, Vinke PC, Navis G, Corpeleijn E, Dekker LH. Ultra-processed food and incident type 2 diabetes: studying the underlying consumption patterns to unravel the health effects of this heterogeneous food category in the prospective Lifelines cohort. BMC Med 2022; 20(1): 7.

23.	Patterson R, McNamara E, Tainio M, et al. Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. Eur J Epidemiol 2018; 33(9): 811-29.

24.	Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2010; 33(2): 414-20.

25.	Colberg SR, Sigal RJ, Yardley JE, et al. Physical Activity/Exercise and Diabetes: A Position Statement of the American Diabetes Association. Diabetes Care 2016; 39(11): 2065-79.

26.	Kivela J, Wikstrom K, Virtanen E, et al. Obtaining evidence base for the development of Feel4Diabetes intervention to prevent type 2 diabetes- a narrative literature review. BMC Endocr Disord 2020; 20(Suppl 1): 140.

27.	Hu T, Mills KT, Yao L, et al. Effects of low-carbohydrate diets versus low-fat diets on metabolic risk factors: a meta-analysis of randomized controlled clinical trials. Am J Epidemiol 2012; 176 Suppl 7(Suppl 7): S44-54.

28.	Schulze MB, Hoffmann K, Manson JE, et al. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. Am J Clin Nutr 2005; 82(3): 675-84; quiz 714-5.

29.	Duan MJ, Dekker LH, Carrero JJ, Navis G. Blood lipids-related dietary patterns derived from reduced rank regression are associated with incident type 2 diabetes. Clin Nutr 2021; 40(7): 4712-9.

30.	Batis C, Mendez MA, Gordon-Larsen P, Sotres-Alvarez D, Adair L, Popkin B. Using both principal component analysis and reduced rank regression to study dietary patterns and diabetes in Chinese adults. Public Health Nutr 2016; 19(2): 195-203.

31.	Michels KB, Schulze MB. Can dietary patterns help us detect diet-disease associations? Nutr Res Rev 2005; 18(2): 241-8.

32.   Larcher S, Benhamou PY, Pepin JL, Borel AL. Sleep habits and diabetes. Diabetes Metab 2015; 41(4): 263-71.

33.   Herd P, Goesling B, House JS. Socioeconomic Position and Health: The Differential Effects of Education versus Income on the Onset versus Progression of Health Problems. J Health Soc Behav 2007; 48(3): 223-38.

34.   Duan M-JF, Zhu Y, Dekker LH, et al. Effects of Education and Income on Incident Type 2 Diabetes and Cardiovascular Diseases: a Dutch Prospective Study. J Gen Intern Med 2022: s11606-022-07548-8.

35.   Stringhini S, Tabak AG, Akbaraly TN, et al. Contribution of modifiable risk factors to social inequalities in type 2 diabetes: prospective Whitehall II cohort study. The BMJ 2012; 345: e5452.

36.   McGuinness D, McGlynn LM, Johnson PC, et al. Socio-economic status is associated with epigenetic differences in the pSoBid cohort. Int J Epidemiol 2012; 41(1): 151-60.

37.   Zhu Y, Duan M-J, Riphagen IJ, et al. Separate and combined effects of individual and neighbourhood socio-economic disadvantage on health-related lifestyle risk factors: a multilevel analysis. Int J Epidemiol 2021; 50(6): 1959-69.

38.   WHO. Healthy, prosperous lives for all: the European Health Equity Status Report Executive Summary. Copenhagen, Denmark, 2019.

39.   Pan X-R, Li G-W, Hu Y-H, et al. Effects of Diet and Exercise in Preventing NIDDM in People With Impaired Glucose Tolerance: The Da Qing IGT and Diabetes Study. Diabetes Care 1997; 20(4): 537-44.

40.   Gong Q, Zhang P, Wang J, et al. Morbidity and mortality after lifestyle intervention for people with impaired glucose tolerance: 30-year results of the Da Qing Diabetes Prevention Outcome Study. Lancet Diabetes Endocrinol 2019; 7(6): 452-61.

41.   Lindström J, Peltonen M, Eriksson JG, et al. Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). Diabetologia 2013; 56(2): 284-93.

42.   The Diabetes Prevention Program Research Group. The Influence of Age on the Effects of Lifestyle Modification and Metformin in Prevention of Diabetes. J Gerontol A Biol Sci Med Sci 2006; 61(10): 1075-81.

43.   The Diabetes Prevention Program Research Group. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. The Lancet 2009; 374(9702): 1677-86.

44.   Zhang Y, Pan X-F, Chen J, et al. Combined lifestyle factors and risk of incident type 2 diabetes and prognosis among individuals with type 2 diabetes: a systematic review and meta-analysis of prospective cohort studies. Diabetologia 2020; 63(1): 21-33.

45.   Duan MJ, Dekker LH, Carrero JJ, Navis G. Lifestyle patterns and incident type 2 diabetes in the Dutch lifelines cohort study. Prev Med Rep 2022; 30: 102012.

46.   Meader N, King K, Moe-Byrne T, et al. A systematic review on the clustering and co-occurrence of multiple risk behaviours. BMC Public Health 2016; 16(1): 657.

47.   Noble N, Paul C, Turon H, Oldmeadow C. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. Prev Med 2015; 81: 16-41.

48.   Ding D, Rogers K, van der Ploeg H, Stamatakis E, Bauman AE. Traditional and Emerging Lifestyle Risk Behaviors and All-Cause Mortality in Middle-Aged and Older Adults: Evidence from a Large Population-Based Australian Cohort. PLOS Med 2015; 12(12): e1001917.

49.   Hendryx M, Dinh P, Chow A, et al. Lifestyle and Psychosocial Patterns and Diabetes Incidence Among Women with and Without Obesity: a Prospective Latent Class Analysis. Prev Sci 2020; 21(6): 850-60.

50.   van Etten S, Crielaard L, Muilwijk M, et al. Lifestyle clusters related to type 2 diabetes and diabetes risk in a multi-ethnic population: The HELIUS study. Prev Med 2020; 137: 106141.

51.   Gijsbers L, Ding EL, Malik VS, de Goede J, Geleijnse JM, Soedamah-Muthu SS. Consumption of dairy foods and diabetes incidence: a dose-response meta-analysis of observational studies. Am J Clin Nutr 2016; 103(4): 1111-24.

52.   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.

53.   Tuomilehto J, Schwarz P, Lindstrom J. Long-term benefits from lifestyle interventions for type 2 diabetes prevention: time to expand the efforts. Diabetes Care 2011; 34 Suppl 2: S210-4.

54.   Van Stappen V, Latomme J, Cardon G, et al. Barriers from Multiple Perspectives Towards Physical Activity, Sedentary Behaviour, Physical Activity and Dietary Habits When Living in Low Socio-Economic Areas in Europe. The Feel4Diabetes Study. Int J Environ Res Public Health 2018; 15(12): 2840.

55.   Van Stappen V, Cardon G, De Craemer M, et al. The effect of a cluster-randomized controlled trial on lifestyle behaviors among families at risk for developing type 2 diabetes across Europe: the Feel4Diabetes-study. Int J Behav Nutr Phys Act 2021; 18(1): 86.

56.   Dekker LH, Rijnks RH, Strijker D, Navis GJ. A spatial analysis of dietary patterns in a large representative population in the north of The Netherlands – the Lifelines cohort study. Int J Behav Nutr Phys Act 2017; 14(1): 166.

57.   Rongen A, Robroek SJW, van Ginkel W, Lindeboom D, Pet M, Burdorf A. How needs and preferences of employees influence participation in health promotion programs: a six-month follow-up study. BMC Public Health 2014; 14(1): 1277.

58.   Agachi E, Bijmolt THA, Mierau JO, van Ittersum K. Adoption of the Website and Mobile App of a Preventive Health Program Across Neighborhoods With Different Socioeconomic Conditions in the Netherlands: Longitudinal Study. JMIR Hum Factors 2022; 9(1): e32112.

59.   Foster HME, Celis-Morales CA, Nicholl BI, et al. The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: a prospective analysis of the UK Biobank cohort. Lancet Public Health 2018; 3(12): e576-e85.

60.   Vinke PC, Navis G, Kromhout D, Corpeleijn E. Socio-economic disparities in the association of diet quality and type 2 diabetes incidence in the Dutch Lifelines cohort. EClinicalMedicine 2020; 19: 100252.

61.   Algren MH, Bak CK, Berg-Beckhoff G, Andersen PT. Health-Risk Behaviour in Deprived Neighbourhoods Compared with Non-Deprived Neighbourhoods: A Systematic Literature Review of Quantitative Observational Studies. PLOS ONE 2015; 10(10): e0139297.

62.   Dekker LH, Rijnks RH, Navis GJ. Regional variation in type 2 diabetes: evidence from 137 820 adults on the role of neighbourhood body mass index. Eur J Public Health 2019; 30(1): 189-94.

63.   van Erpecum C-PL, van Zon SKR, Bültmann U, Smidt N. The association between fast-food outlet proximity and density and Body Mass Index: Findings from 147,027 Lifelines Cohort Study participants. Prev Med 2022; 155: 106915.

64.   Hagenaars L, Waterlander W, den Hertog K, Stronks K. A resilient public health in 2030: #buthow? TSG 2023: 1-4.

65.   Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. PLOS ONE 2015; 10(9): e0137203.

66.   Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol 2014; 44(4): 1172-80.

1

PART 1

LIFESTYLE PATTERNS
and Type 2 Diabetes Risk

CHAPTER 2

Blood lipids-related dietary patterns derived from reduced rank regression are associated with incident type 2 diabetes

Ming-Jie Duan[1]; Louise H. Dekker[1,2,3];

Juan-Jesus Carrero[4]; Gerjan Navis[1]

1. Department of Internal Medicine, University Medical Center Groningen, Groningen, The Netherlands

2. Aletta Jacobs School of Public Health, University of Groningen, Groningen, The Netherlands

3. National Institute for Public Health and the Environment, Bilthoven, The Netherlands

4. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

ABSTRACT

Background & Aims

Blood lipids play a critical role in the pathogenesis of type 2 diabetes, and they are closely related to dietary factors. However, the associations between blood lipids-related dietary patterns and risk of type 2 diabetes are controversial and not fully clear. In this study, we aimed to derive dietary patterns that explained variation in blood lipids and to investigate their associations with incident type 2 diabetes.

Methods

The analysis was based on 39,000 women and 25,777 men participating in the Lifelines cohort study (aged 18-65 years, mean 43.2 and 43.5 years for women and men, respectively). Dietary intake was measured using a 110-item semi-quantitative food frequency questionnaire. Reduced rank regression was used to derive dietary patterns with blood lipids (HDL-cholesterol, LDL-cholesterol, triglycerides, total cholesterol, and total cholesterol:HDL-cholesterol ratio) as response variables for women and men separately. The first dietary pattern identified for each sex was selected because they explained the largest variation in blood lipids. The associations between the identified dietary patterns and incident type 2 diabetes were subsequently investigated using multivariate logistic regression models. All analyses were performed separately for women and men.

Results

During an average follow-up of 43 months, 479 new cases (incidence 0.74%) of type 2 diabetes were identified. Using reduced rank regression, we identified two sex-specific blood lipids-associated dietary patterns characterized by high intake of sugary beverages, added sugar, and low intake of vegetables, fruits, tea, and nuts/seeds. These two sex-specific dietary patterns were similar in food groups but differed in factor loadings of food groups. High dietary pattern scores were associated with increased risk of type 2 diabetes after adjustment for age, total energy intake, body mass index, waist-hip ratio, and blood pressure (ORs for the fifth quintile [Q5] using the first quintile

[Q1] as reference, 1.87 [95% CI 1.23, 2.83] for women [P-trend < 0.001], and 1.72 [95% CI 1.11, 2.66] for men [P-trend = 0.018]). The associations were attenuated but remained significant after further adjustment for lifestyle and socio-economic factors.

Conclusions

Dietary patterns associated with adverse blood lipids are associated with incidence of type 2 diabetes. The present study provides new insights into optimizing blood lipids for the prevention of type 2 diabetes through dietary approaches.

Abbreviations

BMI – Body mass index
FFQ – Food frequency questionnaire
RRR – Reduced rank regression
WHR – Waist-hip ratio

Introduction

Type 2 diabetes is a major global health problem that leads to considerable morbidity, mortality, and economic burden.[1] As a modifiable risk factor, diet is crucial to the prevention of type 2 diabetes. Suboptimal diet characterized by insufficient intake of whole grains, nuts/seeds, and fruits is estimated to be responsible for 338,714 deaths and 24 million disability-adjusted life-years for type 2 diabetes globally.[2] While intervention strategies emphasize weight loss by limiting dietary energy intake and increasing physical activity, dietary factors may contribute to the risk of type 2 diabetes through mechanisms independent of weight change.[3-6]

Accumulating evidence indicates that blood lipids play a critical role in the pathogenesis of type 2 diabetes, such as inflammation and dysfunction of pancreatic beta cells.[7,8] Meanwhile, several prospective studies have shown that serum HDL-cholesterol and triglycerides are both independent predictors of the development of type 2 diabetes.[9-11] Approaches to optimize blood lipids include medication use and dietary modification. Although statins may effectively lower serum LDL-cholesterol level and inhibit internal synthesis of cholesterol, they are associated with increased risk of type 2 diabetes.[12] Dietary modification, therefore, may be prioritized as a better intervention target. Several dietary factors, such as saturated fatty acids and simple carbohydrates, have been found to have long-term detrimental effects on blood lipid profiles, independent of weight change.[13,14] However, these dietary factors, if considering their overall effects as dietary patterns, have rarely been examined in relation to blood lipids and simultaneously to type 2 diabetes risk.

Dietary pattern analysis can better capture the "real-world" complexities of eating habits of the general population, because this method accounts for the synergistic and interactive effects of different nutrients and food groups.[15] Reduced rank regression (RRR) can identify dietary patterns that are specifically associated with a set of pre-defined response variables. These response variables are usually biomarkers, and ideally, they are hypothesized intermediate outcomes of a certain disease. In brief, the extracted dietary pattern scores from RRR are linear functions of food intake that

concurrently maximize the explained variation of response variables. Compared with pure data-driven methods of dietary pattern analysis (e.g. principal component analysis, which derives empirical dietary patterns), this combination of *a priori* knowledge with *a posteriori* data-driven statistical technique endows RRR etiological and biological advantages, which allows investigation into certain diet-disease pathways of interest.[16,17]

Previous studies investigating the effects of dietary patterns on type 2 diabetes using RRR mainly focused on pathways of glucose homeostasis (e.g. HOMA-IR) and inflammatory biomarkers (e.g. C-reactive protein).[4,18] One recent RRR study found that a dietary pattern related to plasma circulating fatty acids was associated with risk of type 2 diabetes.[19] However, it is still not fully clear how blood lipids-related dietary patterns, that specifically reflect lipid metabolisms, may influence the risk of type 2 diabetes. Therefore, in this study, we aimed to identify dietary patterns using RRR that are exclusively associated with 5 blood lipid markers (HDL-cholesterol, LDL-cholesterol, triglycerides, total cholesterol, and total cholesterol:HDL-cholesterol ratio); and subsequently, to examine the associations between these dietary patterns and incident type 2 diabetes.

Materials and Methods

Study Population and General Exclusion Criteria

Lifelines is a multi-disciplinary prospective population-based cohort study that applies in a unique three-generation design to study the health and health-related behaviors of 167,729 people living in the north of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical, and psychological factors that contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics. After baseline assessment, two sets of follow-up questionnaires (both including questions for diabetes) were issued to participants at approximately 1.5-year intervals. A general second assessment was conducted approximately 5 years after the baseline assessment. Comprehensive physical examinations, biobanking, and questionnaires

were conducted at baseline and the second assessment (see **Supplementary Figure S1**). Before study entry, a signed informed consent form was obtained from each participant. The Lifelines cohort study was conducted according to the principles of the Declaration of Helsinki and in accordance with the research code of the University Medical Center Groningen (UMCG). The Lifelines study was approved by the medical ethical committee of the UMCG, The Netherlands. More detailed information about the Lifelines cohort study can be found elsewhere.[20,21]

In this study, participants aged 18-65 years at baseline with valid dietary intake data and blood data were included. At baseline, participants with cancer and liver cirrhosis, as well as participants who took lipid modifying agents (ATC code C10), corticosteroids for systemic use (ATC code H02), and anabolic steroids (ATC code A14A) were excluded from the analysis.[22] In addition, at baseline, non-fasting participants and participants with unreliable energy intake level (assessed by the Goldberg method, see **Supplementary File S1**) were excluded.[23] Moreover, participants with any kinds of diabetes or prediabetes at baseline were also excluded from the study. The ascertainment of diabetes at baseline were based on: (1) self-report questionnaires; (2) fasting glucose $\geq$ 7.0 mmol/L or HbA$_{1c}$ $\geq$ 48 mmol/mol (6.5%) according to the 2006 WHO diabetes diagnostic criteria;[24] and (3) medication use on glucose lowering agents (ATC code A10).[22] Prediabetes (impaired fasting glucose) was based on the same WHO criteria as fasting glucose ranged from 6.1 mmol/L to 6.9 mmol/L.[24]

For the two follow-ups and the second assessment, participants without any information on either self-reported diabetes or blood data on fasting glucose or HbA$_{1c}$ were not included in the analysis. Participants who reported having type 1 diabetes or gestational diabetes at the two follow-ups and the second assessment were excluded from the analysis. Participants at baseline without valid anthropometric data were also excluded from the analysis.

After all exclusions, in total 64,777 participants (39,000 women and 25,777 men) were included in the final analysis. Detailed flow charts with number of exclusions at each step can be found in **Supplementary Figure S2**.

Ascertainment of Incident Type 2 Diabetes

Incidence of type 2 diabetes was assessed by self-report questionnaires at the two follow-ups and the second assessment, as well as the blood test at the second assessment. Participants were considered having type 2 diabetes if they met either one of the following criteria: (1) self-reported newly developed type 2 diabetes since last visit; (2) fasting glucose ≥ 7.0 mmol/L; or (3) HbA$_{1c}$ ≥ 48 mmol/mol (6.5%).[24]

Clinical Measurements

Blood samples were collected at baseline and the second assessment. The samples were collected by venipuncture in a fasting state between 8 and 10 am, and were further transferred to the Lifelines central laboratory for analysis. Serum levels of glucose, HbA$_{1c}$, HDL-cholesterol, LDL-cholesterol, triglycerides, and total cholesterol were subsequently analyzed (**Supplementary File S1**). Total cholesterol:HDL-cholesterol ratio was calculated by dividing total cholesterol by HDL-cholesterol, both in mmol/L. Abnormal blood lipids were defined as: (1) HDL-cholesterol < 1.03 mmol/L for men or <1.30 mmol/L for women; (2) LDL-cholesterol ≥ 4.1 mmol/L; and (3) triglycerides ≥ 1.70 mmol/L.[25,26]

Baseline measurements of blood pressure and anthropometry (weight, height, and waist and hip circumferences) were made by trained research staff following standardized protocols (online **Supplementary File S1**). Anthropometric measurements were performed without shoes and heavy clothing. Body mass index (BMI) was calculated as weight in kilograms divided by the square of height in meters; and waist-hip ratio (WHR) was calculated as waist circumference divided by hip circumference, both in centimeters. Hypertension status was defined by meeting either one of the following criteria: (1) hypertensive medication use (ATC codes C02, C03, C07, C08, C09);[22] (2) systolic blood pressure ≥ 130 mmHg or diastolic blood pressure ≥ 80 mmHg;[27] and (3) self-reported hypertension from the questionnaires.

Assessment of Other Baseline Covariates

Age, smoking habits, income level, and educational level were assessed by self-

administered questionnaires. Smoking habits were categorized as non-, ex-, and current smoker. Income level (household net income per month) was further categorized into 4 categories: (1) < €1000; (2) €1000-€2000; (3) €2000-€3000; and (4) > €3000. Highest educational level achieved was categorized into 3 categories: (1) junior general secondary education or lower; (2) secondary vocational education and senior general secondary education; and (3) higher vocational education or university. Physical activity level was assessed by the validated Short QUestionnaire to ASsess Health-enhancing physical activity (SQUASH). Moderate to vigorous physical activities were defined as metabolic equivalent (MET) ≥ 4 (aged 18–55 years) or ≥ 3 (aged >55 years).[28] The physical activity covariate was derived by dividing total minutes per week of moderate to vigorous physical activities (if the value was not zero) into quartiles in each sex (labelled as 1, 2, 3, and 4), and if the value was zero then the covariate was coded to 0.

Dietary Assessment

Dietary intake was assessed by a self-administered semi-quantitative food frequency questionnaire (FFQ) at baseline. The FFQ aimed to assess the habitual intake of 110 food items (including alcoholic drinks) during the last month (4 weeks) and was designed based on the validated Dutch FFQ.[29] For 46 main food items, information of frequency (options ranged from "not in this month", "1 day per month" to "6-7 days per week") and quantity (in units or specified portion size) of consumption was asked. For another 37 sub-items (e.g., cream cheese), information of frequency ("never", "sometimes", "often", and "always") of consumption was asked. Responses were converted to food intake in grams per day. Macro- and micro-nutrients intake was calculated based on the FFQ according to the 2006 Dutch Food Composition Table (NEVO).[30] Energy misreporting and extreme values were not originally handled and corrected during data processing of the FFQs. The Goldberg cut-off was applied to evaluate the reliability of participants' energy intake level (**Supplementary File S1**).[23]

Dietary Pattern Analysis

Dietary patterns were derived by reduced rank regression (RRR) first introduced by

Hoffmann et al.[16] Before the RRR procedure, 110 single food items were combined into 50 food groups based on their similarities of nutrient profiles and habitual culinary practices (**Supplementary Table S1**). These 50 food groups (in grams per day) except for alcoholic drinks were then applied as predictor variables in RRR. Alcohol intake was not included but further adjusted as a separate covariate, since alcohol may exert distinct effects on the development of type 2 diabetes.

Considering differences in lipid metabolisms as well as the consumption level of certain food groups between women and men, we split our data by sex and conducted RRR and further analyses on women and men, separately. Blood lipids measured at baseline (HDL-cholesterol, LDL-cholesterol, triglycerides, total cholesterol, and total cholesterol:HDL-cholesterol ratio) were chosen as response variables to derive dietary patterns. All response variables were log transformed before the RRR procedure to improve normality. In order to reduce the confounding effects on derived dietary patterns from blood lipids that may be influenced by body composition and age, each of the log-transformed blood lipid indicators was further adjusted for BMI and age before the RRR procedure with linear regression models. Sensitivity analysis was performed by repeating the RRR procedure 4 times on random half sample for each sex. Dietary pattern analysis was performed on SAS (version 9.4; SAS Institute, Cary, NC) by SAS PLS procedure with method = RRR option.[16]

For each sex, five dietary patterns (equal to the number of response variables) were identified. Only the first dietary pattern for each sex was kept for further analyses, since they explained the biggest variance of the response variables (women 1.61%, men 1.60%) and total food groups (women 3.61%, men 3.47%). In order to improve the clinical relevance and interpretation of these dietary patterns, we selected food groups with absolute factor loadings ≥ 0.15 to construct simplified dietary pattern scores while retaining the weight (factor loading) of each selected food group. The simplified dietary pattern scores were then divided into quintiles for further analyses. As a separate covariate, intake of alcoholic drinks, if the value was not zero, was divided into sex-specific quartiles (labelled as 1, 2, 3, and 4). If the value was zero, then the covariate was coded to 0.

Statistical Analysis

Associations between simplified dietary pattern scores and incident type 2 diabetes were estimated using multivariate logistic regression models and the results were shown as ORs. In four steps, the models were adjusted for: (1) age and total energy intake; (2) BMI, WHR, and blood pressure; (3) intake of alcoholic drinks, smoking habits, and physical activity level; and (4) income level and educational level. Tests for linear trend (P-trend) was assessed by assigning the median value to each quintile and treating this as a continuous variable. Multiple imputation by chained equations was performed to deal with missing data for income level (proportion of missing for women 16.4% and men 11.6%) and physical activity level (proportion of missing for women 5.8% and men 8.3%).[31] Sensitivity analyses were performed by complete case analyses and including a missing class into the models for these two variables. All analyses were conducted separately for women and men. Statistical analyses for risk of type 2 diabetes were performed on Stata (version 13.1; StataCorp, College Station, TX). P values < 0.05 were considered to represent significant results.

Results

Using RRR, we derived two sex-specific simplified dietary pattern scores that were negatively associated with HDL-cholesterol, but positively associated with LDL-cholesterol, triglycerides, total cholesterol, and total cholesterol:HDL-cholesterol ratio (**Supplementary Table S2**). For both women and men, these dietary patterns were unhealthy, and were characterized by high intake of sugary beverages, added sugar, juice, and low intake of tea, fruits, vegetables, nuts/seeds, cereals, and low fat fermented unsweetened dairy products. Additionally, the female dietary pattern was characterized by low intake of fatty fish, other fish (such as squid), high fat cheese, and eggs; and the male dietary pattern was characterized by high intake of coffee and savory snacks, and low intake of chocolate spreads and bread products (**Table 1**). These simplified dietary pattern scores explained 90.4% and 96.3% of the original dietary pattern scores for women and men, respectively. Among these food groups, added

sugar (women 21.2%, men 26.3%), sugary beverages (women 19.0%), and fruits (men 18.9%) explained the most of the original dietary pattern scores (**Supplementary File S1**).

Table 2 and **Table 3** show the baseline characteristics for each sex across quintiles of the simplified dietary pattern scores. In general, with increasing quintiles of simplified dietary pattern scores, participants tended to be younger, be less physically active, smoke more, have less education, earn less income, and have worse blood lipid profiles. Intake of alcoholic drinks was different between women and men, as with increasing quintiles of simplified dietary pattern scores, men tended to consume more alcohol while women tended to consume less.

Table 4 shows the associations between quintiles of simplified dietary pattern scores and incident type 2 diabetes. Among 64,777 participants included in the analysis, we identified 258 female (incidence 0.66%) and 221 male (incidence 0.86%) cases of type 2 diabetes during an average follow-up of 43 months. Significant associations between these simplified dietary pattern scores and incident type 2 diabetes were observed, after adjustment for age, total energy intake, BMI, WHR, blood pressure, alcoholic drinks intake, and physical activity level (Q5 vs Q1, OR 1.66 [95%CI 1.08, 2.53], P-trend = 0.004 for women; and Q5 vs Q1, OR 1.64 [95%CI 1.04, 2.58], P-trend = 0.033 for men). Further adjustment for educational level and income level attenuated the associations in women. Regarding sensitivity analyses for missing data, compared with results after multiple imputation, including a missing class into the models for physical activity level and income level as well as complete case analyses yielded stronger associations and larger confidence intervals but did not show any substantial differences (data not shown).

Table 1. Intake of food groups from the simplified dietary patterns with absolute factor loading ≥ 0.15[a,b]

Food groups	Factor loading	Quintiles of female simplified dietary pattern scores (n = 39,000)					
		Q1	Q2	Q3	Q4	Q5	Total
High intake							
Sugary beverages	0.40	0 (0, 13)	0 (0, 26)	6 (0, 42)	26 (0, 90)	134 (42, 271)	13 (0, 70)
Added sugar	0.38	0 (0, 1)	0 (0, 1)	0 (0, 3)	1 (0, 9)	12 (1, 27)	0 (0, 7)
Juice	0.20	13 (0, 54)	21 (5, 54)	21 (5, 54)	27 (11, 107)	54 (13, 139)	27 (5, 96)
Low intake							
Tea	-0.27	465 (232, 697)	348 (223, 465)	232 (116, 465)	232 (80, 348)	116 (36, 232)	232 (89, 465)
Vegetables	-0.26	161 (110, 186)	112 (76, 161)	108 (75, 113)	75 (62, 112)	75 (42, 109)	109 (74, 149)
Nuts/seeds	-0.26	13 (6, 25)	8 (4, 16)	7 (3, 13)	6 (2, 11)	5 (1, 10)	7 (3, 14)
Fatty fish	-0.25	6 (3, 14)	3 (0, 6)	2 (0, 5)	0 (0, 3)	0 (0, 3)	2 (0, 6)
Fruits	-0.24	220 (152, 330)	220 (85, 220)	110 (76, 220)	76 (42, 152)	42 (17, 110)	110 (42, 220)
Cereals	-0.19	6 (0, 21)	1 (0, 10)	0 (0, 6)	0 (0, 3)	0 (0, 1)	0 (0, 8)
Other fish	-0.18	0 (0, 4)	0 (0, 2)	0 (0, 1)	0 (0, 0)	0 (0, 0)	0 (0, 2)
High fat cheese	-0.17	20 (8, 39)	17 (7, 33)	16 (6, 30)	14 (5, 26)	12 (5, 24)	16 (6, 30)
Low fat fermented unsweetened dairy products	-0.17	15 (0, 65)	0 (0, 46)	0 (0, 28)	0 (0, 18)	0 (0, 0)	0 (0, 31)
Eggs	-0.15	14 (7, 18)	7 (7, 18)	7 (4, 18)	7 (4, 18)	7 (4, 14)	7 (4, 18)

Table 1. Continued.

Food groups	Factor loading	Quintiles of male simplified dietary pattern scores (n = 25,777)					
		Q1	Q2	Q3	Q4	Q5	Total
High intake							
Added sugar	0.42	0 (0, 9)	1 (0, 15)	4 (0, 19)	10 (0, 26)	30 (12, 47)	7 (0, 25)
Coffee	0.25	348 (232, 581)	465 (348, 581)	465 (348, 697)	581 (348, 697)	581 (465, 813)	465 (348, 697)
Sugary beverages	0.22	14 (0, 52)	26 (0, 94)	42 (4, 126)	63 (13, 156)	141 (42, 282)	42 (5, 136)
Juice	0.18	27 (5, 96)	27 (11, 96)	43 (11, 107)	43 (11, 107)	43 (11, 139)	40 (11, 107)
Savory snacks	0.17	8 (3, 15)	11 (5, 19)	13 (7, 22)	17 (9, 25)	21 (12, 35)	13 (7, 23)
Low intake							
Fruits	-0.40	220 (152, 330)	152 (76, 220)	85 (42, 152)	76 (21, 110)	34 (10, 76)	85 (42, 220)
Vegetables	-0.26	114 (106, 164)	111 (74, 149)	103 (62, 113)	75 (42, 112)	74 (41, 110)	104 (62, 114)
Cereals	-0.25	5 (0, 23)	0 (0, 12)	0 (0, 5)	0 (0, 2)	0 (0, 0)	0 (0, 6)
Tea	-0.24	232 (116, 465)	161 (45, 241)	89 (18, 232)	45 (11, 161)	22 (0, 89)	89 (18, 232)
Chocolate spreads	-0.22	19 (5, 29)	14 (3, 28)	11 (2, 20)	5 (1, 16)	4 (0, 14)	11 (2, 28)
Bread and bread products	-0.18	193 (141, 249)	172 (133, 216)	164 (129, 206)	152 (112, 201)	140 (102, 193)	165 (126, 209)
Nuts/seeds	-0.15	15 (7, 30)	12 (6, 23)	11 (5, 21)	9 (4, 18)	8 (3, 17)	11 (5, 21)
Low fat fermented unsweetened dairy products	-0.15	0 (0, 46)	0 (0, 20)	0 (0, 9)	0 (0, 0)	0 (0, 0)	0 (0, 11)

[a] Data for food intake are expressed as median (interquartile) in grams/day.
[b] For all food groups P-trend < 0.001.

2

Table 2. Baseline characteristics according to quintiles of female simplified dietary pattern scores[a,b]

Characteristics	Quintiles of female simplified dietary pattern scores (n = 39,000)					
	Q1	Q2	Q3	Q4	Q5	Total
Demographics & lifestyle characteristics						
Age, years	47.2±10.4	45.6±10.6	44.0±10.7	41.9±10.7	37.5±11.1	43.2±11.2
Moderate to vigorous physical activity, min/week	400 (180, 930)	340 (150, 820)	300 (120, 750)	270 (105, 660)	240 (80, 645)	300 (120, 760)
Current smoker, %	10.0	12.1	15.1	19.8	29.9	17.4
University degree, %	43.6	35.6	30.3	26.2	19.8	31.1
High income (> €3000/month), %	35.6	33.1	29.8	26.2	18.6	28.6
Biomarkers						
BMI, kg/m^2	25.0±4.0	25.5±4.1	25.6±4.3	25.8±4.6	25.1±4.6	25.4±4.4
WHR	0.859±0.068	0.861±0.069	0.864±0.070	0.866±0.071	0.863±0.072	0.862±0.070
Fasting glucose, mmol/L	4.78±0.43	4.78±0.43	4.78±0.43	4.76±0.42	4.74±0.40	4.77±0.42
HbA$_{1c}$, mmol/mol	36.78±3.18	36.69±3.21	36.65±3.24	36.43±3.23	36.15±3.25	36.54±3.23
HbA$_{1c}$, %	5.51±0.29	5.50±0.29	5.49±0.30	5.47±0.30	5.45±0.30	5.48±0.30
Hypertension, %	33.1	34.8	34.5	33.0	29.2	32.9

Table 2. Continued.

Characteristics	Quintiles of female simplified dietary pattern scores ($n = 39,000$)					
	Q1	Q2	Q3	Q4	Q5	Total
Blood lipids						
HDL-cholesterol, mmol/L	1.75±0.41	1.68±0.40	1.65±0.39	1.59±0.37	1.53±0.36	1.64±0.39
HDL-cholesterol < 1.30 mmol/L, %	9.2	12.4	14.1	17.5	23.4	15.3
LDL-cholesterol, mmol/L	3.16±0.92	3.16±0.89	3.15±0.91	3.08±0.87	3.00±0.85	3.11±0.89
LDL-cholesterol ≥ 4.1 mmol/L, %	16.2	15.6	15.7	13.0	11.3	14.4
Triglycerides, mmol/L	0.91±0.46	0.96±0.49	0.97±0.50	0.99±0.58	1.01±0.53	0.97±0.51
Triglycerides ≥ 1.70 mmol/L, %	5.4	7.3	7.0	7.8	8.9	7.3
Total cholesterol, mmol/L	5.18±1.01	5.13±0.98	5.09±1.00	4.97±0.95	4.85±0.92	5.05±0.98
Total cholesterol:HDL-cholesterol ratio	3.09±0.87	3.19±0.91	3.24±0.93	3.28±0.95	3.34±1.00	3.23±0.94
Nutrient intake						
Total energy intake, kJ/day	8046±2075	7782±1879	7581±1828	7590±1912	8096±2054	7878±1964
Fat energy contribution, %	35.5±5.7	35.5±4.9	35.5±4.7	35.6±4.5	34.6±4.7	35.3±4.9
Alcoholic drinks intake, g/day	42 (9, 97)	36 (7, 89)	28 (4, 75)	20 (0, 71)	13 (0, 50)	27 (2, 76)

[a] Data are expressed as unadjusted mean ± standard deviation, or median (interquartile), or percentage as appropriate.
[b] Tests for significant differences in baseline characteristics across quintiles were performed using Kruskal-Wallis test or χ^2 test for proportion as appropriate; P < 0.001 for all baseline characteristics.

Table 3. Baseline characteristics according to quintiles of male simplified dietary pattern scores[a,b]

Characteristics	Quintiles of male simplified dietary pattern scores (n = 25,777)					
	Q1	Q2	Q3	Q4	Q5	Total
Demographics & lifestyle characteristics						
Age, years	45.6±11.1	45.0±11.0	43.9±10.8	42.5±10.8	40.5±10.3	43.5±11.0
Moderate to vigorous physical activity, min/week	470 (213, 1310)	420 (180, 1200)	360 (150, 1080)	355 (120, 1080)	300 (90, 1200)	385 (150, 1195)
Current smoker, %	8.7	12.4	18.0	23.3	40.9	20.6
University degree, %	45.5	39.5	35.8	30.7	22.8	34.9
High income (> €3000/month), %	39.2	38.2	36.9	32.6	27.4	34.8
Biomarkers						
BMI, kg/m^2	25.4±3.2	26.0±3.3	26.4±3.4	26.4±3.5	26.0±3.6	26.0±3.4
WHR	0.942±0.064	0.953±0.064	0.959±0.065	0.960±0.067	0.959±0.066	0.955±0.066
Fasting glucose, mmol/L	4.94±0.42	4.98±0.41	5.01±0.42	5.02±0.41	5.04±0.41	5.00±0.41
HbA$_{1c}$, mmol/mol	36.56±3.10	36.58±3.09	36.61±3.18	36.50±3.23	36.75±3.14	36.60±3.15
HbA$_{1c}$, %	5.49±0.28	5.49±0.28	5.49±0.29	5.48±0.30	5.50±0.29	5.49±0.29
Hypertension, %	50.4	55.2	57.1	57.8	57.6	55.6

Table 3. Continued.

Characteristics	Quintiles of male simplified dietary pattern scores (n = 25,777)					
	Q1	Q2	Q3	Q4	Q5	Total
Blood lipids						
HDL-cholesterol, mmol/L	1.38±0.33	1.35±0.31	1.32±0.31	1.31±0.30	1.28±0.31	1.33±0.31
HDL-cholesterol < 1.03 mmol/L, %	14.2	15.3	18.4	18.8	23.9	18.1
LDL-cholesterol, mmol/L	3.36±0.86	3.44±0.87	3.47±0.86	3.49±0.88	3.49±0.92	3.45±0.88
LDL-cholesterol ≥ 4.1 mmol/L, %	20.5	23.9	23.9	25.3	26.6	24.0
Triglycerides, mmol/L	1.18±0.74	1.27±0.78	1.36±0.89	1.41±0.95	1.48±1.16	1.34±0.92
Triglycerides ≥ 1.70 mmol/L, %	15.3	19.0	22.2	24.1	27.0	21.5
Total cholesterol, mmol/L	5.09±0.95	5.19±0.96	5.23±0.96	5.25±0.98	5.26±1.03	5.20±0.98
Total cholesterol:HDL-cholesterol ratio	3.90±1.33	4.02±1.14	4.15±1.18	4.22±1.24	4.36±1.37	4.13±1.26
Nutrient intake						
Total energy intake, kJ/day	10736±2753	10037±2548	9807±2527	9778±2644	10565±2870	10184±2700
Fat energy contribution, %	34.5±5.1	35.5±4.6	35.9±4.6	36.1±4.8	35.6±5.0	35.5±4.9
Alcoholic drinks intake, g/day	80 (25, 161)	95 (36, 193)	107 (38, 214)	120 (43, 249)	143 (43, 286)	106 (36, 215)

[a] Data are expressed as unadjusted mean ± standard deviation, or median (interquartile), or percentage as appropriate.
[b] Tests for significant differences in baseline characteristics across quintiles were performed using Kruskal-Wallis test or χ^2 test for proportion as appropriate; P < 0.001 for all baseline characteristics except for HbA_{1c} (P = 0.005).

Table 4. Risk of incident type 2 diabetes according to quintiles of simplified dietary pattern scores

| | Quintiles of simplified dietary pattern scores | | | | | |
	Q1	Q2	Q3	Q4	Q5	P-trend
Women						
Cases/population	39/7800	42/7800	51/7800	66/7800	60/7800	
Incidence, %	0.50	0.54	0.65	0.85	0.77	
Model 1[a]	1.00	1.13 (0.73, 1.75)	1.46 (0.96, 2.21)	2.08 (1.39, 3.11)	2.37 (1.56, 3.60)	<0.001
Model 2[b]	1.00	1.02 (0.66, 1.59)	1.23 (0.81, 1.88)	1.64 (1.09, 2.47)	1.87 (1.23, 2.83)	<0.001
Model 3[c]	1.00	0.97 (0.63, 1.52)	1.18 (0.77, 1.80)	1.52 (1.00, 2.31)	1.66 (1.08, 2.53)	0.004
Model 4[d]	1.00	0.97 (0.62, 1.51)	1.16 (0.76, 1.79)	1.46 (0.95, 2.25)	1.57 (1.01, 2.44)	0.012
Men						
Cases/population	34/5155	40/5155	45/5156	45/5155	57/5156	
Incidence, %	0.66	0.78	0.87	0.87	1.11	
Model 1[a]	1.00	1.21 (0.76, 1.92)	1.44 (0.93, 2.25)	1.55 (0.98, 2.43)	2.18 (1.41, 3.38)	<0.001
Model 2[b]	1.00	1.07 (0.67, 1.70)	1.15 (0.73, 1.81)	1.19 (0.75, 1.89)	1.72 (1.11, 2.66)	0.018
Model 3[c]	1.00	1.10 (0.69, 1.76)	1.18 (0.74, 1.87)	1.20 (0.76, 1.92)	1.64 (1.04, 2.58)	0.033
Model 4[d]	1.00	1.11 (0.69, 1.77)	1.19 (0.74, 1.89)	1.21 (0.75, 1.93)	1.65 (1.04, 2.62)	0.034

[a] Model 1: OR (95% CI) derived from multivariate logistic regression models adjusted for age and total energy intake, women (n = 39,000), men (n = 25,777).

[b] Model 2: OR (95% CI) derived from multivariate logistic regression models adjusted for Model 1 covariates plus BMI, WHR, and blood pressure, women (n = 38,989), men (n = 25,772).

[c] Model 3: OR (95% CI) derived from multivariate logistic regression models adjusted for Model 2 covariates plus alcoholic drinks intake, smoking status, and physical activity level, women (n = 38,648), men (n = 25,536).

[d] Model 4: OR (95% CI) derived from multivariate logistic regression models adjusted for Model 3 covariates plus income level and educational level, women (n = 38,539), men (n = 25,458).

2

Discussion

We identified two sex-specific dietary patterns that explained the variation in blood lipids and were significantly associated with incident type 2 diabetes in a large population-based sample. Our findings support the hypothesis that dietary factors increase the risk of type 2 diabetes through altered blood lipid profiles, and provide important insights in optimizing blood lipids for the prevention of type 2 diabetes through dietary approaches.

For both women and men, our identified dietary patterns were characterized by high consumption of sugary beverages, added sugar, and juice; and low consumption of vegetables, fruits, tea, cereals, nuts/seeds, and low fat fermented unsweetened dairy products. Our identified dietary patterns share some similarities with previous studies using different statistical methods (e.g., cluster analysis) or other intermediate response variables. In general, dietary patterns that have been found to be associated with higher risk of type 2 diabetes were characterized by high intake of sugary beverages, red meat, processed meat, and refined grains; and low intake of vegetables, fruits, and cereals.[4,5,18,32-34] Compared with those previous similar studies, it is noteworthy that there are some unique features in our identified dietary patterns. First, red meat and processed meat did not contribute substantially to our identified dietary patterns. In addition, food groups such as added sugar, tea, nuts/seeds, juice, low fat fermented unsweetened dairy products, and fatty fish (women) were unique in our dietary patterns. Our results indicate that these unique food groups may be essential to a diet that influences risk of type 2 diabetes through a distinct metabolic pathway characterized by altered blood lipid profiles.

In our study, the detrimental effects of the identified dietary patterns on blood lipids and risk of type 2 diabetes are plausible, especially considering those aforementioned unique food groups. The relationships of dietary simple sugars with adverse blood lipids, obesity, the metabolic syndrome, and type 2 diabetes have been well established; although evidence regarding juice and type 2 diabetes is less conclusive, which is probably due to the various types of juice consumed.[35] In our study, we did

observe that juice was associated with blood lipids in a similar manner as added sugar and sugary beverages (**Supplementary Table S2**). In previous studies, fatty fish and nuts/seeds were found to improve blood lipids,[36,37] which is consistent with our findings. Notwithstanding the inconclusive evidence, it is previously speculated that the protective effects of fatty fish and nuts/seeds on type 2 diabetes were largely attributable to their richness in mono- and polyunsaturated fatty acids (e.g., DHA and EPA in fatty fish) that may ameliorate inflammation and increase insulin sensitivity.[38] In addition, nuts/seeds are good sources of fiber and vegetable protein, which may improve post-prandial glycemic control.[37] Nonetheless, previous RRR studies, which used inflammation or insulin-resistance related biomarkers as response variables, did not identify these two food groups as important contributors to their dietary patterns.[4,5,18,34,39] Moreover, higher consumption of several food groups – such as vegetables, fruits, and whole grain products, which were commonly identified as part of the prudent/healthy diet – were also found to be associated with lower risk of type 2 diabetes in previous studies. These food groups may act through various different metabolic pathways in relation to the development of type 2 diabetes.[3-5,18,32] Finally, it is also conceivable that the detrimental effects of the identified dietary patterns were partly attributable to the synergy between individual food groups as well as nutrients, in addition to each of their own independent health effects.[15]

It should be noted that the high factor loadings of some food groups may indicate other food groups with which they were habitually consumed together. In the male dietary pattern, chocolate spread was highly correlated with bread intake (Spearman correlation coefficient 0.337, P < 0.001) and was negatively loaded on the dietary pattern. A cautious interpretation is warranted, since these correlated food groups may not be causally associated with the disease outcome. There is rather limited evidence concerning the health effects of chocolate spread on risk of type 2 diabetes, although the high concentration of fats and free sugar makes chocolate spread very unlikely to be beneficial.[30,35]

For both women and men, adjustments for BMI, WHR, and blood pressure attenuated the associations between dietary patterns and incident type 2 diabetes by approximately

21% (regarding ORs for Q5), but the associations remained significant. This suggests that the identified dietary patterns may also be partly related to metabolic syndromes. On the other hand, it seems in men that the associations between the identified dietary patterns and incident type 2 diabetes were little explained by lifestyle behaviors and socioeconomic characteristics (approximately 4% attenuation in OR for Q5); whereas for women, adjustments for these covariates further attenuated the association by approximately 16% (OR for Q5).

Our research possesses several unique strengths. First, the large sample size from the Lifelines cohort study, even with our strict exclusion criteria, enables a strong statistical power and good representativeness of the general population.[20] Second, the exclusion of participants with impaired glucose tolerance at baseline helps confirm our hypotheses in a healthy general population-based setting and minimizes the chance of reverse causation. Our results may therefore be more informative for helping formulating dietary recommendations for the general population. In addition, compared with previous similar RRR studies, our identified dietary patterns were less confounded by age, body composition, and relevant medication use, since all the response variables were pre-adjusted by those two biological factors before the RRR procedure and participants who took lipid modifying agents were excluded from the study. Although this practice may lead to smaller total explained variation in response variables, we still observed clear trends for all five blood lipid markers across the quintiles of dietary pattern scores, which supports that our identified dietary patterns reflected the difference of blood lipids within the population studied.

There are also several limitations that should be noted. First, the Lifelines cohort study is a relatively young cohort with less than 5-year follow-up. Therefore, the long-term dietary effects on type 2 diabetes might not appear, especially considering our included participants were free from prediabetes at baseline. The design of the study also does not allow us to adjust for follow-up time in the analyses. Second, the lack of endpoint medication data could result in misidentification of type 2 diabetes cases. However, this limitation may be largely resolved as most of the cases were identified by the objective blood test in a fasting state at the second follow-up. Third, the Lifelines FFQ

at baseline was conceptually limited as it could not distinguish several food groups. For instance, white bread and wholegrain bread do not exert similar metabolic effects, but we could only analyze them as one food group. Nevertheless, wholegrain bread was estimated to account for approximately 70% of the bread consumption in The Netherlands, which limits the potential bias caused by the possible misclassification.[40] This may also partly explain that bread was found to be a protective dietary factor in the male dietary pattern. In addition, we are currently unable to investigate to what extent participants might change their diet during the follow-up and how this could potentially influence our results. Finally, because the Lifelines cohort study mainly consists of participants residing in the northern Netherlands, it may not be possible to extrapolate our findings to other population groups. Previous study has found that the associations between similar dietary patterns and fasting glucose differed by ethnicity.[39] Although, by calculating the simplified dietary pattern scores, the issue of generalizability of the identified dietary patterns may be partly resolved.[41]

Blood lipids have been shown to be important risk factors for the development of type 2 diabetes, including several pathophysiological pathways, such as inflammation and pancreatic beta cell dysfunction.[7-11] To our knowledge, this is the first study that exclusively used five blood lipid markers as response variables to derive dietary patterns using RRR, and subsequently examined their associations with risk of type 2 diabetes. RRR can be a powerful tool for dietary pattern analysis if the response variables represent a specific diet-disease pathway of interest. The derived dietary patterns by RRR do not maximize the variation in habitual food consumption but explain the largest variation in the hypothesized intermediate risk factors (response variables).[17] Our results support this statement and confirm that dietary factors may affect risk of type 2 diabetes partly through altered blood lipid profiles. More specifically, for instance, we found that comparing the highest quintile (Q5) to the lowest quintile (Q1) of dietary pattern scores, triglycerides was 10.0% and 20.3% higher in women and men, respectively; while HDL-cholesterol was 12.6% and 7.2% lower in women and men, respectively. However, no clear trend was observed for other biomarkers, such as BMI and HbA$_{1c}$. Adjustments for metabolic factors also only partly attenuated the

associations between the identified dietary patterns and risk of type 2 diabetes.

Previous RRR studies on type 2 diabetes have examined various sets of response variables, including markers of glucose homeostasis, insulin resistance, and inflammation.[4,5,18,32] The dietary patterns identified from those studies mostly reported strong associations with risk of type 2 diabetes. This is probably due to that those biomarkers are more directly linked to the later stage of pathogenesis of type 2 diabetes. In addition, it is also conceivable that the difference of the study population may explain the relatively weaker associations in our study. As per our inclusion criteria, our population at baseline was free from impaired glucose tolerance and did not use lipid modifying agents, being thus intrinsically at low risk for type 2 diabetes. Nevertheless, this pre-selection of the study population may help reduce the chance of reverse causation as participants with high risk of developing type 2 diabetes may alter their dietary behaviors before the entry of the study.

Conclusions

In conclusion, we identified dietary patterns characterized by high intake of sugary beverages, added sugar, and juice; and low intake of fruits, vegetables, nuts/seeds, cereals, and tea. These identified dietary patterns specifically reflect the variation in blood lipids and were significantly associated with increased risk of type 2 diabetes. In other words, our results indicate that this diet-disease association was possibly mediated by diet-induced alterations in blood lipid profiles. Our findings provide important insights in optimizing blood lipids for the prevention of type 2 diabetes through dietary modifications.

Reference

1.   Yach D, Stuckler D, Brownell KD. Epidemiologic and economic consequences of the global epidemics of obesity and diabetes. Nat Med 2006; 12(1): 62-6.

2.   GBD Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019; 393(10184): 1958-72.

3.   Nettleton JA, Steffen LM, Ni H, Liu K, Jacobs DR, Jr. Dietary patterns and risk of incident type 2 diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA). Diabetes Care 2008; 31(9): 1777-82.

4.   Schulze MB, Hoffmann K, Manson JE, et al. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. Am J Clin Nutr 2005; 82(3): 675-84; quiz 714-5.

5.   Liese AD, Weis KE, Schulz M, Tooze JA. Food intake patterns associated with incident type 2 diabetes: the Insulin Resistance Atherosclerosis Study. Diabetes Care 2009; 32(2): 263-8.

6.   Steyn NP, Mann J, Bennett PH, et al. Diet, nutrition and the prevention of type 2 diabetes. Public Health Nutr 2004; 7(1A): 147-65.

7.   Rutti S, Ehses JA, Sibler RA, et al. Low- and high-density lipoproteins modulate function, apoptosis, and proliferation of primary human and murine pancreatic beta-cells. Endocrinology 2009; 150(10): 4521-30.

8.   Kruit JK, Brunham LR, Verchere CB, Hayden MR. HDL and LDL cholesterol significantly influence beta-cell function in type 2 diabetes mellitus. Curr Opin Lipidol 2010; 21(3): 178-85.

9.   Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW. Two risk-scoring systems for predicting incident diabetes mellitus in U.S. adults age 45 to 64 years. Ann Intern Med 2009; 150(11): 741-51.

10.   Schmidt MI, Duncan BB, Bang H, et al. Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study. Diabetes Care 2005; 28(8): 2013-8.

11.   Chien K, Cai T, Hsu H, et al. A prediction model for type 2 diabetes risk among Chinese people. Diabetologia 2009; 52(3): 443-50.

12.   Sattar N, Preiss D, Murray HM, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. Lancet 2010; 375(9716): 735-42.

13.   Sonestedt E, Hellstrand S, Drake I, et al. Diet Quality and Change in Blood Lipids during 16 Years of Follow-up and Their Interaction with Genetic Risk for Dyslipidemia. Nutrients 2016; 8(5).

14.   Berglund L, Brunzell JD, Goldberg AC, et al. Evaluation and treatment of hypertriglyceridemia: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2012; 97(9): 2969-89.

15.   Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 2002; 13(1): 3-9.

16.   Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U, Boeing H. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. Am J Epidemiol 2004; 159(10): 935-44.

17.   Weikert C, Schulze MB. Evaluating dietary patterns: the role of reduced rank regression. Curr Opin Clin Nutr Metab Care 2016; 19(5): 341-6.

18.   McNaughton SA, Mishra GD, Brunner EJ. Dietary patterns, insulin resistance, and incidence of type 2 diabetes in the Whitehall II Study. Diabetes Care 2008; 31(7): 1343-8.

19.   Seah JYH, Ong CN, Koh WP, Yuan JM, van Dam RM. A Dietary Pattern Derived from Reduced Rank Regression and Fatty Acid Biomarkers Is Associated with Lower Risk of Type 2 Diabetes and Coronary Artery Disease in Chinese Adults. J Nutr 2019; 149(11): 2001-10.

20.   Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. PLOS ONE 2015; 10(9): e0137203.

21.   Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol 2015; 44(4): 1172-80.

22.   WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health. ATC/DDD Index 2020. Oslo, Norway; 2020.

23.   Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. Int J Obes Relat Metab Disord 2000; 24(9): 1119-30.

24.   WHO, International Diabetes Federation. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: report of a WHO/IDF consultation. Genève, Switzerland: WHO, 2006.

25.   van Vliet-Ostaptchouk JV, Nuotio ML, Slagter SN, et al. The prevalence of metabolic syndrome and metabolically healthy obesity in Europe: a collaborative analysis of ten large cohort studies. BMC Endocr Disord 2014; 14: 9.

26.   Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). JAMA 2001; 285(19): 2486-97.

27.   Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018; 71(6): 1269-324.

28.   Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003; 56(12): 1163-9.

29.   Streppel MT, de Vries JH, Meijboom S, et al. Relative validity of the food frequency questionnaire used to assess dietary intake in the Leiden Longevity Study. Nutr J 2013; 12: 75.

30.   RIVM. Dutch food composition database (NEVO). Bilthoven, The Netherlands; 2006.

31.   van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999; 18(6): 681-94.

32.   Imamura F, Lichtenstein AH, Dallal GE, Meigs JB, Jacques PF. Generalizability of dietary

patterns associated with incidence of type 2 diabetes mellitus. Am J Clin Nutr 2009; 90(4): 1075-83.

33.   Pastorino S, Richards M, Pierce M, Ambrosini GL. A high-fat, high-glycaemic index, low-fibre dietary pattern is prospectively associated with type 2 diabetes in a British birth cohort. Br J Nutr 2016; 115(9): 1632-42.

34.   Heidemann C, Hoffmann K, Spranger J, et al. A dietary pattern protective against type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)--Potsdam Study cohort. Diabetologia 2005; 48(6): 1126-34.

35.   Bray GA, Popkin BM. Dietary sugar and body weight: have we reached a crisis in the epidemic of obesity and diabetes?: health be damned! Pour on the sugar. Diabetes Care 2014; 37(4): 950-6.

36.   He K. Fish, long-chain omega-3 polyunsaturated fatty acids and prevention of cardiovascular disease--eat fish or take fish oil supplement? Prog Cardiovasc Dis 2009; 52(2): 95-114.

37.   Kendall CW, Josse AR, Esfahani A, Jenkins DJ. Nuts, metabolic syndrome and diabetes. Br J Nutr 2010; 104(4): 465-73.

38.   Wallin A, Di Giuseppe D, Orsini N, Akesson A, Forouhi NG, Wolk A. Fish consumption and frying of fish in relation to type 2 diabetes incidence: a prospective cohort study of Swedish men. Eur J Nutr 2017; 56(2): 843-52.

39.   Dekker LH, van Dam RM, Snijder MB, et al. Comparable Dietary Patterns Describe Dietary Behavior across Ethnic Groups in the Netherlands, but Different Elements in the Diet Are Associated with Glycated Hemoglobin and Fasting Glucose Concentrations. J Nutr 2015; 145(8): 1884-91.

40.   van Rossum CTM, Fransen HP, Verkaik-Kloosterman J, Buurma-Rethans EJM, Ocké MC. Dutch National Food Consumption Survey 2007-2010. Bilthoven, The Netherlands: RIVM, 2011.

41.   Schulze MB, Hoffmann K, Kroke A, Boeing H. An approach to construct simplified measures of dietary patterns from exploratory factor analysis. Br J Nutr 2003; 89(3): 409-19.

Additional Information

Contributions

MD, LHD, and GJN designed the study. MD analyzed the data and drafted the manuscript. LHD, JJC, and GJN contributed to the discussion and critically reviewed/edited the manuscript. MD has primary responsibility for the final content. All authors approved the final content of the manuscript.

Supplementary Data

Online **Supplementary Files** can be accessed via
doi.org/10.1016/j.clnu.2021.04.046, or by scanning the QR code.
Article PMID: 34237698.
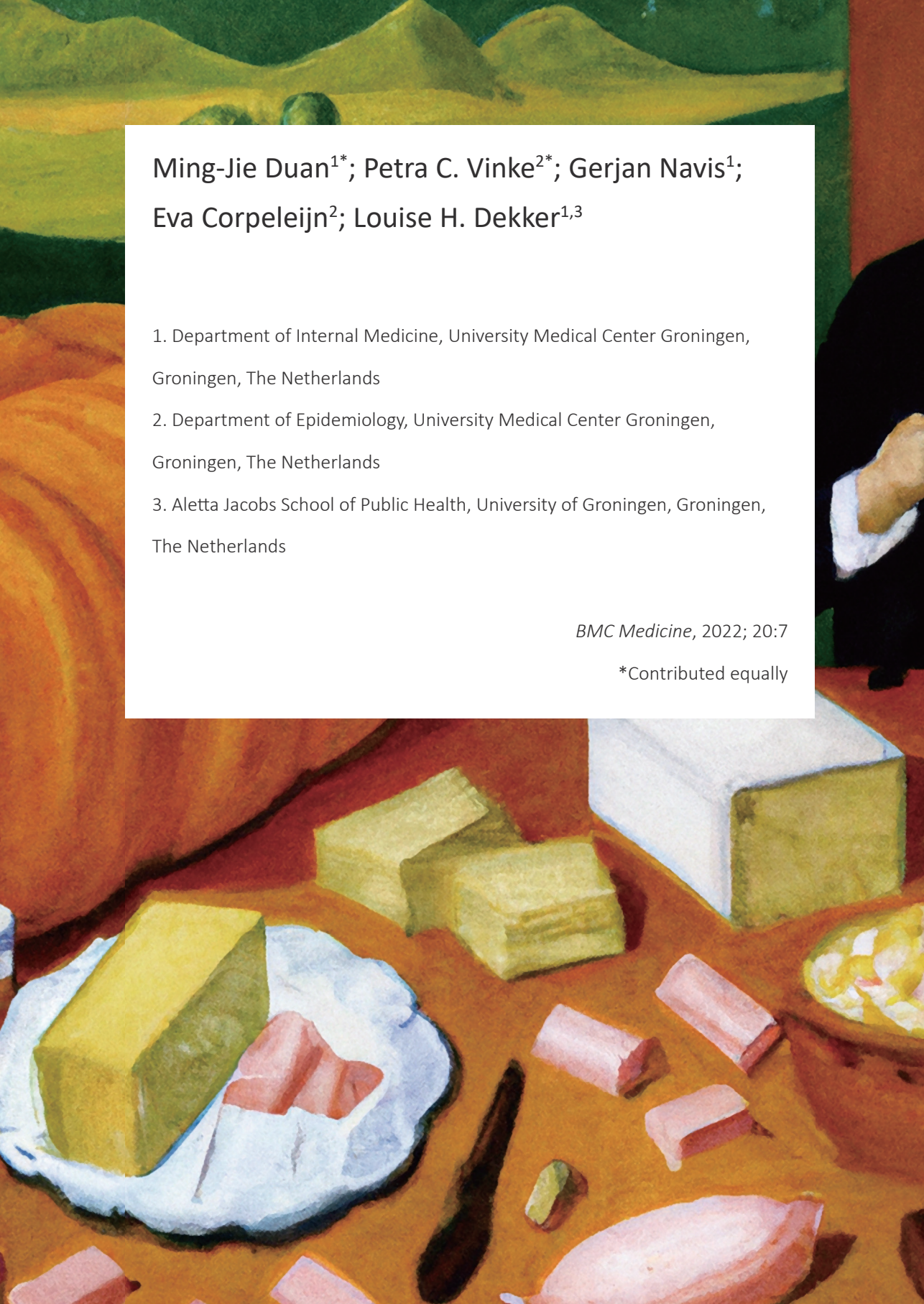
Ultra-processed food and incident type 2 diabetes: studying the underlying consumption patterns to unravel the health effects of this heterogeneous food category in the prospective Lifelines cohort

CHAPTER 3

Ming-Jie Duan[1*]; Petra C. Vinke[2*]; Gerjan Navis[1];
Eva Corpeleijn[2]; Louise H. Dekker[1,3]

1. Department of Internal Medicine, University Medical Center Groningen,
Groningen, The Netherlands
2. Department of Epidemiology, University Medical Center Groningen,
Groningen, The Netherlands
3. Aletta Jacobs School of Public Health, University of Groningen, Groningen,
The Netherlands

*Contributed equally

ABSTRACT

Background

The overall consumption of ultra-processed food (UPF) has previously been associated with type 2 diabetes. However, due to the substantial heterogeneity of this food category, in terms of their nutritional composition and product type, it remains unclear whether previous results apply to all underlying consumption patterns of UPF.

Methods

Of 70,421 participants (35-70 years, 58.6% women) from the Lifelines cohort study, dietary intake was assessed with a food frequency questionnaire. UPF was identified according to the NOVA classification. Principal component analysis (PCA) was performed to derive UPF consumption patterns. The associations of UPF and adherence to UPF consumption patterns with incidence of type 2 diabetes were studied with logistic regression analyses adjusted for age, sex, diet quality, energy intake, alcohol intake, physical activity, TV watching time, smoking status, and educational level.

Results

During a median follow-up of 41 months, a 10% increment in UPF consumption was associated with a 25% higher risk of developing type 2 diabetes (1128 cases; OR 1.25 [95% CI 1.16, 1.34]). PCA revealed four habitual UPF consumption patterns. A pattern high in cold savory snacks (OR 1.16 [95% CI 1.09, 1.22]) and a pattern high in warm savory snacks (OR 1.15 [95% CI 1.08, 1.21]) were associated with an increased risk of incident type 2 diabetes; a pattern high in traditional Dutch cuisine was not associated with type 2 diabetes incidence (OR 1.05 [95% CI 0.97, 1.14]), while a pattern high in sweet snacks and pastries was inversely associated with type 2 diabetes incidence (OR 0.82 [95% CI 0.76, 0.89]).

Conclusions

The heterogeneity of UPF as a general food category is reflected by the discrepancy in associations between four distinct UPF consumption patterns and incident type 2

diabetes. For better public health prevention, research is encouraged to further clarify how different UPF consumption patterns are related to type 2 diabetes.

3

Abbreviations

FFQ – Food frequency questionnaire

LLDS – Lifelines diet score

MVPA – Moderate-to-vigorous physical activity

PCA – Principal component analysis

UPF – Ultra-processed food

Background

The magnitude of the worldwide burden of diabetes continues to grow. It is estimated that 578 million people will be living with diabetes globally by the year 2030, approximately 90% of which will be type 2 diabetes.[1] Abundant evidence has shown that adherence to a healthy diet (such as the Mediterranean diet) is crucial to the prevention of type 2 diabetes.[2] However, these dietary patterns studied generally focused on conventional food groups such as fruits and vegetables.[3-5] Recent studies show that higher intake of ultra-processed food (UPF) is associated with higher risk of type 2 diabetes.[6-9] However, UPF forms a highly heterogeneous food category, especially in terms of their nutritional composition, product types, and contribution to a habitual diet. It is therefore unclear whether previous results that identify total intake of UPF as a single risk factor for type 2 diabetes apply to all underlying consumption patterns that fall under this "umbrella-term".

Research on UPF has been facilitated by the development of the NOVA classification. The NOVA classification is a frequently used method to categorize food and drinks based on the nature, extent, and purpose of food processing. The NOVA classification comprises four categories, ranging from un-processed/minimally processed food to UPF.[10-12] According to the NOVA classification, UPF is mostly formulated from food substances and industrial ingredients that undergo a series of chemical and physical manufacturing processes. The resulting food products are often pre-packed, contain little or no intact (un-processed) food, and are considered microbiologically safe, convenient, and palatable.[10-12]

Since the intake of UPF has substantially increased in most parts of the world over the past decades,[13] there is an increasing interest in the potential health impacts of UPF. Prospective cohort studies on the associations between UPF and health so far mostly focused on total intake of UPF. These prospective cohort studies found that higher intake of UPF was associated with higher risks of obesity,[14-16] cardiovascular diseases,[17-19] cancer,[20] mortality,[21-23] the metabolic syndrome,[24] and type 2 diabetes.[6-9] Associations established from these studies underline the fact that UPF is not neglectable when

studying dietary effects on disease outcomes.

However, an often overlooked virtue of UPF is that it forms a highly heterogeneous food category. Food products considered as UPF are heterogeneous with respect to their nutritional composition, as well as their contribution to a habitual diet, and the context in which they are consumed.[25] For example, according to the frequently used NOVA classification,[10-12] UPF includes pre-packaged bread, a staple food item which in many cultures is consumed with main meals; as well as cakes or fast food, which are consumed more occasionally. Therefore, results from previous studies analyzing UPF as one single food group may not apply to all underlying consumption patterns that fall within this food group. Scientific evidence so far may therefore not be sufficient to formulate evidence-based guidelines and health policies regarding UPF in the battle against type 2 diabetes.

In this study, we first aimed to assess the association between overall UPF intake and incident type 2 diabetes. More importantly, we aimed to identify underlying consumption patterns of UPF and to investigate how they were related to incident type 2 diabetes in a large cohort of Dutch adults.

Methods

Cohort Design and Study Population

The Lifelines cohort study is a multidisciplinary prospective population-based cohort study that applies a unique three-generation design, the health and health-related behaviors of 167,729 persons living in the north of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical, and psychological factors, which contribute to health and disease of the general population.

Participants were included in the study between 2006 and 2013. So far, four follow-up assessment rounds took place, i.e., T1=baseline, median (interquartile) months to follow-up rounds: T2=13 (13-15), T3=25 (23-28), and T4=44 (35-51). Comprehensive

physical examinations, biobanking, and questionnaires were conducted at T1 and T4, and follow-up questionnaires (including questions for diabetes status) were issued to participants at T2 and T3. The timeline of data collection of the Lifelines cohort study is presented in **Additional file 1: Fig. S1**. Before study entry, a signed informed consent form was obtained from each participant. The Lifelines study is conducted according to the principles of the Declaration of Helsinki and approved by the Medical Ethics Committee of the University Medical Center Groningen, The Netherlands (approval number 2007/152). The overall design and rationale of the study have been described in detail elsewhere.[26,27]

Participants aged between 35 and 70 years who were free of diabetes at baseline, and for whom valid dietary intake data was available were included in this study. The ascertainment of prevalent diabetes cases at baseline was based on (1) self-report questionnaires, (2) fasting glucose ≥ 7.0 mmol/L, (3) HbA$_{1c}$ ≥ 48 mmol/mol (6.5%),[28] and (4) medication use on glucose-lowering agents (ATC code A10).[29] Dietary intake data was considered unreliable when the ratio between reported energy intake and basal metabolic rate (calculated with the Schofield equation)[30] was below 0.50 or above 2.75 (based on the considerations by Goldberg).[31] Moreoever, participants for whom only baseline data was available, or who reported the development of type 1 diabetes or gestational diabetes during the follow-ups, were excluded. In total, 70,421 participants (41,243 women, 29,178 men) were included in the analysis (**Additional file 1: Fig. S2**).

Data Collection

Ascertainment of Incident Type 2 Diabetes

Incident type 2 diabetes was assessed by self-report questionnaires at the two follow-ups (T2, from 2011 to 2015; and T3, from 2012 to 2016) and the second assessment (T4, from 2014 to 2018). Additionally, blood glucose and HbA$_{1c}$ measurements were available at the second assessment (T4). Participants were considered an incident case if they met one of the following criteria: (1) self-reported newly developed type 2 diabetes since last time they filled out a questionnaire, (2) fasting glucose ≥ 7.0 mmol/L, or (3) HbA$_{1c}$ ≥ 48 mmol/mol (6.5%).[28] However, data on prescribed medication was

not available during follow-ups and the precise time of diabetes diagnosis was not documented.

Clinical Measurements

Blood samples were collected by venipuncture in a fasting state between 8 and 10 am and were further transferred to the Lifelines central laboratory for analysis. Serum levels of glucose and HbA_{1c} were subsequently analyzed. Anthropometric measurements were made by trained research staff following standardized protocols. These measurements were performed without shoes and heavy clothing. BMI was calculated as weight in kilograms divided by the square of height in meters.

Dietary Assessment

At baseline, dietary consumption was assessed using a validated 110-item semi-quantitative food frequency questionnaire (FFQ), which was designed to assess the food consumption (including alcohol) over the previous month.[32] The questionnaire assessed the frequency of consumption and portion sizes, the latter of which were estimated by fixed portion sizes (e.g., slices of bread, pieces of fruit) and commonly used household measures (e.g., cups, spoons). For insight into the overall diet quality, the food-based Lifelines Diet Score (LLDS) was calculated. This score ranks the relative intake of nine food groups with positive health effects (vegetables, fruit, whole grain products, legumes/nuts, fish, oils/soft margarines, unsweetened dairy, coffee, and tea) and three food groups with negative health effects (red/processed meat, butter/hard margarines, and sugar-sweetened beverages). The development of this score has been described in detail elsewhere.[33]

Categorizing the Degree of Food Processing – The NOVA Classification

The NOVA classification was used to categorize all 110 food items into the four proposed categories: (1) un-processed or minimally processed food (e.g., fresh vegetables/fruits, unprocessed meat), (2) processed culinary ingredients (e.g., butter/oil for cooking, sugar, salt), (3) processed food (e.g., canned vegetables/fish, fruits in syrup), and (4) ultra-processed food (e.g., processed meat, soft drinks).[11,12] The proportion (weight

ratio, %) of intake of UPF in the total weight of food and beverages consumed per day was calculated and was then divided into sex-specific quartiles for further analyses. Using weight ratio of UPF intake accounts for the food that does not provide energy (e.g., artificially sweetened beverages) as well as non-nutritional factors (e.g., additives, by-products during processing). The categorization of the items was verified by four of the authors and can be found in **Additional file 1: Table S1**.

Assessment of Other Baseline Covariates

Age, smoking status, TV watching time, and educational level were assessed by self-administered questionnaires. Smoking status was categorized as never, former, and current smoker. The highest educational level achieved was categorized as (1) low – junior general secondary education or lower (International Standard Classification of Education [ISCED] level 0, 1, or 2); (2) middle – secondary vocational education and senior general secondary education (ISCED level 3 or 4); and (3) high – higher vocational education or university (ISCED level 5 or 6).[34] The validated Short QUestionnaire to ASsess Health-enhancing physical activity (SQUASH) was used to assess physical activity level.[35] From the SQUASH data, leisure time and commuting physical activities, including sports, at moderate (4.0-6.4 MET) to vigorous (≥ 6.5 MET) intensity (non-occupational moderate-to-vigorous physical activity [MVPA]), were calculated in minutes per week.[35] The variable was categorized by dividing participants who reported any non-occupational MVPA into sex-specific quartiles. For participants who reported zero non-occupational MVPA, the categorical variable was coded as 0.

Statistical Analysis

Consumption Patterns of Ultra-processed Food

As UPF is highly heterogeneous on multiple concepts (i.e., nutrient density, nutrient composition, taste, snack or main meal items), it is difficult to create well-founded subgroups. Therefore, instead of using a priori defined subgroups, we used principal component analysis (PCA) to derive underlying consumption patterns of UPF, to obtain real-world insight into the intake of this highly heterogeneous food category. Based on the Scree plot, eigenvalues, and explained variations, four UPF consumption patterns

were selected. Thereafter, the derived components were orthogonally rotated to obtain uncorrelated components to enhance interpretability. We selected food items with absolute factor loadings ≥ 0.20 to construct simplified pattern scores while retaining the weight (factor loading) of each selected food item. The simplified UPF consumption pattern scores (hereafter referred to as UPF consumption patterns) were standardized and then divided into sex-specific quartiles for further analyses. Sensitivity analysis was performed by repeating the PCA procedure 3 times on a random half sample.

Risk of Incident Type 2 Diabetes

Associations between UPF intake (total intake [continuous or sex-specific quartiles] and UPF consumption patterns [continuous or sex-specific quartiles]) with incident type 2 diabetes were estimated with logistic regression models and results were shown as ORs with 95% confidence intervals. In models where UPF intake was included as a continuous variable (weight ratio), ORs regarding a 10% absolute increment of UPF in the total diet were calculated. In four steps, the analyses were adjusted for (1) age and sex; (2) diet quality (LLDS), total energy intake, and alcohol intake; (3) non-occupational MVPA, TV watching time, smoking status, and educational level; and (4) BMI (continuous). This addition of BMI in the last step aimed to investigate the role of this intermediate factor in the association between UPF and type 2 diabetes. Additionally, the possibility of effect modification by sex was tested by including the interaction-term for sex and UPF intake in the models. To account for missing covariates, multiple imputation by chained equations was performed to deal with missing data for non-occupational MVPA (proportion of missing 6.5%), TV watching time (proportion of missing 0.6%), smoking status (proportion of missing 0.6%), and educational level (proportion of missing 0.4%).

We performed several sensitivity analyses to test the robustness of our results. First, analyses were performed using energy-adjusted UPF intake. Second, sensitivity analyses on missing data were performed by complete case analysis. Moreover, we excluded participants who were lost to follow-up after 24 months, in an attempt to address the possible reverse causation caused by short follow-up time.

Post Hoc Analysis

Baseline Diabetes Risk and Ultra-processed Food Consumption Patterns

Individuals' awareness of elevated diabetes risk may have influenced individuals' dietary behaviors at baseline. Therefore, linear regression models were performed to investigate whether type 2 diabetes risk at baseline, as calculated with the PROCAM risk algorithm (**Additional file 1: Table S2**),[36] was associated with the total intake of UPF and distinctive UPF consumption patterns. In the linear regression models, the total intake of UPF or the UPF consumption pattern scores were set as dependent variable one by one. The analyses were additionally adjusted for the same covariates as described above, except for energy intake and BMI. Energy intake was not considered to be a confounding factor, and BMI was not included due to its high correlation with the PROCAM diabetes risk algorithm (Pearson correlation coefficient = 0.835).

Results

Baseline characteristics across quartiles of UPF consumption are shown in **Table 1**. In the total study population, the median contribution of UPF to the total diet was 34.9 weight% (**Additional file 1: Fig. S3**). Of all UPF groups, staple/starchy food and cereals like sliced bread or granola (22.1%), non-cheese dairy products like chocolate milk and ice cream (13.7%), and sugary beverages like lemonade or ice tea (9.7%) contributed most to the overall intake of UPF (median weight% of total UPF, **Additional file 1: Table S3**). In general, with increasing quartiles of UPF consumption, participants tended to be younger, have higher BMI, have lower type 2 diabetes risk scores, be less physically active, have worse overall diet quality, consume less alcohol, smoke less, be less highly educated, and spend more time on watching TV.

Overall Consumption of Ultra-processed Food and Risk of Incident Type 2 Diabetes

Table 2 shows the association between consumption of UPF and the risk of incident type 2 diabetes. Among 70,421 participants included in the analysis, we identified 1128 cases (550 female cases and 578 male cases, **Additional file 1: Fig. S2**) of type

2 diabetes during a median follow-up of 41 months. A significant positive association between the overall consumption of UPF and incident type 2 diabetes was observed. Per 10% absolute increment intake of UPF, participants had 33% higher odds of incident type 2 diabetes (OR 1.33 [95% CI 1.26, 1.41], P < 0.001, model 1, sex and age adjusted). This association remained significant after additional adjustment for diet quality and other covariates (OR 1.25 [95% CI 1.16, 1.34], P < 0.001, model 3). Additional adjustment for BMI further explained part of the association (OR 1.17 [95% CI 1.09, 1.26], model 4). When comparing the highest versus the lowest quartile of UPF consumption, participants in the highest quartile had an 80% higher odds of incident type 2 diabetes (OR Q4 versus Q1 1.80 [95% CI 1.47, 2.20], P-trend < 0.001, model 3). We did not find evidence of effect modification by sex (P-interaction > 0.05). Sensitivity analyses on missing data showed similar results (**Additional file 1: Table S4**). Excluding participants who only had 24 months of follow-up also yielded similar results (**Additional file 1: Table S5**). In addition, results from energy-adjusted intake of UPF were basically unchanged (**Additional file 1: Table S6**).

UPF Consumption Patterns and Incident Type 2 Diabetes

To identify habitual consumption patterns of UPF, we performed PCA analysis and selected four UPF consumption patterns. These four patterns explained 15.5% of the total variance of UPF intake. **Additional file 1: Table S7** shows the factor loadings of UPF products within their consumption patterns. Briefly, these four patterns were (1) warm savory snack pattern, characterized by high intakes of fried snacks, fries, and snack sauce; (2) cold savory snack pattern, characterized by high intakes of cheese, deli meat, and savory spreads for crackers or baguette; (3) traditional Dutch cuisine pattern, characterized by high intakes of main meal items typical for the Dutch culture, such as sliced bread, lunch meat, and gravy; and (4) sweet snack pattern, characterized by high intakes of sweet biscuits/cookies, pastries, and chocolate. Explained variance was highest for the warm savory snack pattern (5.0%) and lowest for the cold savory snack pattern (3.3%). Baseline characteristics across different UPF consumption patterns (highest quartiles) are shown in **Additional file 1: Table S8**. Similar UPF patterns were identified when analyzing the random half sample (**Additional file 1: Table S9**).

Table 1. Baseline characteristics of study participants according to sex-specific quartiles of ultra-processed food consumption (n = 70,421)[a,b]

	Quartiles of ultra-processed food consumption				
	First (n = 17,604)	Second (n = 17,606)	Third (n = 17,606)	Fourth (n = 17,606)	Total (n = 70,421)
Age, years	52.3±9.1	50.2±8.8	48.3±8.4	45.7±7.6	49.1±8.8
Sex, %					
Women	58.6	58.6	58.6	58.6	58.6
Men	41.4	41.4	41.4	41.4	41.4
Ultra-processed food intake, weight%	23.7 (20.3, 26.0)	31.6 (29.0, 34.1)	38.4 (35.6, 40.5)	48.7 (45.2, 53.9)	34.9 (28.1, 42.7)
Lifeline diet score	28.6±5.1	25.4±4.7	22.8±4.6	19.2±4.7	24.0±5.9
Total energy intake, kcal/day	1811±520	2032±543	2150±579	2261±647	2063±598
Total alcohol intake, grams/day	6.2 (1.4, 12.1)	5.8 (1.3, 11.4)	4.4 (1.0, 10.4)	2.9 (0.4, 9.9)	4.7 (0.9, 11.2)
Fasting glucose, mmol/L	4.96±0.51	4.96±0.50	4.96±0.50	4.97±0.51	4.96±0.50
HbA_{1c}, %	5.56±0.30	5.56±0.30	5.54±0.30	5.54±0.30	5.55±0.30
HbA_{1c}, mmol/mol	37.3±3.2	37.2±3.2	37.1±3.3	37.0±3.3	37.2±3.3
BMI, kg/m^2	25.6±3.8	25.6±3.9	26.2±4.0	26.7±4.5	26.2±4.1
Highest tertile of PROCAM diabetes risk algorithm, %	37.2	33.2	31.9	30.4	33.2
MVPA, minutes/week[c]	240 (90, 420)	210 (80, 380)	180 (60, 360)	150 (60, 330)	190 (60, 365)

Table 1. Continued.

	Quartiles of ultra-processed food consumption				
	First (*n* = 17,604)	Second (*n* = 17,606)	Third (*n* = 17,606)	Fourth (*n* = 17,606)	Total (*n* = 70,421)
Educational level, %					
Low	28.4	29.5	30.5	33.1	30.4
Middle	34.8	37.7	40.1	42.8	38.9
High	36.2	32.4	29.0	23.7	30.3
Smoking status, %					
Never	39.8	43.6	45.4	48.7	44.4
Former	44.2	40.0	36.8	30.6	37.9
Current	15.4	15.8	17.4	20.0	17.2
TV watching time, hours/day	2.4±1.3	2.4±1.3	2.5±1.2	2.6±1.4	2.5±1.3

[a] Data are expressed as unadjusted mean ± standard deviation for age, Lifelines diet score (no unit), total energy intake, fasting glucose, HbA_{1c}, BMI, and TV watching time; data are expressed as median (interquartile) for ultra-processed food intake (weight percentage), total alcohol intake, and MVPA; data are expressed as observed percentage for sex, highest tertile of PROCAM diabetes risk algorithm, educational level, and smoking status.

[b] Tests for significant differences in baseline characteristics (except for sex) across quartiles of ultra-processed food consumption were performed using Kruskal-Wallis test or χ^2 test for proportion, as appropriate; P < 0.001 for all baseline characteristics except for fasting glucose P = 0.019.

[c] MVPA denotes non-occupational moderate-to-vigorous physical activity level.

3

Table 2. Associations between consumption of ultra-processed food and incident type 2 diabetes

	Quartiles of ultra-processed food consumption				P-trend	Continuous[a]	P value
	First	Second	Third	Fourth			
Cases/population	255/17,604	247/17,606	272/17,605	354/17,606		1128/70,421	
Incidence, %	1.4	1.4	1.5	2.0		1.6	
Model 1[b]	1	1.11 (0.93, 1.33)	1.39 (1.17, 1.65)	2.17 (1.83, 2.58)	<0.001	1.33 (1.26, 1.41)	<0.001
Model 2[c]	1	1.08 (0.90, 1.30)	1.30 (1.07, 1.57)	1.87 (1.52, 2.30)	<0.001	1.27 (1.18, 1.36)	<0.001
Model 3[d]	1	1.08 (0.90, 1.30)	1.28 (1.06, 1.55)	1.80 (1.47, 2.20)	<0.001	1.25 (1.16, 1.34)	<0.001
Model 4[e]	1	1.04 (0.87, 1.26)	1.20 (0.99, 1.45)	1.56 (1.27, 1.92)	<0.001	1.17 (1.09, 1.26)	<0.001

[a] Continuous model indicates OR (95% CI) for an absolute increment of 10% consumption of ultra-processed food in the total diet.
[b] Model 1: OR (95% CI) derived from multivariate logistic regression models adjusted for age and sex, n = 70,421.
[c] Model 2: OR (95% CI) derived from multivariate logistic regression models adjusted for Model 1 covariates plus Lifelines diet score, total energy intake, and alcohol intake, n = 70,421.
[d] Model 3: OR (95% CI) derived from multivariate logistic regression models adjusted for Model 2 covariates plus smoking status, educational level, non-occupational moderate-to-vigorous physical activity level, and TV watching time, n = 70,418.
[e] Model 4: OR (95% CI) derived from multivariate logistic regression models adjusted for Model 3 covariates plus BMI, n = 70,403.

Associations between UPF habitual consumption patterns and incident type 2 diabetes are shown in **Table 3**. For the ORs treating consumption pattern scores as a continuous variable, the warm savory snack pattern (OR 1.15 [95% CI 1.08, 1.21], P < 0.001) and the cold savory snack pattern (OR 1.16 [95% CI 1.09, 1.22], P < 0.001) were positively associated with incident type 2 diabetes (model 3). For the traditional Dutch cuisine pattern, no significant association was found with incident type 2 diabetes (OR 1.05 [95% CI 0.97, 1.14], P = 0.207, model 3). Oppositely, higher adherence to the sweet snack pattern was negatively associated with incident type 2 diabetes (OR 0.82 [95% CI 0.76, 0.89], P < 0.001, model 3). Results were consistent when comparing the highest quartile with the lowest quartile of the consumption pattern scores. Additional adjustment for baseline BMI (model 4) led to minor attenuation of all associations, except for the warm savory snack pattern. For the latter, the ORs were moderately attenuated and became insignificant, but were still positively associated with higher risk of incident type 2 diabetes (OR 1.07 [95%CI 1.00, 1.14], P = 0.057; OR Q4 versus Q1 1.17 [95%CI 0.96, 1.44], P-trend = 0.097). Sensitivity analysis on missing data (complete case analysis) yielded similar results (**Additional file 1: Table S4**). Results are basically unchanged when excluding participants who were lost to follow-up after 24 months (**Additional file 1: Table S5**).

Baseline Diabetes Risk and Ultra-processed Food Consumption Patterns

To explore how diet may be dependent on the baseline health condition, the estimated diabetes risk score at baseline was related to the total intake of UPF and four UPF consumption patterns (**Table 4**). The results showed that baseline type 2 diabetes risk was positively associated with the total UPF intake, as well as the warm savory snack pattern and the cold savory snack pattern, but negatively associated with the traditional Dutch cuisine pattern and the sweet snack pattern. The strongest association was found for the sweet snack pattern (β = -0.104 [95% CI -0.113, -0.094], P < 0.001), which indicates that those with high diabetes risk scores at baseline had lower adherence to the sweet UPF pattern. Results from complete case analysis are basically unchanged (**Additional file 1: Table S10**).

Table 3. Associations between ultra-processed food consumption pattern scores and incident type 2 diabetes

Consumption patterns scores	Models	Quartiles of consumption pattern scores of ultra-processed food				P-trend	Continuous	P value
		First	Second	Third	Fourth			
Warm savory snack pattern	Cases/population	291/17,561	272/17,649	273/17,605	292/17,606			
	Model 1[a]	1	1.15 (0.97, 1.37)	1.42 (1.19, 1.69)	1.82 (1.52, 2.18)	<0.001	1.22 (1.16, 1.27)	<0.001
	Model 2[b]	1	1.13 (0.95, 1.34)	1.36 (1.14, 1.64)	1.74 (1.43, 2.12)	<0.001	1.22 (1.16, 1.29)	<0.001
	Model 3[c]	1	1.07 (0.90, 1.27)	1.23 (1.02, 1.48)	1.43 (1.17, 1.75)	<0.001	1.15 (1.08, 1.21)	<0.001
	Model 4[d]	1	1.02 (0.86, 1.22)	1.11 (0.92, 1.34)	1.17 (0.96, 1.44)	0.097	1.07 (1.00, 1.14)	0.057
Traditional Dutch cuisine pattern	Cases/population	276/17,605	282/17,605	301/17,605	269/17,606			
	Model 1[a]	1	1.01 (0.85, 1.19)	1.11 (0.94, 1.31)	1.06 (0.89, 1.26)	0.332	1.03 (0.97, 1.10)	0.330
	Model 2[b]	1	1.06 (0.89, 1.26)	1.19 (0.99, 1.43)	1.15 (0.94, 1.42)	0.113	1.07 (0.99, 1.15)	0.101
	Model 3[c]	1	1.05 (0.88, 1.25)	1.16 (0.96, 1.39)	1.11 (0.90, 1.37)	0.192	1.05 (0.97, 1.14)	0.207
	Model 4[d]	1	1.03 (0.87, 1.23)	1.11 (0.92, 1.33)	1.07 (0.87, 1.31)	0.411	1.03 (0.95, 1.11)	0.476

Table 3. Continued.

Consumption patterns scores	Models	Quartiles of consumption pattern scores of ultra-processed food						
		First	Second	Third	Fourth	P-trend	Continuous	P value
Sweet snack pattern	Cases/population	400/17,605	273/17,605	231/17,605	224/17,606			
	Model 1[a]	1	0.68 (0.58, 0.79)	0.58 (0.49, 0.68)	0.60 (0.50, 0.70)	<0.001	0.82 (0.76, 0.89)	<0.001
	Model 2[b]	1	0.66 (0.57, 0.78)	0.55 (0.47, 0.66)	0.53 (0.44, 0.64)	<0.001	0.79 (0.72, 0.86)	<0.001
	Model 3[c]	1	0.69 (0.59, 0.81)	0.60 (0.50, 0.71)	0.59 (0.49, 0.71)	<0.001	0.82 (0.76, 0.89)	<0.001
	Model 4[d]	1	0.75 (0.64, 0.88)	0.68 (0.57, 0.81)	0.69 (0.57, 0.84)	<0.001	0.87 (0.80, 0.94)	0.001
Cold savory snack pattern	Cases/population	292/17,605	266/17,605	289/17,605	281/17,606			
	Model 1[a]	1	0.98 (0.83, 1.16)	1.10 (0.93, 1.30)	1.08 (0.92, 1.28)	0.188	1.09 (1.02, 1.15)	0.007
	Model 2[b]	1	1.04 (0.88, 1.24)	1.20 (1.01, 1.43)	1.22 (1.03, 1.46)	0.010	1.13 (1.06, 1.20)	<0.001
	Model 3[c]	1	1.07 (0.91, 1.27)	1.27 (1.07, 1.51)	1.33 (1.12, 1.59)	<0.001	1.16 (1.09, 1.22)	<0.001
	Model 4[d]	1	1.04 (0.87, 1.23)	1.20 (1.00, 1.42)	1.20 (1.00, 1.44)	0.020	1.11 (1.04, 1.18)	0.001

[a] Model 1: OR (95% CI) derived from multivariate logistic regression models adjusted for age and sex, n = 70,421.

[b] Model 2: OR (95% CI) derived from multivariate logistic regression models adjusted for Model 1 covariates plus Lifelines diet score, total energy intake, and alcohol intake, n = 70,421.

[c] Model 3: OR (95% CI) derived from multivariate logistic regression models adjusted for Model 2 covariates plus smoking status, educational level, non-occupational moderate-to-vigorous physical activity level, and TV watching time, n = 70,418.

[d] Model 4: OR (95% CI) derived from multivariate logistic regression models adjusted for Model 3 covariates plus BMI, n = 70,403.

Table 4. Associations of ultra-processed food intake and its consumption patterns with type 2 diabetes risk and education level at baseline[a]

Ultra-processed food consumption (patterns)	Standardized beta-coefficients[b]	
	Type 2 diabetes risk	High education (low education as reference)
Total ultra-processed food intake	0.052 (0.045, 0.059)	-0.073 (-0.089,-0.058)
Warm savory snack pattern	0.091 (0.082, 0.101)	-0.212 (-0.230,-0.194)
Traditional Dutch cuisine pattern	-0.032 (-0.041,-0.023)	-0.278 (-0.298,-0.258)
Sweet snack pattern	-0.104 (-0.113,-0.094)	0.138 (0.118, 0.158)
Cold savory snack pattern	0.041 (0.032, 0.050)	0.274 (0.254, 0.294)

[a] Type 2 diabetes risk at baseline was assessed by PROCAM diabetes risk algorithm (Supplementary Table S2). This table has been slightly adapted from the published version by adding the beta-coefficients for education (high education compared with low education).
[b] Standardized beta-coefficients (95% CI) derived from multivariate linear regression models adjusted for age, sex, Lifelines diet score, alcohol intake, smoking status, non-occupational moderate-to-vigorous physical activity level, and TV watching time, all P values < 0.001, n = 70,085.

3

Discussion

In this large population-based cohort study, the overall consumption of UPF was associated with a higher risk of type 2 diabetes, independent of overall diet quality and energy intake. We illustrated the importance of considering the heterogeneity of UPF when studying its health effects, as associations with incident type 2 diabetes varied across different patterns of UPF consumption. A positive association with incident type 2 diabetes was found for both warm savory snack and cold savory snack UPF consumption patterns, while a negative association was found for sweet snack UPF pattern. On the other hand, the absence of a clear association between diabetes risk and the traditional Dutch cuisine UPF pattern, which was high in main meal food items, suggests that not all types of UPF are necessarily detrimental to health.

Over the past few years, scientific interests and public awareness on UPF have risen substantially.[11,37] So far, four studies have investigated the association of UPF with type 2 diabetes.[6-9] Our results provide an independent confirmation of the association between UPF intake and incident type 2 diabetes in a different population setting. When comparing our results to those from the French NutriNet-Santé cohort and the UK Biobank cohort in which similar methods were used, the extent to which UPF contributed to the habitual diet differed considerably. The mean weight percentage of UPF in the diet was 35.9% in this Dutch cohort, versus 15.4% in the French cohort and 22.1% in the British cohort.[6,7] Nevertheless, the reported hazard ratio of 1.15 in the previous French study and 1.12 in the UK Biobank study, regarding each 10 percent increment in the proportion of UPF in the diet, was comparable to our OR of 1.17 in our fourth model, adjusting for comparable potential confounding factors.[6,7] In addition, it is noteworthy that in all three studies, associations were independent of the overall diet quality as well as total energy intake. This consolidates the potential role of UPF as an independent dietary factor in the development of type 2 diabetes. More importantly, it emphasizes that eating an otherwise healthy diet may not fully compensate for the detrimental effects of UPF.

Notwithstanding the high heterogeneity among different types of UPF, previous

studies on the health consequences of UPF mainly focused on its overall intake. To our knowledge, the current study is the first that investigated the relation of overall intake and consumption patterns of UPF with incident type 2 diabetes in a large population-based sample. Our findings emphasize that it is crucial to consider various habitual UPF consumption patterns and their unique food groups when studying their health effects. In line with overall UPF intake, both the warm savory snack and the cold savory snack UPF patterns were associated with higher risks of type 2 diabetes. Results deviated for the traditional Dutch cuisine pattern and the sweet snack pattern, as the associations with type 2 diabetes were absent for the first, and inverse for the latter. The absence of an association for the traditional Dutch cuisine pattern illustrates that the detrimental effects of UPF may not be solely due to the degree of food processing. As UPF forms a highly heterogeneous food category, it is also important to consider their nutritional quality.[38] For instance, a key food product in the traditional Dutch cuisine pattern was sliced bread. Despite mostly being ultra-processed, approximately 70% of the sliced bread consumed in The Netherlands is brown bread (made with a mixture of whole-wheat and white flour) or whole-wheat bread, and therefore often high in fiber and micronutrients. Higher intake of fiber and whole-wheat products was found to be associated with lower risk of type 2 diabetes.[39] On the other hand, the UPF products identified in these two savory snack patterns are generally high in salt and fat and are often energy dense. It is conceivable that they may increase diabetes risk through metabolic disturbances, such as elevated blood pressure and lipid abnormality.[40,41] Therefore, a cautious interpretation of the health effects of UPF is warranted. More specifically, their effects on health may be determined by more than the level of food processing alone, which makes that not all types of UPF are necessarily detrimental to health.

Despite remaining statistically significant, our observation that estimates for the associations between UPF intake and type 2 diabetes were clearly attenuated when additionally adjusting for BMI, illustrating that BMI plays a role in the studied association. This role, however, may be two-fold, as BMI may be both a confounding and a mediating factor. Individuals with higher baseline BMI appeared to have higher

total UPF intake, as well as a higher risk of type 2 diabetes, showing its confounding property. However, since previous prospective studies have illustrated that UPF is a risk factor for obesity,[14-16] higher intake of UPF may also increase type 2 diabetes risk through an increase in body weight, which illustrates the potential mediating role of BMI in the associations studied. However, we also could not rule out the possibility of residual confounding, even in our analysis various covariates (including demographic, lifestyle, and socioeconomic factors) were adjusted. Future studies, preferably in the form of randomized controlled trials, are required to help disentangle the role of BMI in the relationship between UPF and health.

Our finding that higher adherence to the sweet snack UPF pattern was associated with lower risk of incident type 2 diabetes was counterintuitive. Previous evidence indicates that the intake of dietary sugar from food and beverages was associated with weight gain and obesity, and may thus contribute to the risk of type 2 diabetes.[42,43] Nonetheless, a study in EPIC (European Prospective Investigation into Cancer and Nutrition) also found that non-consumers of cakes and cookies had a higher risk of type 2 diabetes.[44] To assess for possible reverse causation, we performed a post hoc analysis to evaluate whether individuals' baseline type 2 diabetes risk score was involved in this unexpected finding. As shown, a higher PROCAM diabetes risk score at baseline was associated with lower adherence to the sweet snack UPF pattern. Those with a high risk of type 2 diabetes could have been made aware of their situation through opportunistic screening by general practitioners, public health campaigns, or family history of the disease. Hence, awareness of high type 2 diabetes risk may have driven participants to avoid products that are high in sugar.

From a public health point of view, this can be perceived as a positive message, suggesting that public health initiatives to inform the public on the importance of a healthy diet in the prevention of chronic diseases, such as type 2 diabetes, did come across. In addition, the fact that the inverse association with baseline type 2 diabetes risk observed for the sweet snack pattern may be related to the layman's term for type 2 diabetes, which is "sugar disease" in Dutch and several other languages. Although there is still some scientific uncertainty as to whether all types of sugar intake are

associated with risk of type 2 diabetes,[45-48] limiting the consumption of energy dense, sugar-rich foods will be likely to benefit health, not only by reducing the risk of diabetes, but obesity and cardiovascular diseases as well.[48] Furthermore, it is worth noticing that the adherence to both two savory UPF patterns was higher among individuals with higher diabetes risk scores at baseline, and both patterns were also associated with a higher risk of incident type 2 diabetes. Future research, preferably in the form of randomized controlled trials, is needed to confirm the detrimental effects on glucose homeostasis of both sugary and savory UPF items. A subsequent challenge would then be to create further awareness that it is not only sugary products, but also other kinds of UPF, which may be associated with higher diabetes risk. This could also bear relevance to prevention strategies not only by recommendations for health behaviors, but also by recommendations for product reformulation.[49]

Strengths of this study include the large sample size, which yields a strong statistical power. In addition, our study is the first that thoroughly investigated the habitual consumption patterns of UPF using PCA. The empirical consumption patterns identified reflected not only nutritional properties of UPF, but also its behavioral drivers, which provide a distinct added value over the nutritional information of the NOVA classification and strengthen the real-world robustness of the results of this study.[50] Nevertheless, it should be noted that the four consumption patterns analyzed in total only explained 15.5% of the total variance of UPF intake, which inevitably left a certain proportion of the UPF consumption pattern information uncaptured. It is conceivable that this seemingly low explained variance is attributed to our large study sample size. Secondly, this may also be attributed to the fact that we did not apply massive food groupings in dietary pattern analysis (i.e., combining the intake of several food products into one, such as treating all sorts of cheeses as one single food product),[50-52] which in fact facilitates our study objective for disentangling the consumption patterns of this highly heterogeneous food category. On the other hand, our 15.5% explained variance is comparable with previous studies using PCA and did offer us informative insights into the real-world eating patterns, especially considering our cohort setting.[50-52] We encourage future studies to further explore the UPF consumption patterns in a

different population setting.

Furthermore, there are also several other limitations that should be noted. First, the FFQ used in this study was designed to assess the intake of major food groups, energy, and macronutrients. The aim of assessing energy intake resulted in good coverage of energy dense food, including many kinds of UPF.[32] However, since the FFQ was not designed to assess the intake of UPF, questions asked in the FFQ covered food items with varying levels of processing, inevitably leading to some misclassification. Second, misclassification could also occur in the ascertainment of type 2 diabetes cases, since at T2 and T3 only self-reported data was available. However, as most cases were identified by objective laboratory measurements at T4, this limitation is not expected to influence our results. Third, the exact time of diagnosis of diabetes cases was not collected in the Lifelines study, which unfortunately reduced the suitability of our data for survival analyses. Nevertheless, considering the low event rate and the relatively short follow-up time, logistic regression analysis may provide similar estimates for the effect sizes.[53,54] We therefore used logistic regression analysis instead. Furthermore, the use of self-reported questionnaires such as FFQ might lead to misreporting due to social desirability or recall bias. Finally, we illustrated that some reverse causation could be involved in the results of this study, despite our prospective design.

Conclusions

In conclusion, this study illustrated that the heterogeneity of UPF as a general food category is also reflected by the discrepancy in associations of four distinct UPF consumption patterns and incident type 2 diabetes. The positive associations of the warm savory snack and the cold savory snack UPF consumption patterns with incident type 2 diabetes suggest that savory UPF may be a suitable target for future public health initiatives for type 2 diabetes prevention. More importantly, since UPF consumption was associated with type 2 diabetes risk independent of overall diet quality, eating an otherwise healthy diet may not fully compensate for the detrimental effects of UPF. Therefore, in addition to promoting consumption of healthy food products, active

discouragement of unhealthy food products such as savory UPF should be considered as part of diabetes prevention strategies. In addition, considering the intricate role of BMI in the relationships between UPF and health, it is of equal importance to consider weight management in public health promotion, in addition to the discouragement of UPF consumption. Further research on UPF subgroups and its underlying consumption patterns is encouraged to allow a better understanding of the health effects of this highly heterogeneous food category, which will also facilitate the integration of UPF into dietary assessment tools and recommendations.

Reference

1.   Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. Diabetes Res Clin Pract 2019; 157: 107843.

2.   Ardisson Korat AV, Willett WC, Hu FB. Diet, lifestyle, and genetic risk factors for type 2 diabetes: a review from the Nurses' Health Study, Nurses' Health Study 2, and Health Professionals' Follow-up Study. Curr Nutr Rep 2014; 3(4): 345-54.

3.   Koloverou E, Esposito K, Giugliano D, Panagiotakos D. The effect of Mediterranean diet on the development of type 2 diabetes mellitus: a meta-analysis of 10 prospective studies and 136,846 participants. Metabolism 2014; 63(7): 903-11.

4.   Vinke PC, Navis G, Kromhout D, Corpeleijn E. Socio-economic disparities in the association of diet quality and type 2 diabetes incidence in the Dutch Lifelines cohort. EClinicalMedicine 2020; 19: 100252.

5.   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.

6.   Srour B, Fezeu LK, Kesse-Guyot E, et al. Ultraprocessed Food Consumption and Risk of Type 2 Diabetes Among Participants of the NutriNet-Santé Prospective Cohort. JAMA Intern Med 2020; 180(2): 283-91.

7.   Levy RB, Rauber F, Chang K, et al. Ultra-processed food consumption and type 2 diabetes incidence: A prospective cohort study. Clin Nutr 2021; 40(5): 3608-14.

8.   Nardocci M, Polsky JY, Moubarac JC. Consumption of ultra-processed foods is associated with obesity, diabetes and hypertension in Canadian adults. Can J Public Health 2021; 112(3): 421-9.

9.   Llavero-Valero M, Escalada-San Martin J, Martinez-Gonzalez MA, Basterra-Gortari FJ, de la Fuente-Arrillaga C, Bes-Rastrollo M. Ultra-processed foods and type-2 diabetes risk in the SUN project: A prospective cohort study. Clin Nutr 2021; 40(5): 2817-24.

10.   Monteiro CA. Nutrition and health. The issue is not food, nor nutrients, so much as processing. Public Health Nutr 2009; 12(5): 729-31.

11.   Monteiro CA, Cannon G, Levy RB, et al. Ultra-processed foods: what they are and how to identify them. Public Health Nutr 2019; 22(5): 936-41.

12.   Monteiro CA, Cannon G, Levy R, et al. NOVA. The star shines bright. World Nutrition 2016; 7(1-3): 28-38.

13.   Vandevijvere S, Jaacks LM, Monteiro CA, et al. Global trends in ultraprocessed food and drink product sales and their association with adult body mass index trajectories. Obes Rev 2019; 20 Suppl 2: 10-9.

14.   Cordova R, Kliemann N, Huybrechts I, et al. Consumption of ultra-processed foods associated with weight gain and obesity in adults: A multi-national cohort study. Clin Nutr 2021;

40(9): 5079-88.

15.   Rauber F, Chang K, Vamos EP, et al. Ultra-processed food consumption and risk of obesity: a prospective cohort study of UK Biobank. Eur J Nutr 2021; 60(4): 2169-80.

16.   Li M, Shi Z. Ultra-Processed Food Consumption Associated with Overweight/Obesity among Chinese Adults-Results from China Health and Nutrition Survey 1997-2011. Nutrients 2021; 13(8).

17.   Srour B, Fezeu LK, Kesse-Guyot E, et al. Ultra-processed food intake and risk of cardiovascular disease: prospective cohort study (NutriNet-Sante). BMJ 2019; 365: l1451.

18.   Du S, Kim H, Rebholz CM. Higher Ultra-Processed Food Consumption Is Associated with Increased Risk of Incident Coronary Artery Disease in the Atherosclerosis Risk in Communities Study. J Nutr 2021; 151(12): 3746-54.

19.   Bonaccio M, Di Castelnuovo A, Costanzo S, et al. Ultra-processed food consumption is associated with increased risk of all-cause and cardiovascular mortality in the Moli-sani Study. Am J Clin Nutr 2021; 113(2): 446-55.

20.   Fiolet T, Srour B, Sellem L, et al. Consumption of ultra-processed foods and cancer risk: results from NutriNet-Sante prospective cohort. BMJ 2018; 360: k322.

21.   Rico-Campa A, Martinez-Gonzalez MA, Alvarez-Alvarez I, et al. Association between consumption of ultra-processed foods and all cause mortality: SUN prospective cohort study. BMJ 2019; 365: l1949.

22.   Romero Ferreiro C, Martin-Arriscado Arroba C, Cancelas Navia P, Lora Pablos D, Gomez de la Camara A. Ultra-processed food intake and all-cause mortality: DRECE cohort study. Public Health Nutr 2021: 1-10.

23.   Zhong GC, Gu HT, Peng Y, et al. Association of ultra-processed food consumption with cardiovascular mortality in the US population: long-term results from a large prospective multicenter study. Int J Behav Nutr Phys Act 2021; 18(1): 21.

24.   Martinez Steele E, Juul F, Neri D, Rauber F, Monteiro CA. Dietary share of ultra-processed foods and metabolic syndrome in the US adult population. Prev Med 2019; 125: 40-8.

25.   Fardet A, Rock E. Ultra-processed foods: A new holistic paradigm? Trends in Food Science & Technology 2019; 93: 174-84.

26.   Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol 2015; 44(4): 1172-80.

27.   Stolk RP, Rosmalen JG, Postma DS, et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. Eur J Epidemiol 2008; 23(1): 67-74.

28.   WHO, International Diabetes Federation. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: report of a WHO/IDF consultation. Genève, Switzerland: WHO, 2006.

29.   WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health. ATC/DDD Index 2020. Oslo, Norway; 2020.

30.   Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. Hum Nutr Clin Nutr 1985; 39 Suppl 1: 5-41.

31. Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. Int J Obes Relat Metab Disord 2000; 24(9): 1119-30.

32. Siebelink E, Geelen A, de Vries JH. Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. Br J Nutr 2011; 106(2): 274-81.

33. Vinke PC, Corpeleijn E, Dekker LH, Jacobs DR, Jr., Navis G, Kromhout D. Development of the food-based Lifelines Diet Score (LLDS) and its application in 129,369 Lifelines participants. Eur J Clin Nutr 2018; 72(8): 1111-9.

34. UNESCO. International Standard Classification of Education. 1997. https://www.unesco.org/en/education.

35. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003; 56(12): 1163-9.

36. von Eckardstein A, Schulte H, Assmann G. Risk for diabetes mellitus in middle-aged Caucasian male participants of the PROCAM study: implications for the definition of impaired fasting glucose by the American Diabetes Association. Prospective Cardiovascular Munster. J Clin Endocrinol Metab 2000; 85(9): 3101-8.

37. Gibney MJ. Ultra-Processed Foods: Definitions and Policy Issues. Curr Dev Nutr 2019; 3(2): nzy077.

38. Elizabeth L, Machado P, Zinocker M, Baker P, Lawrence M. Ultra-Processed Foods and Health Outcomes: A Narrative Review. Nutrients 2020; 12(7).

39. 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.

40. Juul F, Vaidean G, Parekh N. Ultra-processed Foods and Cardiovascular Diseases: Potential Mechanisms of Action. Adv Nutr 2021; 12(5): 1673-80.

41. Almoraie NM, Saqaan R, Alharthi R, Alamoudi A, Badh L, Shatwan IM. Snacking patterns throughout the life span: potential implications on health. Nutr Res 2021; 91: 81-94.

42. Hu FB, Malik VS. Sugar-sweetened beverages and risk of obesity and type 2 diabetes: epidemiologic evidence. Physiol Behav 2010; 100(1): 47-54.

43. Te Morenga L, Mallard S, Mann J. Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. BMJ 2012; 346: e7492.

44. Buijsse B, Boeing H, Drogan D, et al. Consumption of fatty foods and incident type 2 diabetes in populations from eight European countries. Eur J Clin Nutr 2015; 69(4): 455-61.

45. Macdonald IA. A review of recent evidence relating to sugars, insulin resistance and diabetes. Eur J Nutr 2016; 55(Suppl 2): 17-23.

46. Wang J, Light K, Henderson M, et al. Consumption of added sugars from liquid but not solid sources predicts impaired glucose homeostasis and insulin resistance among youth at risk of obesity. J Nutr 2014; 144(1): 81-6.

47.   Janket SJ, Manson JE, Sesso H, Buring JE, Liu S. A prospective study of sugar intake and risk of type 2 diabetes in women. Diabetes Care 2003; 26(4): 1008-15.

48.   Malik VS, Popkin BM, Bray GA, Despres JP, Hu FB. Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk. Circulation 2010; 121(11): 1356-64.

49.   Popkin BM, Barquera S, Corvalan C, et al. Towards unified and impactful policies to reduce ultra-processed food consumption and promote healthier eating. Lancet Diabetes Endocrinol 2021; 9(7): 462-70.

50.   Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 2002; 13(1): 3-9.

51.   Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. Nutr Rev 2004; 62(5): 177-203.

52.   Dekker LH, Rijnks RH, Strijker D, Navis GJ. A spatial analysis of dietary patterns in a large representative population in the north of The Netherlands- the Lifelines cohort study. Int J Behav Nutr Phys Act 2017; 14(1): 166.

53.   Green MS, Symons MJ. A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. J Chronic Dis 1983; 36(10): 715-23.

54.   Ngwa JS, Cabral HJ, Cheng DM, et al. A comparison of time dependent Cox regression, pooled logistic regression and cross sectional pooling with simulations and an application to the Framingham Heart Study. BMC Med Res Methodol 2016; 16(1): 148.

Additional Information

Contributions

LHD, MJD, PCV, and EC designed and conceptualized the study. MJD analyzed the data. PCV and MD drafted the manuscript. EC, GN, and LHD contributed to the discussion and critically reviewed/edited the manuscript. MJD and PCV have primary responsibilities for the final content. All authors approved the final content of the manuscript.

Supplementary Data

Online **Additional Files** (supplementary data) can be accessed via
doi.org/10.1186/s12916-021-02200-4, or by scanning the QR code.
Article PMID: 35022060.

CHAPTER 4

1. Department of Internal Medicine, University Medical Center Groningen, Groningen, The Netherlands

2. National Institute for Public Health and the Environment, Bilthoven, The Netherlands

3. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Lifestyle patterns and incident type 2 diabetes in the Dutch Lifelines cohort study

Ming-Jie Duan[1]; Louise H. Dekker[1,2];

Juan-Jesus Carrero[3]; Gerjan Navis[1]

ABSTRACT

Objectives

We aimed to identify the underlying subgroups of the population characterized by distinct lifestyle patterns, and to investigate the associations between lifestyle patterns and risk of incident type 2 diabetes.

Methods

Using data from the Dutch Lifelines cohort study, latent class analysis was performed to derive lifestyle patterns on five lifestyle factors, i.e., smoking, diet quality, TV watching time, physical activity level, and risk drinking. Associations between lifestyle patterns and incident type 2 diabetes were estimated.

Results

Among 61,869 participants analyzed, we identified 900 cases of type 2 diabetes during follow-up (205,696 person-years; incidence rate 4.38 per 1000 person-years). Five lifestyle pattern groups were identified. Using the "healthy lifestyle group" as reference, the "unhealthy lifestyle group" had the highest risk for type 2 diabetes (HR 1.51 [95%CI 1.24, 1.85]), followed by the "poor diet and low physical activity group" (HR 1.26 [95%CI 1.03, 1.55]). The "risk drinker group" and the "couch potato group" (characterized by excessive TV watching) showed no significantly elevated risk. These models were adjusted for age, sex, total energy intake, education, BMI, family history of diabetes, and blood glucose level at baseline.

Conclusions

Our study shows that lifestyle factors tended to cluster in unique behavioral patterns within the heterogeneous population. These lifestyle patterns were differentially associated with incident type 2 diabetes. Our findings support the relevance of considering lifestyle patterns in type 2 diabetes prevention. Tailored prevention strategies that target multiple lifestyle risk factors for different lifestyle pattern groups may optimize the effectiveness of diabetes prevention at the population level.

Abbreviations

BIC-LL – Bayesian information criterion with log likelihood for the number of parameters adjusted

FFQ – Food frequency questionnaire

LCA – Latent class analysis

LLDS – Lifelines diet score

MVPA – Moderate-to-vigorous physical activity

PAF – Population attributable fraction

SQUASH – Short QUestionnaire to ASsess Health-enhancing physical activity

4

Introduction

Type 2 diabetes is a major public health challenge that leads to considerable morbidity, mortality, and economic burden.[1] Lifestyle is crucial to the prevention of type 2 diabetes. Adherence to a combination of healthy lifestyle factors – healthy diet, avoiding smoking, vigorous physical activity – is found to substantially lower the risk of developing type 2 diabetes.[2-4]

For studying the relationships between lifestyle factors and type 2 diabetes, a single lifestyle factor approach has been widely applied. Studies have also examined the combined effects of lifestyle factors, such as using an unweighted lifestyle score, but they do not take account of the distribution of lifestyle factors in the population.[3] Prior studies have implicated that lifestyle factors often co-occur in behavioral patterns and may have interdependent effects on health.[5-15] Better methodological approaches are therefore needed to understand the complexities of lifestyle factors and their associations with health.

For type 2 diabetes prevention, current evidence supports the relevance of targeting multiple lifestyle risk factors simultaneously.[14-16] It is therefore essential to have a clear understanding of the clustering of lifestyle risk factors of the target populations. However, to date the knowledge basis is lacking. Specifically, only three studies have identified lifestyle patterns in the Dutch population, and only one of them further studied their associations with risk of type 2 diabetes.[6,9,17] There is considerably less knowledge about the relevance of lifestyle patterns for type 2 diabetes prevention in the general population.

Previous studies on lifestyle patterns mainly included smoking, alcohol consumption, physical activity level, and fruit and vegetable intake.[5-11,17] However, those identified lifestyle patterns may not fully represent the overall lifestyle risk profiles. While fruit and vegetable intake is an important indicator of diet,[18] overall diet quality, commonly assessed by diet scores, may better represent the overall dietary "risk profile" of the target populations.[19] Moreover, high TV watching time, as an emerging lifestyle risk factor representing sedentary behavior, has been found to be a risk factor for type

2 diabetes and mortality, independent of moderate-to-vigorous physical activity (MVPA),[20] while it has never been included in lifestyle pattern analysis. Therefore, incorporating overall diet quality and TV watching time in lifestyle pattern analysis will provide more information on the clinical relevance of lifestyle patterns.

Using a large Dutch population cohort, we aimed to reveal how lifestyle factors cluster within populations, i.e., the diverse lifestyle risk patterns of the population, and subsequently, to investigate the prospective associations between lifestyle patterns and incident type 2 diabetes. The analysis focused on four traditional and one emerging lifestyle factors, including overall diet quality,[2,19,21] physical activity,[22] smoking,[23] risk drinking,[24] and TV watching time.[20] These lifestyle factors included are common in the general population. Having a clear understanding of how these common lifestyle factors cluster and how different lifestyle clusters affect type 2 diabetes risk will facilitate the design of effective prevention strategies at population level.

Methods

Study Design and Population

The Lifelines cohort study is a multidisciplinary prospective population-based cohort study that applies a unique three-generation design to study the health and health-related behaviors of 167,729 persons living in the north of The Netherlands. Before study entry, a signed informed consent form was obtained from each participant. The Lifelines study is conducted according to the principles of the Declaration of Helsinki and approved by the Medical Ethics Committee of the University Medical Center Groningen, The Netherlands. The overall design and rationale of the study have been described in detail elsewhere.[25,26]

Participants were included in the study between 2006 and 2013. So far, four assessment rounds took place, including baseline assessment (T1) and three follow-ups (T2-T4). Comprehensive physical examinations, biobanking, and questionnaires were conducted at T1 and T4. Follow-up questionnaires were issued to participants at

T2, T3, and T4.

Participants aged between 35 and 65 years who were free of diabetes at baseline, and for whom lifestyle data was available were included in this study. Participants who had no follow-up data, or who reported the development of type 1 diabetes or gestational diabetes during follow-up were excluded. In total, 61,869 participants were included in the analysis (**Supplementary Fig. S1**).

Ascertainment of Incident Type 2 Diabetes

Incident type 2 diabetes was assessed by self-report questionnaires during follow-up at T2, T3, and T4, as well as blood glucose and HbA_{1c} measurements at T4. Blood measurements are not available at T2 and T3. Participants were considered an incident case if they met one of the following criteria: (1) self-reported newly developed type 2 diabetes since last time they filled out a questionnaire; (2) fasting blood glucose ≥ 7.0 mmol/L; or (3) HbA_{1c} ≥ 48 mmol/mol (6.5 %).[27]

Clinical Measurements

Blood samples were collected by venipuncture in a fasting state, and were further transferred to the Lifelines central laboratory for analysis. Serum levels of glucose and HbA_{1c} were subsequently analyzed. Anthropometry was measured by trained research staff following standardized protocols. These measurements were performed without shoes and heavy clothing. Family history of diabetes was assessed by self-administered questionnaires. Participants were considered having a family history of diabetes if they reported having a first-degree relative (i.e., parent, sibling, or child) ever being diagnosed with type 2 diabetes.

Assessment of Lifestyle Factors and Sociodemographic Covariates

Age, smoking status, TV watching time per day, and education were assessed by self-administered questionnaires. Highest education achieved was categorized as: (1) low – junior general secondary education or lower; (2) middle – secondary vocational education and senior general secondary education; and (3) high – higher vocational education or university.[28]

Habitual physical activity level of a normal week was assessed by the Short QUestionnaire to ASsess Health-enhancing physical activity (SQUASH). The SQUASH was pre-structured into four domains: commuting, leisure time, household, and occupational activities. For each reported activity, frequency (days per week) and duration (average time per day) were asked. From the SQUASH data, non-occupational moderate-to-vigorous physical activity (MVPA), including commuting and sports (if ≥ 4.0 MET), was calculated in minutes per week. The SQUASH has been validated in the general population using objective accelerometer measurements for a 2-week period.[29]

Dietary intake was assessed by a semi-quantitative self-administered food frequency questionnaire (FFQ). The FFQ aimed to assess the habitual intake of 110 food items (including alcohol) during the past 4 weeks. For 46 main food items (such as bread and milk), frequency of consumption was indicated as 'not in this month' or in days per week or month, including the amount (in units or specified portion sizes) consumed each time. The FFQ also included 37 questions on intake of sub-items (such as different types of cheese) for which frequency was specified as never, sometimes, often, and always. The FFQ was designed based on the validated Dutch FFQ.[30] In brief, the intake of the food items and the energy intake have been tested and validated against three 24-h dietary recalls and actual energy intake in controlled feeding trials, respectively.[31,32] The Lifelines Diet Score (LLDS) was calculated to evaluate the relative diet quality of each participant.[19]

Statistical Analysis

Lifestyle Pattern Analysis with Latent Class Analysis

Lifestyle patterns were derived using latent class analysis (LCA). LCA is a latent variable mixture model that relates a set of observed indicators (i.e., lifestyle variables) to a set of latent variables (i.e., lifestyle pattern classes).[33] LCA enables the analysis and interpretation of higher-order interactions among lifestyle factors, which overcomes the issue of collinearity between lifestyle factors.[34,35]

The LCA output mainly consists of two parts. The first part is the posterior class probability, which estimates the probability of an individual belonging to each latent

class given the individual's observed response on the measured indicators. Each participant was assigned to the lifestyle pattern group for which they had the highest posterior class probability. A number of mutually exclusive lifestyle pattern groups would thus be identified. The second part is the class-specific response probability, which estimates the likelihood that an individual, who belongs to a particular latent class, adheres to a certain measured indicator, such as the probability of being a never smoker.[33]

Since LCA requires that items are measured categorically, we further defined lifestyle factors into risky versus non-risky categories based on evidence, resulting in nine indicators. The interpretation of the results also becomes clearer when lifestyle factors are categorized into risky versus non-risky groups. Specifically, smoking status, i.e., never, former, and current smoker, was treated as three dummy variables. Alcohol intake was categorized as risk drinking (>15 g alcohol/day) versus non-risk drinking (≤15 g alcohol/day).[36] This amount was approximated to one drink per day. TV watching time was categorized as excessive TV watching (highest sex-specific tertile) versus non-excessive TV watching (other tertiles). LLDS was divided into sex-specific tertiles. Physical activity level was categorized as whether the participants met the Dutch recommendation for physical activity level, i.e., ≥150 min non-occupational MVPA per week.[37]

A series of latent class models were examined with three through nine classes. We selected the best-fitting latent class solution based on Bayesian information criterion with log likelihood for the number of parameters adjusted (BIC-LL). BIC-LL is a model goodness-of-fit index, for which a lower value is preferred.[38] We also considered other model goodness-of-fit indices (**Supplementary Table S1**), as well as the interpretability of the identified lifestyle patterns.[33] LCA was performed with LatentGOLD (version 5.0.0.14260; Statistical Innovations Inc., Belmont, MA, USA).[39]

Risk of Incident Type 2 Diabetes

Associations between lifestyle patterns and incident type 2 diabetes were estimated using Cox proportional hazards regression models. Non-diabetes cases were censored at the last time-point, for which data was available. Additionally, all participants were

censored after 60 months. Analyses were adjusted in a stepwise manner for (1) age, sex, and total energy intake; (2) education; (3) BMI; (4) family history of diabetes; and (5) blood glucose level at baseline. Proportional hazards assumption was assessed by calculating the Schoenfeld residuals and by performing Cox regression models with time-dependent covariates. Potential effect modification was evaluated for age, sex, BMI, education, and family history of diabetes. Analyses were repeated excluding participants who had less than 12-month follow-up, in an attempt to address possible reverse causation caused by short follow-up time. For comparisons, we additionally tested the associations of incident type 2 diabetes with each lifestyle risk factor separately. Statistical analyses for calculating the risk of type 2 diabetes were performed on Stata (version 13.1; StataCorp, College Station, TX, USA).

To obtain insights into the lifestyle-related diabetes disease burden, namely the fraction of cases preventable if having a healthy lifestyle profile, we calculated the adjusted population attributable fraction (PAF) based on the odds ratios estimated using logistic regression models adjusting for the abovementioned Cox proportional hazards model covariates. The calculation of PAFs was performed using punaf package in Stata, as described by Newson.[40]

Results

Lifestyle Patterns

After examining models with three through nine latent classes, we selected a 5-latent class model (five lifestyle patterns) since it offered the lowest BIC-LL value (best model fit) and the best subjective interpretability. Most of the other model goodness-of-fit indices also showed their best values at the 5-latent class model solution. **Supplementary Table S1** shows the detailed model goodness-of-fit indices for all models tested.

Fig. 1 and **Supplementary Table S2** show the estimated probabilities of adhering to lifestyle factors for lifestyle patterns identified. The first pattern was named the

"healthy lifestyle group" (n = 27,413, 44.3 %), as it was characterized by moderate to low probabilities across all lifestyle risk factors. The second pattern was designated as the "poor diet and low physical activity group" (n = 13,846, 22.4 %), because it was characterized primarily by moderate to high probabilities of poor diet quality (lowest tertile of LLDS) and insufficient physical activity. The third pattern was labelled the "unhealthy lifestyle group" (n = 12,031, 19.5 %), since it was characterized by moderate to low probabilities of risk drinking and former smoker, but moderate to high probabilities across all other lifestyle risk factors. The fourth pattern was named the "couch potato group" (n = 4726, 7.6 %). Persons in this pattern had moderate to high probabilities of excessive TV watching and also notably former smoker, but they had moderate to low probabilities elsewhere. The fifth pattern was labelled the "risk drinker group" (n = 3853, 6.2 %), as persons in this pattern mainly had very high probability of risk drinking and moderate to high probability of former smoker.

Baseline Characteristics

Baseline characteristics for each lifestyle pattern group are shown in **Table 1**. Participants from the "poor diet and low physical activity group" and the "unhealthy lifestyle group" tended to be younger, while participants from the latter group and the "couch potato group" tended to be less educated. In total, there were 59.6 % female participants included in the analysis, whereas there were more male participants (61.1 %) in the "risk drinker group". Clinical biomarkers showed diverse distributions among different groups. The "couch potato group" had the highest prevalence of family history of diabetes (10.2 %).
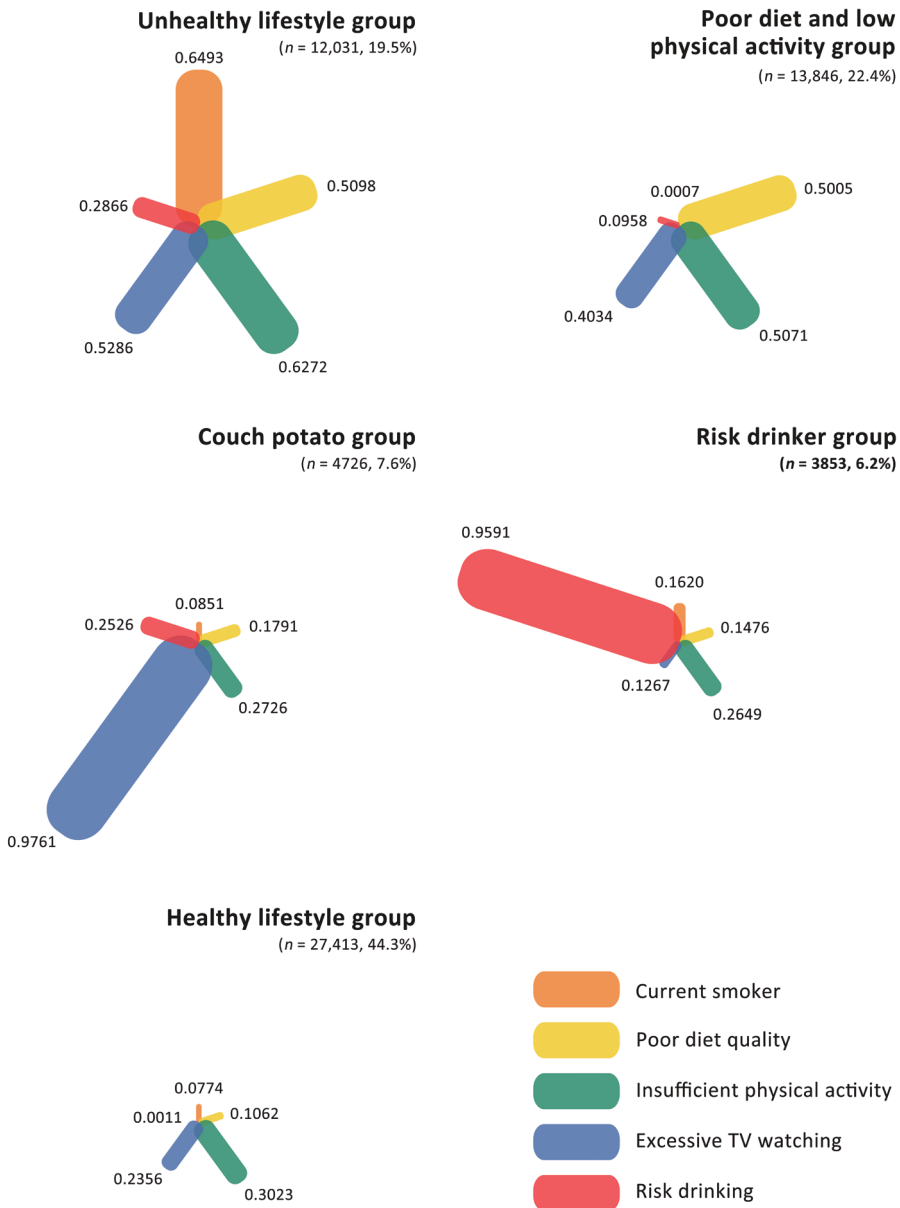
Unhealthy lifestyle group
(*n* = 12,031, 19.5%)

0.6493
0.5098
0.2866
0.5286
0.6272

Poor diet and low physical activity group
(*n* = 13,846, 22.4%)

0.0007
0.5005
0.0958
0.4034
0.5071

Couch potato group
(*n* = 4726, 7.6%)

0.0851
0.2526
0.1791
0.2726
0.9761

Risk drinker group
(*n* = 3853, 6.2%)

0.9591
0.1620
0.1476
0.1267
0.2649

Healthy lifestyle group
(*n* = 27,413, 44.3%)

0.0774
0.0011
0.1062
0.2356
0.3023

- Current smoker
- Poor diet quality
- Insufficient physical activity
- Excessive TV watching
- Risk drinking

Fig. 1. Estimated probabilities of adhering to examined lifestyle risk factors for each identified lifestyle pattern[a]

[a] The adapted spider charts show the estimated probabilities of adhering to the examined lifestyle risk factors according to each lifestyle pattern, in which the width and the length of each bar was proportionately illustrated according to the values of the estimated probabilities that are displayed next to each bar.

Table 1. Baseline characteristics according to lifestyle pattern groups[a]

	Healthy lifestyle group	Poor diet and low physical activity group	Unhealthy lifestyle group	Couch potato group	Risk drinker group	Total
Number of participants	27,413	13,846	12,031	4726	3853	61,869
Class size, %	44.3	22.4	19.5	7.6	6.2	100
Socio-demographic characteristics						
Age, years	48.8±8.0	45.6±7.2	47.5±7.5	51.8±8.1	50.5±7.8	48.2±7.9
Sex – women, %	63.0	61.8	57.1	56.9	38.9	59.6
Education, %						
Low	23.9	26.3	40.9	42.3	19.3	28.8
Middle	38.2	43.8	41.0	37.0	33.6	39.6
High	37.8	29.6	17.9	20.2	46.8	31.3
Clinical biomarkers						
Fasting glucose, mmol/L	4.91±0.49	4.93±0.49	5.00±0.52	5.07±0.52	5.05±0.51	4.95±0.50
HbA_{1c}, %	5.53±0.29	5.51±0.30	5.58±0.29	5.56±0.30	5.52±0.29	5.54±0.30
HbA_{1c}, mmol/mol	36.96±3.22	36.71±3.23	37.50±3.21	37.30±3.30	36.84±3.22	37.03±3.24
BMI, kg/m^2	25.8±3.9	26.4±4.4	26.5±4.3	26.9±3.9	25.8±3.3	26.1±4.1
Family history of diabetes, %	8.6	8.9	9.3	10.2	7.0	8.8

Table 1. Continued.

Lifestyle factors	Healthy lifestyle group	Poor diet and low physical activity group	Unhealthy lifestyle group	Couch potato group	Risk drinker group	Total
Lifestyle factors						
Total energy intake, kcal/day	1944±540	2158±598	2150±632	2109±589	2218±607	2062±590
Lifeline diet score	27.8±4.8	19.9±3.9	19.9±4.6	22.9±4.9	25.9±4.9	24.0±5.9
Lowest tertile, %	4.5	56.2	57.3	29.8	9.5	28.6
Middle tertile, %	31.4	43.8	36.8	52.1	46.2	37.7
Highest tertile, %	64.2	0	5.9	18.1	44.3	33.7
Risk drinking, %	0	8.5	27.6	39.1	100	16.5
Alcohol intake, g/day	3.3 (0.9, 6.9)	2.4 (0, 6.7)	6.6 (1.5, 16.4)	9.5 (2.6, 17.9)	18.2 (17.3, 25.4)	4.5 (0.9, 11.0)
Meeting physical recommendation (150 min/week MVPA)	73.9	44.4	22.7	94.4	79.1	59.2
MVPA, min/week	250 (135, 420)	120 (50, 280)	60 (0, 130)	305 (210, 495)	270 (160, 450)	180 (60, 360)
Excessive TV watching (highest tertile), %	20.9	46.3	61.2	100	0	39.1
TV watching time, hours/day	2.0±1.1	2.6±1.3	3.0±1.4	3.7±0.9	1.6±0.7	2.4±1.3
Smoking, %						
Never	46.2	100	0	9.6	21.9	44.9
Former	45.4	0	32.8	90.4	64.1	37.4
Current	8.4	0	67.2	0	14.0	17.7

[a] Data are expressed as unadjusted mean ± standard deviation for age, fasting glucose, HbA_{1c}, body mass index (BMI), total energy intake, Lifelines diet score (0-48 no unit), and TV watching time; data are expressed as median (interquartile) for alcohol intake and non-occupational moderate-to-vigorous physical activity level (MVPA); data are expressed as actual observed values for other variables.

4

Table 2. Associations between lifestyle pattern groups and incident type 2 diabetes[a]

	Healthy lifestyle group	Poor diet and low physical activity group	Unhealthy lifestyle group	Couch potato group	Risk drinker group
Cases/Population	321 / 27,413	187 / 13,846	255 / 12,031	81 / 4726	56 / 3853
Incidence, %	1.17	1.35	2.12	1.71	1.45
Incidence rates, per 1000 person-years	3.51	4.09	6.42	5.17	4.25
Model 1	1.00 (ref)	1.46 (1.22, 1.75)	1.99 (1.68, 2.34)	1.28 (1.00, 1.64)	0.99 (0.74, 1.32)
Model 2		1.40 (1.17, 1.69)	1.80 (1.51, 2.13)	1.18 (0.92, 1.51)	1.05 (0.79, 1.39)
Model 3		1.22 (1.01, 1.47)	1.64 (1.38, 1.94)	1.06 (0.83, 1.36)	1.09 (0.82, 1.45)
Model 4		1.21 (1.00, 1.46)	1.63 (1.37, 1.93)	1.05 (0.82, 1.35)	1.11 (0.83, 1.47)
Model 5		1.26 (1.03, 1.55)	1.51 (1.24, 1.85)	0.98 (0.76, 1.25)	1.03 (0.77, 1.39)

[a] All models: HRs (95 % CI) derived from multivariate Cox proportional hazards models. Model 1 was adjusted for age, sex, and total energy intake, n = 61,869; model 2 was adjusted for model 1 covariates plus education, n = 61,714; model 3 was adjusted for model 2 covariates plus BMI, n = 61,714; model 4 was adjusted for model 3 covariates plus family history of diabetes, n = 61,714; model 5 was adjusted for model 4 covariates plus blood glucose level at baseline, n = 61,512.

Risk of Incident Type 2 Diabetes

Table 2 shows the associations between different lifestyle pattern groups and risks of incident type 2 diabetes. Among 61,869 participants included in the analysis, we identified 900 cases of type 2 diabetes during follow-up (205,696 person-years; median [interquartile] follow-up time, 41 [29–50] months; incidence rate 4.38 per 1000 person-years). The incidence rates of type 2 diabetes ranged from 3.51 per 1000 person-years for the "healthy lifestyle group" to 6.42 per 1000 person-years for the "unhealthy lifestyle group". In the fully adjusted model (model 5) using the "healthy lifestyle group" as the low risk reference group, the "risk drinker group" (HR 1.03 [95 %CI 0.77, 1.39]) and the "couch potato group" (HR 0.98 [95 %CI 0.76, 1.25]) were not associated with incident type 2 diabetes, whereas the "poor diet and low physical activity group" (HR 1.26 [95 %CI 1.03, 1.55]) and the "unhealthy lifestyle group" (HR 1.51 [95 %CI 1.24, 1.85]) had significantly higher risks of incident type 2 diabetes. **Supplementary Table S3** shows the associations using the "unhealthy lifestyle group" as reference. Statistically, the associations between lifestyle pattern groups and risks of incident type 2 diabetes were not significantly modified by age, sex, BMI, education, and family history of diabetes (all P-interaction > 0.05). Results were basically unchanged when excluding participants who had less than 12-month follow-up (**Supplementary Table S4**). **Supplementary Table S5** presents the PAFs for each lifestyle pattern group using the "healthy lifestyle group" as reference. **Supplementary Table S6** shows the associations between single lifestyle factors and incident type 2 diabetes.

Discussion

There are two main findings of our study. First, using a large population-based sample, we identified five lifestyle patterns. Second, we found that different combinations of lifestyle risk factors, as manifested in lifestyle patterns, were differentially associated with risk of developing type 2 diabetes.

Lifestyle Patterns and Risk of Incident Type 2 Diabetes

There is robust evidence showing that avoiding risky lifestyle behaviors is effective in the prevention of type 2 diabetes.[3,4] For example, an Iranian study found that a higher healthy lifestyle score, characterized by no smoking, normal body weight, vigorous physical activity, and healthy diet, was associated with up to 75% lower risk of type 2 diabetes, independent of multiple confounders.[4] The current analysis extends previous knowledge by considering multiple co-occurring lifestyle risk factors simultaneously in the form of real-life lifestyle patterns in the general population. We are aware of only two other studies that have applied a lifestyle pattern approach when predicting the risk of type 2 diabetes. One study from the US Women's Health Initiative cohort found that the "poor diet and low exercise pattern" and the "high multiple lifestyle and psychosocial risks pattern" were associated with higher risks of incident type 2 diabetes.[5] Likewise, the Dutch HELIUS cohort study of a multi-ethnic population reported unhealthy lifestyle patterns were associated with higher risks of developing type 2 diabetes.[6] Despite the differences in risk factors and patterns considered that preclude direct comparisons between previous evidence and our results, taken together, these findings support an important role of lifestyle patterns in the development of type 2 diabetes.

The classic approach of studying single lifestyle factors usually assumes independent effects between each lifestyle factor, but does not account for their interrelations.[3-7,10-15] Although further investigation is warranted, we did observe that the risks related to different lifestyle patterns were neither additive nor proportionate to the number of risk factors present, especially compared with the effect sizes when studying each lifestyle factor separately (**Supplementary Table S6**). Notably, the "couch potato group" was not associated with risk of type 2 diabetes, especially after adjustment for BMI. This counterintuitive finding suggests that BMI may play an important role in the studied associations for participants from this lifestyle pattern group. As such, the average effects estimated for a single lifestyle risk factor may not be accurate for a substantial proportion of the study population. Alternatively, a lifestyle pattern may therefore be a proxy for an underlying behavioral variable that is not measured, but nevertheless relevant.

Methodological Considerations

Our study was conducted in a single cohort, albeit large. Accordingly, the generalizability and reproducibility of the current lifestyle pattern analysis require further substantiation from independent cohorts. Various lifestyle patterns have been identified but in limited number of studies. At least partly, this is due to the heterogeneity of the source data, namely, numbers and categorization of lifestyle factors in different studies. Nevertheless, true differences in lifestyle patterns may exist between different populations. Analysis of differences and similarities in lifestyle patterns between populations would be highly relevant for identifying generic as well as specific patterns. So far, patterns primarily characterized by minimal risk behaviors, maximal risk behaviors, and poor diet combined with low physical activity were commonly identified. Patterns characterized by risk drinking generally showed large variations in its coexisting lifestyle risk factors across studies, which may be partly attributed to the lack of an evidence-based definition for that (Davis et al., 2019, Hendryx et al., 2020, Luo et al., 2021, Noble et al., 2015, van Etten et al., 2020, Watts et al., 2016).[5-7,10-15] Using a normalized lifestyle evaluation scheme may therefore benefit the reproducibility and generalizability of the identified patterns to other populations.

Implications for Public Health Prevention

In our analysis, participants from the "healthy lifestyle group" formed the largest group (44.3 %), although conspicuously their lifestyles were still not entirely optimal. Nevertheless, our analysis on lifestyle-related disease burden did show that substantial public health benefits could be obtained. For instance approximately one third of the diabetes cases in the "unhealthy lifestyle group" could be preventable, if participants in this group had the same lifestyle pattern as the "healthy lifestyle group" (**Supplementary Table S5**).

Current evidence supports the relevance of targeting multiple lifestyle risk factors simultaneously.[14,15] Although certain efforts in diabetes prevention have been made on improving diet quality and physical activity, other lifestyle risk factors and within-population heterogeneity in the distribution of lifestyle factors have often been

overlooked.[41] As observed in our population, lifestyle factors may coexist with each other in a counterintuitive manner. The "couch potato group", characterized by excessive TV watching, also had the highest level of non-occupational MVPA. The differential risks found for each lifestyle pattern group also further emphasize the importance and relevance of considering different lifestyle patterns when designing lifestyle programs, rather than adopting the generic one-size-fits-all approach.

Strengths and Limitations

Strengths of our study include a large sample size and the availability of data on TV watching time as an emerging lifestyle factor. Sensitivity analyses ensured the robustness of our findings. We exclusively studied lifestyle risk factors without conflation of lifestyle with its health outcomes (e.g., obesity status). However, a number of limitations are worth mentioning. First, over-reporting of healthy lifestyle behaviors due to social-desirability is possible.[42] Nevertheless, in our study this over-reporting might mainly compromise the discrimination power of the identification of lifestyle clusters. Second, possible changes in lifestyle behaviors might be relevant but were not assessed. Third, as the Lifelines cohort mainly consists of participants in the northern Netherlands, it might not be possible to extrapolate our results to other population groups. Furthermore, in LCA analysis, the assignment of lifestyle pattern group for individuals was based on their highest posterior probability class membership, which unfortunately cannot account for the uncertainty of the classification. Finally, we could not analyze the potential impacts of lost to follow-up (23.0 %) among eligible participants. Nonetheless, the baseline characteristics of those who had no follow-up data were comparable with the study population, except for some minor differences (**Supplementary Table S7**). Simulation studies suggested that such attrition bias may only have limited influences on estimates of associations in cohort studies (Howe et al., 2013, Peters et al., 2012).[43,44]

Conclusions

In conclusion, focusing on five lifestyle factors, namely smoking, overall diet quality, TV watching time, physical activity, and risk drinking, we identified five groups of individuals with different lifestyle patterns using a data-driven approach in a large population-based sample. These five lifestyle patterns were differentially associated with risk of developing type 2 diabetes. The clustering of lifestyle risk factors extends previous knowledge that those lifestyle factors tend to cluster, particularly in behavioral patterns within a general and heterogeneous population. Our findings pave the way for a more effective strategy for public health prevention for type 2 diabetes through targeting multiple lifestyle risk factors simultaneously.

4

Reference

1.   Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022; 183: 109119.

2.   Duan MJ, Dekker LH, Carrero JJ, Navis G. Blood lipids-related dietary patterns derived from reduced rank regression are associated with incident type 2 diabetes. Clin Nutr 2021; 40(7): 4712-9.

3.   Zhang Y, Pan X-F, Chen J, et al. Combined lifestyle factors and risk of incident type 2 diabetes and prognosis among individuals with type 2 diabetes: a systematic review and meta-analysis of prospective cohort studies. Diabetologia 2020; 63(1): 21-33.

4.   Farhadnejad H, Teymoori F, Asghari G, Mokhtari E, Mirmiran P, Azizi F. The higher adherence to a healthy lifestyle score is associated with a decreased risk of type 2 diabetes in Iranian adults. BMC Endocr Disord 2022; 22(1): 42.

5.   Hendryx M, Dinh P, Chow A, et al. Lifestyle and psychosocial patterns and diabetes incidence among women with and without obesity: a prospective latent class analysis. Prev Sci 2020; 21(6): 850-60.

6.   van Etten S, Crielaard L, Muilwijk M, et al. Lifestyle clusters related to type 2 diabetes and diabetes risk in a multi-ethnic population: The HELIUS study. Prev Med 2020; 137: 106141.

7.   Watts P, Buck D, Netuveli G, Renton A. Clustering of lifestyle risk behaviours among residents of forty deprived neighbourhoods in London: lessons for targeting public health interventions. J Public Health (Oxf) 2016; 38(2): 308-15.

8.   Poortinga W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. Prev Med 2007; 44(2): 124-8.

9.   Hofstetter H, Dusseldorp E, van Empelen P, Paulussen TW. A primer on the use of cluster analysis or factor analysis to assess co-occurrence of risk behaviors. Prev Med 2014; 67: 141-6.

10.   Luo J, Dinh P, Hendryx M, Li W, Robinson J, Margolis KL. Risk Patterns and Mortality in Postmenopausal Women Using Latent Class Analysis. Am J Prev Med 2021; 61(5): e225-e33.

11.   Davis JS, Banfield E, Lee HY, Peng HL, Chang S, Wood AC. Lifestyle behavior patterns and mortality among adults in the NHANES 1988-1994 population: A latent profile analysis. Prev Med 2019; 120: 131-9.

12.   Morris LJ, D'Este C, Sargent-Cox K, Anstey KJ. Concurrent lifestyle risk factors: Clusters and determinants in an Australian sample. Prev Med 2016; 84: 1-5.

13.   Ding D, Rogers K, van der Ploeg H, Stamatakis E, Bauman AE. Traditional and Emerging Lifestyle Risk Behaviors and All-Cause Mortality in Middle-Aged and Older Adults: Evidence from a Large Population-Based Australian Cohort. PLOS MED 2015; 12(12): e1001917.

14.   Meader N, King K, Moe-Byrne T, et al. A systematic review on the clustering and co-occurrence of multiple risk behaviours. BMC Public Health 2016; 16: 657.

15.   Noble N, Paul C, Turon H, Oldmeadow C. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity

('SNAP') health risk factors. Prev Med 2015; 81: 16-41.

16.   Tuomilehto J, Schwarz P, Lindstrom J. Long-term benefits from lifestyle interventions for type 2 diabetes prevention: time to expand the efforts. Diabetes care 2011; 34 Suppl 2: S210-4.

17.   de Vries H, van 't Riet J, Spigt M, et al. Clusters of lifestyle behaviors: results from the Dutch SMILE study. Prev Med 2008; 46(3): 203-8.

18.   Halvorsen RE, Elvestad M, Molin M, Aune D. Fruit and vegetable consumption and the risk of type 2 diabetes: a systematic review and dose–response meta-analysis of prospective studies. BMJ Nutr Prev Health 2021; 4(2): 519-31.

19.   Vinke PC, Corpeleijn E, Dekker LH, Jacobs DR, Jr., Navis G, Kromhout D. Development of the food-based Lifelines Diet Score (LLDS) and its application in 129,369 Lifelines participants. Eur J Clin Nutr 2018; 72(8): 1111-9.

20.   Patterson R, McNamara E, Tainio M, et al. Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. Eur J Epidemiol 2018; 33(9): 811-29.

21.   Maghsoudi Z, Ghiasvand R, Salehi-Abargouei A. Empirically derived dietary patterns and incident type 2 diabetes mellitus: a systematic review and meta-analysis on prospective observational studies. Public Health Nutr 2016; 19(2): 230-41.

22.   Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ. Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. Eur J Epidemiol 2015; 30(7): 529-42.

23.   Pan A, Wang Y, Talaei M, Hu FB, Wu T. Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis. Lancet Diabetes Endocrinol 2015; 3(12): 958-67.

24.   Knott C, Bell S, Britton A. Alcohol consumption and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis of more than 1.9 million individuals from 38 observational studies. Diabetes care 2015; 38(9): 1804-12.

25.   Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol 2015; 44(4): 1172-80.

26.   Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. PLOS ONE 2015; 10(9): e0137203.

27.   WHO, International Diabetes Federation (IDF). Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. Genève, Switzerland: WHO, 2006.

28.   UNESCO. International Standard Classification of Education ISCED 2011. Montréal, Canada: UNESCO Institute for Statistics; 2012.

29.   Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003; 56(12): 1163-9.

30.   Molag ML, de Vries JH, Duif N, et al. Selecting informative food items for compiling food-

4

frequency questionnaires: comparison of procedures. Br J Nutr 2010; 104(3): 446-56.

31.   Streppel MT, de Vries JH, Meijboom S, et al. Relative validity of the food frequency questionnaire used to assess dietary intake in the Leiden Longevity Study. Nutr J 2013; 12: 75.

32.   Siebelink E, Geelen A, de Vries JH. Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. Br J Nutr 2011; 106(2): 274-81.

33.   Hagenaars JA, McCutcheon AL. Applied latent class analysis. Cambridge, England, UK: Cambridge University Press; 2002.

34.   Lanza ST, Rhoades BL. Latent class analysis: an alternative perspective on subgroup analysis in prevention and treatment. Prev Sci 2013; 14(2): 157-68.

35.   Rabe-Hesketh S, Skrondal A. Classical latent variable models for medical research. Stat Methods Med Res 2008; 17(1): 5-32.

36.   Ding C, O'Neill D, Bell S, Stamatakis E, Britton A. Association of alcohol consumption with morbidity and mortality in patients with cardiovascular disease: original data and meta-analysis of 48,423 men and women. BMC Med 2021; 19(1): 167.

37.   Weggemans RM, Backx FJG, Borghouts L, et al. The 2017 Dutch Physical Activity Guidelines. Int J Behav Nutr Phys Act 2018; 15(1): 58.

38.   Nylund KL, Asparouhov T, Muthén BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Structural Equation Modeling: A Multidisciplinary Journal 2007; 14(4): 535-69.

39.   Vermunt JK, Magidson J. LATENT GOLD 4.0 User's Guide. Belmont, Massachusetts, USA: Statistical Innovations Inc.; 2005.

40.   Newson RB. Attributable and Unattributable Risks and Fractions and other Scenario Comparisons. The Stata Journal 2013; 13(4): 672-98.

41.   Kivela J, Wikstrom K, Virtanen E, et al. Obtaining evidence base for the development of Feel4Diabetes intervention to prevent type 2 diabetes- a narrative literature review. BMC Endocr Disord 2020; 20(Suppl 1): 140.

42.   Newell SA, Girgis A, Sanson-Fisher RW, Savolainen NJ. The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: a critical review. Am J Prev Med 1999; 17(3): 211-29.

43.   Howe LD, Tilling K, Galobardes B, Lawlor DA. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. Epidemiology 2013; 24(1): 1-9.

44.   Peters SA, Bots ML, den Ruijter HM, et al. Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. J Clin Epidemiol 2012; 65(6): 686-95.

Additional Information

Contributions

MJD: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Visualization, Project administration. LHD: Conceptualization, Writing – review & editing, Supervision, Project administration. JJC: Writing – review & editing. GN: Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Supplementary Data

Online **Supplementary Files** can be accessed via
doi.org/10.1016/j.pmedr.2022.102012, or by scanning the QR code.
Article PMID: 36237838.

4

UNDERLYING FACTORS of Lifestyle Factors

2
PART

CHAPTER 5

Ming-Jie Duan[1]; Louise H. Dekker[1,2];

Juan-Jesus Carrero[3]; Gerjan Navis[1]

1. Department of Internal Medicine, University Medical Center Groningen, Groningen, The Netherlands

2. National Institute for Public Health and the Environment, Bilthoven, The Netherlands

3. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Using structural equation modeling to untangle pathways of risk factors associated with incident type 2 diabetes: the Lifelines cohort study

ABSTRACT

Background

Risk factors for type 2 diabetes are multifaceted and interrelated. Unraveling the complex pathways of modifiable risk factors related to incident type 2 diabetes will help prioritize prevention targets.

Methods

The current analysis extended a previously proposed conceptual model by Bardenheier et al on prediabetes with a cross-sectional design.[1] The model described the pathways of four aspects of modifiable risk factors in relation to incident type 2 diabetes, including socioeconomic status (income and education); lifestyle behaviors (diet quality, physical activity, TV watching, smoking, risk drinking, and unhealthy sleep duration); clinical markers (HDL-cholesterol, triglycerides, BMI, and waist circumference); and blood pressure. We performed structural equation modeling to test this conceptual model using a prospective population-based sample of 68,649 participants (35-80 years) from the Lifelines cohort study.

Results

During a median follow-up of 41 months, 1124 new cases of type 2 diabetes were identified (incidence 1.6%). The best-fitting model indicated that among all modifiable risk factors included, waist circumference had the biggest direct effect on type 2 diabetes (standardized β-coefficient 0.214), followed by HDL-cholesterol (standardized β-coefficient -0.134). Less TV watching and more physical activity were found to play an important role in improving clinical markers that were directly associated with type 2 diabetes. Education had the biggest positive effects on all lifestyle behaviors except for unhealthy sleep duration.

Conclusions

Our analysis provides evidence to support that structural equation modeling enables a holistic assessment of the interplay of type 2 diabetes risk factors, which not only

allows the estimation of their total effects but also prioritization of prevention targets. Regarding the current guideline for diabetes prevention, waist management in addition to BMI control (clinical level), as well as less TV watching in addition to more physical activity (behavioral level), may provide additional public health benefits. Better education would be the main societal goal for the prevention of type 2 diabetes.

5

Abbreviations

CFI – Comparative fit index

FFQ – Food frequency questionnaire

LLDS – Lifelines diet score

MVPA – Moderate-to-vigorous physical activity

RMSEA – Root mean square error of approximation

SEM – Structural equation modeling

SRMR – Standardized root mean square residual

TLI – Tucker-Lewis index

Introduction

The development of type 2 diabetes is multifactorial. Besides inherited traits and age, various modifiable risk factors have been identified. Among clinical risk factors, obesity has been found to be one of the strongest risk factors for type 2 diabetes. It has been suggested that excess body fat, especially visceral fat, is central to the pathogenesis of insulin resistance.[2,3] Prospective cohort studies also found abnormal blood lipid profile, such as low HDL-cholesterol and high triglycerides, to be a strong predictor for the development of type 2 diabetes.[4-6] For lifestyle behaviors, both interventions and observational studies have demonstrated that poor diet,[7,8] physical inactivity,[9,10] and smoking[11] may contribute to the risk of type 2 diabetes independent of weight change. Observational studies have also established that risk drinking is associated with high risk of type 2 diabetes.[12] In addition, emerging lifestyle risk factors, such as excessive TV watching and unhealthy sleep duration,[13-15] have potential as new type 2 diabetes prevention targets. After controlling for the aforementioned risk factors, socioeconomic status, such as low education and insufficient income, has been found to be associated with higher risk of type 2 diabetes.[16-18] We present a more extensive summary of evidence in **Supplementary Table 1**.

In diabetes research, conventional approaches for risk identification often apply traditional regression models, in which the net effects of risk factors are estimated under the assumption of an independent direct effect on diabetes status. However, some risk factors may act as mediators (e.g., obesity, blood lipids) or mainly exert indirect effects (e.g., education, income).[1,19] The lack of insight into their holistic interrelationships has led to the fragmentation of evidence and development of unfocused prevention programs. More specifically, obesity and abnormal blood lipids are largely attributed to unhealthy lifestyle behaviors, whereas all are strongly influenced by socioeconomic status. These factors, in turn, collectively form several hypothesized intersecting pathways that lead to the eventual development of type 2 diabetes.[16-18,20,21] Socioeconomic status is thus considered the overarching upstream determinant of type 2 diabetes for its significant effects on proximal (or downstream) risk factors. Likewise, lifestyle behaviors are the upstream determinants of clinical

disorders such as obesity.[22] In terms of primary prevention, it would be highly useful to understand the relatedness of a broad range of risk factors, so that aiming at prioritized risk factor targets and their most influential upstream determinants would optimize the effectiveness of diabetes prevention at population level.

To this purpose, we aimed to analyze a conceptual model (originally proposed by Bardenheier et al on prevalent prediabetes),[1,19] including multiple modifiable risk factors and their interrelationships for type 2 diabetes (**Fig. 1**). We extended the original conceptual model with 4 important lifestyle behaviors, i.e., TV watching,[13,14] smoking,[11] sleep duration,[15] and risk drinking.[12] We examined this model by structural equation modeling (SEM) using data from the Lifelines cohort study, focusing on incident type 2 diabetes as outcome. SEM is a multivariate statistical technique that allows the quantification of multiple intersecting pathways (yielding path coefficients) within a conceptual model simultaneously. Untangling the pathways of these risk factors may provide the additional evidence needed to develop better prevention strategies by identifying the most crucial pathways as priority prevention targets.

Methods

Study Design of the Lifelines Cohort Study

The Lifelines study is a multi-disciplinary prospective general population-based cohort study that applies in a unique three-generation design to study the health and health-related behaviors of 167,729 people living in the north of The Netherlands. The Lifelines cohort study was established from year 2006 to 2013. Detailed information regarding recruitment strategy and the representativeness of the Lifelines study population are shown in **Supplementary Text 1**.[23,24]

Four assessment rounds have taken place: T1-baseline assessment (year 2007 to 2014) and three follow-ups, i.e., T2, T3, and T4. Comprehensive physical examinations, biobanking, and questionnaires were conducted at T1 and T4 (**Supplementary Fig. 1**). The Lifelines study was conducted according to the principles of the Declaration of

Helsinki and was approved by the medical ethical committee of the University Medical Center Groningen, The Netherlands (approval number 2007/152). All participants gave written informed consent to participate in the study.

Study Population and Exclusion Criteria

In this study, participants between the ages of 35 and 80 years who were free of diabetes at baseline from the Lifelines cohort study were included. We further excluded participants if (1) they were diagnosed with cancer or renal failure before enrollment; (2) they were pregnant at baseline; (3) they developed type 1 diabetes or gestational diabetes during follow-ups; (4) they had no available follow-up data; and (5) they had unreliable dietary intake data. Dietary intake data was considered unreliable when the ratio between reported energy intake and basal metabolic rate, calculated with the Schofield equation,[25] was below 0.50 or above 2.75, based on the considerations of Goldberg.[26] Furthermore, except for physical activity and income, participants with missing data on other variables (missing less than 1%) were excluded. This led to an additional exclusion of 1.7% of the study population. In this study, multiple imputation was used to deal with missing data.[27] This additional exclusion aimed to avoid massive imputation and was not expected to have major impacts on our results. After applying exclusion criteria, in total 68,649 participants (40,121 women and 28,528 men) were included in the analysis. **Supplementary Fig. 2** shows the study flow chart.

Clinical Measurements

Blood samples were collected by venipuncture in a fasting state between 8 and 10 am. Serum levels of glucose, HbA_{1c}, HDL-cholesterol, and triglycerides were subsequently analyzed. Baseline measurements of blood pressure and anthropometry were made by trained research staff following standardized protocols. Anthropometric measurements were performed without shoes and heavy clothing. Participants were considered having hypertension at baseline if they (1) used hypertensive medication (ATC codes C02, C03, C07, C08, and C09);[28] (2) had systolic blood pressure ≥ 140 mmHg; or (3) had diastolic blood pressure ≥ 90 mmHg.[29] Detailed information for clinical measurements is available in **Supplementary Text 2**.
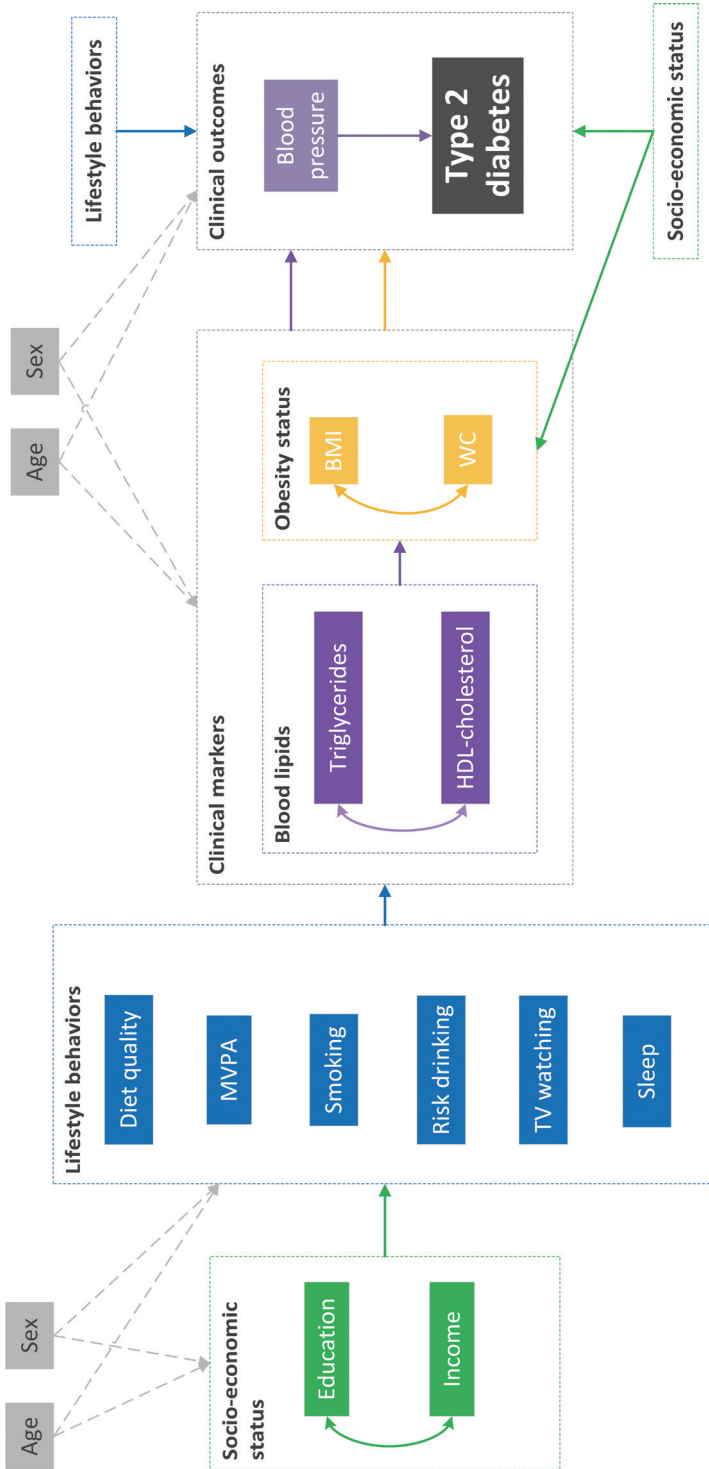
Fig. 1. Conceptual model illustrating pathways of risk factors to incident type 2 diabetes[a]

[a] MVPA denotes non-occupational moderate-to-vigorous physical activity; WC denotes waist circumference; and sleep denotes unhealthy sleep duration (versus healthy sleep duration). Straight line with one arrowhead denotes a direct effect (e.g., income to MVPA), and curved line with double arrowheads denotes a correlation term (e.g., triglycerides and HDL-cholesterol). For easy reading, several factors are repeated at different locations with different pathways depicted, but they do not differ from their identical others (e.g., education and income [socioeconomic status]).

5

Assessment of Lifestyle and Socioeconomic Covariates

Age, education level, income level, smoking status, sleep duration, TV watching time, and physical activity level were assessed by self-administered questionnaires. Age at baseline was calculated from date of birth in the questionnaire. Highest education level achieved was categorized according to the International Standard Classification of Education (ISCED): (1) low – level 0, 1, or 2; (2) middle – level 3 or 4; and (3) high – level 5 or 6.[30] Income was based on monthly household net income and was categorized as < 1000, 1000–2000, 2000–3000, and > 3000 euro/month. Smoking status was categorized as never, former, and current smoker. Unhealthy sleep duration was defined as sleep time less than 6 or more than 9 h per day.[15] Average TV watching time per day was asked in hours plus minutes. Physical activity level was assessed by the validated Short QUestionnaire to ASsess Health-enhancing physical activity (SQUASH),[31] from which non-occupational moderate-to-vigorous physical activity (MVPA), including commuting and sports (both if ≥ 4.0 MET), was calculated in minutes per week, and was further divided into sex-specific quartiles (if not zero) or coded to zero.[31,32]

Dietary intake was assessed using a semi-quantitative self-administered food frequency questionnaire (FFQ), which was aimed to assess the habitual intake of 110 food items (including alcohol) during the last month and was designed based on the validated Dutch FFQ.[33] The questionnaire assessed the frequency of consumption and portion sizes. The latter was estimated using fixed portion sizes (e.g., slices of bread, pieces of fruit) and commonly used household measures (e.g., cups, spoons). The food-based Lifelines Diet Score (LLDS) was calculated to evaluate the diet quality of each participant. More specifically, this score ranks the relative intake of nine food groups with positive health effects (vegetables, fruit, whole grain products, legumes/nuts, fish, oils/soft margarines, unsweetened dairy, coffee, and tea) and three food groups with negative health effects (red/processed meat, butter/hard margarines, and sugar-sweetened beverages). The development of this score is described in detail elsewhere.[34] Risk drinking was defined as consuming more than 15 g of alcohol per day, which was approximated to one drink per day.

Ascertainment of Incident Type 2 Diabetes

Incident type 2 diabetes was assessed by self-report questionnaires (T2, T3, and T4) and blood test (T4). Participants were considered an incident case if they met either of the following criteria: (1) self-reported newly developed type 2 diabetes from last available questionnaire; (2) had fasting glucose $\geq$ 7.0 mmol/L; or (3) had $HbA_{1c} \geq 48$ mmol/mol (6.5%).[35]

The Conceptual Model

Fig. 1 illustrates the conceptual model that connects modifiable risk factors with incident type 2 diabetes and with each other, in which they are grouped into four different levels, i.e., socioeconomic status (education and income), lifestyle behaviors (diet quality [LLDS], non-occupational MVPA, smoking status, TV watching time, unhealthy sleep duration, and risk drinking), clinical markers (triglycerides, HDL-cholesterol, BMI, and waist circumference), and clinical outcomes (blood pressure and incident type 2 diabetes).

The original conceptual model was first proposed by Bardenheier et al on prevalent prediabetes.[1,19] We extended the original model by adding four modifiable lifestyle behaviors (smoking, TV watching, risk drinking, and unhealthy sleep duration) and adapting several pathways based on previous evidence (**Supplementary Table 1**). Specifically, we hypothesized that (**Fig. 1**) (1) socioeconomic status had direct effects on lifestyle behaviors; (2) lifestyle behaviors had direct effects on clinical markers; (3) blood lipids (HDL-cholesterol and triglycerides) had direct effects on obesity status (BMI and waist circumference); (4) blood pressure had direct effect on incident type 2 diabetes; and (5) clinical markers had direct effects on clinical outcomes. In the conceptual model, we also allowed direct effects from socioeconomic status and lifestyle behaviors on obesity status and clinical outcomes, because there might be unobserved mediators along the causal pathways. Age and sex, as two strong unmodifiable risk factors for type 2 diabetes, were also included in the conceptual model and were hypothesized to have direct effects on all other factors. In total, the conceptual model yielded 96 hypothesized paths and 3 correlations between the measurement errors of variables.

5

Statistical Analysis

We used structural equation modeling (SEM) to examine our conceptual model (**Fig. 1**). SEM analysis is chiefly a confirmatory statistical technique to test if the hypothesized model is correctly specified and supported by the data observed, rather than generating new hypothesis.[27] Because the hypothesized model consisted of ordered categorical variables (e.g., income), we used the estimation method – weighted least square with mean and variance adjustment.[36] The WLSMV is suggested to be the most suitable estimator in SEM if the model tested contains multiple binary or ordered endogenous categorical variables.[36] Additionally, we estimated the associations between each included risk factor and incident type 2 diabetes using logistic regression model as a conventional approach for risk identification.

In order to improve and evaluate model fit, the following aspects were considered. First, we referred to the model fit indices calculated from the SEM output, i.e., comparative fit index (CFI), standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA), and Tucker-Lewis index (TLI). We did not purely rely on the commonly used cut-offs of these fit indices as the absolute criteria.[37] Additionally, we performed sensitivity analyses using other estimators to cross-check the model fit. Second, modification indices, which are based on chi-square statistics indicating the changes in model's goodness-of-fit if an omitted path was added, were also used as reference for adjustments of particular paths.[27]

Missing data for income (proportion of missing 15.3%) and non-occupational MVPA (proportion of missing 6.4%) were imputed with chained equation creating 25 imputed datasets,[38] from which results were pooled according to the Rubin's rule.[39]

In order to ensure the robustness of our results, we performed several sensitivity analyses. Detailed methods and results are discussed in **Supplementary Text 3**.

We used Stata (version 13.1) for data management and descriptive data analyses, and R Studio (version 1.1.383) with lavaan package (version 0.6-5; Y. Rosseel) for SEM analysis.[40] Multiple imputation was performed with mice package (version 3.8.0; S. van Buuren et al.) in R Studio,[41] and results from the imputed datasets were pooled

with semTools package (version 0.5-2; T.D. Jorgensen et al.) in R Studio.[42] Statistical significance was considered if P value < 0.05.

Results

Descriptive Statistics

Among 68,649 participants (aged 35-80 years) included in the analysis, we identified 1124 type 2 diabetes cases (incidence 1.6%) after a median follow-up of 41 months. Compared with participants who did not develop type 2 diabetes throughout the study, those who developed type 2 diabetes tended to be older and male, have less education and lower income at baseline, engage in negative lifestyle behaviors, and have poorer clinical markers (**Table 1**).

Structural Equation Model

The best-fit model (**Fig. 2**; CFI 0.981, TLI 0.949, RMSEA 0.032, SRMR 0.023) was achieved after we made adjustments to our original hypothesized model (**Fig. 1**; CFI 0.953, TLI 0.774, RMSEA 0.068, SRMR 0.039). The model fit indices of the best-fit model indicated that the hypothesized model was well supported by the observed data (cut-offs commonly considered for a good model fit: CFI > 0.090, TLI > 0.090, RMSEA < 0.080, and SRMR < 0.060). In brief, we dropped paths that did not yield significant estimates. Based on modification indices (mi), we further added two correlation paths between smoking status and risk drinking (mi = 2444.854), and between non-occupational MVPA and LLDS (mi = 869.306). Additionally, several paths (e.g., TV watching to incident type 2 diabetes) were dropped because results from sensitivity analyses showed substantial changes in path coefficients, which suggested that these estimates were not robust. We present details of stepwise adjustments and reasons for changes in **Supplementary Table 2**.

Fig. 2 presents the best-fit hypothesized model with standardized path coefficients. Paths related to age and sex are not shown in **Fig. 2** but available in **Supplementary Table 3**. Among all modifiable risk factors included in the conceptual model (standardized

β-coefficients are given in parentheses), waist circumference (0.214) had the strongest direct effect on type 2 diabetes, followed by HDL-cholesterol (-0.134), triglycerides (0.096), income (-0.074), blood pressure (0.055), diet quality (-0.045), and smoking (0.035). Except for unhealthy sleep duration, education showed larger positive effects than income on all lifestyle behaviors. All included lifestyle behaviors were significantly associated with clinical markers, among which non-occupational MVPA, smoking, and TV watching yielded larger effect sizes. Risk drinking and smoking showed mixed effects on metabolic profiles. Almost all factors received strong direct effects from age and sex. In addition, correlations were found between BMI and waist circumference, between education and income, between triglycerides and HDL-cholesterol, between smoking status and risk drinking, and between diet quality and non-occupational MVPA.

For more information, please see **Supplementary Table 3**, which shows all standardized and unstandardized coefficients with standard errors for all paths.

Supplementary Table 4 shows the results of logistic regression model as a conventional approach for risk identification. The strongest effects were found for income group > 3000 euro/month (-0.405), waist circumference (0.386), sex (women compared with men, 0.355), and HDL-cholesterol (-0.339).

Results from sensitivity analyses showed consistent results, which indicated our estimates are robust. Compared with the main analysis, some variations were found when replacing incident type 2 diabetes by fasting glucose and HbA_{1c} measured at T4. Detailed discussions of sensitivity analyses are presented in **Supplementary Text 3**.
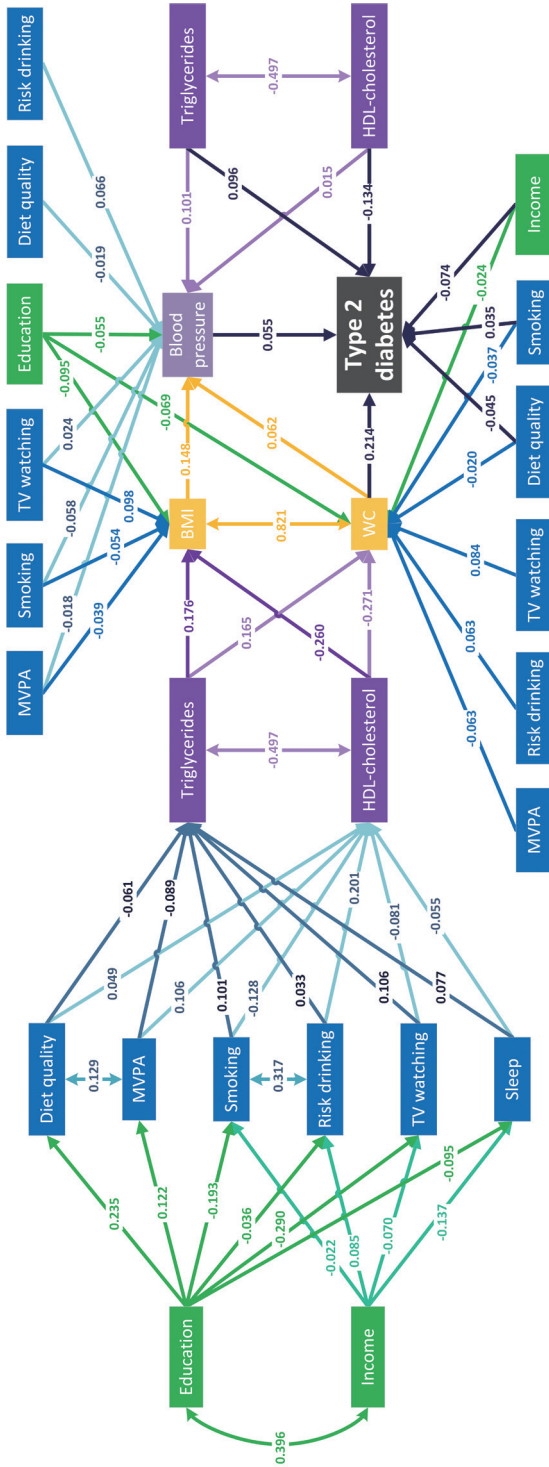
Fig. 2. Quantified best-fit conceptual model illustrating pathways of risk factors to incident type 2 diabetes[a]

[a] MVPA denotes non-occupational moderate-to-vigorous physical activity; WC denotes waist circumference; and sleep denotes unhealthy sleep duration (versus healthy sleep duration). Straight line with one arrowhead denotes a direct effect (e.g., income to MVPA), and straight or curved line with double arrowheads denotes a correlation term (e.g., triglycerides and HDL-cholesterol). For easy reading, several factors are repeated at different locations with different pathways depicted, but they do not differ from their identical others (e.g., education and income [socioeconomic status]). Sample size tested for the conceptual model, n=68,649. Tests for significance: P value<0.001 for all path coefficients except for HDL-cholesterol to blood pressure (P value=0.002) and smoking to incident type 2 diabetes (P value=0.012). The model was adjusted for sex and age.

5

Table 1. Baseline characteristics by diabetes status[a]			
Characteristics	Total (*n* = 68,649)	Type 2 diabetes (*n* = 1124)	Non-diabetes (*n* = 67,525)
Age, years	49.7±9.5	54.8±10.0	49.6±9.4
Sex, %			
Women	58.4	49.1	58.6
Men	41.6	50.1	41.4
Fasting glucose, mmol/L	4.97±0.50	5.81±0.65	4.95±0.48
HbA_{1c}, mmol/mol	37.31±3.27	41.55±3.49	37.24±3.22
HbA_{1c}, %	5.55±0.30	5.94±0.32	5.55±0.29
Triglycerides, mmol/L	1.19±0.80	1.77±1.54	1.18±0.78
HDL-cholesterol, mmol/L	1.53±0.41	1.30±0.37	1.53±0.41
BMI, kg/m^2	26.2±4.0	29.6±4.7	26.1±4.0
Underweight (<18.5), %	0.4	0.1	0.4
Normal (18.5-24.9)	41.5	13.4	41.9
Overweight (25.0-29.9), %	43.0	45.6	43.0
Obese (>30.0), %	15.1	40.8	14.7
Waist circumference, cm	91.0±11.7	101.5±12.1	90.8±11.6
Large waist circumference[b], %	34.2	66.6	33.6
Hypertension, %	28.8	59.4	28.3
Systolic blood pressure, mmHg	126.4±15.5	134.7±16.0	126.3±15.4
Diastolic blood pressure, mmHg	74.9±9.4	77.8±10.0	74.8±9.4

Characteristics	Total (n = 68,649)	Type 2 diabetes (n = 1124)	Non-diabetes (n = 67,525)
Table 1. Continued.			
Lowest tertile of Lifelines diet score, %	28.6	32.1	28.6
Alcohol intake, grams/day	4.57 (0.89, 11.11)	3.79 (0.52, 12.25)	4.64 (0.89, 11.09)
Risk drinking (>15 grams/day), %	16.7	20.3	16.6
Non-occupational MVPA, minutes/week[c]	190 (65, 370)	160 (30, 360)	190 (70, 370)
Smoking status, %			
Never	44.6	33.7	44.8
Former	38.5	46.7	38.3
Current	16.9	19.6	16.8
TV watching time, hours/day	2.5±1.3	3.0±1.5	2.5±1.3
Sleep duration, hours/day	7.42±0.85	7.42±0.96	7.42±0.85
Having unhealthy sleep duration (<6 or >9 hours/day), %	2.97	5.42	2.93
Education, %			
Low	31.2	46.9	30.9
Middle	38.7	33.1	38.7
High	30.2	20.0	30.4
Income (euro/month), %[d]			
< 1000	3.0	5.0	3.0
1000-2000	18.5	26.2	18.3
2000-3000	30.2	30.3	30.2
> 3000	33.0	24.0	33.2

[a] Data are expressed as unadjusted mean ± standard deviation for age, fasting glucose, HbA$_{1c}$, triglycerides, HDL-cholesterol, BMI, waist circumference, systolic blood pressure, diastolic blood pressure, TV watching time, and sleep duration; Data are expressed as median (interquartile) for non-occupational MVPA and alcohol intake; Data are expressed as observed percentage for sex, obesity status, large waist circumference, hypertension, lowest tertile of Lifelines diet score, risk drinking, smoking status, having unhealthy sleep duration, education, and income.

[b] Large waist circumference is defined as waist circumference >102 cm (40 inches) in men and >88 cm (35 inches) in women.

[c] Non-occupational MVPA denotes non-occupational moderate-to-vigorous physical activity level. The percentages of missing data were: total 6.4%, type 2 diabetes cases 8.8% and non-diabetes cases 6.4%.

[d] For income level, the percentages of missing data were: total 15.3%, type 2 diabetes cases 14.6% and non-diabetes cases 15.3%.

Discussion

This study is the first that examined a broad range of key modifiable risk factors simultaneously in relation to incident type 2 diabetes using SEM. Our analysis quantified the complex pathways of these concomitant risk factors on the subsequent risk of developing type 2 diabetes, which provides valuable insights into the identification of priority prevention targets. Our results further extend knowledge of previous similar studies on prevalent prediabetes and prevalent type 2 diabetes by incorporating four important lifestyle behavioral factors, i.e., smoking, TV watching, risk drinking, and unhealthy sleep duration.[1,19]

Interrelationships of Risk Factors

There are several key findings. First, of the two obesity indicators examined, large waist circumference was found to have a strong direct effect on type 2 diabetes. Our results highlight the importance of waist management, in addition to BMI control, for diabetes prevention in both clinical practice and public health interventions.[2,3] Second, blood lipids, assessed as a higher level of HDL-cholesterol and a lower level of triglycerides, had critical direct effects on lowering diabetes risk. Additionally, healthier lifestyle behaviors, especially watching less TV and engaging in more non-occupational MVPA, indirectly and favorably affected diabetes risk through the mediation of clinical markers (i.e., blood lipids and obesity status), indicating their equal importance in diabetes prevention.

For socioeconomic status, our analysis dissected the differential effects between education and income, showing that low education, rather than insufficient income, is the major upstream determinant of unhealthy lifestyle behaviors. In the context of The Netherlands, where the level of income inequality is relatively low, the effect of lower income on lifestyle behaviors may not predominantly be due to less access to healthy lifestyle resources. Instead, it is suggested that self-perceived control, attitudes, and social norms towards adopting a healthier lifestyle are more restrained among those with lower education.[43] Programs promoting healthy lifestyle should be complemented by additional elements to help people with lower education.[44,45]

It is noteworthy that we observed direct effects of education on obesity status, as well as of income, diet quality, and smoking on type 2 diabetes. A cautious interpretation is warranted, as it cannot be excluded that the observed direct effects are in fact due to other, but unobserved, existing mediators or confounders, such as neighborhood deprivation (distal environmental factors) and chronic inflammation (proximal clinical biomarkers).[21,46,47]

Identification of Priority Prevention Targets

In terms of primary prevention, this simultaneous quantification of multiple risk factors and their intersecting pathways puts scattered evidence together and enables the identification of key upstream prevention targets for type 2 diabetes. Public health programs on these targets may have the potential to address as much of the broader risk profile as possible, particularly for those proximal clinical markers, for which pharmacological interventions may often be needed. Based on our results, (1) reducing large waist circumference may be prioritized as a main clinical target for diabetes prevention; (2) less TV watching time and more physical activity may be the main behavioral targets; and (3) better education may be the main societal target. Future studies are encouraged to examine the conceptual model in other populations.

It should be noted that the prevalence of type 2 diabetes at baseline in our population from the northern Netherlands (4.5%) is comparable to the average of upper-middle-income countries (5.6%), but lower than the average of high-income countries (7.9%).[48] Regarding incidence, 1.6% of our study sample developed type 2 diabetes after a median follow-up of 41 months (230,259 person-years), which is translated into an incidence rate of 4.9 per 1000 person-years. In the literature, we found a wide range of incidence across different countries and cohorts, ranging from 2.6 per 1000 person-years in the UK Biobank study to 11.4 per 1000 person-years in the American Multi-Ethnic Study of Atherosclerosis.[49,50] Despite the differences in cohort design and methodology that preclude direct comparisons, this high prevalence and incidence of type 2 diabetes worldwide call for us researchers to further work on curbing this global pandemic, especially by adopting innovative approaches to further build the evidence

5

base for the design of more effective public health programs (for detailed data, please see **Supplementary Table 5**).

Strengths and Limitations

Conventional approaches for risk identification commonly estimate the total net effects of risk factors, but leave their interrelationships masked. We further illustrated this by comparing the results between using SEM and logistic regression model (**Supplementary Table 4**). More specifically, SEM clearly elucidated the extent to which education impacted on risk of type 2 diabetes through the mediation of lifestyle behaviors, while such information is unavailable in results from logistic regression models. Using SEM also avoids possible multiple testing of significance if each mediation pathway was modelled separately.

In our conceptual model, we did not develop latent variables as in previous similar studies.[1,19] Instead, we used single aggregate measures for diet and physical activity, and additionally added a correlation term between income and education. For diet and physical activity, our selected indicators are evidence-based and easy to apply to evaluation at population level.[32,34] However, for latent variables, indicators were usually arbitrarily selected that were specific to the study population, which may limit their generalizability. Nevertheless, we acknowledge that constructing a latent variable for lifestyle factors may help reduce measurement errors. For effects of socioeconomic status, we clearly illustrated that the effects of income and education were different along the pathways to type 2 diabetes.

Our study also has some limitations. Even though we constructed the model in a prospective setting, the hypothesized pathways from socioeconomic status to clinical biomarkers are still of cross-sectional nature, although the lifestyle questionnaires were collected before the clinical measurements, and socioeconomic status was unlikely to change throughout the study period. An alternative conceptual model is also possible, even if model fit indices and sensitivity analyses indicate that our final model was well supported by the data observed. In addition, as the Lifelines cohort mainly consists of local Dutch participants, it may not be possible to extrapolate our results to other

populations. Another limitation of this study is that misclassification could occur in the ascertainment of type 2 diabetes cases, since at T2 and T3 only self-reported data was available. We also regrettably do not have data on medication use during follow-ups to validate self-reported diagnosis of type 2 diabetes. However, as most cases were identified by objective laboratory measurements at T4, this limitation is unlikely to have introduced severe bias in our results. A final concern is that we regrettably could not analyze the potential impacts of lost to follow-up (23.2%) among eligible participants. Such attrition could affect our estimation, specifically for the pathways directly linked to type 2 diabetes status. Nonetheless, the baseline characteristics of those who had no follow-up data were comparable with the study population, except for some minor differences in education level (**Supplementary Table 6**). Simulation studies have shown that such attrition bias may only have limited influences on estimates of associations in regression analysis.[51,52]

Conclusions

This prospective study examined modifiable risk factors as a system in relation to incident type 2 diabetes through integrated pathways in a large population-based cohort. Quantifying the pathways of those modifiable risk factors using SEM may be a useful tool for the prioritization of prevention targets. Primary prevention strategies targeting proximal clinical risk factors should be complemented with public health initiatives that simultaneously address their corresponding upstream determinants. Regarding the current guideline for diabetes prevention, waist management in addition to BMI control (clinical level), as well as less TV watching in addition to more physical activity (behavioral level), may provide additional public health benefits. Better education would be the main societal goal for the prevention of type 2 diabetes.

Reference

1.   Bardenheier BH, Bullard KM, Caspersen CJ, Cheng YJ, Gregg EW, Geiss LS. A novel use of structural equation models to examine factors associated with prediabetes among adults aged 50 years and older: National Health and Nutrition Examination Survey 2001-2006. Diabetes Care 2013; 36(9): 2655-62.

2.   Lee DH, Keum N, Hu FB, et al. Comparison of the association of predicted fat mass, body mass index, and other obesity indicators with type 2 diabetes risk: two large prospective studies in US men and women. Eur J Epidemiol 2018; 33(11): 1113-23.

3.   Neeland IJ, Ross R, Despres JP, et al. Visceral and ectopic fat, atherosclerosis, and cardiometabolic disease: a position statement. Lancet Diabetes Endocrinol 2019; 7(9): 715-25.

4.   Despres JP, Lemieux I. Abdominal obesity and metabolic syndrome. Nature 2006; 444(7121): 881-7.

5.   Kruit JK, Brunham LR, Verchere CB, Hayden MR. HDL and LDL cholesterol significantly influence beta-cell function in type 2 diabetes mellitus. Curr Opin Lipidol 2010; 21(3): 178-85.

6.   von Eckardstein A, Widmann C. High-density lipoprotein, beta cells, and diabetes. Cardiovasc Res 2014; 103(3): 384-94.

7.   Maghsoudi Z, Ghiasvand R, Salehi-Abargouei A. Empirically derived dietary patterns and incident type 2 diabetes mellitus: a systematic review and meta-analysis on prospective observational studies. Public Health Nutr 2016; 19(2): 230-41.

8.   Schulze MB, Hoffmann K, Manson JE, et al. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. Am J Clin Nutr 2005; 82(3): 675-84; quiz 714-5.

9.   Astrup A. Healthy lifestyles in Europe: prevention of obesity and type II diabetes by diet and physical activity. Public Health Nutr 2001; 4(2B): 499-515.

10.   Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ. Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. Eur J Epidemiol 2015; 30(7): 529-42.

11.   Pan A, Wang Y, Talaei M, Hu FB, Wu T. Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis. Lancet Diabetes Endocrinol 2015; 3(12): 958-67.

12.   Knott C, Bell S, Britton A. Alcohol Consumption and the Risk of Type 2 Diabetes: A Systematic Review and Dose-Response Meta-analysis of More Than 1.9 Million Individuals From 38 Observational Studies. Diabetes Care 2015; 38(9): 1804-12.

13.   Llavero-Valero M, Escalada San Martin J, Martinez-Gonzalez MA, et al. Promoting exercise, reducing sedentarism or both for diabetes prevention: The "Seguimiento Universidad De Navarra" (SUN) cohort. Nutr Metab Cardiovasc Dis 2021; 31(2): 411-9.

14.   Patterson R, McNamara E, Tainio M, et al. Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. Eur J Epidemiol 2018; 33(9): 811-29.

15.   Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2010; 33(2): 414-20.

16.   Foster HME, Celis-Morales CA, Nicholl BI, et al. The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: a prospective analysis of the UK Biobank cohort. Lancet Public Health 2018; 3(12): e576-e85.

17.   Maty SC, Everson-Rose SA, Haan MN, Raghunathan TE, Kaplan GA. Education, income, occupation, and the 34-year incidence (1965-99) of Type 2 diabetes in the Alameda County Study. Int J Epidemiol 2005; 34(6): 1274-81.

18.   Vinke PC, Navis G, Kromhout D, Corpeleijn E. Socio-economic disparities in the association of diet quality and type 2 diabetes incidence in the Dutch Lifelines cohort. EClinicalMedicine 2020; 19: 100252.

19.   Roman-Urrestarazu A, Ali FM, Reka H, Renwick MJ, Roman GD, Mossialos E. Structural equation model for estimating risk factors in type 2 diabetes mellitus in a Middle Eastern setting: evidence from the STEPS Qatar. BMJ Open Diabetes Res Care 2016; 4(1): e000231.

20.   Duan MJ, Dekker LH, Carrero JJ, Navis G. Blood lipids-related dietary patterns derived from reduced rank regression are associated with incident type 2 diabetes. Clin Nutr 2021; 40(7): 4712-9.

21.   Zhu Y, Duan MJ, Riphagen IJ, et al. Separate and combined effects of individual and neighbourhood socio-economic disadvantage on health-related lifestyle risk factors: a multilevel analysis. Int J Epidemiol 2022; 50(6): 1959-69.

22.   Lakerveld J, Mackenbach J. The Upstream Determinants of Adult Obesity. Obes Facts 2017; 10(3): 216-22.

23.   Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol 2015; 44(4): 1172-80.

24.   Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. PLOS ONE 2015; 10(9): e0137203.

25.   Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. Hum Nutr Clin Nutr 1985; 39 Suppl 1: 5-41.

26.   Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. Int J Obes Relat Metab Disord 2000; 24(9): 1119-30.

27.   Kline RB. Principles and practice of structural equation modeling (4th ed.). New York, NY, USA: Guilford Press; 2016.

28.   WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health. ATC/DDD Index 2020. Oslo, Norway; 2020.

29.   Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European

Society of Hypertension. J Hypertens 2018; 36(10): 1953-2041.

30.   UNESCO. International Standard Classification of Education. 1997. https://www.unesco.org/en/education.

31.   Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003; 56(12): 1163-9.

32.   Byambasukh O, Snieder H, Corpeleijn E. Relation Between Leisure Time, Commuting, and Occupational Physical Activity With Blood Pressure in 125 402 Adults: The Lifelines Cohort. J Am Heart Assoc 2020; 9(4): e014313.

33.   Streppel MT, de Vries JH, Meijboom S, et al. Relative validity of the food frequency questionnaire used to assess dietary intake in the Leiden Longevity Study. Nutr J 2013; 12: 75.

34.   Vinke PC, Corpeleijn E, Dekker LH, Jacobs DR, Jr., Navis G, Kromhout D. Development of the food-based Lifelines Diet Score (LLDS) and its application in 129,369 Lifelines participants. Eur J Clin Nutr 2018; 72(8): 1111-9.

35.   American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. Diabetes Care 2020; 43(Suppl 1): S14-S31.

36.   Muthén B, du Toit SHC, Spisic D. Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes, 1997.

37.   Xia Y, Yang Y. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. Behav Res Methods 2019; 51(1): 409-28.

38.   van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999; 18(6): 681-94.

39.   Li K-HL, Meng X-L, Raghunathan TE, Rubin DB. Significance levels from repeated p-values with multiply-imputed data. Statistica Sinica 1991; 1: 65-92.

40.   Rosseel Y. Lavaan: An R package for structural equation modeling and more. Version 0.5-12 (BETA). Journal of Statistical Software 2012; 48(2): 1-36.

41.   van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software; 45(3): 1-67.

42.   Jorgensen TD, Pornprasertmanit S, Schoemann AM, et al. Package 'semTools'. 2019.

43.   Stronks K, van de Mheen HD, Looman CW, Mackenbach JP. Cultural, material, and psychosocial correlates of the socioeconomic gradient in smoking behavior among adults. Prev Med 1997; 26(5 Pt 1): 754-66.

44.   Ball K, Abbott G, Cleland V, et al. Resilience to obesity among socioeconomically disadvantaged women: the READI study. Int J Obes (Lond) 2012; 36(6): 855-65.

45.   van der Lucht F, Polder JJ. Towards better health : The Dutch 2010 Public Health Status and Forecasts Report. Bilthoven, The Netherlands: RIVM, 2010.

46.   Dekker LH, Rijnks RH, Navis GJ. Regional variation in type 2 diabetes: evidence from 137

820 adults on the role of neighbourhood body mass index. Eur J Public Health 2020; 30(1): 189-94.

47.		Kivimaki M, Vahtera J, Tabak AG, et al. Neighbourhood socioeconomic disadvantage, risk factors, and diabetes from childhood to middle age in the Young Finns Study: a cohort study. Lancet Public Health 2018; 3(8): e365-e73.

48.		Institute for Health Metrics and Evaluation. Global Health Data Exchange (GHDx) query tool. Global Burden of Diseases, Injuries, and Risk Factors Study. 2021.

49.		Levy RB, Rauber F, Chang K, et al. Ultra-processed food consumption and type 2 diabetes incidence: A prospective cohort study. Clin Nutr 2021; 40(5): 3608-14.

50.		Joseph JJ, Echouffo-Tcheugui JB, Golden SH, et al. Physical activity, sedentary behaviors and the incidence of type 2 diabetes mellitus: the Multi-Ethnic Study of Atherosclerosis (MESA). BMJ Open Diabetes Res Care 2016; 4(1): e000185.

51.		Howe LD, Tilling K, Galobardes B, Lawlor DA. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. Epidemiology 2013; 24(1): 1-9.

52.		Peters SA, Bots ML, den Ruijter HM, et al. Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. J Clin Epidemiol 2012; 65(6): 686-95.

5

Additional Information

Contributions

MJD, LHD, and GN designed the study. MJD analyzed the data and drafted the manuscript. LHD, JJC, and GN contributed to the discussion and critically reviewed/ edited the manuscript. MJD has primary responsibility for the final content. All authors approved the final content of the manuscript.

Supplementary Data

Online **Supplementary Files** can be accessed via
doi.org/10.1007/s11121-022-01357-5, or by scanning the QR code.
Article PMID: 35230614.

5

CHAPTER 6

Ming-Jie F. Duan[1*]; Yinjie Zhu[1*]; Louise H. Dekker[1,2,3]; Jochen O. Mierau[2,4]; Eva Corpeleijn[5]; Stephan J.L. Bakker[1]; Gerjan Navis[1]

Effects of education and income on incident type 2 diabetes and cardiovascular diseases: a Dutch prospective study

1. Department of Internal Medicine, University Medical Center Groningen, Groningen, The Netherlands

2. Aletta Jacobs School of Public Health, University of Groningen, Groningen, The Netherlands

3. National Institute for Public Health and the Environment, Bilthoven, The Netherlands

4. Faculty of Economics and Business, University of Groningen, The Netherlands

5. Department of Epidemiology, University Medical Center Groningen, Groningen, The Netherlands

*Joint first authors

ABSTRACT

Background

Education and income, as two primary socioeconomic indicators, are often used interchangeably in health research. However, there is a lack of clear distinction between these two indicators concerning their associations with health.

Objective

This study aimed to investigate the separate and combined effects of education and income in relation to incident type 2 diabetes and cardiovascular diseases in the general population.

Design and Participants

Participants aged between 30 and 65 years from the prospective Dutch Lifelines cohort study were included. Two sub-cohorts were subsequently created, including 83,759 and 91,083 participants for a type 2 diabetes cohort and a cardiovascular diseases cohort, respectively.

Main Measures

Education and income level were assessed by self-report questionnaires. The outcomes were incident type 2 diabetes and cardiovascular diseases (defined as the earliest non-fatal cardiovascular event).

Key Results

A total of 1228 new cases of type 2 diabetes (incidence 1.5%) and 3286 (incidence 3.6%) new cases of cardiovascular diseases were identified, after a median follow-up of 43 and 44 months, respectively. Low education and low income (<1000 euro/month) were both positively associated with a higher risk of incident type 2 diabetes (OR 1.24 [95%CI 1.04-1.48] and OR 1.71 [95%CI 1.30-2.26], respectively); and with a higher risk of incident cardiovascular diseases (OR 1.15 [95%CI 1.04-1.28] and OR 1.24 [95%CI 1.02-1.52], respectively); independent of age, sex, lifestyle factors, BMI, clinical biomarkers, comorbid conditions at baseline, and each other. Results from the

combined associations of education and income showed that within each education group, a higher income was associated with better health; and similarly, a higher education was associated with better health within each income group, except for the low-income group.

Conclusions

Education and income were both independently associated with incident type 2 diabetes and cardiovascular diseases. The combined associations of these two socioeconomic indicators revealed that within each education or income level, substantial health disparities existed across strata of the other socioeconomic indicator. Education and income are two equally important socioeconomic indicators in health, and should be considered simultaneously in health research and policymaking.

6

Abbreviations

BMI – Body mass index

FFQ – Food frequency questionnaire

LLDS – Lifelines diet score

MVPA – Moderate-to-vigorous physical activity

SES – Socioeconomic status

Introduction

Health disparities related to non-communicable diseases persist across socioeconomic strata. Abundant evidence has demonstrated that people with low socioeconomic status (SES) are disproportionately affected by higher risks of all-cause mortality, the metabolic syndrome, type 2 diabetes, and cardiovascular diseases.[1-4] It has been suggested that limited access to health and health-care resources, chronic stress, unhealthy lifestyle, and exposure to pollutants were found to play an important role in explaining the adverse health outcomes associated with low SES.[5]

Education and income are two primary components of SES. However, a clear distinction between these two socioeconomic indicators is often lacking.[3,6,7] Many studies on health disparities only considered one of them,[7] while some other studies focused on an aggregate measure of SES derived from multivariate statistics.[8,9] Research often made references to one indicator to corroborate findings into the other. There is increasing awareness that education and income should not be used interchangeably, since they capture different dimensions of health-related resources and may impact on health through different pathways.[6,7] As the main upstream determinants of health outcomes, research is needed to clarify the differences between education and income concerning their associations with health outcomes.

In public health practice, the inconsistent use of education and income may result in inaccurate identification of socioeconomically vulnerable groups, since people do not always hold a matching socioeconomic position.[8] It has been suggested that having such status inconsistency carries its own health risks. However, to date, only a few studies have explored such health disparities within different socioeconomic strata.[10-13] It is therefore also important to assess how different combinations of education and income levels are associated with health outcomes.

Therefore, using a large Dutch population cohort, this study aimed to evaluate the effects of education and income – separately and jointly using a combined indicator – on incident type 2 diabetes and cardiovascular diseases. Specifically, this study aimed to address how education and income may contribute to the short-term inequities in

these two health outcomes.

Methods

Cohort Design

The Lifelines cohort is a multidisciplinary prospective population-based cohort study that uses a unique three-generation design to study the health and health-related behaviors of 167,729 persons living in the north of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, and physical factors, which contribute to health and disease of the general population. Before study entry, a signed informed consent form was obtained from each participant. The Lifelines study is conducted according to the principles of the Declaration of Helsinki and approved by the Medical Ethics Committee of the University Medical Center Groningen, The Netherlands. The overall design and rationale of the study have been described in detail elsewhere.[14,15]

After the baseline assessment (T1, years 2007 to 2013), all participants were invited for new rounds of assessments approximately every 5 years. In between assessments, follow-up questionnaires were completed approximately once every 1.5-2.5 years (**Supplementary Figure S1**). The current analysis used data from the baseline assessment T1 and the second assessment T4, as well as the two follow-ups (T2 and T3) in between. Currently, the third round of assessment is on-going. Comprehensive physical examinations, biobanking of blood and urine, and questionnaires were conducted at T1 and T4. Follow-up questionnaires for status of type 2 diabetes and cardiovascular diseases were issued to participants at T2, T3, and T4.

Study Population

For this study, we included all participants aged between 30 and 65 years. We subsequently created two sub-cohorts from those included participants, with one for type 2 diabetes and the other for cardiovascular diseases. For the diabetes cohort, we included participants who were free of diabetes at baseline, and further excluded

participants who had no follow-up data to determine status of diabetes. We also excluded participants who reported the development of type 1 diabetes or gestational diabetes during the follow-ups. For the cardiovascular diseases cohort, we included participants who were free of cardiovascular diseases at baseline, and further excluded participants who had no follow-up data to determine status of cardiovascular diseases. Participants who had less than 1 year of follow-up after baseline were also excluded. In order to avoid massive imputation, we additionally excluded participants who had no available data on education level and BMI at baseline for both sub-cohorts. This led to an additional exclusion of approximately 0.5% of the study population, which was not expected to influence our results. In total, 83,759 and 91,083 participants were included and analyzed in the diabetes cohort and the cardiovascular diseases cohort, respectively. **Supplementary Figure S2** shows the study flow chart.

Data Collection

Ascertainment of Incident Type 2 Diabetes and Cardiovascular Diseases

Incident type 2 diabetes and cardiovascular diseases were assessed by self-report questionnaires at the two follow-ups (T2 and T3) and the second assessment (T4). Additionally, we assessed incident cases based on blood measurements and pathology on electrocardiograms, which were available at the second assessment (T4). For type 2 diabetes, an incident case was considered as fasting blood glucose ≥ 7.0 mmol/L or $HbA_{1c} \geq 6.5\%$.[16] For cardiovascular diseases, the primary outcome was defined as the earliest non-fatal major cardiovascular event, including stroke (ischemic and hemorrhagic), myocardial infarction, heart failure, percutaneous transluminal coronary angioplasty surgery, and coronary artery bypass grafting surgery.[17] Secondary outcome was a composite of death from any cause and non-fatal major cardiovascular event as described above. However, data on prescribed medication was not available during follow-ups. Data of medical records, causes of death, and the precise time of diagnosis were also not available.

Assessment of Education and Income Levels

Education and income levels were assessed by self-report questionnaires

(**Supplementary Table S1**). Highest education level was categorized according to the International Standard Classification of Education (ISCED): (1) low (level 0, 1, or 2); (2) middle (level 3 or 4); and (3) high (level 5 or 6).[18] Income level was based on monthly household net income and was categorized as (1) low (<1000 euro/month); (2) lower-middle (1000–2000 euro/month); (3) upper-middle (2000–3000 euro/month); (3) high (>3000 euro/month); and (4) do not know/prefer not to answer.

Clinical Measurements

Blood samples were collected by venipuncture in a fasting state and serum levels of glucose, HbA_{1c}, HDL-cholesterol, total cholesterol, and triglycerides were analyzed. Measurements of blood pressure, 12-lead electrocardiograms, and anthropometry were made by trained research staff following standardized protocols. These measurements were performed without shoes and heavy clothing. Body mass index (BMI) was calculated as weight in kilograms divided by the square of height in meters. Hypertension status was defined as (1) hypertensive medication use (ATC codes C02, C03, C07, C08, C09); (2) systolic blood pressure ≥ 140 mmHg; or (3) diastolic blood pressure ≥ 90 mmHg.[19]

Assessment of Other Baseline Covariates

Age, smoking status (never, former, and current), and TV watching time were assessed by self-administered questionnaires. Physical activity level was assessed by the validated Short QUestionnaire to ASsess Health-enhancing physical activity (SQUASH), from which non-occupational moderate-to-vigorous physical activities (MVPA) were calculated in minutes per week.[20] Dietary intake was assessed by a validated 110-item semi-quantitative self-administered food frequency questionnaire (FFQ).[21] Macro- and micro-nutrients intake was calculated from the FFQ data according to the 2011 Dutch Food Composition Table (NEVO).[22] The Lifelines Diet Score (LLDS) was calculated to evaluate the relative diet quality of each participant. The development of the LLDS has been described in detail elsewhere.[23]

Statistical Analysis

Associations of income and education with incident type 2 diabetes and cardiovascular diseases were estimated by logistic regression models, and results were shown as odds ratios (ORs) with 95% confidence intervals. For evaluation of the separate effects of education and income, these two socioeconomic indicators were singly and mutually adjusted in the models. An interaction term between education and income was also fit into the model to test possible effect modification. For evaluation of the combined effects of education and income, these two socioeconomic indicators were combined into twelve groups, e.g., a group of participants had high education and lower-middle income. The associations of these combined groups of education and income with incident type 2 diabetes and cardiovascular diseases were subsequently estimated. For all estimations, models were adjusted in a two-step manner: (1) basic model: age and sex; (2) multivariate model: age and sex from basic model, plus lifestyle behaviors (smoking status, TV watching time, non-occupational MVPA, total energy intake, LLDS, and alcohol intake), BMI, and clinical biomarkers (HDL-cholesterol, triglycerides, and blood pressure). For cardiovascular diseases, in the multivariate model, we additionally adjusted total cholesterol level and comorbid conditions at baseline (atrial fibrillation and diabetes). In all models, age was adjusted as a categorical variable, i.e., 30-39, 40-49, 50-59, and 60-65 years. Before estimation, values of HDL-cholesterol, triglycerides, BMI, total energy intake, total cholesterol, and blood pressure were log-transformed to improve normality. We also assessed the associations with adjustments for different domains of modifiable risk factors separately. Additionally, we determined the contribution of each modifiable risk factor in explaining the associations of income and education with incident type 2 diabetes and cardiovascular diseases by calculating the percentage of attenuation in the ORs after additional adjustment for another modifiable risk factor, in comparison to the previous reference model, namely $100\% \times (OR_{ref} - OR_{new})/OR_{ref}$.

Multiple imputation by chained equations was performed (creating 25 imputed datasets) to deal with missing data for income level (including both missing values and participants who responded "do not know" or "prefer not to answer"), LLDS, total

energy intake, alcohol intake, non-occupational MVPA, and smoking status.[24] These variables all had missing data more than 1%. All statistical analyses were conducted using Stata (version 13.1; StataCorp, College Station, TX).

Sensitivity Analysis

First, we repeated our analysis without imputation of income level for those who responded "do not know" or "prefer not to answer". Instead, we recoded them as a single category in the income variable. Second, we evaluated the potential effect modifications by sex, age, unemployment status, comorbid conditions at baseline (cancer and cardiovascular diseases), and diabetes status at baseline and during follow-ups (for cardiovascular diseases), by additionally including an interaction term with education or income in the model. Third, for cardiovascular diseases, we performed a separate analysis, in which we adjusted the SCORE2 risk prediction algorithms according to the European Society of Cardiology.[25] For type 2 diabetes, we additionally analyzed a composite outcome of incident type 2 diabetes and death from any cause, to gain insights into how death events during follow-ups may influence the results.

Results

Baseline characteristics are shown in **Table 1** and **Supplementary Table S2**. Approximately 28%, 40%, and 32% of participants reported having low, middle, and high education, respectively. For household net income, approximately 3%, 18%, 30%, and 33% of participants reported having low (<1000 euro), lower-middle (1000–2000 euro), upper-middle (2000–3000 euro), and high (>3000 euro) level of income per month, respectively; approximately 15% of participants did not disclose their income level. These numbers were comparable with the national-level data in The Netherlands, e.g., approximately 10% low-income households, and approximately 28% and 30% of the population had high and low education, respectively.[26,27] With increasing education level, participants tended to be younger and have higher income. In general, lifestyle behaviors, BMI, and clinical biomarkers were also socioeconomically patterned with more favorable conditions among people who had higher education level. Baseline

characteristics across income levels showed similar socioeconomic patterns.

Frequency measures of incidences of type 2 diabetes and cardiovascular diseases across education and income levels are shown in **Table 2**. Among 83,759 participants included in the type 2 diabetes cohort, we identified 1228 cases of type 2 diabetes (incidence 1.5%) during a median follow-up of 43 months. Among 91,083 participants included in the cardiovascular diseases cohort, we identified 3286 cases of cardiovascular diseases (non-fatal cardiovascular events) during a median follow-up of 44 months. Additionally, a total of 1127 deaths were recorded during the follow-up in the cardiovascular diseases cohort. With decreasing education or income levels, incidences of type 2 diabetes and cardiovascular diseases increased. **Supplementary Table S3** shows the frequency measures of incidences of type 2 diabetes and cardiovascular diseases among different combinations of education and income levels.

Separate associations of education and income with incident type 2 diabetes and cardiovascular diseases are shown in **Tables 3**, **Table 4** and **Supplementary Table S4**. Low education and low income were both positively associated with higher risks of type 2 diabetes and cardiovascular diseases after adjustment for age and sex (basic model). The mutual adjustment between education and income only moderately attenuated those associations. Additional adjustment for other covariates attenuated those associations as well. In the mutually adjusted multivariate model, participants with low education had 24% (OR 1.24 [95%CI 1.04-1.48]) and 15% (OR 1.15 [95%CI 1.04-1.28]) higher odds of incident type 2 diabetes and cardiovascular diseases, respectively; participants with low income had 71% (OR 1.71 [95%CI 1.30-2.26]) and 24% (OR 1.24 [95%CI 1.02-1.52]) higher odds of incident type 2 diabetes and cardiovascular diseases, respectively, using high education and high income as reference group as appropriate. Multiplicative interactive effects between education and income were absent: OR-interaction 1.01 (95%CI 0.91-1.12) and OR-interaction 0.97 (95%CI 0.91-1.04) in multivariate models for type 2 diabetes and cardiovascular diseases, respectively. For cardiovascular diseases, similar associations were observed for the secondary composite outcome, including both non-fatal cardiovascular event and death from any cause (**Supplementary Table S5**).

Joint associations of education and income with incident type 2 diabetes and cardiovascular diseases are shown in **Fig. 1** and **Supplementary Table S6**. In general, gradients of associations across education and income levels were observed after adjustment for age and sex. Further adjustment for other covariates substantially attenuated these associations. For cardiovascular diseases, gradients of associations were weakened after adjustments for these risk factors. In the multivariate model, participants who had high education and low income had the highest risks for incident type 2 diabetes (OR 3.04 [95%CI 1.52-6.05]) and cardiovascular diseases (OR 1.85 [95%CI 1.18-2.91]), followed by participants who had low education and low income, i.e., OR 2.24 [95%CI 1.54-3.25] for type 2 diabetes and OR 1.46 [95%CI 1.11-1.92] for cardiovascular diseases, using participants who had high education and high income as reference.

6

Percentages of attenuation in ORs across each education and income group are shown in **Supplementary Table S4**. When education and income were simultaneously controlled, adjustment for modifiable risk factors at baseline in total explained 33.1% and 15.2% of the associations of education with type 2 diabetes and cardiovascular diseases, respectively; for income, in total 23.5% and 7.7% of the associations were explained for type 2 diabetes and cardiovascular diseases, respectively. Adjustment for lifestyle behaviors explained more socioeconomic variations than other modifiable risk factors. Additional adjustments for clinical biomarkers and comorbid conditions at baseline (for cardiovascular diseases) showed no clear effects on explaining these socioeconomic variations.

Sensitivity analyses in general yielded similar results compared with the main analyses. **Supplementary Table S7** presents the results of the analysis by including the responses of "do not know" or "prefer not to answer" for income as a single category. Large variations in the risks of incident type 2 diabetes and cardiovascular diseases were found in this income group across education levels. Furthermore, no significant multiplicative interactive effects were found for sex, age, unemployment status, comorbid conditions at baseline, and diabetes status at baseline and during the follow-up (for cardiovascular diseases), with education and income (**Supplementary Table**

S8). For cardiovascular diseases, results were basically unchanged when adjusting for the SCORE2 risk prediction algorithms according to the European Society of Cardiology (**Supplementary Table S9**). For type 2 diabetes, using a composite outcome of incident type 2 diabetes and death from any cause yielded a smaller effect size for education (low education: OR 1.11 [95%CI 0.98-1.25]) but a stronger effect size for income (low income: OR 1.85 [95%CI 1.51-2.27]), compared with the main results (**Supplementary Table S10**).

Table 1. Baseline characteristics of study participants[a]		
	Type 2 diabetes cohort	**Cardiovascular diseases cohort**
Population	83,759	91,083
Cases	1228	3286
Incidence, %	1.5	3.6
Follow-up time, months		
Median	43	44
Interquartile	31-53	34-54
Range	13-123	13-131
Education, %		
Low	28.3	28.1
Middle	40.0	39.9
High	31.7	31.9
Household net income, %		
Low	3.1	3.2
Lower-middle	18.3	18.3
Upper-middle	30.1	29.6
High	33.2	33.3
No response or missing	15.4	15.6

Table 1. Continued.		
	Type 2 diabetes cohort	**Cardiovascular diseases cohort**
Age, years	46.2±8.8	46.1±8.8
Women, %	58.7	59.0
Lifelines diet score	24.0±5.9	24.0±5.9
Total energy intake, kcal/day	2081±604	2078±605
Total alcohol intake, grams/day	4.1 (0.9, 10.5)	4.0 (0.9, 10.4)
TV watching time, hours/day	2.4±1.3	2.4±1.3
Non-occupational MVPA, minutes/week	180 (60, 360)	180 (60, 360)
Smoking status, %:		
Never	45.0	44.9
Former	35.3	34.8
Current	18.5	18.6
BMI, kg/m^2	26.1±4.1	26.2±4.2
Fasting glucose, mmol/L	4.94±0.50	5.01±0.80
HbA$_{1c}$, %	5.52±0.30	5.55±0.42
Triglycerides, mmol/L	1.18±0.80	1.19±0.82
HDL-cholesterol, mmol/L	1.51±0.40	1.50±0.40
Total cholesterol, mmol/L	5.17±0.97	5.16±0.98
Hypertension, %	24.6	24.2
Systolic blood pressure, mmHg	125.3±14.8	125.2±14.9
Diastolic blood pressure, mmHg	74.5±9.3	74.4±9.4
Diabetes at baseline, %		3.2
Atrial fibrillation at baseline, %		0.6
[a] Data are expressed as unadjusted mean ± standard deviation for age, Lifelines diet score (no unit, ranging from 0 to 48), total energy intake, TV watching time, BMI, fasting glucose, HbA$_{1c}$, systolic blood pressure, diastolic blood pressure, triglycerides, HDL-cholesterol, and total cholesterol; Data are expressed as median (interquartile) for total alcohol intake and non-occupational MVPA; Data are expressed as observed percentage for education level, household net income level, sex (women), smoking status, hypertension, diabetes at baseline, and atrial fibrillation at baseline.		

6

Table 2. Frequency measures of incident type 2 diabetes (a) and cardiovascular diseases (b) across education and income levels

(a) Type 2 diabetes	Cases/population	Incidence, %	Risk difference, %[a]	Proportion of cases, %[b]
Education				
Low	548 / 23,679	2.3	1.4	44.6
Middle	439 / 33,527	1.3	0.4	35.7
High	241 / 26,553	0.9	Ref.	19.6
Income				
Low	74 / 2606	2.8	1.8	6.0
Lower-middle	297 / 15,288	1.9	0.9	24.2
Upper-middle	363 / 25,175	1.4	0.4	29.6
High	298 / 27,819	1.1	Ref.	24.3
No response / missing	196 / 12,871	1.5	0.4	16.0
Total	1228 / 83,759	1.5		100

[a] Risk difference was calculated by subtracting the incidence in the reference group from the incidence in the group of interests.
[b] Proportion of cases was calculated by dividing number of cases in the group of interests by total number of cases.

Table 2. Continued.

(b) Cardiovascular diseases	Cases/population	Incidence, %	Risk difference, %[a]	Proportion of cases, %[b]
Education				
Low	1170 / 25,636	4.6	1.7	35.6
Middle	1291 / 36,381	3.5	0.7	39.3
High	825 / 29,066	2.8	Ref.	25.1
Income				
Low	127 / 2928	4.3	1.2	3.9
Lower-middle	647 / 16,632	3.9	0.7	19.7
Upper-middle	996 / 26,999	3.7	0.5	30.3
High	959 / 30,323	3.2	Ref.	29.2
No response / missing	557 / 14,201	3.9	0.8	17.0
Total	3286 / 91,083	3.6		100

[a] Risk difference was calculated by subtracting the incidence in the reference group from the incidence in the group of interests.
[b] Proportion of cases was calculated by dividing number of cases in the group of interests by total number of cases.

Table 3. Separate associations of education and income with incident type 2 diabetes (a) and cardiovascular diseases (b)

(a) Type 2 diabetes	Basic model[a]		Multivariate model[b]	
	Singly adjusted[c]	Mutually adjusted[d]	Singly adjusted[c]	Mutually adjusted[d]
Education				
Low	2.17 (1.86-2.53)	1.85 (1.57-2.19)	1.33 (1.12-1.58)	1.24 (1.04-1.48)
Middle	1.48 (1.26-1.73)	1.37 (1.16-1.61)	1.11 (0.94-1.31)	1.07 (0.90-1.27)
High		1.00 (ref)		
Income				
Low	2.76 (2.13-3.58)	2.24 (1.71-2.92)	1.82 (1.38-2.39)	1.71 (1.30-2.26)
Lower-middle	1.76 (1.49-2.06)	1.44 (1.22-1.71)	1.30 (1.10-1.54)	1.23 (1.03-1.46)
Upper-middle	1.34 (1.15-1.56)	1.17 (1.00-1.37)	1.11 (0.95-1.29)	1.07 (0.91-1.25)
High		1.00 (ref)		

[a] Basic model: OR (95% CI) derived from multivariate logistic regression models adjusted for age and sex, n = 83,759 for education-singly adjusted model, and n = 83,381 for income-singly adjusted model and mutually adjusted model.
[b] Multivariate model: OR (95% CI) derived from multivariate logistic regression models adjusted for basic model covariates plus BMI, smoking status, TV watching time, non-occupational MVPA, total energy intake, LLDS, alcohol intake, HDL-cholesterol, triglycerides, and blood pressure, n = 82,908 for education-singly adjusted model, and n = 82,722 for income-singly adjusted model and mutually adjusted model.
[c] Singly adjusted: models were adjusted for education and income separately.
[d] Mutually adjusted: models were adjusted for education and income simultaneously.

Table 3. Continued.

(b) Cardiovascular diseases	Basic model[a]		Multivariate model[b]	
	Singly adjusted[c]	Mutually adjusted[d]	Singly adjusted[c]	Mutually adjusted[d]
Education				
Low	1.42 (1.29-1.55)	1.36 (1.23-1.50)	1.17 (1.06-1.30)	1.15 (1.04-1.28)
Middle	1.31 (1.19-1.43)	1.27 (1.16-1.40)	1.18 (1.07-1.29)	1.16 (1.06-1.28)
High		1.00 (ref)		
Income				
Low	1.50 (1.23-1.81)	1.35 (1.11-1.64)	1.29 (1.06-1.58)	1.24 (1.02-1.52)
Lower-middle	1.22 (1.10-1.35)	1.10 (0.99-1.23)	1.08 (0.97-1.21)	1.04 (0.93-1.17)
Upper-middle	1.15 (1.05-1.26)	1.07 (0.97-1.18)	1.06 (0.96-1.16)	1.02 (0.93-1.13)
High		1.00 (ref)		

[a] Basic model: OR (95% CI) derived from multivariate logistic regression models adjusted for age and sex, n = 91,083 for education-singly adjusted model, and n = 90,531 for income-singly adjusted model and mutually adjusted model.

[b] Multivariate model: OR (95% CI) derived from multivariate logistic regression models adjusted for basic model covariates plus BMI, smoking status, TV watching time, non-occupational MVPA, total energy intake, LLDS, alcohol intake, HDL-cholesterol, triglycerides, total cholesterol, blood pressure, diabetes at baseline, and atrial fibrillation at baseline, n = 89,473 for education-singly adjusted model, and n = 89,251 for income-singly adjusted model and mutually adjusted model.

[c] Singly adjusted: models were adjusted for education and income separately.

[d] Mutually adjusted: models were adjusted for education and income simultaneously.
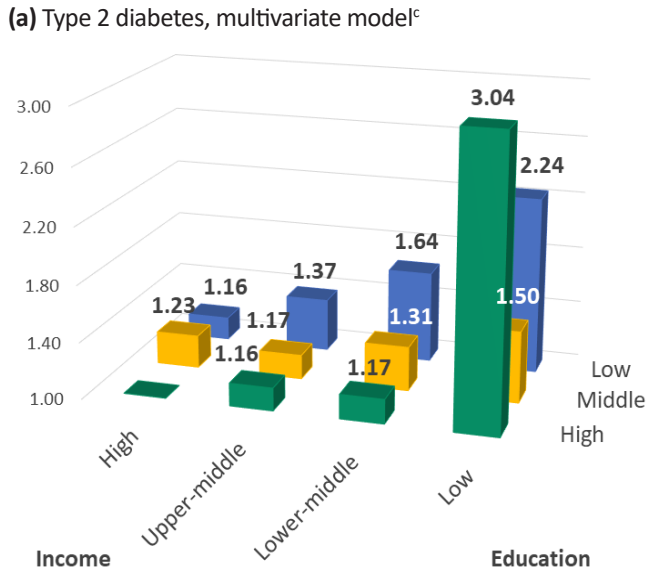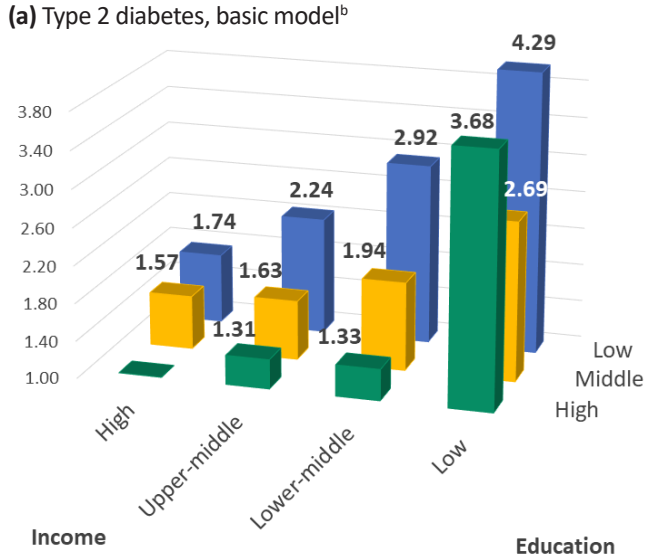
6

(a) Type 2 diabetes, basic model[b]

(a) Type 2 diabetes, multivariate model[c]

Fig. 1. Joint associations of education and income with incident type 2 diabetes (a) and cardiovascular diseases (b)[a]

[a] Figures are shown according to each education and income level, using high education and high income (>3000 euro/month) group as low risk reference (OR = 1.00).

[b] Basic model: OR derived from multivariate logistic regression models adjusted for age and sex, n = 83,381 and n =90,531 for type 2 diabetes and cardiovascular diseases, respectively.

(b) Cardiovascular diseases, basic model[b]

Income

Education
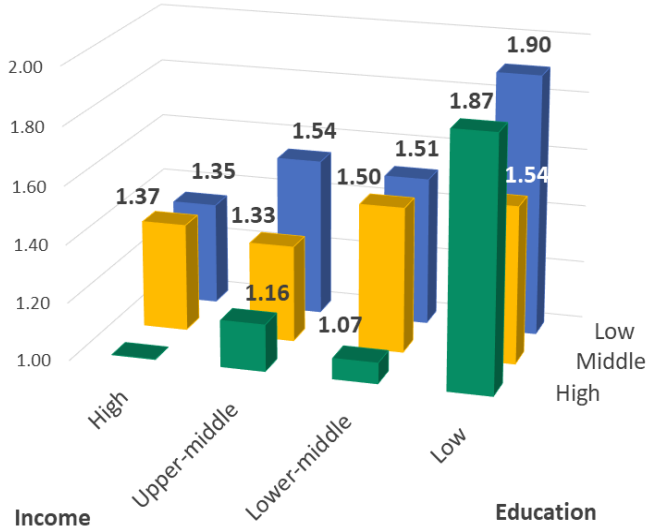
(b) Cardiovascular diseases, multivariate model[c]

Income

Education
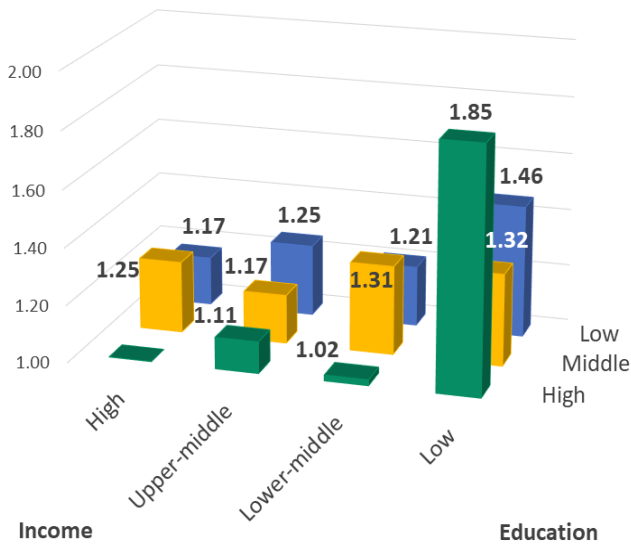
[c] Multivariate model: OR derived from multivariate logistic regression models adjusted for basic model covariates plus BMI, smoking status, TV watching time, non-occupational MVPA, total energy intake, LLDS, alcohol intake, HDL-cholesterol, triglycerides, blood pressure, n = 82,722 and n = 89,251 for type 2 diabetes and cardiovascular diseases, respectively; for cardiovascular diseases, total cholesterol, diabetes at baseline, atrial fibrillation at baseline were additionally adjusted in the multivariate model.

Discussion

The primary objective of this study was to investigate the effects of education and income simultaneously on incident type 2 diabetes and cardiovascular diseases. Specifically, this study was directed at assessing the short-term inequities in incident type 2 diabetes and cardiovascular diseases. Using this large population-based cohort sample, we found that low education and low income were both independently but also differentially associated with higher risks of these two health outcomes. In addition, results from the combined associations of education and income revealed substantial health disparities in these two health outcomes across education and income levels.

In general, our results are consistent with previous similar studies on prevalent type 2 diabetes in a German cohort and incident type 2 diabetes in a US community-based cohort.[7,28] Our results on cardiovascular diseases are also comparable with an Italian cohort.[29] Our analyses thus provide the very important additional evidence demonstrating the independent associations of education and income with incident type 2 diabetes and cardiovascular diseases in a European setting. With a broader perspective, we also found that our results were partly in line with studies conducted in different geographical and socioeconomic settings, despite the differences in study design and methodology that preclude direct comparisons. For example, a systematic review reported that low or middle education and low income were associated with higher risks of cardiovascular outcomes in US and European settings, while the effects of education were absent in Asian settings.[30] On the other hand, a global study found that low education was a strong predictor for cardiovascular diseases in all 20 countries analyzed, while wealth showed no or weak associations.[31] For type 2 diabetes, results seemed to be more consistent, as a meta-analysis found that both education and income were associated with a higher risk of developing type 2 diabetes, irrespective of different geographical settings.[32] In brief, our study further underlines the broader notion that it is important to consider and prioritize education and income as two indispensable socioeconomic dimensions when addressing health disparities, irrespective of geographical and socioeconomic settings.[4]

The independent associations of education and income with health highlight that these two socioeconomic indicators are of equal importance and should both be considered in health research. Our findings support the hypothesis suggesting that education and income may impact on health through different causal processes, as they provide different dimensions of resources in relation to health.[1,3,6,7,28,33-37] More specifically, education determines one's non-material resources such as knowledge, skills, and self-efficacy that help individuals ease their barriers to be more receptive to health messages and transfer those messages into health behaviors. Such improvements in cognitive functioning associated with higher education level were argued to be the major driver in delaying the onset of non-communicable diseases. On the other hand, income reflects one's material resources in regard to health, such as healthy food, health services, and leisure time activities.[6,38-40] In line with these theoretical assumptions, we did observe that lifestyle behaviors explained a considerable proportion of the associations for both socioeconomic indicators.

Results of the combined associations of education and income indicated that their effects on health were likely to be additive. As we observed within each education or income level, substantial health disparities existed across strata of the other socioeconomic indicator. More specifically, we showed that within different education groups, a higher income was associated with a better health; similarly, within different income groups, a higher education was associated with a better health. Differences in modifiable risk factors did not fully annul these excessive risks. We further illustrate this with status inconsistency, that is, people having discrepant socioeconomic positions in two or more of these ranking indicators. For example, we observed that participants who had high income but low education were worse off regarding their health outcomes, compared with those who had a matching socioeconomic position (i.e., high income and high education). And such status inconsistency-related health disparities were prevalent across almost every education and income level in our study sample. Previous studies have shown that status inconsistency between education and occupational class carried higher health risks.[12,13,41] Our findings thus provide further support of this in the dimensions of education and income. It should be noted that

we observed some non-linear associations especially after adjustment for modifiable risk factors. As these non-linear associations appeared gradually with the stepwise adjustments, we were unable to clearly specify the causes of this counterintuitive finding.

In our study sample, approximately 5% of the study population had extreme status inconsistency; not surprisingly, education and income were only moderately correlated (Spearman's correlation coefficient = 0.34, **Supplementary Table S3**). We therefore further emphasize the importance and necessity of considering both education and income. Especially in public health programs (such as healthy eating campaigns or diabetes screening) where the effectiveness and outreach are often compromised among people who are socioeconomically disadvantaged, additional attention and support should be given to those not only having low income, but also having low education.[42] Additionally, further understanding of such within-group differences may lead the way towards the design of policies that do not require the adjustment of socioeconomic characteristics that are generally fixed such as education. Indeed, in our case, income supports individuals who had low education, which may contribute to their health even after their education level has been attained.

Of all modifiable risk factors examined, lifestyle behaviors and BMI contributed the most to the socioeconomic gradients, while additional adjustment for clinical biomarkers did not further explain those health disparities. The higher risk for cardiovascular diseases conferred by poor socioeconomic status also appeared to be independent of diabetes status. It is noteworthy that after accounting for all modifiable risk factors, a large proportion of the risks for type 2 diabetes and cardiovascular diseases were still left unexplained. Previous studies have found that the most socioeconomically deprived individuals had disproportionately higher risks for type 2 diabetes and cardiovascular diseases, even if they practiced the healthiest lifestyle.[2,43] These results indicate that even though people who have low socioeconomic status may benefit from lifestyle interventions and obesity control, their excessively higher risks of developing type 2 diabetes and cardiovascular diseases may still be preserved. Particularly, as our study was based in the context of The Netherlands, a developed country with a high

coverage of government-subsidized public education system and well-structured social security system, the persistent socioeconomic patterning in health inequities observed may not be addressed only by extensive public health interventions, but also through institutional and structural changes with support in all socioeconomic dimensions simultaneously.[6]

Strengths of this study include the large sample size, which allows the investigation of joint associations of education and income with sufficient statistical power. Secondly, we also conducted sensitivity analyses supporting the robustness of our findings. On the other hand, several limitations should be noted. Because of the intrinsic limitation of the Lifelines questionnaire, we are unable to translate household net income level into individual equivalent disposable income level. Since the Lifelines cohort study was established in The Netherlands, a country with a well-developed welfare system, it may not be possible to extrapolate our results to other population groups in another setting. We are also unable to assess the possible changes in participants' education and income level. However, education is considered to be very stable over the entire adult life. Similarly, income in The Netherlands is also relatively stable because of the organization of the Dutch labor market (e.g., wide-spread use of collective wage bargaining as well as generous unemployment insurance). We therefore do not expect dramatic changes in participants' socioeconomic status during follow-up.[44] Another limitation is that the resolution in time, regarding the time of diagnosis of type 2 diabetes and cardiovascular diseases, was limited in the Lifelines dataset, hence limiting the suitability of the data for survival analysis. Nevertheless, considering the low event rate, moderate effect sizes, and the relatively short follow-up time, logistic regression models may provide similar estimates for the effect sizes. We therefore used logistic regression models instead.[45,46] Furthermore, misclassification could occur in the ascertainment of cases of type 2 diabetes and cardiovascular diseases, as at T2 and T3 only self-reported data was available. Data on participants' medical records and causes of death were not available in the Lifelines study. Natriuretic peptide measurements, echocardiography, and coronary imaging were not performed in the Lifelines study. For type 2 diabetes, however, we consider this lack of medical

6

records is not expected to substantially influence our results, since at T4 most new cases were identified by objective laboratory measurements, which is a strength of our study. For cardiovascular diseases, we cross-checked new self-reported cases at T4 with electrocardiographic results. Finally, approximately 18% of the study population was excluded because of loss of follow-up. However, we do not expect this attrition to substantially influence our results. We did not observe substantial differences in the baseline characteristics between the included participants and those who had no follow-up data (**Supplementary Table S11**), although there seemed to be fewer participants having high education or high income among those who had no follow-up information; participants who had no follow-up data also appeared to smoke more. A simulation study found that loss to follow-up (<50%) may lead to minor underestimation on the estimates of socioeconomic inequities in cohort studies.[47] This suggests if full information was available, our estimation would be even more pronounced, despite the clear gradients of associations that have already been revealed in our results.

Conclusions

In conclusion, our results showed that education and income were both independently and also differentially associated with incident type 2 diabetes and cardiovascular diseases. Additionally, by analyzing the effects of education and income using a combined indicator, substantial health disparities were observed within socioeconomic groups. These findings suggest that education and income are two equally indispensable socioeconomic indicators in health, and should both be considered in health research and policymaking.

Reference

1.   Bonaccio M, Di Castelnuovo A, Costanzo S, et al. Interaction between education and income on the risk of all-cause mortality: prospective results from the MOLI-SANI study. Int J Public Health 2016; 61(7): 765-76.

2.   Vinke PC, Navis G, Kromhout D, Corpeleijn E. Socio-economic disparities in the association of diet quality and type 2 diabetes incidence in the Dutch Lifelines cohort. EClinicalMedicine 2020; 19: 100252.

3.   Blanquet M, Legrand A, Pelissier A, Mourgues C. Socio-economics status and metabolic syndrome: A meta-analysis. Diabetes Metab Syndr 2019; 13(3): 1805-12.

4.   Schultz WM, Kelli HM, Lisko JC, et al. Socioeconomic status and cardiovascular outcomes: challenges and interventions. Circulation 2018; 137(20): 2166-78.

5.   Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health Aff (Millwood) 2002; 21(2): 60-76.

6.   Herd P, Goesling B, House JS. Socioeconomic position and health: the differential effects of education versus income on the onset versus progression of health problems. J Health Soc Behav 2007; 48(3): 223-38.

7.   Geyer S, Hemström Ö, Peter R, Vågerö D. Education, income, and occupational class cannot be used interchangeably in social epidemiology. Empirical evidence against a common practice. J Epidemiol Community Health 2006; 60(9): 804-10.

8.   Sharker MAY, Nasser M, Abedin J, Arnold BF, Luby SP. The risk of misclassifying subjects within principal component based asset index. Emerg Themes Epidemiol 2014; 11(1): 6.

9.   Zhu Y, Duan MJ, Riphagen IJ, et al. Separate and combined effects of individual and neighbourhood socio-economic disadvantage on health-related lifestyle risk factors: a multilevel analysis. Int J Epidemiol 2022; 50(6): 1959-69.

10.   Ferrer RL, Palmer R. Variations in health status within and between socioeconomic strata. J Epidemiol Community Health 2004; 58(5): 381-7.

11.   de Boer WIJ, Dekker LH, Koning RH, Navis GJ, Mierau JO. How are lifestyle factors associated with socioeconomic differences in health care costs? Evidence from full population data in the Netherlands. Prev Med 2020; 130: 105929.

12.   Honjo K, Iso H, Inoue M, Sawada N, Tsugane S. Socioeconomic status inconsistency and risk of stroke among Japanese middle-aged women. Stroke 2014; 45(9): 2592-8.

13.   Braig S, Peter R, Nagel G, Hermann S, Rohrmann S, Linseisen J. The impact of social status inconsistency on cardiovascular risk factors, myocardial infarction and stroke in the EPIC-Heidelberg cohort. BMC Public Health 2011; 11: 104.

14.   Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol 2014; 44(4): 1172-80.

15.   Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. PLOS ONE 2015; 10(9): e0137203.

6

16.   WHO, International Diabetes Federation. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: report of a WHO/IDF consultation. Genève, Switzerland: WHO, 2006.

17.   van der Ende MY, Hartman MH, Hagemeijer Y, et al. The Lifelines Cohort Study: prevalence and treatment of cardiovascular disease and risk factors. Int J Cardiol 2017; 228: 495-500.

18.   UNESCO. International Standard Classification of Education. 1997.

19.   Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension. J Hypertens 2018; 36(10): 1953-2041.

20.   Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003; 56(12): 1163-9.

21.   Streppel MT, de Vries JH, Meijboom S, et al. Relative validity of the food frequency questionnaire used to assess dietary intake in the Leiden Longevity Study. Nutr J 2013; 12: 75.

22.   RIVM. Dutch food composition database (NEVO). Bilthoven, The Netherlands; 2006.

23.   Vinke PC, Corpeleijn E, Dekker LH, Jacobs DR, Jr., Navis G, Kromhout D. Development of the food-based Lifelines Diet Score (LLDS) and its application in 129,369 Lifelines participants. Eur J Clin Nutr 2018; 72(8): 1111-9.

24.   van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999; 18(6): 681-94.

25.   SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J 2021; 42(25): 2439-54.

26.   Statistics Netherlands (CBS). Trends in The Netherlands 2017. https://longreads.cbs.nl/trends17-eng/labour_and_income/figures/income_and_expenditure/ (accessed Feb 10th 2022).

27.   Statistics Netherlands (CBS). Dutch population better educated. 2013. https://www.cbs.nl/en-gb/news/2013/40/dutch-population-better-educated (accessed Feb 10th 2022).

28.   Maty SC, Everson-Rose SA, Haan MN, Raghunathan TE, Kaplan GA. Education, income, occupation, and the 34-year incidence (1965-99) of Type 2 diabetes in the Alameda County Study. Int J Epidemiol 2005; 34(6): 1274-81.

29.   Petrelli A, Gnavi R, Marinacci C, Costa G. Socioeconomic inequalities in coronary heart disease in Italy: a multilevel population-based study. Soc Sci Med 2006; 63(2): 446-56.

30.   Khaing W, Vallibhakara SA, Attia J, McEvoy M, Thakkinstian A. Effects of education and income on cardiovascular outcomes: A systematic review and meta-analysis. Eur J Prev Cardiol. 2017; 24(10): 1032-42.

31.   Rosengren A, Smyth A, Rangarajan S, et al. Socioeconomic status and risk of cardiovascular disease in 20 low-income, middle-income, and high-income countries: the Prospective Urban Rural Epidemiologic (PURE) study. Lancet Glob Health 2019; 7(6): e748-e60.

32.   Agardh E, Allebeck P, Hallqvist J, Moradi T, Sidorchuk A. Type 2 diabetes incidence and

socio-economic position: a systematic review and meta-analysis. Int J Epidemiol 2011; 40(3): 804-18.

33.   Fitzpatrick T, Rosella LC, Calzavara A, et al. Looking beyond income and education: socioeconomic status gradients among future high-cost users of health care. Am J Prev Med 2015; 49(2): 161-71.

34.   Leinonen T, Martikainen P, Lahelma E. Interrelationships between education, occupational social class, and income as determinants of disability retirement. Scand J Public Health 2012; 40(2): 157-66.

35.   Veenstra G, Vanzella-Yang A. Does household income mediate the association between education and health in Canada? Scand J Public Health 2020; 49(8): 857-64.

36.   Allen K, McFarland M. How are income and education related to the prevention and management of diabetes? J Aging Health 2020; 32(9): 1063-74.

37.   Hoveling LA, Liefbroer AC, Bültmann U, Smidt N. Understanding socioeconomic differences in incident metabolic syndrome among adults: What is the mediating role of health behaviours? Prev Med 2021; 148: 106537.

38.   Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 1). J Epidemiol Community Health 2006; 60(1): 7-12.

39.   Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 2). J Epidemiol Community Health 2006; 60(2): 95-101.

40.   Mirowsky J, Ross CE. Education, health, and the default American lifestyle. J Health Soc Behav 2015; 56(3): 297-306.

41.   Faresjö T, Svärdsudd K, Tibblin G. The concept of status incongruence revisited: a 22-year follow-up of mortality for middle-aged men. Scand J Soc Med 1997; 25(1): 28-32.

42.   Kivela J, Wikstrom K, Virtanen E, et al. Obtaining evidence base for the development of Feel4Diabetes intervention to prevent type 2 diabetes- a narrative literature review. BMC Endocr Disord 2020; 20(Suppl 1): 140.

43.   Foster HME, Celis-Morales CA, Nicholl BI, et al. The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: a prospective analysis of the UK Biobank cohort. Lancet Public Health 2018; 3(12): e576-e85.

44.   de Boer WIJ, Mierau JO, Schoemaker J, Viluma L, Koning RH, Lifelines Corona Research I. The impact of the Covid-19 crisis on socioeconomic differences in physical activity behavior: Evidence from the Lifelines COVID-19 cohort study. Prev Med 2021; 153: 106823.

45.   Annesi I, Moreau T, Lellouch J. Efficiency of the logistic regression and Cox proportional hazards models in longitudinal studies. Stat Med 1989; 8(12): 1515-21.

46.   Green MS, Symons MJ. A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. J Chronic Dis 1983; 36(10): 715-23.

47.   Howe LD, Tilling K, Galobardes B, Lawlor DA. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. Epidemiology 2013; 24(1): 1-9.

6

Additional Information

Contributions

MJD conceptualized the study. MJD, YZ, LHD, and GN designed the study. MJD analyzed the data and drafted the manuscript. YZ organized the supplementary data. YZ, LHD, JOM, EC, SJLB, and GN contributed to the discussion and critically reviewed/edited the manuscript. LHD and GN supervised the project. MJD and YZ have primary responsibility for the final content. All authors approved the final content of the manuscript.

Supplementary Data

Online **Supplementary Files** can be accessed via
doi.org/10.1007/s11606-022-07548-8, or by scanning the QR code.
Article PMID: 35419742.

6

Separate and combined effects of individual and neighborhood socioeconomic disadvantage on health-related lifestyle risk factors: a multilevel analysis

1. Department of Internal Medicine, University Medical Center Groningen, Groningen, The Netherlands

2. Department of Laboratory Medicine, University Medical Center Groningen, Groningen, The Netherlands

3. Faculty of Economics and Business, University of Groningen, The Netherlands

4. Aletta Jacobs School of Public Health, University of Groningen, Groningen, The Netherlands

5. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

*Joint first authors

CHAPTER 7

Yinjie Zhu[1*]; Ming-Jie Duan[1*]; Ineke J. Riphagen[2];
Isidor Minovic[2]; Jochen O. Mierau[3,4]; Juan-Jesus
Carrero[5]; Stephan J.L. Bakker[1]; Gerjan Navis[1];
Louise H. Dekker[1,4]

ABSTRACT

Background

Socioeconomic disadvantage at both individual and neighborhood levels has been found to be associated with single lifestyle risk factors. However, it is unknown to what extent their combined effects contribute to a broad lifestyle profile. We aimed to (i) investigate the associations of individual socioeconomic disadvantage (ISED) and neighborhood socioeconomic disadvantage (NSED) in relation to an extended score of health-related lifestyle risk factors (i.e., lifestyle risk index); and to (ii) investigate whether NSED modified the association between ISED and the lifestyle risk index.

Methods

Of 77,244 participants [median age (interquartile): 46 (40-53) years] from the Lifelines cohort study in the north of The Netherlands, we calculated a lifestyle risk index by scoring the lifestyle risk factors including smoking status, alcohol consumption, diet quality, physical activity, TV-watching time, and sleep time. A higher lifestyle risk index was indicative of an unhealthier lifestyle. Composite scores of ISED and NSED based on a variety of socioeconomic indicators were calculated separately. Linear mixed-effect models were used to examine the association of ISED and NSED with the lifestyle risk index and to investigate whether NSED modified the association between ISED and the lifestyle risk index by including an interaction term between ISED and NSED.

Results

Both ISED and NSED were associated with an unhealthier lifestyle, because ISED and NSED were both positively associated with the lifestyle risk index, i.e., ISED beta-coefficient quartile 4 (Q4) versus Q1 0.64 (95%CI 0.62-0.66), $P < 0.001$, and NSED beta-coefficient quintile 5 (Q5) versus Q1 0.17 (95%CI 0.14-0.21), $P < 0.001$, after adjustment for age, sex, and body mass index (BMI). In addition, a positive interaction was found between NSED and ISED on the lifestyle risk index (beta-coefficient 0.016 [95% CI 0.011-0.021], P-interaction < 0.001), which indicated that NSED modified the association between ISED and the lifestyle risk index, i.e., the gradient of the

associations across all ISED quartiles (Q4 versus Q1) was steeper among participants residing in the most disadvantaged neighborhoods compared with those who resided in the less disadvantaged neighborhoods.

Conclusions

Our findings suggest that public health initiatives addressing lifestyle-related socioeconomic health differences should not only target individuals, but also consider neighborhood factors.

7

Abbreviations

BMI – Body mass index

ISED – Individual socioeconomic disadvantage

LLDS – Lifeline diet score

MVPA – Moderate-to-vigorous physical activity

NSED – Neighborhood socioeconomic disadvantage

PCA – Principal component analysis

Introduction

Lifestyle risk factors are key to the prevention of non-communicable diseases. Abundant epidemiological studies have demonstrated that socioeconomic differences bear a considerable impact on lifestyle risk factors;[1,2] i.e., individuals who are more socioeconomically disadvantaged are more likely to have an unhealthy lifestyle (e.g., poor diet, smoking, less physical activity).[3-7] However, variations within individual socioeconomic strata remain. Meanwhile, studies have also suggested that neighborhood socioeconomic disadvantage (NSED), as an important contextual factor, has an independent effect on individual-level lifestyle risk factors.[8-19] More insights into the socioeconomic disadvantage from different ecological levels are needed at the same time to better understand the mechanisms behind socioeconomically patterned lifestyle and health inequalities.

Studies on both smoking and drinking habits have suggested an interaction between individual socioeconomic disadvantage (ISED) and NSED,[4,19] showing that NSED had disproportionate effects across different ISED strata on lifestyle behaviors. More precisely, the impact of NSED has been found to be greater for those who were more socioeconomically disadvantaged.[20,21] It has been suggested that less socioeconomically disadvantaged individuals may be protected by their individual resources from NSED, whereas more socioeconomically disadvantaged individuals may be more dependent on neighborhood resources.[22] However, those previous studies have only examined single and traditional lifestyle risk factors; whereas a broader range of a combination of lifestyle factors, including emerging lifestyle factors, has rarely been studied in this context for their relationships with the combined effects of ISED and NSED.[23-25]

To our knowledge, it is still not clear whether NSED modifies the effect of ISED on a broader lifestyle risk profile. Therefore, this study aimed to investigate (i) the separate and combined effects of ISED and NSED on a combination of health-related lifestyle risk factors (i.e., lifestyle risk index); and (ii) whether NSED modifies the association between ISED and the lifestyle risk index.

Methods

Study Design and Participants

The Lifelines cohort study is a multidisciplinary prospective population-based cohort study that applies in a unique three-generation design, studying the health and health-related behaviors of 167,729 persons living in The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, sociodemographic, behavioral, physical, and psychological factors that contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics. Before study entry, a signed informed consent form was obtained from each participant. Adult participants (≥18 years old) were asked to complete several self-administered questionnaires regarding various aspects, including demographics, socioeconomic status, and lifestyle. A detailed description of the Lifelines cohort study can be found elsewhere.[26,27] For the current study, 77,244 participants from the Lifelines cohort study aged between 31 and 69 years who had available and reliable data on demographics, NSED, ISED, and lifestyle were included in the analysis (**Supplementary Fig. S1**). The Lifelines study is conducted according to the principles of the Declaration of Helsinki and approved by the Medical Ethics Committee of the University Medical Center Groningen, The Netherlands.

NSED and ISED

A neighborhood socioeconomic disadvantage (NSED) score was derived from principal component analysis (PCA) to summarize three NSED indicators. These indicators included percentage of the population with the highest 20% income, percentage of the population with the lowest 20% income, and percentage of the population receiving social benefits. NSED data were obtained from the Neighborhood Statistics (year 2011) from the Statistics Netherlands (CBS), which is in accordance with the Lifelines baseline assessment. Neighborhoods with <10 inhabitants were excluded and each neighborhood was identified by a unique neighborhood code. The first component from PCA was selected to form the NSED score (**Supplementary Description**). The derived NSED score was subsequently divided into quintiles, with higher quintiles

7

indicating more disadvantaged neighborhoods.

An individual socioeconomic disadvantage (ISED) score was calculated using factor analysis of mixed data (FAMD) to summarize four ISED variables at baseline: education, income, status of social benefits, and unemployment status. Since information on education and income was not available for all participants (education 0.31% and income 14.7%), multiple imputation was conducted with FAMD analysis (**Supplementary Description**). The highest education level achieved was categorized as: (1) low – junior general secondary education or lower (International Standard Classification of Education [ISCED] level 0, 1 or 2); (2) middle – secondary vocational education and senior general secondary education (ISCED level 3 or 4); and (3) high – higher vocational education or university (ISCED level 5 or 6).[28] Income level was categorized as: (1) <1000 euro/month; (2) 1000-2000 euro/month; (3) 2000-3000 euro/month; and (4) >3000 euro/month. Welfare and unemployment status were both binary variables obtained from questions "I am on national assistance benefit" and "I am unemployed/looking for a job", respectively. The ISED score was subsequently categorized into quartiles, with higher quartiles indicating more disadvantaged individuals.

Lifestyle Risk Index and Demographics

Six lifestyle factors (i.e., smoking status, alcohol consumption, diet quality, physical activity, TV-watching time, and sleep time) were selected to form the lifestyle risk index. Smoking status was categorized into never, former, and current smoker. Alcohol intake and dietary consumption were derived from an externally validated 110-item semi-quantitative food-frequency questionnaire (FFQ) that assessed food consumption over the past month.[29] Heavy drinking was defined as >40 or >20 g per day alcohol consumption for men and women, respectively.[30] The Lifelines Diet Score (LLDS) was calculated to assess the overall diet quality. This score ranks the relative intake of nine food groups with positive health effects (vegetables, fruit, whole-grain products, legumes/nuts, fish, oils/soft margarines, unsweetened dairy, coffee, and tea) and three food groups with negative health effects (red/processed meat, butter/hard margarines, and sugar-sweetened beverages). The development of this score is described in detail

elsewhere.[31] Non-occupational moderate-to-vigorous physical activity (MVPA) was calculated in minutes per week from the validated Short QUestionnaire to ASsess Health-enhancing physical activity (SQUASH) data, which incorporated leisure-time and commuting physical activities, including sports, at moderate (4.0-6.4 metabolic equivalent of task [MET]) to vigorous (≥6.5 MET) intensity.[32] TV-watching time and sleep time were recorded in hours per day.

The lifestyle risk index was based on former publications from the 45 and Up Study[23] and the UK Biobank cohort.[25] Each lifestyle factor was categorized into a dichotomized variable (point 0 indicated healthy and point 1 indicated unhealthy). Participants were assigned one point for each unhealthy lifestyle factor (current smoker, heavy drinker, lowest two quintiles of LLDS, <75 min/week of vigorous physical activity or <150 min/week of moderate physical activity or less than the equivalent combination of MVPA, ≥4 h/day of TV-watching, and <7 or >9 h of sleep time per day). Points were summed to create an unweighted index ranging from 0 to 6 for each participant, for which a higher index indicated an unhealthier lifestyle. In sensitivity analyses, the lifestyle risk index was further classified into three categories: participants who scored 0 or 1 were classified as having the least unhealthy lifestyle; and those who scored 2 or 3 were classified as having a moderately unhealthy lifestyle; and those who scored 4, 5 or 6 were classified as having the most unhealthy lifestyle (**Supplementary Table S1**).

Statistical Analysis

Nominal variables are presented as frequencies or percentage (%). Continuous variables are shown as mean ± standard deviation or median plus interquartile range.

We analyzed the associations of ISED and NSED with the lifestyle risk index using linear mixed-effect models. Each neighborhood was treated as a single unit in our study (the median number of participants per neighborhood was 101 [interquartile 39-213]) and the corresponding neighborhood code was treated as a random intercept in all linear mixed-effect models. First, we investigated the associations of ISED or NSED in relation to the lifestyle risk index (0-6, ordinal variable). ISED and NSED were first entered into the model separately (model 1) and then combined and adjusted for potential

confounders (model 2 – model 1 plus age and sex; model 3 – model 2 plus BMI). Second, we investigated whether NSED modified the association between ISED and the lifestyle risk index (model 4). Interactions between ISED and NSED on the lifestyle risk index were tested by treating ISED and NSED as continuous variables, and by fitting an interaction term between the two variables (i.e., ISED by NSED). We further stratified our analyses with participants in the least socioeconomically disadvantaged quartile and who resided in the least socioeconomically disadvantaged neighborhoods as the reference group. When an interaction was observed, additional linear-regression analyses were performed stratified by NSED and ISED, respectively.

Sensitivity analyses included models with single measures of ISED (education or income). Additional sensitivity analyses included treating the lifestyle risk index as a categorical variable and using those six single lifestyle factors from the lifestyle risk index as the outcome separately. Sensitivity analyses with additional adjustment for neighborhood-level education (percentage of participants with low education) collected from the Lifelines cohort was also conducted because the neighborhood-level education information was unavailable in the CBS Neighborhood Statistics. All statistical analyses were conducted using Stata (version 13.1, StataCorp, College Station, TX, USA) or RStudio version 3.5.2 (version 3.5.2, RStudio, PBC, Boston, MA, USA).

Results

Of the 77,244 participants included in this study, 49,879 (64.6%) had the least unhealthy lifestyle (0 or 1 unhealthy lifestyle factor), 24,604 (31.9%) had a moderately unhealthy lifestyle (2 or 3 unhealthy lifestyle factors), whereas only 2760 (3.6%) had the most unhealthy lifestyle (4, 5, or 6 unhealthy lifestyle factors) (**Supplementary Table S1**). With increasing ISED quartiles, participants were more likely to have a higher lifestyle risk index (**Supplementary Fig. S2**), have a higher BMI, be female, and be older (**Table 1**). Moreover, the least socioeconomically disadvantaged individuals were more likely to reside in the least disadvantaged neighborhoods (**Table 1**), although the correlation coefficient was weak between ISED and NSED (r = 0.19, P < 0.001, **Supplementary**

Table S2).

Multilevel modeling results are shown in **Table 2**. ISED (beta-coefficient Q4 versus Q1 0.58 [95% CI 0.56-0.60], P < 0.001) and NSED (beta-coefficient Q5 versus Q1 0.32 [95% CI 0.28-0.36], P < 0.001) were positively associated with the lifestyle risk index (model 1, **Table 2**). According to the linear mixed-effect models, the magnitude of the associations (beta-coefficient) for participants who were in Q4 and Q2 of ISED were 0.64 (95% CI 0.62-0.66, P < 0.001) and 0.27 (95% CI 0.25-0.30, P < 0.001), compared with the reference Q1 ISED group, respectively. A positive interaction was found between NSED and ISED on the lifestyle risk index (beta-coefficient 0.016 [95% CI 0.011-0.021], P-interaction < 0.001) (**Table 2**); and the association between ISED and the lifestyle risk index was steeper for those who resided in a more disadvantaged neighborhood (**Fig. 1**). Because of the positive interaction between ISED and NSED, analyses were repeated and stratified by NSED quintiles (**Table 3**). The results showed that the strength of the adjusted associations between ISED and the lifestyle risk index was stronger for the most disadvantaged neighborhood quintile (Q5). In this quintile of NSED (Q5), the estimated beta-coefficient was 0.81 (95% CI 0.76-0.87, P < 0.001, model 2) for those who were the most individually socioeconomically disadvantaged, which was higher compared with individuals who were less individually socioeconomically disadvantaged. In the least disadvantaged neighborhoods, the association magnitude was 0.58 (95% CI 0.54-0.63, P < 0.001, model 2) for participants who were in the highest ISED quartile (**Table 3**), compared with those in the lowest ISED quartile. Additional adjustment for BMI (model 3) only slightly attenuated the associations at all ISED or NSED levels. When treating participants in the lowest ISED quartile as well as the lowest NSED quintile as the reference group, the likelihood of having a higher lifestyle risk index was higher across all NSED levels among participants who were the most individually socioeconomically disadvantaged, compared with those who were the least socioeconomically disadvantaged (**Fig. 1**). Furthermore, the gradient of the association across ISED levels (Q4 versus Q1) was larger for participants who resided in the most disadvantaged neighborhood compared with those residing in the least disadvantaged neighborhood (**Fig. 1** and **Supplementary Fig. S3**).

Table 1. Characteristics of individuals at different individual socioeconomic disadvantage (ISED) levels

	Total	Individual socioeconomic disadvantage			
		Q1 (least disadvantaged)	Q2	Q3	Q4 (most disadvantaged)
	n = 77,244	n = 19,854	n = 14,501	n = 19,880	n = 23,009
Sex, men %	41.4	48.0	40.7	38.4	38.7
Age, years (interquartile)	46 (40-53)	44 (38-51)	46 (40-51)	45 (39-50)	49 (43-58)
Neighborhood socioeconomic disadvantage, n (%)					
Q1 (least disadvantaged)	16,358 (21.2)	6299 (31.7)	3539 (24.4)	3463 (17.4)	3057 (13.3)
Q2	15,829 (20.5)	4271 (21.5)	3238 (22.3)	4068 (20.5)	4252 (18.4)
Q3	15,773 (20.4)	3582 (18.1)	2799 (19.3)	4473 (22.5)	4919 (21.4)
Q4	15,298 (19.8)	3095 (15.6)	2697 (18.6)	4217 (21.2)	5289 (23.0)
Q5 (most disadvantaged)	13,986 (18.1)	2607 (13.1)	2228 (15.4)	3659 (18.4)	5492 (23.9)
BMI, kg/m², n (%)	26.4±4.3	25.5±3.8	26.1±4.0	26.5±4.4	27.1±4.6
Underweight	381 (0.5)	101 (0.5)	53 (0.4)	96 (0.5)	131 (0.6)
Normal weight	31,570 (40.9)	9813 (49.4)	6055 (41.8)	7960 (40.0)	7742 (33.7)
Overweight	32,104 (41.6)	7751 (39.1)	6258 (43.1)	8178 (41.1)	9917 (43.1)
Obese	13,163 (17.0)	2182 (11.0)	2129 (14.7)	3645 (18.4)	5207 (22.6)
Education, %[a]					
High	30.1	99.9	17.1	0	4.1
Middle	39.3	0	59.2	99.6	8.5
Low	30.3	0	23.0	0	87.3
Income (euro/month), %[a]					
<1000	3.5	0	0	0	11.6
1000-2000	19.7	0	17.1	27.9	31.3
2000-3000	30.1	27.7	0.3	50.7	33.2
>3000	32.0	62.7	82.2	0	1.6

Table 1. Continued.

	Total	Individual socioeconomic disadvantage			
		Q1 (least disadvantaged)	Q2	Q3	Q4 (most disadvantaged)
	n = 77,244	*n* = 19,854	*n* = 14,501	*n* = 19,880	*n* = 23,009
Receiving welfare, %	1.2	0	0	0	4.2
Unemployed, %	3.6	0	0	0	12.2
Lifestyle behaviors					
Lifelines diet score	24.3±6.0	25.4±5.7	24.4±5.9	23.9±5.8	23.8±6.1
Alcohol intake, g/day	7.2±9.0	7.8±8.4	7.8±9.0	6.5±8.5	7.0±9.8
Heavy drinker, n (%)	2032 (2.6)	432 (2.2)	437 (3.0)	439 (2.2)	724 (3.2)
TV-watching time, hours/day	2.5±1.3	2.1±1.0	2.5±1.1	2.6±1.2	3.1±1.4
≥ 4 hours/day, n (%)	14,347 (18.6)	1496 (7.5)	2054 (14.1)	3503 (17.6)	7294 (31.7)
Sleep time, hours/day	7.4±0.9	7.4±0.8	7.4±0.8	7.4±0.9	7.5±1.0
< 7 or > 9 hours/day, n (%)	12,463 (16.1)	2954 (14.9)	2266 (15.63)	3063 (15.4)	4180 (18.1)
Smoking status, n (%)					
Current	15,516 (20.1)	2522 (12.7)	2935 (20.2)	4104 (20.6)	5955 (25.9)
Former	27,584 (35.7)	6611 (33.3)	5295 (36.5)	6780 (34.1)	8898 (38.7)
Never	34,144 (44.2)	10,721 (54.0)	6271 (43.3)	8996 (45.3)	8156 (35.4)

7

Table 1. Continued.

| | | Individual socioeconomic disadvantage | | | |
	Total *n* = 77,244	Q1 (least disadvantaged) *n* = 19,854	Q2 *n* = 14,501	Q3 *n* = 19,880	Q4 (most disadvantaged) *n* = 23,009
Moderate-to-vigorous physical activity level, min/week (interquartile)	180 (60-360)	210 (90-360)	180 (60-360)	180 (60-360)	180 (60-360)
Below physical activity recommendation, n (%)	20,674 (26.8)	4185 (21.1)	3767 (26.0)	5637 (28.4)	7085 (30.8)
Lifestyle risk index, n (%)					
0 best	22,807 (29.5)	7983(40.2)	4549 (31.4)	5471 (27.5)	4804 (20.9)
1	27,072 (35.1)	7283(36.7)	5097 (35.1)	7145 (36.0)	7547 (32.8)
2	17,091 (22.1)	3346 (16.9)	3115 (21.5)	4613 (23.2)	6017 (26.2)
3	7513 (9.7)	1002 (5.0)	1313 (9.1)	1970 (9.9)	3228 (14.0)
4	2338(3.0)	208 (1.0)	366 (2.5)	596 (3.0)	1168 (5.1)
5	396 (0.5)	32(0.2)	60 (0.4)	80 (0.4)	224 (1.0)
6 worst	27(0.03)	0	1 (0)	5 (0)	21 (0.1)

[a] Percentage of missing data: education (0.3%), income (14.7%).

Supplementary analyses have shown that the relative risk ratio to be in the most unhealthy lifestyle category (i.e., lifestyle risk index higher than 3) among the participants from the highest quartile of ISED was 8.23 (95% CI 7.13-9.49, P < 0.001) times higher than those in the lowest ISED quartile (**Supplementary Table S3**). The neighborhood-disadvantage level was also positively associated with lifestyle risk index categories, with participants residing in the most disadvantaged neighborhoods having a 1.84 (95% CI 1.62-2.10, P < 0.001) times higher relative risk ratio of being in the most unhealthy lifestyle category compared with those who lived in the least disadvantaged neighborhoods (**Supplementary Table S3**). Because of the positive interaction between ISED and NSED, analyses were repeated and stratified by ISED quartiles (**Supplementary Table S4**). The magnitude of the adjusted association between NSED and the lifestyle risk index was the highest among participants who were the most individually socioeconomically disadvantaged. Sensitivity analyses using only education or income as an indicator for ISED (**Supplementary Tables S5** and **S6**) as well as categorizing the lifestyle risk index into three classes as the outcome (**Supplementary Table S3**) showed the same pattern as our main results. Individuals who had the lowest income or education and who resided in the most disadvantaged neighborhoods had the highest likelihood of having a higher lifestyle risk index. Moreover, the patterns of interactions between NSED and education or income were also similar to the patterns between NSED and ISED (**Supplementary Tables S5** and **S6**). However, some variations were shown for alcohol intake and MVPA, when each lifestyle factor was tested separately in the same model (**Supplementary Tables S7** and **S8**).

7

Table 2. Independent associations of individual socioeconomic disadvantage (ISED) and neighborhood socioeconomic disadvantage (NSED) with the lifestyle risk index[a]

	Model 1		Model 2	
	Beta-coefficients (95% CI)	P-trend	Beta-coefficients (95% CI)	P-trend
ISED				
Q4 (most disadvantaged)	0.58 (0.56-0.60)	<0.001	0.66 (0.64-0.68)	<0.001
Q3	0.32 (0.30-0.35)		0.35 (0.33-0.38)	
Q2	0.25 (0.23-0.27)		0.28 (0.26-0.30)	
Q1 (least disadvantaged)	Ref		Ref	
Random effect, estimate	0.018 (0.015-0.022)			
ICC	0.015 (0.013-0.019)			
NSED				
Q5 (most disadvantaged)	0.32 (0.28-0.36)	<0.001	0.18 (0.15-0.22)	<0.001
Q4	0.28 (0.24-0.33)		0.16 (0.12-0.20)	
Q3	0.20 (0.16-0.24)		0.10 (0.07-0.14)	
Q2	0.13 (0.09-0.17)		0.06 (0.02-0.09)	
Q1 (least disadvantaged)	Ref		Ref	
Random effect, estimate	0.020 (0.016-0.024)		0.010 (0.008-0.013)	
ICC	0.016 (0.013-0.020)		0.009 (0.007-0.011)	

Table 2. Continued.				
	Model 3		**Model 4**	
	Beta-coefficients (95% CI)	**P-trend**	**Beta-coefficients (95% CI)**	**P value**
ISED				
Q4 (most disadvantaged)	0.64 (0.62-0.66)	<0.001		
Q3	0.34 (0.32-0.36)			
Q2	0.27 (0.25-0.30)			
Q1 (least disadvantaged)	Ref			
Random effect, estimate				
ICC				
NSED				
Q5 (most disadvantaged)	0.17 (0.14-0.21)	<0.001		
Q4	0.15 (0.11-0.18)			
Q3	0.09 (0.06-0.13)			
Q2	0.05 (0.02-0.08)			
Q1 (least disadvantaged)	Ref			
Random effect, estimate	0.009 (0.007-0.011)		0.008 (0.007-0.011)	
ICC	0.008 (0.006-0.010)		0.008 (0.006-0.010)	
Interaction: ISED×NSED			0.016 (0.011-0.021)	<0.001

[a] Models were adjusted for: model 1, ISED or NSED singly adjusted; model 2, ISED, NSED, age, and sex; model 3, model 2 covariates plus BMI; model 4, model 3 covariates plus ISED×NSED; ICC, intraclass correlation coefficient.

7

Table 3. Associations between individual socioeconomic disadvantage (ISED) and the lifestyle risk index stratified by neighborhood socioeconomic disadvantage (NSED)[a]

	Model 1		Model 2		Model 3	
	Beta-coefficients (95% CI)	P-trend	Beta-coefficients (95% CI)	P-trend	Beta-coefficients (95% CI)	P-trend
NSED Q5 (most disadvantaged)						
ISED Q4 (most disadvantaged)	0.72 (0.67-0.77)	<0.001	0.81 (0.76-0.87)	<0.001	0.79 (0.74-0.85)	<0.001
ISED Q3	0.42 (0.37-0.48)		0.46 (0.40-0.51)		0.44 (0.39-0.50)	
ISED Q2	0.30 (0.23-0.36)		0.33 (0.27-0.39)		0.32 (0.26-0.38)	
ISED Q1 (least disadvantaged)	Ref		Ref		Ref	
NSED Q4						
ISED Q4 (most disadvantaged)	0.62 (0.57-0.67)	<0.001	0.71 (0.66-0.76)	<0.001	0.68 (0.64-0.73)	<0.001
ISED Q3	0.37 (0.32-0.42)		0.41 (0.35-0.46)		0.39 (0.34-0.44)	
ISED Q2	0.25 (0.19-0.31)		0.28 (0.22-0.33)		0.27 (0.21-0.32)	
ISED Q1 (least disadvantaged)	Ref		Ref		Ref	
NSED Q3						
ISED Q4 (most disadvantaged)	0.55 (0.51-0.60)	<0.001	0.65 (0.60-0.70)	<0.001	0.61 (0.57-0.66)	<0.001
ISED Q3	0.30 (0.26-0.35)		0.33 (0.28-0.38)		0.31 (0.26-0.36)	
ISED Q2	0.25 (0.20-0.31)		0.28 (0.23-0.33)		0.26 (0.21-0.32)	
ISED Q1 (least disadvantaged)	Ref		Ref		Ref	

Table 3. Associations between individual socioeconomic disadvantage (ISED) and the lifestyle risk index stratified by neighborhood socioeconomic disadvantage (NSED)[a]

	Model 1		Model 2		Model 3	
	Beta-coefficients (95% CI)	P-trend	Beta-coefficients (95% CI)	P-trend	Beta-coefficients (95% CI)	P-trend
NSED Q2						
ISED Q4 (most disadvantaged)	0.55 (0.51-0.59)	<0.001	0.64 (0.60-0.69)	<0.001	0.62 (0.57-0.66)	<0.001
ISED Q3	0.32 (0.27-0.36)		0.35 (0.31-0.39)		0.33 (0.29-0.38)	
ISED Q2	0.24 (0.19-0.29)		0.27 (0.22-0.32)		0.26 (0.21-0.21)	
ISED Q1 (least disadvantaged)	Ref		Ref		Ref	
NSED Q1 (least disadvantaged)						
ISED Q4 (most disadvantaged)	0.50 (0.46-0.55)	<0.001	0.58 (0.54-0.63)	<0.001	0.55 (0.50-0.59)	<0.001
ISED Q3	0.29 (0.25-0.33)		0.33 (0.28-0.37)		0.30 (0.26-0.34)	
ISED Q2	0.28 (0.24-0.32)		0.31 (0.27-0.36)		0.30 (0.26-0.34)	
ISED Q1 (least disadvantaged)	Ref		Ref		Ref	

[a] Models were adjusted for: model 1, ISED; model 2, model 1 covariates plus age and sex; model 3, model 2 covariates plus BMI.
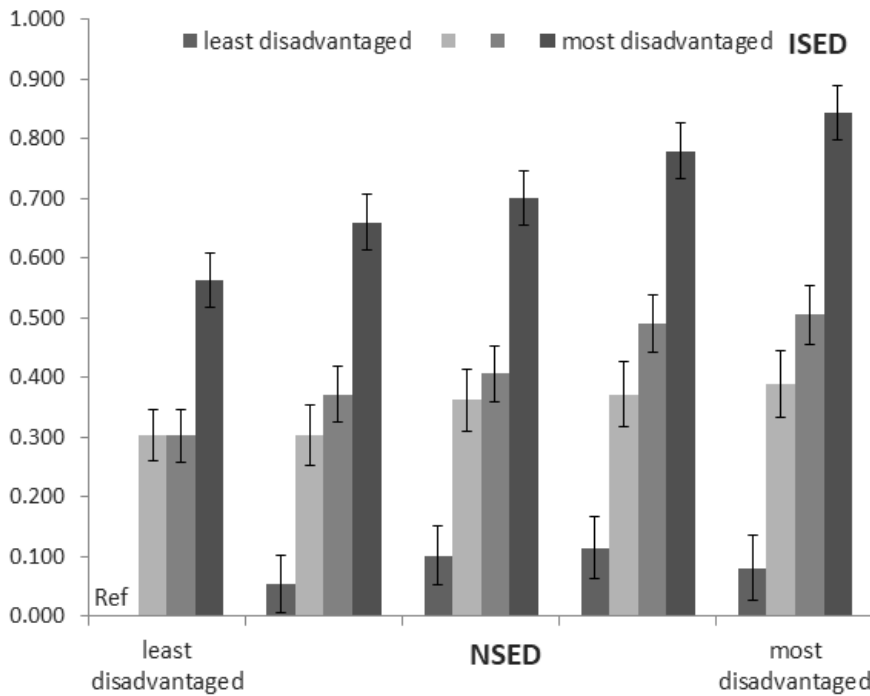
7

Fig. 1. Mixed-model coefficients of the joint associations of individual socioeconomic disadvantage (ISED) and neighborhood socioeconomic disadvantage (NSED) with the lifestyle risk index[a]

[a] The models were adjusted for age, sex, and BMI; reference group: least socioeconomically disadvantaged individuals and neighborhoods; random effect of neighborhood estimate (beta-coefficient 0.008 [95%CI 0.006-0.011]); intraclass correlation coefficient 0.008 (95%CI 0.006-0.010).

Discussion

In this large population-based study, we found that both ISED and NSED were positively associated with the lifestyle risk index. More importantly, the association between ISED and the lifestyle risk index was positively modified by NSED. Subgroup analyses revealed that the gradient of the association between ISED and the lifestyle risk index was steeper for those living in the most disadvantaged neighborhoods.

To our knowledge, the current study is the first to simultaneously investigate the relationship of ISED, NSED, and their interactive effects with an index of a broad range of lifestyle risk factors. Our study extends previous knowledge by demonstrating that the higher vulnerability of practicing an unhealthy lifestyle for individuals residing in socioeconomically disadvantaged neighborhoods applies to a wider range of lifestyle factors than previously understood, including both traditional and emerging lifestyle factors such as TV-watching time and sleep time. Our findings are partly consistent with previous studies showing that NSED was associated with a higher chance of having more unhealthy lifestyle factors net of ISED.[15] The only systematic review of 22 studies found that a higher level of NSED was consistently associated with smoking and physical inactivity independently of ISED, whereas evidence of fruit/vegetable intake and excessive alcohol consumption was ambiguous.[15] In the present study, we focused on a composite lifestyle risk index, rather than studying a single lifestyle factor. There are two major considerations for that. First, previous evidence suggests that lifestyle risk factors tended to cluster in different patterns within the population.[23] Studying the effects of NSED on a single lifestyle factor could lead to inaccurate estimates, as their coexisting lifestyle risk factors are not simultaneously accounted for. Second, single lifestyle risk factor cannot fully capture one's overall lifestyle risk profile, as those lifestyle factors were found to have synergistic risk contributions to one's health outcomes.[23-25]

The underpinning mechanisms of the steeper gradient associations between ISED and the lifestyle risk index across NSED strata may be explained by several socio-health theoretical models, i.e., the double-jeopardy model, fundamental-cause theory, and

collective-resources model.[20,22,33] In general, those three models all emphasize that individuals who are more socioeconomically disadvantaged will be particularly worse off if they live in a disadvantaged neighborhood, because (i) they originally have fewer individual health resources, and (ii) when living in a neighborhood with fewer health resources, their health is expected to be worsened more if one is already disadvantaged, compared with their less disadvantaged neighbors. On the contrary, individuals who are less disadvantaged will be less affected by neighborhood disadvantage, as they are always able to get access to health resources and depend less on their residing neighborhoods. From another point of view, in addition to unfavorable resources, previous evidence also suggests that the neighborhood may serve as a social platform for the spread of certain health beliefs and social norms.[34,35] For those facing an unfavorable social environment as well as limited resources, individuals who are less disadvantaged may be more resilient and resistant to such negative factors because of their higher level of self-perceived control and knowledge that enables them to avoid such unhealthy lifestyle behaviors.[36,37]

Our findings of the steeper gradient association between ISED and the unhealthy lifestyle risk index for individuals living in disadvantaged neighborhoods provide two important public health implications. First, while conducting lifestyle interventions with a focus on addressing individual-level socioeconomic inequalities, it is of equal importance to consider the socioeconomic inequalities originating from the living neighborhood, particularly with additional support for those who are of low individual socioeconomic status. As the basic single census unit, neighborhoods also provide a geographically tangible platform for conducting such public health interventions, which thus may help to improve the reach of health programs for those vulnerable groups.[38] Second, given the concrete evidence that lifestyle factors are the most important modifiable behavioral risk factors for the prevention of non-communicable diseases,[39] public health initiatives directed towards disadvantaged neighborhoods, in terms of both physical and social resources, may have the potential to achieve substantial public health benefits and ameliorate the persistent health inequalities within society.[40]

The strengths of this study include the relatively immobile physical and social

environment of the study population, thus limiting the potential influences of the fast-changing environment and population mobility on individual lifestyle factors. In fact, we only observed approximately 10% of the total participants who moved between 2011 (baseline) and 2016 (second follow-up). Furthermore, our study is the first to thoroughly investigate the extent to which NSED modified the association between ISED and a spectrum of unhealthy lifestyle factors. We also conducted numerous sensitivity analyses supporting the robustness of our findings. Nevertheless, there are also limitations. First, neighborhood-level education data were not available from the CBS Neighborhood Statistics used for the construction of the NSED score. Thus, our estimated neighborhood effects might have missed the potential influences of neighborhood educational level. However, sensitivity analysis with additional adjustment for neighborhood-level education (percentage of participants with low education) did not materially change the results (**Supplementary Table S9**). Second, there might be some misclassifications of the unweighted lifestyle risk index in more disadvantaged groups because of social desirability bias. Thus, the proportion of individuals with high lifestyle risk index might be underestimated; although the distribution of the lifestyle risk index is comparable to a previous study.[25] In fact, misclassification of lifestyle in a more disadvantaged group would flatten the association between ISED and the lifestyle risk index, indicating that the associations would be even more pronounced with an accurate classification. In addition, we are not able to provide more detailed information about smoking status such as the period of cessation and the number of cigarettes, because the quality of the data in this part of the questionnaire was unfortunately insufficient due to missing data. Third, the Lifelines cohort is a single cohort study from a region with a population of predominantly European descent (>99%). The Netherlands is a country with a well-developed social-security system. This may limit the generalizability of the results to populations of another ethnicity and in a different social context. Fourth, participants with missing lifestyle factors (25.4%) and NSED (13.7%) were excluded from the current study, which could possibly introduce selection bias. However, the characteristics of the excluded participants did not differ substantially from those of the study population; still, participants with missing lifestyle or NSED data were more likely to report low or

7

missing income data (**Supplementary Table S10**). Finally, no causal inferences should be drawn from our findings given the cross-sectional nature of our study, although additional adjustment for BMI may to some extent help to reduce the potential bias caused by reverse causation, as individuals with high BMI might alter their lifestyle factors before the entry of the study.

Conclusions

In conclusion, this study illustrates that NSED, in addition to ISED, was associated with a higher likelihood of practicing an unhealthy lifestyle. More importantly, the association between ISED and the lifestyle risk index was positively modified by NSED. In other words, the gradient of the association between ISED and the lifestyle risk index was steeper for individuals living in the most disadvantaged neighborhoods. These findings suggest that public health initiatives addressing lifestyle-related socioeconomic health differences should not only target individual lifestyles, but also consider neighborhood factors, in particular providing more health resources and social opportunities for those socioeconomically disadvantaged neighborhoods.

Reference

1.   Kivimaki M, Batty GD, Pentti J, et al. Association between socioeconomic status and the development of mental and physical health conditions in adulthood: a multi-cohort study. Lancet Public Health 2020; 5(3): e140-e9.

2.   Mackenbach JP, Stirbu I, Roskam AJ, et al. Socioeconomic inequalities in health in 22 European countries. N Engl J Med 2008; 358(23): 2468-81.

3.   Beenackers MA, Kamphuis CB, Giskes K, et al. Socioeconomic inequalities in occupational, leisure-time, and transport related physical activity among European adults: a systematic review. Int J Behav Nutr Phys Act 2012; 9: 116.

4.   Mathur C, Erickson DJ, Stigler MH, Forster JL, Finnegan JR, Jr. Individual and neighborhood socioeconomic status effects on adolescent smoking: a multilevel cohort-sequential latent growth analysis. Am J Public Health 2013; 103(3): 543-8.

5.   Hulshof KF, Brussaard JH, Kruizinga AG, Telman J, Lowik MR. Socio-economic status, dietary intake and 10 y trends: the Dutch National Food Consumption Survey. Eur J Clin Nutr 2003; 57(1): 128-37.

6.   Thorpe MG, Milte CM, Crawford D, McNaughton SA. Education and lifestyle predict change in dietary patterns and diet quality of adults 55 years and over. Nutr J 2019; 18(1): 67.

7.   Andrade-Gomez E, Garcia-Esquinas E, Ortola R, Martinez-Gomez D, Rodriguez-Artalejo F. Watching TV has a distinct sociodemographic and lifestyle profile compared with other sedentary behaviors: A nationwide population-based study. PLOS ONE 2017; 12(12): e0188836.

8.   Xiao Q, Hale L. Neighborhood socioeconomic status, sleep duration, and napping in middle-to-old aged US men and women. Sleep 2018; 41(7).

9.   Shishehbor MH, Gordon-Larsen P, Kiefe CI, Litaker D. Association of neighborhood socioeconomic status with physical fitness in healthy young adults: the Coronary Artery Risk Development in Young Adults (CARDIA) study. Am Heart J 2008; 155(4): 699-705.

10.   Compernolle S, Oppert JM, Mackenbach JD, et al. Mediating role of energy-balance related behaviors in the association of neighborhood socio-economic status and residential area density with BMI: The SPOTLIGHT study. Prev Med 2016; 86: 84-91.

11.   Cambron C, Kosterman R, Hawkins JD. Neighborhood Poverty Increases Risk for Cigarette Smoking From Age 30 to 39. Ann Behav Med 2019; 53(9): 858-64.

12.   Ball K, Lamb KE, Costa C, et al. Neighbourhood socioeconomic disadvantage and fruit and vegetable consumption: a seven countries comparison. Int J Behav Nutr Phys Act 2015; 12: 68.

13.   Dubowitz T, Heron M, Bird CE, et al. Neighborhood socioeconomic status and fruit and vegetable intake among whites, blacks, and Mexican Americans in the United States. Am J Clin Nutr 2008; 87(6): 1883-91.

14.   Calling S, Ohlsson H, Sundquist J, Sundquist K, Kendler KS. Socioeconomic status and alcohol use disorders across the lifespan: A co-relative control study. PLOS ONE 2019; 14(10): e0224127.

7

15.   Algren MH, Bak CK, Berg-Beckhoff G, Andersen PT. Health-Risk Behaviour in Deprived Neighbourhoods Compared with Non-Deprived Neighbourhoods: A Systematic Literature Review of Quantitative Observational Studies. PLOS ONE 2015; 10(10): e0139297.

16.   Xiao Q, Keadle SK, Berrigan D, Matthews CE. A prospective investigation of neighborhood socioeconomic deprivation and physical activity and sedentary behavior in older adults. Prev Med 2018; 111: 14-20.

17.   Moss JL, Xiao Q, Matthews CE. Patterns of cancer-related health behaviors among middle-aged and older adults: Individual- and area-level socioeconomic disparities. Prev Med 2018; 115: 31-8.

18.   Miranda JJ, Carrillo-Larco RM, Gilman RH, et al. Patterns and Determinants of Physical Inactivity in Rural and Urban Areas in Peru: A Population-Based Study. J Phys Act Health 2016; 13(6): 654-62.

19.   Chuang YC, Li YS, Wu YH, Chao HJ. A multilevel analysis of neighborhood and individual effects on individual smoking and drinking in Taiwan. BMC Public Health 2007; 7: 151.

20.   Boylan JM, Robert SA. Neighborhood SES is particularly important to the cardiovascular health of low SES individuals. Soc Sci Med 2017; 188: 60-8.

21.   Winkleby M, Cubbin C, Ahn D. Effect of cross-level interaction between individual and neighborhood socioeconomic status on adult mortality rates. Am J Public Health 2006; 96(12): 2145-53.

22.   Stafford M, Marmot M. Neighbourhood deprivation and health: does it affect us all equally? Int J Epidemiol 2003; 32(3): 357-66.

23.   Ding D, Rogers K, van der Ploeg H, Stamatakis E, Bauman AE. Traditional and Emerging Lifestyle Risk Behaviors and All-Cause Mortality in Middle-Aged and Older Adults: Evidence from a Large Population-Based Australian Cohort. PLOS MED 2015; 12(12): e1001917.

24.   Watts P, Buck D, Netuveli G, Renton A. Clustering of lifestyle risk behaviours among residents of forty deprived neighbourhoods in London: lessons for targeting public health interventions. J Public Health (Oxf) 2016; 38(2): 308-15.

25.   Foster HME, Celis-Morales CA, Nicholl BI, et al. The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: a prospective analysis of the UK Biobank cohort. Lancet Public Health 2018; 3(12): e576-e85.

26.   Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol 2015; 44(4): 1172-80.

27.   Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. PLOS ONE 2015; 10(9): e0137203.

28.   UNESCO. International Standard Classification of Education (ISCED). 1997. https://www.unesco.org/en/education.

29.   Siebelink E, Geelen A, de Vries JH. Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. Br J Nutr 2011; 106(2): 274-81.

30.   Dawson DA. Defining risk drinking. Alcohol Res Health 2011; 34(2): 144-56.

31.   Vinke PC, Corpeleijn E, Dekker LH, Jacobs DR, Jr., Navis G, Kromhout D. Development of the food-based Lifelines Diet Score (LLDS) and its application in 129,369 Lifelines participants. Eur J Clin Nutr 2018; 72(8): 1111-9.

32.   Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003; 56(12): 1163-9.

33.   Robert SA. SOCIOECONOMIC POSITION AND HEALTH: The Independent Contribution of Community Socioeconomic Context. Annu Rev Sociol 1999; 25(1): 489-516.

34.   Karasek D, Ahern J, Galea S. Social norms, collective efficacy, and smoking cessation in urban neighborhoods. Am J Public Health 2012; 102(2): 343-51.

35.   National Research Council. Critical Perspectives on Racial and Ethnic Differences in Health in Late Life. Washington, DC: The National Academies Press; 2004.

36.   Stronks K, van de Mheen HD, Looman CW, Mackenbach JP. Cultural, material, and psychosocial correlates of the socioeconomic gradient in smoking behavior among adults. Prev Med 1997; 26(5 Pt 1): 754-66.

37.   Ball K, Abbott G, Cleland V, et al. Resilience to obesity among socioeconomically disadvantaged women: the READI study. Int J Obes (Lond) 2012; 36(6): 855-65.

38.   Ribeiro AI. Public health: why study neighborhoods? Porto Biomed J 2018; 3(1): e16.

39.   GBD Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016; 388(10053): 1659-724.

40.   Ludwig J, Sanbonmatsu L, Gennetian L, et al. Neighborhoods, obesity, and diabetes--a randomized social experiment. N Engl J Med 2011; 365(16): 1509-19.

7

Additional Information

Acknowledgements

Funding

Contributions

YZ, MJD, and GN designed the study. LHD conceptualized the study. YZ conducted data cleaning/statistical analysis and drafted the manuscript. MJD participated in data cleaning, drafted the discussion, and revised the manuscript. LHD, GN, IJR, IM, JOM, JJC, and SJLB critically reviewed/edited the manuscript. LHD and GN supervised the project.

Supplementary Data

Online **Supplementary Files** can be accessed via
doi.org/10.1093/ije/dyab079, or by scanning the QR code.
Article PMID: 34999857.

7

General Discussion
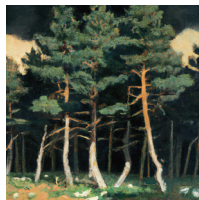# CHAPTER 8

Aims of This Thesis

Summary of Main Findings

Methodological Considerations and Research Limitations

Implications for Public Health Practice and Policymaking

Future Perspectives

Overall Conclusions

Aims of This Thesis

The **aim of this thesis** is to provide an empirical evidence base for better type 2 diabetes prevention in the general population. Specifically, this thesis focuses on lifestyle patterns (**part 1**) and underlying factors of lifestyle factors (**part 2**):

Part 1 – Lifestyle Patterns describes dietary and lifestyle patterns – how multiple co-occurring dietary or lifestyle factors cluster – in the general population, and investigates the associations between these dietary and lifestyle patterns with risk of incident type 2 diabetes. The relevance of dietary and lifestyle patterns in the development and prevention of type 2 diabetes is extensively discussed.

Part 2 – Underlying Factors investigates the relationships of lifestyle factors with their underlying factors, with a special focus on individual socioeconomic status and neighborhood socioeconomic status. How these factors collectively affect type 2 diabetes risk is subsequently investigated.

The following parts in the **General Discussion** include a summary of scientific findings, methodological considerations, implications for public health practice and policymaking, and the overall conclusions of this thesis.

Summary of Main Findings

Part 1 – Lifestyle Patterns and Type 2 Diabetes Risk

Main Findings

♦ Lifestyle and dietary factors cluster in the population, in the form of lifestyle and dietary patterns (**chapter 2**, **3**, and **4**).

♦ Different dietary patterns are differentially associated with risk of incident type 2 diabetes (**chapter 2** and **3**).

♦ Different combinations of lifestyle factors, as manifested in lifestyle patterns, are differentially associated with risk of incident type 2 diabetes. The collective effects of different lifestyle factors may not be simply additive (**chapter 4**).

Implications for Research

● Consider dietary and lifestyle patterns in lifestyle research: (1) the development of type 2 diabetes is not due to single dietary and lifestyle factors; (2) there are interdependent effects among different dietary and lifestyle factors in the development of type 2 diabetes.

Implications for Public Health and Policymaking

⚘ Focus on lifestyle and dietary patterns as intervention targets:

(1) enable assessing the overall lifestyle risk profiles of the target population – targeting multiple co-occurring lifestyle factors;

(2) allow personalized lifestyle interventions at the aggregate level in the general population – tailored intervention strategies based on the lifestyle features of lifestyle pattern groups;

(3) allow identifying priority lifestyle intervention targets for different lifestyle pattern groups.

8

Part 2 – Underlying Factors of Lifestyle Factors

Main Findings

♦ Education and income are independently and differentially associated with incident type 2 diabetes (**chapter 6**).

♦ Education has bigger positive effects on healthy lifestyle factors compared with income (**chapter 5**).

♦ Within each education or income level, substantial health disparities exist across strata of the other indicator (**chapter 6**).

♦ After adjustment for age, sex, lifestyle factors, obesity status, clinical biomarkers, and family history of diabetes, a large proportion of the associations (more than 60%) of education and income with incident type 2 diabetes is unexplained (**chapter 6**).

♦ Individual and neighborhood socioeconomic disadvantage is associated with a higher chance of having unhealthy lifestyle factors (**chapter 7**).

♦ Neighborhood socioeconomic disadvantage positively modifies the association between individual socioeconomic disadvantage and the lifestyle risk index. This indicates that disadvantaged individuals are disproportionately affected by neighborhood socioeconomic disadvantage, as they have a disproportionately higher chance of having unhealthy lifestyle if they live in disadvantaged neighborhoods. Neighborhood socioeconomic disadvantage is an amplifier of the detrimental effects of individual socioeconomic disadvantage on unhealthy lifestyle factors (**chapter 7**).

Implications for Research

● Consider both education and income in research on health inequalities.

● Study the potential mechanisms of unexplained individual socioeconomic inequalities in health.

● Study the mechanisms of how contextual (neighborhood) factors affect health.

● Consider the interplay among lifestyle factors, individual socioeconomic status, contextual (neighborhood) factors, and their collective effects on health.

Implications for Public Health and Policymaking

⚛ Tackling socioeconomic inequalities in lifestyle and health:

(1) targeting and improving underlying factors of lifestyle factors, including (but not limited to the ones investigated in this thesis) education, income, and contextual (neighborhood) factors;

(2) additional support for socioeconomically-disadvantaged groups;

(3) improving education level of the general population and ameliorating income inequalities using generic measures are the two fundamental strategies.

8

Chapter-specific Findings

Chapter 2: Blood lipids-related dietary patterns and incident type 2 diabetes

Findings: dietary patterns that explain the variation in blood lipids are associated with higher risk of incident type 2 diabetes. These dietary patterns are characterized by a high intake of sugary beverages, added sugar, and juice; and a low intake of fruits, vegetables, nuts/seeds, cereals, and tea.

Implications: it is possible to optimize blood lipid profiles to lower the risk of type 2 diabetes through changes in dietary patterns.

Chapter 3: Ultra-processed food and incident type 2 diabetes: studying the underlying consumption patterns

Findings:

(1) a higher overall intake of ultra-processed food is associated with higher risk of type 2 diabetes, independent of the overall background diet quality. This indicates that eating an otherwise healthy diet may not fully compensate for the detrimental effects of ultra-processed food.

(2) discrepancies in the associations are found for different ultra-processed food consumption patterns. Ultra-processed food consumption patterns characterized by high consumption of cold savory snacks and warm savory snacks are associated with higher risk of incident type 2 diabetes. The traditional Dutch cuisine pattern is not associated with type 2 diabetes risk. The pattern characterized by high consumption of sweet snacks is inversely associated with type 2 diabetes risk. The counterintuitive finding of the sweet snack pattern may be due to reverse causation, as adherence to this pattern is lower for people with a high risk of diabetes at baseline.

Implications: in addition to promoting the consumption of healthy food, food-based dietary guidelines and dietary interventions are recommended to limit the consumption of unhealthy ultra-processed food products, specifically savory snacks.

Chapter 5: Using structural equation modeling to untangle pathways of risk factors associated with incident type 2 diabetes

<u>Findings:</u> Interrelationships of modifiable risk factors related to type 2 diabetes can be quantified by analyzing a conceptual model including multiple intersecting pathways of these risk factors. Among all modifiable risk factors analyzed, waist circumference has the biggest direct effect on the development of type 2 diabetes, followed by HDL-cholesterol, and triglycerides. Less TV watching (as an emerging lifestyle factor) and more physical activity play an important role in improving clinical markers that are directly associated with type 2 diabetes. Education has the biggest positive effects on lifestyle factors. Income shows direct effects on type 2 diabetes status that are not explained by lifestyle factors, obesity status, and clinical biomarkers, suggesting that health inequalities are influenced by other uncaptured factors such as social insecurity.

<u>Implications:</u> Analyzing a conceptual model including multiple pathways of risk factors related to type 2 diabetes allows to compare the relative contribution of these risk factors, which can support the prioritization of prevention targets at the population level. Regarding the current guidelines for diabetes prevention, waist management in addition to BMI control (clinical level), as well as less TV watching in addition to more physical activity (lifestyle level), may provide additional public health benefits. Better education may improve lifestyle and would be the main societal goal for the prevention of type 2 diabetes. Focusing on these priority prevention targets may have the potential to address as much of the broader risk profiles as possible. The effects of income that are not explained by the broad array of lifestyle factors documented here suggest that ameliorating income inequalities and improving social security stability and coverage may have the potential to reduce health inequalities.

8

Methodological Considerations and Research Limitations

General Design – The Lifelines Cohort Study

Studies in this thesis are based on the Lifelines cohort study. The Lifelines cohort study is a multi-disciplinary prospective general population-based cohort study, which applies a unique three-generation design to study the health and health-related behaviors of 167,729 people living in the north of The Netherlands. The Lifelines cohort study employs a broad range of investigative procedures in collecting and assessing the biomedical, socio-demographic, behavioral, and physical factors, which contribute to the health and disease of the general population.[1-3] For this thesis, the large sample size, prospective design, and the wide spectrum of factors collected enable detailed and comprehensive investigations into the relationships among lifestyle factors, underlying factors of lifestyle factors, and health outcomes.

Operationalization of Socioeconomic Indicators

This thesis studies two primary socioeconomic indicators, i.e., education and income. For the former, the highest education level achieved is used and is further categorized according to the International Standard Classification of Education (ISCED): (1) low (level 0, 1, or 2) – junior general secondary education or lower; (2) middle (level 3 or 4) – secondary vocational education and senior general secondary education; and (3) high (level 5 or 6) – higher vocational education or university.[4] For the latter, the monthly household net income level and the categorization (original choices) from the questionnaire are used: (1) low (<1000 euro/month); (2) lower-middle (1000–2000 euro/month); (3) upper-middle (2000–3000 euro/month); (3) high (>3000 euro/month); and (4) do not know/prefer not to answer.[1]

With the aforementioned definitions of education and income, the results of this thesis show a clear picture of the socioeconomic gradients in health. The choice of investigating these two definitions is partly due to pragmatic reasons, as they were directly asked by the Lifelines questionnaires and thus easily available. The clear cut-offs for categorization may facilitate the interpretation of the findings.

However, it is not difficult to notice that education and income have been operationalized in various ways in research, which may also be largely attributed to pragmatic reasons such as data availability. Education and income can both be measured and defined as continuous variables. A more complex situation may arise when income needs to be distinguished between individual and household levels. While this thesis underscores that education and income should be simultaneously considered when studying socioeconomic inequalities in health, it is likewise worth considering different definitions of each socioeconomic indicator. A limitation of this thesis is that the potential impacts of different definitions of socioeconomic indicators on health are not considered.

Different definitions of education and income may reflect different theoretical links between socioeconomic status and health inequalities, and may therefore inform substantially different public health practice and policies.[5] For example, household income may better capture actual family material conditions than individual income. People – approximately 60% of women in the Dutch labor market – working part-time,[6] although they may have a lower individual income, may still have adequate economic resources to invest in health. When further analyzing such situations, it is also necessary to consider other socio-demographic indicators, such as partnership status, occupational class, and family power discrepancies, which are also not included in this thesis. In addition, socioeconomic status may change over the life course, with income peaks at middle age and valleys during young adulthood and after retirement.[5] The potential influences of cohort effects could also affect the results, as early-acquired education in older generations may devaluate over time with longer education received in younger generations.[7]

There is much less knowledge about how different definitions of education and income may affect the estimation of health inequalities. Empirical evidence from the Finnish national administrative data shows that different definitions of income substantially influenced the level of mortality inequality, and the period trends of mortality inequality across income definitions differed between men and women.[5] Such empirical evidence on healthy lifestyle inequalities appears to be a sleepy backwater that needs substantial

research development.

From a broader perspective, when interpreting results based on different definitions of education and income, it is also necessary to consider the geographical and socioeconomic settings of the study. The Lifelines cohort study is established in The Netherlands – a high-income country with well-developed infrastructure. People who have been categorized in our studies as having low socioeconomic status in general still have living conditions that are better than many people in middle- or low-income countries. Hence, using the same definition of education or income across different countries may yield different results, and may also provide substantially different public health implications.

In our studies, income has not been categorized further for analysis. This may prevent introducing more uncertainties in statistical analysis, especially considering that income was asked in categories in the Lifelines questionnaire based on a rather wide thousand euro unit. Another study investigating socioeconomic inequalities in metabolic syndrome in the Lifelines cohort categorizes education and income into years of education and household equivalent income, respectively. Results from that study show that more education and higher occupational prestige, but not higher income, are associated with lower risk of incident metabolic syndrome, with lifestyle factors being the strongest mediators of these associations; while lifestyle factors do not mediate the association between income and incident metabolic syndrome.[8] Despite different health outcomes focused that preclude direct comparisons, different definitions of income may affect the estimation of health inequalities. However, the public health implications of comparing different definitions used among studies in the Lifelines cohort are not self-evident.

Substantial work has been done on revealing socioeconomic inequalities in health. Most studies (including the ones in this thesis) are explorative and hypothesis-generating. Hence, they may not require a hypothesis on the suitability of certain definitions of a socioeconomic indicator, as the main merit of this type of study is to pave the way towards hypothesis-driven studies.[5,9] Accordingly, for future research, it is important

to formulate research questions and hypotheses in advance, including the relevance of certain definitions of a socioeconomic indicator. This will not only allow testing potential mechanisms of socioeconomic inequalities in health, but will also provide guidance on the pros and cons of specific indicators for research and practice in public health.

Changes in Lifestyle Patterns and Socioeconomic Status

Using the Lifelines cohort, this thesis studies the prospective relationships of lifestyle factors and socioeconomic status with incident type 2 diabetes. The longitudinal design aids in studying the potential causal associations among these factors. Despite this prospective design, all lifestyle factors and socioeconomic status were measured at baseline. Potential changes of these factors during follow-up are not considered in this thesis but they may influence clinical outcomes. It has been reported that dietary pattern trajectories are longitudinally associated with changes in HbA_{1c},[10] risk of obesity,[11] and mortality,[12] whereas evidence on trajectories of overall lifestyle patterns is substantially lacking. Lifestyle factors can vary over time and the life course. Changes of lifestyle factors may also be influenced by interdependent changes in individual socioeconomic status and contextual factors, that collectively shape health outcomes.[13] Other issues concerning such changes include the cumulative exposures to different lifestyle factors and socioeconomic positions and the sequence of these exposures over the life course. Until 2022, participants from the Lifelines cohort study have been followed up for 16 years, and there will still be at least 14 years of follow-up.[2] Questions related to changes in lifestyle patterns and socioeconomic status may be investigated in the near future with new data being available.

Representativeness of the Study Population and Generalizability of Results

Research from this thesis is conducted in the Dutch Lifelines cohort study. Notwithstanding a large sample size, results from a single cohort may not be generalizable to other populations. Differences in socio-demographic and lifestyle factors have been identified by comparing the Lifelines participants with the population of the north of The Netherlands. For example, the Lifelines participants were more often female,

8

middle-aged, and married. Immigrants and individuals with low education or without a paid job were also underrepresented in the Lifelines cohort at baseline. Despite these differences, the Lifelines population is found to be broadly representative of the population of the north of The Netherlands, especially for lifestyle factors and the prevalence of chronic diseases.[3]

For the associations between lifestyle factors and type 2 diabetes studied in this thesis, the risk of selection bias is low. However, over-reporting of healthy lifestyle factors due to social desirability could flatten the associations, especially among socioeconomically-disadvantaged populations.[14] Validation against lifestyle questionnaires using objective measurements (such as accelerometer for physical activity level, or 24-h urine collections for specific dietary factors) may provide insights into the level of misreporting.[14-16] For the associations of socioeconomic status with health outcomes, the underrepresentation of socioeconomically-disadvantaged populations could lead to underestimation of socioeconomic inequalities. It is generally unclear about the mechanisms of such underrepresentation, while qualitative conclusions about the direction and the approximate magnitude of health inequalities may hold the same.[17]

For lifestyle pattern analysis, the issue of reproducibility and generalizability has been frequently questioned, since patterns identified in one population and their associations with disease outcomes may not necessarily be able to be replicated in other population groups.[18-21] This issue is however not further studied in this thesis. Using data from the InterConnect project – a data-sharing platform including data from 25 cohorts that enables cross-cohort analyses without pooling data, researchers have found that none of the associations between type 2 diabetes and healthy dietary patterns identified in one cohort can be replicated in the others, but this does not apply to unhealthy dietary patterns.[20] No study has investigated the generalizability of the overall lifestyle pattern. Inter-cohort differences in socio-demographic backgrounds, lifestyle factors, confounders adjusted, and definitions in lifestyle factors may all compromise the generalizability of lifestyle patterns. Some additional methodological considerations of lifestyle patterns have been discussed in **chapter 4**. For dietary patterns, it may be more relevant to focus on major contributors among food groups to a pattern and

allow descriptive comparisons across patterns identified in different cohorts.

As illustrated in this thesis, one of the core objectives of lifestyle pattern analysis is to describe the clustering of multiple co-occurring lifestyle factors, which facilitates the design of better-targeted interventions as opposed to the one-size-fits-all approaches. True differences in lifestyle patterns may exist between different populations because of different cultures, geographical locations, and social environments. Accordingly, seeking unified lifestyle patterns across different populations is less relevant to be a goal in public health prevention. However, the generalizability of lifestyle patterns does remain relevant at a higher level of abstraction. More precisely, if interventions based on lifestyle patterns for type 2 diabetes prevention are found to be more effective, it is expected that this approach can also work effectively in different populations, although the contents of lifestyle patterns may be different across different populations.

8

Implications for Public Health Practice and Policymaking

Lifestyle Patterns as Intervention Target

Personalized Intervention at Intermediate Population Level

For type 2 diabetes prevention, current evidence supports the relevance of targeting multiple lifestyle risk factors simultaneously through a personalized approach.[21-27] This poses a very ambitious challenge for the design of lifestyle intervention programs at the population level. First, it underscores the need to assess the overall lifestyle profiles of the target populations, rather than a single lifestyle factor. Although certain efforts have been made on improving physical activity level and increasing fruit and vegetable intake, little attention in public health practice has been given to other lifestyle factors and the distribution of lifestyle factors within the target populations.[28-33] Second, it is questionable whether a fully personalized approach is feasible at the population level. In the context of primary prevention, full personalization, with tailor-made lifestyle interventions for each individual, can be highly effective; but this strategy may not be feasible for the broader general population due to its high-cost and extensive labor in need. The generic approach, on the other hand, usually performs poorly on discriminating personal habits and needs, which thereby compromises the effectiveness of lifestyle intervention programs.

Findings from this thesis suggest that **lifestyle patterns may be the intervention target**, that have the potential to reconcile the aforementioned "conflicts of interests" between the generic (one-size-fits-all) approach and the strictly personalized approach. Lifestyle pattern analysis provides a way to segment the population into several sub-group populations by analyzing multiple lifestyle factors at the aggregate level. As illustrated in **chapter 2**, **3**, and **4**, various dietary and lifestyle patterns have been identified, varying according to specific dietary and lifestyle factors concerned.[21,34,35] Within each pattern, people have similar dietary or lifestyle habits. These lifestyle patterns are differentially associated with type 2 diabetes risk, demonstrating each of their clinical relevance. Based on the lifestyle behavioral features, tailored lifestyle interventions can be designed and concentrated for each lifestyle pattern group, where the largest

health gains can be achieved (improved effectiveness) with less labor and cost. Through this approach, **personalized intervention at an intermediate population level** can be enabled.

Interestingly, such a strategy, that targets people at an intermediate population level, has been successfully applied in marketing during the past decades, referred to as targeting "personas". A persona refers to a group of people with similar habits, needs, values, and perspectives, covering the whole spectrum of consumer behaviors.[36] Some personas have been identified in marketing practice such as "young hedonists", "traditional elderly", and "multicultural diversifier".[37,38] Tailored novel products and marketing messages have been designed and provided to different sub-group populations, which allows personalized marketing in the general population. Likewise, focusing on dietary and lifestyle patterns as intervention targets may also help translate the needs, habits, and preferences into the evidence base and new opportunities for the design of lifestyle intervention programs, as these factors are often the "blind spots" that are not directly visible and comprehensible for policymakers, researchers, and public health workers.

However, it needs to be acknowledged that lifestyle patterns may not fully capture the population-level variation in lifestyle factors. Population-level heterogeneity in lifestyle factors naturally exists, as every single person is unique in their lifestyle choices. Lifestyle pattern analysis (especially approaches used in this thesis) therefore aims to identify sub-group populations in which people have lifestyle behaviors that are most similar to each other. Population-level heterogeneity in lifestyle factors has been explained and reduced to an intermediate level that is feasible for public health interventions within reach and with current technology. Uncaptured lifestyle patterns are possible, but major lifestyle patterns that have the biggest population-wide public health relevance and benefit are retained.

Potential Intervention Strategies Based on Lifestyle Patterns Identified in This Thesis

The findings from **chapter 4** support the feasibility of designing targeted approaches for groups with different lifestyle patterns. Specifically, efforts can be made on improving

diet quality and physical activity level for the "poor diet and low physical activity group"; reducing alcohol intake should be prioritized for the "risk drinker group"; people from the "couch potato group" may benefit from limiting TV watching (sedentary time); for the "unhealthy lifestyle group", accounting for approximately 20% of the population, extensive lifestyle interventions may be implemented with prioritization on smoking cessation and increasing physical activity level. Lifestyle interventions on smoking cessation, however, would not meet the lifestyle intervention needs of the "poor diet and low physical activity group" and the "couch potato group". Similar approaches can also be applied to ultra-processed food consumption patterns (**chapter 3**), such as limiting savory ultra-processed food intake for people with the highest adherence to the two savory snack patterns. In **chapter 2**, the identified dietary patterns reflect the variation in blood lipid profiles in the population. Based on the features of the identified dietary patterns (such as high consumption of sugary beverages and added sugar, and low consumption of vegetables and nuts/seeds), tailored dietary interventions can be designed focusing on people with the highest adherence to these dietary patterns, which may optimize the blood lipid profiles of the target population. Conspicuously, the validity of such a targeted approach requires empirical evaluation, which should include both the effectiveness and efficacy of the lifestyle modification, as well as the evaluation of its possible health benefits.

Facilitating Cooperation with Target Groups

Another merit of focusing on lifestyle and dietary patterns as intervention targets is that it allows flexibility and personal preferences in making choices of lifestyle and dietary changes. To encourage healthy lifestyle and dietary choices, it is not necessary for individuals to make drastic changes, such as removing whole food groups from their diet or strictly following another dietary pattern. Instead, within lifestyle and dietary patterns, people can combine small changes in a variety of flexible ways that meet their preferences, health needs, habits, and cultural traditions. Lifestyle intervention programs work effectively with cooperation with the target groups.[39] Nutrition and lifestyle interventions can thus facilitate collective decision making and prioritization in achieving lifestyle changes based on their lifestyle features, which in turn may increase

the compliance and effectiveness of lifestyle interventions.[39-41] As the majority of the lifestyle factors have been found to follow a dose-response relationship as to the risk of developing type 2 diabetes,[20,42-45] it is conceivable that such small changes across lifestyle or dietary patterns may have the potential to achieve substantial public health benefits,[46] although further studies are warranted. As shown in **chapter 2** and **3**, the estimates of associations between some dietary pattern scores and type 2 diabetes risk suggest dose-response relationships.

Ultra-processed Food Consumption Patterns

The public health relevance of ultra-processed food consumption patterns is worth a discussion. As an emerging risk factor, ultra-processed food (at least the concept) is not yet included in most of the dietary guidelines worldwide and has rarely been considered in public health interventions.[35] Although ultra-processed food has often been studied as a single risk factor, the results from **chapter 3** show that different consumption patterns of ultra-processed food may have different health consequences. This finding raises the important question for the development of dietary guidelines: what kinds of ultra-processed food should be restricted? Industrially-produced brown bread? Or fruit yogurt? Or "bitterballen"? Based on the results from **chapter 3**, particularly, ultra-processed savory snacks should be limited. Furthermore, the associations between ultra-processed food consumption patterns and incident type 2 diabetes may also be influenced by people's diabetes risk at baseline as well as other socio-demographic features such as age and education. Possible reverse causation is noted for the "sweet snack pattern". For public health interventions for ultra-processed food, it is important to consider different aspects of nutrition using an integrated approach. Future studies are encouraged to further explore this topic.

8

Box 2. Describing dietary and lifestyle patterns according to research and prevention objectives

One interesting fact from the findings of **part 1** is that lifestyle patterns can be derived in various ways, with different definitions and statistical methods, which largely depend on the research questions. This shows an exceptional merit of lifestyle pattern analysis, as it is highly flexible, that can be tailored to the needs of specific research and prevention objectives whilst preserving real-world insights into the clustering of lifestyle factors. In this thesis, the ultra-processed food consumption patterns show a case of dissecting one risk factor into several "pattern-based" risk profiles (**chapter 2**). The overall lifestyle pattern approach (**chapter 4**) combines different co-occurring lifestyle factors, which enables the description of the overall lifestyle risk profiles of the target populations that cannot be directly measured. While the former explores the internal heterogeneity of a single lifestyle factor, the latter aims to provide a stronger observational base for the clustering of several lifestyle factors within the population. On the other hand, the blood lipids-related dietary patterns are integrated with a hypothesis (**chapter 3**), which tests a specific diet-disease pathway. There are also supporting examples from other studies. For instance, by including co-occurring risk factors that may be related to lifestyle factors, such as psychosocial problems, joint lifestyle-risk factor patterns have been identified in some studies.[47,48] These joint lifestyle-risk factor patterns may guide the design of lifestyle interventions, that can be strengthened by additional pharmacological treatments when necessary.

Underlying Factors as Key to Boost Changes

Health education on lifestyle and type 2 diabetes prevention does not automatically lead to individual lifestyle behavior changes.[49,50] To boost changes towards healthy lifestyle for all, findings from this thesis highlight the importance of **considering and targeting underlying factors of lifestyle factors** in public health practice and policymaking. These underlying factors (as investigated in this thesis) include not only **individual socioeconomic status (education and income)**, but also **neighborhood socioeconomic status (contextual factors)**.

The effects of individual socioeconomic status on health have been well acknowledged. However, it often remains equivocal in public health practice and policy documents on which indicator should be focused.[5,51-54] It has been clearly shown in this thesis (**chapter 5** and **6**) that low education and low income affect health differentially and they are represented (to some extent overlapping) by different population groups.[54,55] This is further illustrated in **chapter 5** that education is found to have stronger direct effects on lifestyle factors compared with income. To boost changes towards healthy lifestyle, additional support should be provided for people with low education, while improving the education level of the general population should be prioritized as the long-term societal goal.

Equally important as improving individual socioeconomic status, neighborhood socioeconomic status (as an important contextual factor) is another piece of the puzzle to boost changes in lifestyle behaviors.[40,56-68] **Chapter 7** shows that neighborhood is associated with the lifestyle risk index, independent of individual socioeconomic disadvantage. When conducting lifestyle interventions, it is important to focus on disadvantaged neighborhoods with additional support for people with low individual socioeconomic status.

Furthermore, the findings from **chapter 7** support the importance and relevance of improving contextual factors to boost changes towards healthy lifestyle. Contextual factors include factors at regional/local level, such as neighborhood socioeconomic status (investigated in **chapter 7**), neighborhood walkability, safety, availability of

healthy food, fast food stands, sport clubs, and reducing air pollution. From a broader perspective, contextual factors also include generic factors, such as price measures (tobacco and sugar tax), cultural and social norms, commercial and marketing regulations, and health values and opinions from governments.

These contextual factors are closely related to individual lifestyle behaviors. Improving these contextual factors can create opportunities to promote mass lifestyle changes.[61-76] There is empirical evidence suggesting that changes in lifestyle related to changes in contextual factors are not selective but generic for the whole population.[46] Dutch people are proud of their bike tradition and vast network of cycle paths. This integral part of lifestyle is enabled by culture, social attitude, tradition, and the flat landscape, but also changes in contextual factors such as social movements in the 1970s, oil shortage, and later governments investment.[77]

It is conspicuous that healthy lifestyle cannot be simply realized by health education and interventions focusing on individual's lifestyle behaviors. Strategies to improve lifestyle at the population level should prioritize targeting and improving underlying factors of lifestyle factors using collective measures, which go beyond measures that focus on individual behavior changes. Individuals are indeed responsible for their own lifestyle and health. However, such responsibility is based on the prerequisite that relevant factors and opportunities are affordable, abundantly available, and accessible, which can enable people to achieve changes for healthy lifestyle.

Tackling Health Inequalities

In mainstream research and policies, improving individual lifestyle has been considered and practiced as the primary strategy to reduce health inequalities.[78] Such strategies are based on the large body of empirical evidence that unhealthy lifestyle is common in people with low socioeconomic status, which subsequently leads to the linear reasoning: the corollary of healthier lifestyle for socioeconomically-disadvantaged people is improved health, and thus reduction in health inequalities.

However, despite innumerable public health programs that have embraced nearly every aspect of lifestyle interventions over the past decades, we never hear the good news. On the contrary, health inequalities persist and even increase.[9,78,79]

Our observations and the empirical evidence did not lie. It is the misunderstanding and selective interpretation of the evidence that perpetuate myths about the root causes of health inequalities and how to address them. In both **chapter 5** and **6**, a substantial proportion of the socioeconomic gradients in type 2 diabetes (as high as 67%) and cardiovascular diseases (as high as 77%) is left unexplained, after taking into account a wide spectrum of risk factors including lifestyle, clinical biomarkers, obesity status, age, sex, and family history of diabetes. In the Whitehall II cohort study (a population of British civil servants), up to 55% of the socioeconomic gradients in type 2 diabetes (based on occupational class) remained unexplained, even if the long-term exposures (mean follow-up of 14.2 years) to major risk factors have been accounted for.[80] Such a substantial proportion of the unexplained socioeconomic gradients is prevalent in research but has rarely been discussed and studied further. It is highly unlikely to identify novel modifiable risk factors that may have a bigger contribution to type 2 diabetes risk than lifestyle factors and obesity.

This substantial proportion of the unexplained socioeconomic gradients in health is the quantified, irreparable health inequalities for socioeconomically-disadvantaged people compared with their least disadvantaged counterparts, of which the root cause lies in the fundamental inequities in the distribution of socioeconomic resources. Studies have shown that improving lifestyle is not a priority for people living below the poverty

8

line, unless economic resources become available.[72,81,82] This also helps explain the lack of effectiveness of lifestyle interventions focusing on disadvantaged people, since they failed to tackle socioeconomic factors.[78]

More strikingly, socioeconomic inequalities exist in the relationships between lifestyle and health, which further potently refutes the aforementioned linear reasoning that improving lifestyle may narrow health inequalities. In the Lifelines population, people with low education, adhering to high-quality diets, have on average 2 times higher risk of developing type 2 diabetes compared with their highly-educated counterparts. Although eliminating poor diet quality may improve health in all education groups, only 14.8% of cases may be preventable in the low education group, which clearly contrasts to 40.1% and 37.3% of preventable cases in the middle and high education groups, respectively.[83] Similar results have also been observed between lifestyle risk scores and mortality across socioeconomic groups in the UK Biobank cohort.[84] All the growing evidence highlights the disproportionate harms that are associated with lifestyle factors in socioeconomically-disadvantaged populations. The relative health benefit of lifestyle improvements is smaller in disadvantaged people compared with their less disadvantaged counterparts. Disadvantaged people do not choose to be unhealthy.

To tackle health inequalities, findings from this thesis highlight the importance of **improving education of the general population** and **ameliorating societal income inequality** to be the two fundamental strategies. These two fundamental strategies should be applied using generic measures rather than measures that focus on individuals. It has been shown that income instability, poor living condition, and lack of social and human capital are the top three contributors to ill health in European countries.[85] These factors are closely related to socioeconomic factors. Although they are not studied in this thesis, improving these factors, albeit generally not part of health policies but rather part of the broader societal picture, may nevertheless be a strategy with a substantial potential to reduce socioeconomic inequalities in health.

Importantly, **chapter 6** reveals that within each education or income level, substantial health disparities exist across strata of the other socioeconomic indicator. For the

current population, such within-group socioeconomic differences in health may lead the way towards the design of policies that do not require the adjustment of socioeconomic factors that are generally fixed such as education. Income supports individuals irrespective of their education level, which may contribute to their health even after their education has been attained.

Furthermore, **chapter 7** convincingly shows that neighborhood socioeconomic disadvantage positively modifies the association between individual socioeconomic disadvantage and the lifestyle risk index. This indicates that socioeconomically-disadvantaged individuals are disproportionately affected by neighborhood socioeconomic disadvantage, as they have a disproportionately higher chance of having unhealthy lifestyle if they live in disadvantaged neighborhoods. Hence, neighborhood socioeconomic disadvantage is an amplifier of the detrimental effects of individual socioeconomic disadvantage. This "amplifier effect" renders neighborhood-level factors well-suited intervention targets in addressing socioeconomic inequalities in lifestyle. For public health policies, while addressing individual socioeconomic inequalities, it is equally important to improve neighborhood-level factors where individual socioeconomic inequalities in health can develop and thrive.

The previous section discusses the relevance and importance of targeting contextual factors (including but not limited to neighborhood socioeconomic disadvantage investigated in **chapter 7**) for improving lifestyle in the general population. Nevertheless, the mechanisms of how contextual factors and their interplay with individual socioeconomic status affect health inequalities remain poorly understood. Future studies are warranted to explore and better understand the nature and mechanisms of these contextual factors – to better identify the non-individual contextual factors in health risk.[86]

This thesis focuses on education and income-related health inequalities. From a broader perspective, health inequalities should also be addressed as a consequence of the unequal distribution of various forms of resources, such as social capital (social contacts), cultural capital (participation in clubs), and attractiveness and personality

8

capital. These resources are intertwined with each other and are related to even broader contextual factors such as economic growth and climate change, which are much beyond individual lifestyle behavioral choices.[78,87] It is important to acknowledge the complexity of health inequalities and the fundamental inequities in the distribution of resources. Imputing bad health purely to individual responsibility is the ignorance and arrogant disregard for fundamental societal inequities.

In the context of The Netherlands, a developed country with a high coverage of government-subsidized public education system and a well-structured social security system, the persistent socioeconomic patterning in health inequalities observed cannot be addressed only by extensive public health lifestyle interventions that focus on individual behavior changes. Tackling health inequalities requires critical thinking, innovative methods, integrated and generic approaches, and deeper insights into the underlying mechanisms. **Fig. 1** presents the contribution of this thesis findings to the understanding of health inequalities, as well as the knowledge and evidence gaps on the mechanisms of health inequalities identified in this thesis. There is also a need for a better translation of available evidence into public health practice, such as targeting lifestyle patterns studied in this thesis and regulating fast food outlets.[9,22,78,79,88-92] Governments and policymakers should clearly acknowledge the root causes of health inequalities, and reconceptualize health as the outcome of fundamental inequities in resources. Tackling health inequalities should be institutionalized and actions should be taken by all sectors, with science, with determination, and without hesitation.
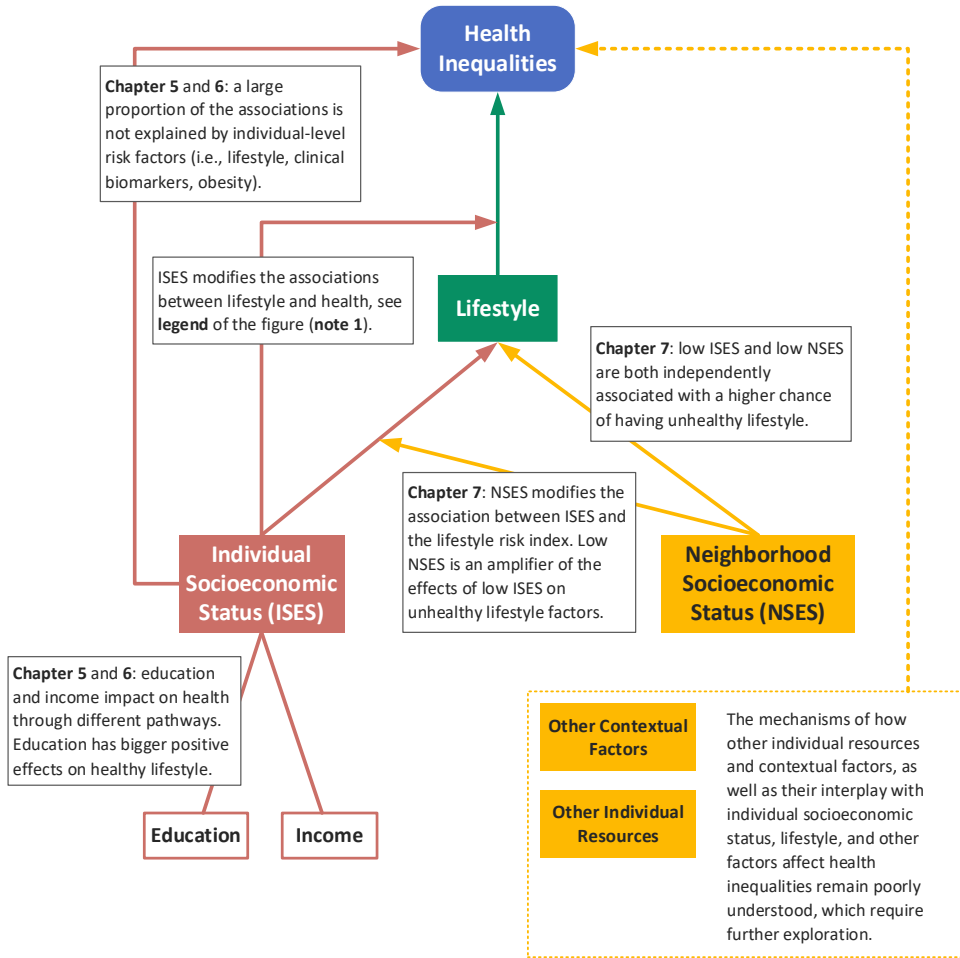
Fig. 1. Illustration of the contribution of this thesis findings to the understanding of health inequalities, as well as the knowledge and evidence gaps on the mechanisms of health inequalities identified in this thesis. Solid lines indicate findings of this thesis and other studies in the Lifelines cohort study. Dotted lines indicate knowledge and evidence gaps that warrant further investigation. The arrowheads indicates the direction of the effects.

Note 1: in this study, socioeconomic inequalities are found to exist in the relationships between diet quality and health. In the Lifelines population, people with low education, adhering to high-quality diets, have on average 2 times higher risk of developing type 2 diabetes compared with their highly-educated counterparts. Although eliminating poor diet quality may improve health in all education groups, only 14.8% of cases may be preventable in the low education group, which clearly contrasts to 40.1% and 37.3% of preventable cases in the middle and high education groups, respectively.[83]

Future Perspectives

Underlying Factors of Lifestyle Patterns

In this thesis we shed new light on underlying factors of lifestyle factors, and support their relevance for health outcomes. Yet, to translate these insights into better prevention strategies, several issues require further exploration. In **chapter 3**, the adherence to ultra-processed food consumption patterns depends not only on diabetes risk at baseline, but also on age and education.[35] While the assessment of robust lifestyle patterns provides strategies to map lifestyle habits at the aggregate level, further analysis of the relevant underlying factors can assist in characterizing lifestyle pattern groups, which can provide guidance on better targeting strategies.

Regional Approaches in Lifestyle Promotion

The prospect of a regional approach for lifestyle promotion is promising. So far, however, the empirical evidence for such approaches in public health practice is sparse. Monitoring and evaluation of previous programs are necessary. How are lifestyle factors related to geographical factors requires further exploration. It has been clearly shown in **chapter 7** that neighborhood-level factors are related to unhealthy lifestyle.[58] Spatial clustering (based on neighborhoods) has been identified for dietary patterns and ultra-processed food consumption using the Lifelines data in the north of The Netherlands.[40,93] These findings provide interesting leads for the design of innovative approaches for lifestyle interventions, hinting at the role of place as an important underlying factor for lifestyle patterns (which is in line with this thesis) and also the role of place as a tangible platform to implement policies and public health programs.

Advanced Techniques for Health Research

Advanced quantitative techniques may aid in better understanding the relationships among lifestyle, underlying factors, and health. This includes not only emerging techniques such as machine learning, but also methods that are often applied in other fields but are less familiar to researchers in the field of nutrition, public health, and epidemiology. For example, social network analysis reveals that obesity and smoking

behavior seem to spread through social ties.[94,95] In addition, current technologies (such as digital e-health) and emerging lifestyle factors (such as social media use) create new opportunities and challenges for improving lifestyle and health.[96-98] Interdisciplinary research is therefore encouraged and also highly relevant for creating evidence that can guide innovative strategies for public health prevention.

Healthy and Sustainable Lifestyle

Our lifestyle is not only related to health, but also has a substantial environmental impact. This is especially prominent for some dietary factors such as meat and dairy products, that are estimated to contribute to 50% of dietary greenhouse gas emissions.[99,100] Future research is needed to define sustainable lifestyle and investigate its health potential. For public health practice and policymaking, it is important to consider sustainability components in lifestyle interventions.

8

Overall Conclusions

Part 1 of this thesis describes several dietary and lifestyle patterns, which contributes to new knowledge about how different dietary and lifestyle factors cluster within the population. These dietary and lifestyle patterns are found to be differentially associated with risk of type 2 diabetes. Better-targeted lifestyle interventions can be designed and enabled by targeting dietary and lifestyle patterns at the population level.

Part 2 of this thesis shows that low individual socioeconomic status (education and income), and low neighborhood socioeconomic status are persistently associated with unhealthy lifestyle factors. Substantial socioeconomic inequalities in health (type 2 diabetes) have been identified, which can only be partly explained by a wide spectrum of factors, including socio-demographic factors, lifestyle factors, obesity status, clinical biomarkers, and family history of diabetes. To tackle inequalities in lifestyle and health, additional support should be provided to socioeconomically-disadvantaged people. Contextual factors – including regional/local factors (such as neighborhood-level factors) and generic factors (such as sugar and tobacco tax) – should be targeted to enable healthy lifestyle changes. Improving education level of the general population and ameliorating income inequalities using generic measures are the two fundamental strategies.
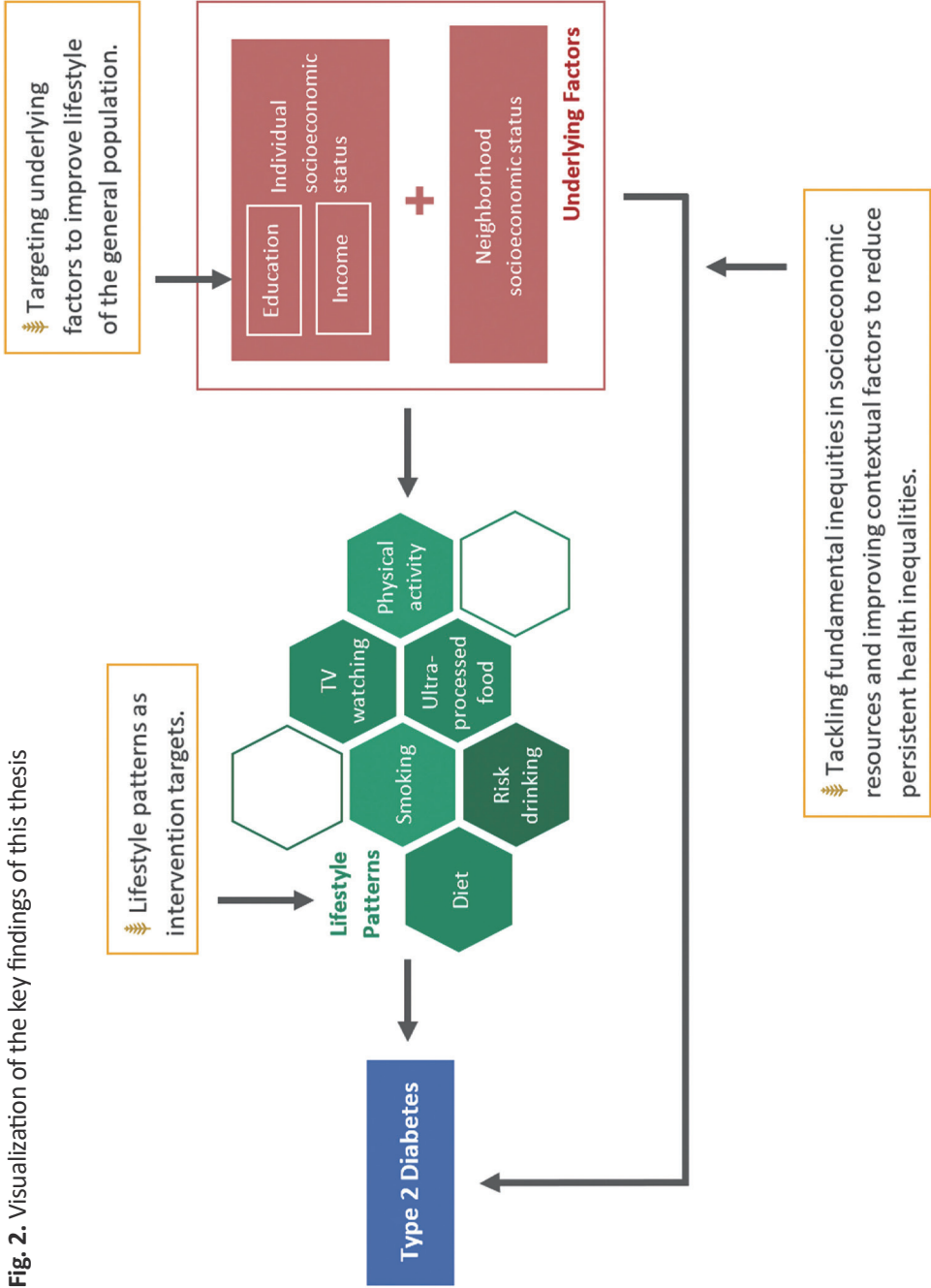
Fig. 2. Visualization of the key findings of this thesis

8

References

1.   Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int J Epidemiol 2015; 44(4): 1172-80.

2.   Sijtsma A, Rienks J, van der Harst P, Navis G, Rosmalen JGM, Dotinga A. Cohort Profile Update: Lifelines, a three-generation cohort study and biobank. Int J Epidemiol 2022; 51(5): e295-e302.

3.   Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. PLOS ONE 2015; 10(9): e0137203.

4.   UNESCO. International Standard Classification of Education (ISCED). 1997.

5.   Shi J, Tarkiainen L, Martikainen P, van Raalte A. The impact of income definitions on mortality inequalities. SSM Popul Health 2021; 15: 100915.

6.   OECD. Part-time and Partly Equal: Gender and Work in the Netherlands, 2019.

7.   Bahk J, Jang SM, Jung-Choi K. Increased breast cancer mortality only in the lower education group: age-period-cohort effect in breast cancer mortality by educational level in South Korea, 1983-2012. Int J Equity Health 2017; 16(1): 56.

8.   Hoveling LA, Liefbroer AC, Bultmann U, Smidt N. Understanding socioeconomic differences in incident metabolic syndrome among adults: What is the mediating role of health behaviours? Prev Med 2021; 148: 106537.

9.   Mackenbach JP. The persistence of health inequalities in modern welfare states: the explanation of a paradox. Soc Sci Med 2012; 75(4): 761-9.

10.   Batis C, Mendez MA, Sotres-Alvarez D, Gordon-Larsen P, Popkin B. Dietary pattern trajectories during 15 years of follow-up and HbA1c, insulin resistance and diabetes prevalence among Chinese adults. J Epidemiol Community Health 2014; 68(8): 773-9.

11.   Pachucki MA. Food pattern analysis over time: unhealthful eating trajectories predict obesity. Int J Obes (Lond) 2012; 36(5): 686-94.

12.   Liu MW, McNaughton SA, He QQ, Leech R. Longitudinal trajectories of diet quality and subsequent mortality among Chinese adults: results from the China health and nutrition survey 1997-2015. Int J Behav Nutr Phys Act 2021; 18(1): 51.

13.   Watt HC, Carson C, Lawlor DA, Patel R, Ebrahim S. Influence of life course socioeconomic position on older women's health behaviors: findings from the British Women's Heart and Health Study. Am J Public Health 2009; 99(2): 320-7.

14.   Newell SA, Girgis A, Sanson-Fisher RW, Savolainen NJ. The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: a critical review. Am J Prev Med 1999; 17(3): 211-29.

15.   Celis-Morales CA, Perez-Bravo F, Ibanez L, Salas C, Bailey ME, Gill JM. Objective vs. self-reported physical activity and sedentary time: effects of measurement method on relationships with risk biomarkers. PLOS ONE 2012; 7(5): e36345.

16.   Hessels NR, Zhu Y, Bakker SJL, de Borst MH, Navis GJ, Riphagen IJ. Low Sodium Intake, Low

Protein Intake, and Excess Mortality in an Older Dutch General Population Cohort: Findings in the Prospective Lifelines-MINUTHE Study. Nutrients 2023; 15(2).

17.   Howe LD, Tilling K, Galobardes B, Lawlor DA. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. Epidemiology 2013; 24(1): 1-9.

18.   Eng JY, Moy FM, Bulgiba A, Rampal S. Consistency and Generalizability of Dietary Patterns in a Multiethnic Working Population. J Acad Nutr Diet 2018; 118(7): 1249-62 e3.

19.   Imamura F, Lichtenstein AH, Dallal GE, Meigs JB, Jacques PF. Generalizability of dietary patterns associated with incidence of type 2 diabetes mellitus. Am J Clin Nutr 2009; 90(4): 1075-83.

20.   Jannasch F, Dietrich S, Bishop TRP, et al. Associations between exploratory dietary patterns and incident type 2 diabetes: a federated meta-analysis of individual participant data from 25 cohort studies. Eur J Nutr 2022; 61(7): 3649-67.

21.   Duan MJ, Dekker LH, Carrero JJ, Navis G. Lifestyle patterns and incident type 2 diabetes in the Dutch lifelines cohort study. Prev Med Rep 2022; 30: 102012.

22.   Kivela J, Wikstrom K, Virtanen E, et al. Obtaining evidence base for the development of Feel4Diabetes intervention to prevent type 2 diabetes- a narrative literature review. BMC Endocr Disord 2020; 20(Suppl 1): 140.

23.   Meader N, King K, Moe-Byrne T, et al. A systematic review on the clustering and co-occurrence of multiple risk behaviours. BMC Public Health 2016; 16: 657.

24.   Noble N, Paul C, Turon H, Oldmeadow C. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. Prev Med 2015; 81: 16-41.

25.   Tuomilehto J, Schwarz P, Lindstrom J. Long-term benefits from lifestyle interventions for type 2 diabetes prevention: time to expand the efforts. Diabetes Care 2011; 34 Suppl 2: S210-4.

26.   Van Stappen V, Cardon G, De Craemer M, et al. The effect of a cluster-randomized controlled trial on lifestyle behaviors among families at risk for developing type 2 diabetes across Europe: the Feel4Diabetes-study. Int J Behav Nutr Phys Act 2021; 18(1): 86.

27.   Van Stappen V, Latomme J, Cardon G, et al. Barriers from Multiple Perspectives Towards Physical Activity, Sedentary Behaviour, Physical Activity and Dietary Habits When Living in Low Socio-Economic Areas in Europe. The Feel4Diabetes Study. Int J Environ Res Public Health 2018; 15(12).

28.   The Diabetes Prevention Program Research Group, Knowler WC, Fowler SE, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. The Lancet 2009; 374(9702): 1677-86.

29.   Pan XR, Li GW, Hu YH, et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. Diabetes Care 1997; 20(4): 537-44.

30.   Lindstrom J, Peltonen M, Eriksson JG, et al. Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). Diabetologia 2013; 56(2): 284-93.

31.   Kyrou I, Tsigos C, Mavrogianni C, et al. Sociodemographic and lifestyle-related risk factors

8

for identifying vulnerable groups for type 2 diabetes: a narrative review with emphasis on data from Europe. BMC Endocr Disord 2020; 20(Suppl 1): 134.

32.   Manios Y, Lambrinou CP, Mavrogianni C, et al. Lifestyle Changes Observed among Adults Participating in a Family- and Community-Based Intervention for Diabetes Prevention in Europe: The 1st Year Results of the Feel4Diabetes-Study. Nutrients 2020; 12(7).

33.   de Hollander E, Hupkens C, van Dorsselaar S, et al. De Leefstijlmonitor: Cijfers voor gezondheidsbeleid. TSG 2022; 100(3): 98-106.

34.   Duan MJ, Dekker LH, Carrero JJ, Navis G. Blood lipids-related dietary patterns derived from reduced rank regression are associated with incident type 2 diabetes. Clin Nutr 2021; 40(7): 4712-9.

35.   Duan MJ, Vinke PC, Navis G, Corpeleijn E, Dekker LH. Ultra-processed food and incident type 2 diabetes: studying the underlying consumption patterns to unravel the health effects of this heterogeneous food category in the prospective Lifelines cohort. BMC Med 2022; 20(1): 7.

36.   Persona (user experience). Nov 7th 2022 ed: Wikipedia.

37.   Southey F. Hipster, athlete, or vegan activist? 'Personas' developed in Norway to help food brands think like consumers. 2019. https://www.foodnavigator.com/Article/2019/12/13/Nofima-develops-personas-to-help-food-brands-think-like-consumer-groups (accessed Dec 14th 2022).

38.   Motivaction. De acht Mentality-milieus. https://www.motivaction.nl/mentality/de-acht-mentality-milieus (accessed Dec 14th 2022).

39.   Mozaffarian D. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity: A Comprehensive Review. Circulation 2016; 133(2): 187-225.

40.   Dekker LH, Rijnks RH, Strijker D, Navis GJ. A spatial analysis of dietary patterns in a large representative population in the north of The Netherlands- the Lifelines cohort study. Int J Behav Nutr Phys Act 2017; 14(1): 166.

41.   2015 Dietary Guidelines Advisory Committee. Scientific Report of the 2015 Dietary Guidelines Advisory Committee. Washington, DC: USDA, 2015.

42.   Knott C, Bell S, Britton A. Alcohol Consumption and the Risk of Type 2 Diabetes: A Systematic Review and Dose-Response Meta-analysis of More Than 1.9 Million Individuals From 38 Observational Studies. Diabetes Care 2015; 38(9): 1804-12.

43.   Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ. Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. Eur J Epidemiol 2015; 30(7): 529-42.

44.   Patterson R, McNamara E, Tainio M, et al. Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. Eur J Epidemiol 2018; 33(9): 811-29.

45.   Pan A, Wang Y, Talaei M, Hu FB, Wu T. Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis. Lancet Diabetes Endocrinol 2015; 3(12): 958-67.

46.   Rose G, Khaw K-T, Marmot M. Rose's Strategy of Preventive Medicine. Oxford University Press; 2008.

47. Luo J, Dinh P, Hendryx M, Li W, Robinson J, Margolis KL. Risk Patterns and Mortality in Postmenopausal Women Using Latent Class Analysis. Am J Prev Med 2021; 61(5): e225-e33.

48. Hendryx M, Dinh P, Chow A, et al. Lifestyle and Psychosocial Patterns and Diabetes Incidence Among Women with and Without Obesity: a Prospective Latent Class Analysis. Prev Sci 2020; 21(6): 850-60.

49. Kelly MP, Barker M. Why is changing health-related behaviour so difficult? Public Health 2016; 136: 109-16.

50. Arlinghaus KR, Johnston CA. Advocating for Behavior Change With Education. Am J Lifestyle Med 2018; 12(2): 113-6.

51. Busch MCM, Schrijvers CTM. Effecten van leefstijlinterventies gericht op lagere sociaaleconomische groepen. Bilthoven: RIVM, 2010.

52. Blanquet M, Legrand A, Pelissier A, Mourgues C. Socio-economics status and metabolic syndrome: A meta-analysis. Diabetes Metab Syndr 2019; 13(3): 1805-12.

53. Geyer S, Hemstrom O, Peter R, Vagero D. Education, income, and occupational class cannot be used interchangeably in social epidemiology. Empirical evidence against a common practice. J Epidemiol Community Health 2006; 60(9): 804-10.

54. Duan MF, Zhu Y, Dekker LH, et al. Effects of Education and Income on Incident Type 2 Diabetes and Cardiovascular Diseases: a Dutch Prospective Study. J Gen Intern Med 2022.

55. Duan MJ, Dekker LH, Carrero JJ, Navis G. Using Structural Equation Modeling to Untangle Pathways of Risk Factors Associated with Incident Type 2 Diabetes: the Lifelines Cohort Study. Prev Sci 2022; 23(7): 1090-100.

56. van Erpecum CL, van Zon SKR, Bultmann U, Smidt N. The association between the presence of fast-food outlets and BMI: the role of neighbourhood socio-economic status, healthy food outlets, and dietary factors. BMC Public Health 2022; 22(1): 1432.

57. Dekker LH, Rijnks RH, Navis GJ. Regional variation in type 2 diabetes: evidence from 137 820 adults on the role of neighbourhood body mass index. Eur J Public Health 2020; 30(1): 189-94.

58. Zhu Y, Duan MJ, Riphagen IJ, et al. Separate and combined effects of individual and neighbourhood socio-economic disadvantage on health-related lifestyle risk factors: a multilevel analysis. Int J Epidemiol 2022; 50(6): 1959-69.

59. Ludwig J, Sanbonmatsu L, Gennetian L, et al. Neighborhoods, obesity, and diabetes--a randomized social experiment. N Engl J Med 2011; 365(16): 1509-19.

60. Algren MH, Bak CK, Berg-Beckhoff G, Andersen PT. Health-Risk Behaviour in Deprived Neighbourhoods Compared with Non-Deprived Neighbourhoods: A Systematic Literature Review of Quantitative Observational Studies. PLOS ONE 2015; 10(10): e0139297.

61. Hajna S, Ross NA, Brazeau AS, Belisle P, Joseph L, Dasgupta K. Associations between neighbourhood walkability and daily steps in adults: a systematic review and meta-analysis. BMC Public Health 2015; 15: 768.

62. Richardson AS, Meyer KA, Howard AG, et al. Multiple pathways from the neighborhood food environment to increased body mass index through dietary behaviors: A structural equation-

based analysis in the CARDIA study. Health Place 2015; 36: 74-87.

63.   McCormack GR, Shiell A. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. Int J Behav Nutr Phys Act 2011; 8: 125.

64.   Christine PJ, Auchincloss AH, Bertoni AG, et al. Longitudinal Associations Between Neighborhood Physical and Social Environments and Incident Type 2 Diabetes Mellitus: The Multi-Ethnic Study of Atherosclerosis (MESA). JAMA Intern Med 2015; 175(8): 1311-20.

65.   Boylan JM, Robert SA. Neighborhood SES is particularly important to the cardiovascular health of low SES individuals. Soc Sci Med 2017; 188: 60-8.

66.   Park S, Zachary WW, Gittelsohn J, Quinn CC, Surkan PJ. Neighborhood Influences on Physical Activity Among Low-Income African American Adults With Type 2 Diabetes Mellitus. Diabetes Educ 2020; 46(2): 181-90.

67.   Strak M, Janssen N, Beelen R, et al. Associations between lifestyle and air pollution exposure: Potential for confounding in large administrative data cohorts. Environ Res 2017; 156: 364-73.

68.   de Boer WIJ, Dekker LH, Koning RH, Navis GJ, Mierau JO. How are lifestyle factors associated with socioeconomic differences in health care costs? Evidence from full population data in the Netherlands. Prev Med 2020; 130: 105929.

69.   Wilkinson AL, Scollo MM, Wakefield MA, Spittal MJ, Chaloupka FJ, Durkin SJ. Smoking prevalence following tobacco tax increases in Australia between 2001 and 2017: an interrupted time-series analysis. Lancet Public Health 2019; 4(12): e618-e27.

70.   Kelly B, Cretikos M, Rogers K, King L. The commercial food landscape: outdoor food advertising around primary schools in Australia. Aust N Z J Public Health 2008; 32(6): 522-8.

71.   Diepeveen S, Ling T, Suhrcke M, Roland M, Marteau TM. Public acceptability of government intervention to change health-related behaviours: a systematic review and narrative synthesis. BMC Public Health 2013; 13: 756.

72.   Teuscher D, Bukman AJ, van Baak MA, Feskens EJ, Renes RJ, Meershoek A. Challenges of a healthy lifestyle for socially disadvantaged people of Dutch, Moroccan and Turkish origin in the Netherlands: a focus group study. Crit Public Health 2015; 25(5): 615-26.

73.   Over EA, Feenstra TL, Hoogenveen RT, Droomers M, Uiters E, van Gelder BM. Tobacco control policies specified according to socioeconomic status: health disparities and cost-effectiveness. Nicotine Tob Res 2014; 16(6): 725-32.

74.   Basu S, Madsen K. Effectiveness and equity of sugar-sweetened beverage taxation. PLOS MED 2017; 14(6): e1002327.

75.   Hashemi N, Sebar B, Harris N. The relationship between cultural capital and lifestyle health behaviours in young people: a systematic review. Public Health 2018; 164: 57-67.

76.   Dixon HG, Scully ML, Wakefield MA, White VM, Crawford DA. The effects of television advertisements for junk food versus nutritious food on children's food attitudes and preferences. Soc Sci Med 2007; 65(7): 1311-23.

77.   Cycling in the Netherlands. Oct 14th 2022 ed: Wikipedia.

78.   Hagenaars L, Waterlander W, den Hertog K, Stronks K. A resilient public health in 2030: #buthow? TSG 2023: 1-4.

79.   Mackenbach JP. Persistence of social inequalities in modern welfare states: Explanation of a paradox. Scand J Public Health 2017; 45(2): 113-20.

80.   Stringhini S, Tabak AG, Akbaraly TN, et al. Contribution of modifiable risk factors to social inequalities in type 2 diabetes: prospective Whitehall II cohort study. BMJ 2012; 345: e5452.

81.   Withall J, Jago R, Fox KR. Why some do but most don't. Barriers and enablers to engaging low-income groups in physical activity programmes: a mixed methods study. BMC Public Health 2011; 11: 507.

82.   Hoveling LA. Unravelling socioeconomic health inequalities: targets for preventing metabolic syndrome and major depressive disorder. Groningen, The Netherlands: University of Groningen; 2022.

83.   Vinke PC, Navis G, Kromhout D, Corpeleijn E. Socio-economic disparities in the association of diet quality and type 2 diabetes incidence in the Dutch Lifelines cohort. EClinicalMedicine 2020; 19: 100252.

84.   Foster HME, Celis-Morales CA, Nicholl BI, et al. The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: a prospective analysis of the UK Biobank cohort. Lancet Public Health 2018; 3(12): e576-e85.

85.   WHO. Healthy, prosperous lives for all: the European Health Equity Status Report Executive Summary. Copenhagen, Denmark, 2019.

86.   Beulens JWJ, Pinho MGM, Abreu TC, et al. Environmental risk factors of type 2 diabetes-an exposome approach. Diabetologia 2022; 65(2): 263-74.

87.   Qi Y, Vrooman JC, Almansa J, Ots P, Brouwer S, Reijneveld SA. A capital-based approach to better understand health inequalities: Theoretical and empirical explorations. SSM Popul Health 2023; 21: 101309.

88.   Marmot M, Bell R. Fair society, healthy lives. Public Health 2012; 126 Suppl 1: S4-S10.

89.   Stronks K, Mackenbach JP. Evaluating the effect of policies and interventions to address inequalities in health: lessons from a Dutch programme. Eur J Public Health 2006; 16(4): 346-53.

90.   Mackenbach JP, Stronks K. The development of a strategy for tackling health inequalities in the Netherlands. Int J Equity Health 2004; 3(1): 11.

91.   Peters D, Harting J, van Oers H, Schuit J, de Vries N, Stronks K. Manifestations of integrated public health policy in Dutch municipalities. Health Promot Int 2016; 31(2): 290-302.

92.   Pineda E, Poelman MP, Aaspollu A, et al. Policy implementation and priorities to create healthy food environments using the Healthy Food Environment Policy Index (Food-EPI): A pooled level analysis across eleven European countries. Lancet Reg Health Eur 2022; 23: 100522.

93.   Cai Q. Novel approaches to assess diet and kidney health risk. Groningen, The Netherlands: University of Groningen; 2021.

8

94.   Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007; 357(4): 370-9.

95.   Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008; 358(21): 2249-58.

96.   Agachi E, Bijmolt THA, Mierau JO, van Ittersum K. Adoption of the Website and Mobile App of a Preventive Health Program Across Neighborhoods With Different Socioeconomic Conditions in the Netherlands: Longitudinal Study. JMIR Hum Factors 2022; 9(1): e32112.

97.   Klassen KM, Borleis ES, Brennan L, Reid M, McCaffrey TA, Lim MS. What People "Like": Analysis of Social Media Strategies Used by Food Industry Brands, Lifestyle Brands, and Health Promotion Organizations on Facebook and Instagram. J Med Internet Res 2018; 20(6): e10227.

98.   De Jans S, Spielvogel I, Naderer B, Hudders L. Digital food marketing to children: How an influencer's lifestyle can stimulate healthy food choices among children. Appetite 2021; 162: 105182.

99.   Eckl MR, Biesbroek S, Van't Veer P, Geleijnse JM. Replacement of Meat with Non-Meat Protein Sources: A Review of the Drivers and Inhibitors in Developed Countries. Nutrients 2021; 13(10).

100.   van Dooren C, Man L, Seves M, Biesbroek S. A Food System Approach for Sustainable Food-Based Dietary Guidelines: An Exploratory Scenario Study on Dutch Animal Food Products. Front Nutr 2021; 8: 712970.

8

English Summary

Nederlandse Samenvatting

中文摘要

Acknowledgements / Dankwoord / 致谢

Description of the Artworks

PhD Portfolio

List of Publications

About the Author

Appendices

ENGLISH SUMMARY

What is type 2 diabetes? How does it harm our health and society?

Type 2 diabetes is a chronic disease characterized by long-term high blood sugar. Prolonged periods of abnormal high blood sugar can result in damages to circulatory, nervous, and immune systems, as well as multiple organs such as the heart, kidneys, and eyes. These damages make individuals with type 2 diabetes more susceptible to developing cardiovascular diseases (such as atherosclerosis), kidney disease, eye diseases (such as retinopathy), and skin infections. For type 2 diabetes patients, these diseases compromise the quality of life, increase the risk of death, and substantially decrease their life expectancy. The prevalence of type 2 diabetes is high, resulting in significant health problems and imposing a heavy socioeconomic burden on society. It is estimated that there were approximately 480 million people having type 2 diabetes globally in 2021. How to prevent type 2 diabetes – curbing its pandemic – in the population is one of the most critical and urgent issues for public health and society.

Unhealthy lifestyle is one of the most important causes (often referred to as risk factors) of type 2 diabetes. Unhealthy lifestyle includes, among other factors, poor diet, smoking, binge drinking, lack of physical activity, and poor sleep quality. Substantial evidence has demonstrated that improving lifestyle can substantially lower the risk of type 2 diabetes. Lifestyle interventions have been prioritized as one of the most important primary prevention strategies for type 2 diabetes in public health practice and policy.

What problems and challenges do we face in the prevention of type 2 diabetes?

However, over the past few decades, large-scale lifestyle intervention programs targeting the general population have achieved limited success. In the general population, lifestyle interventions for type 2 diabetes prevention primarily face the following problems and challenges:

(1) Type 2 diabetes is not solely caused by a single unhealthy lifestyle factor, yet current large-scale lifestyle programs primarily focus on a single one. There is a lack of attention on the combination of multiple unhealthy lifestyle factors – lifestyle patterns.

Additionally, the relationship between lifestyle patterns and the risk of developing diabetes is not yet well understood.

(2) Current lifestyle intervention programs mainly apply a "one-size-fits-all" approach, in which everyone receives the same intervention measures. However, in real life, lifestyle is different between different individuals. A "one-size-fits-all" strategy, therefore, fails to effectively differentiate the needs of different groups for improving lifestyle within the population, which may compromise the effectiveness of lifestyle intervention programs.

(3) Population-level lifestyle programs rarely take into account factors that influence lifestyle (referred to as underlying factors of lifestyle in this thesis), such as socioeconomic status and contextual factors. How these factors collectively affect type 2 diabetes risk are also largely unknown.

In light of these issues and challenges, what research have I conducted?

Part 1 of this thesis described several dietary and lifestyle patterns in the general population. The relationships of these dietary and lifestyle patterns with type 2 diabetes risk were studied. **Part 2** of this thesis investigated the relationships of lifestyle factors with their underlying factors, with a special focus on individual socioeconomic status and neighborhood socioeconomic status. How these factors collectively affect type 2 diabetes risk was subsequently investigated. These studies are based on the Lifelines Cohort, which is a large population cohort in the northern part of the Netherlands.

What are the findings of each chapter?

Chapter 2 describes a dietary pattern. A closer adherence to this dietary pattern is related to a higher risk of type 2 diabetes. This dietary pattern is associated with blood lipids, meaning it may affect the risk of type 2 diabetes by altering blood lipid levels. This dietary pattern is characterized by high consumption of added sugar, sugary beverages, and juice, and low consumption of fruits, vegetables, nuts/seeds, whole grains, and tea.

Chapter 3 shows that eating more ultra-processed food is related to a higher risk of type 2 diabetes. Eating a healthy diet may not compensate for the detrimental effects of ultra-processed food on health. However, it should be noted that not all types of ultra-processed food are related to type 2 diabetes risk. By studying consumption patterns of ultra-processed food, we found that a pattern high in consumption of cold savory snacks (such as cheese, deli meat, and savory spreads for crackers and French bread) and a pattern high in consumption of warm savory snacks (such as fried snacks, fries, and snack sauce) are associated with higher type 2 diabetes risk. However, a pattern high in traditional Dutch cuisine (such as sliced bread, lunch meat, and gravy) is not associated with type 2 diabetes risk, while a pattern high in sweet snacks (such as cookies, cakes, and chocolate) is associated with lower type 2 diabetes risk. This lower risk of the sweet snack pattern may be due to the "reverse causation" in epidemiology. To provide a specific explanation, people with a high risk of diabetes at the start of the study (such as those with a family history of the disease) may try to minimize their consumption of sweet-tasting foods with high sugar content in their daily diet. We have also found evidence that supports the existence of this reverse causation.

Chapter 4 describes several lifestyle patterns in the population and has found differential associations of these lifestyle patterns with risk of type 2 diabetes. Using the "healthy lifestyle pattern" as the reference, the "unhealthy lifestyle pattern" and the "poor diet and low physical activity pattern" are associated with higher risk of type 2 diabetes, whereas the "couch potato pattern" and the "risk drinking pattern" were not associated with type 2 diabetes risk. Interestingly, people from the "couch potato pattern", despite having the longest TV watching time among the entire population, had the highest weekly physical activity level and were mainly non-smokers. People from the "risk drinking pattern" had the lowest TV watching time among the population. These findings suggest that lifestyle factors tend to cluster in unique behavioral patterns within the population. Furthermore, while previous research has shown that each individual unhealthy lifestyle factor is associated with a higher risk of type 2 diabetes when studied separately, the combination of these different lifestyle factors – in the form of lifestyle pattern – may not always correlate with risk of type 2 diabetes. Therefore, it is important to consider the combined effects of different lifestyle factors on health.

Chapter 5 examines a conceptual model that encompasses key risk factors for type 2 diabetes. This model not only incorporates risk factors from various domains, such as socioeconomic status, blood lipids, obesity, and lifestyle factors, but also includes their interrelationships. The findings of this study indicate that larger waist circumference, rather than higher BMI, and poor blood lipids have the biggest impact on increasing type 2 diabetes risk. Among lifestyle factors, reducing TV watching time and increasing physical activity level may yield significant benefits in improving blood lipids and reducing waist circumference. It is worth noting that a higher education level, as opposed to a higher income, is associated with better lifestyle. However, a lower income is directly linked to higher risk of type 2 diabetes, which cannot be accounted for by the aforementioned risk factors. Through the analysis of this conceptual model, we have quantified the interrelationships among the major risk factors for type 2 diabetes within the population. Focusing on the most influential risk factors identified within each domain as priority prevention targets may have the potential to enhance the effectiveness of type 2 diabetes prevention at the population level.

Chapter 6 examines the relationships between two socioeconomic indicators, namely education and income, and the risks of type 2 diabetes and cardiovascular diseases. The results show that lower education and lower income are both associated with higher disease risks, and these associations are independent of each other. Risk factors such as blood lipids, blood pressure, BMI, lifestyle factors, and sex can only partially explain these associations. Additionally, it is observed that individuals with lower education are not necessarily those with lower income, and vice versa. This research further confirms the substantial socioeconomic inequalities in health in the general population. Improving personal lifestyle, blood pressure, and blood lipids alone may not effectively address this health inequality. Furthermore, when conducting health research and formulating health interventions and policies, it is crucial to consider the separate impacts of education and income on health, rather than conflating them together.

Chapter 7 investigates the influence of individual socioeconomic status and neighborhood socioeconomic status on lifestyle. The study has found that lower individual socioeconomic status and lower neighborhood socioeconomic status are both independently associated with poorer lifestyle. Additionally, statistically,

A

there is an interaction between the effects of individual socioeconomic status and neighborhood socioeconomic status on lifestyle. This interaction suggests that residents living in lower socioeconomic status neighborhoods, who also have lower individual socioeconomic status themselves, are more susceptible to the adverse influence from the disadvantaged neighborhood on their lifestyle, compared with their more affluent neighbors. These findings highlight the need to not only consider individual socioeconomic status, but also provide additional health resources and social opportunities for socioeconomically disadvantaged neighborhoods when formulating health programs and policies.

In summary, how do these research findings assist in the development of better type 2 diabetes prevention strategies for the general population?

(1) Focusing on dietary and lifestyle patterns as intervention targets, rather than the generic "one-size-fits-all" approach

Part 1 of this thesis describes several dietary and lifestyle patterns in the general population, wherein each pattern represents individuals with similar dietary or lifestyle habits. These lifestyle patterns are differentially associated with type 2 diabetes risk, demonstrating each of their clinical relevance. The unique characteristics of these dietary and lifestyle patterns make it possible to design tailored dietary and lifestyle interventions for each dietary and lifestyle pattern group, which enable a "personalized" lifestyle intervention approach at an intermediate level of the population. As opposed to the generic "one-size-fits-all" approach, this approach allows a more effective differentiation of lifestyle intervention needs among different groups and facilitates combined interventions targeting multiple unhealthy lifestyle factors. Given these advantages, this pattern-based strategy holds great potential in promoting healthy lifestyle and preventing type 2 diabetes in the general population.

(2) Promoting healthy lifestyle needs to target underlying factors of lifestyle

Promoting healthy lifestyle for the general population cannot be simply achieved through health education and interventions focusing on individuals. The findings in **Part 2** of this thesis emphasize the need to improve the underlying factors of lifestyle factors in public health practice and policy-making. These underlying factors include

not only individual socioeconomic status (such as education and income) but also contextual factors (such as neighborhood socioeconomic status). Regarding individual socioeconomic status, **Chapter 5** indicates that education is a more important underlying factor of lifestyle factors compared with income. Public health policies need to not only provide support for people with low education but also prioritize improving the education level of the general population as a long-term societal goal. As for contextual factors, the results from **Chapter 7** indicate that residing in socioeconomically disadvantaged neighborhoods is unfavorable for residents in terms of lifestyle. This highlights the importance of improving the overall neighborhood environment to promote healthy lifestyle. Practical measures may include enhancing the natural environment, making neighborhood safer, making healthy and fresh food more accessible, adding public exercise and fitness facilities, and reducing noise pollution. From a broader perspective, social norms, government health values, and measures like taxing unhealthy food are also potential contextual factors that may influence lifestyle. It is evident that improving these underlying factors goes beyond individual responsibilities and capabilities. Promoting a healthy lifestyle requires robust collective measures and policies at the public level to drive its realization.

(3) Reducing health inequalities requires addressing fundamental socioeconomic inequalities

Persistent health inequalities exist between socioeconomically-disadvantaged groups and more affluent ones. For type 2 diabetes, socioeconomically disadvantaged groups have higher risk of developing the disease, as well as higher risk of mortality and shorter life expectancy if they developed the disease, compared with the affluent population. Conventionally, the source of this health inequality has been attributed to differences in individual risk factors, with disadvantaged groups more likely to have unhealthy lifestyle. Traditional public health policies and practices have often focused on improving individual risk factors to enhance the health status of the disadvantaged groups, aiming to reduce health inequalities. However, unfortunately, this traditional approach overlooks the fact that health inequalities are not solely caused by differences in individual risk factors. The roots of health inequalities lie in fundamental inequalities between individuals, including the socioeconomic inequalities studied in this thesis. In **Part 2** of this thesis, our findings clearly demonstrate that people with

A

low socioeconomic status still face higher risk of developing type 2 diabetes compared with people with high socioeconomic status, even when accounting for differences in a wide spectrum of individual risk factors. Empirical evidence from other studies has suggested that people with higher socioeconomic status may obtain greater health benefits than those with lower socioeconomic status from lifestyle improvements, potentially leading to bigger health inequalities. My PhD research, along with these other studies, emphasizes the disproportionate harm in health experienced by socioeconomically disadvantaged people. Disadvantaged people do not choose to be unhealthy. Based on these research findings, this thesis underscores the need for two fundamental strategies to reduce health inequalities: improving overall education level and reducing income inequalities in the population, which address the underlying socioeconomic inequalities within the population. Reducing health inequalities cannot be achieved solely by improving individual risk factors; it requires comprehensive and collective social measures and strategies. Furthermore, the findings from **Chapter 6** further highlight the enormous potential of improving income inequalities in reducing health inequalities, especially considering the fact that the majority of the general population whose highest education level have already been attained and are difficult to change.

NEDERLANDSE SAMENVATTING

Wat is diabetes type 2? Wat zijn de schadelijke effecten op onze gezondheid en op de samenleving?

Diabetes type 2 is een chronische ziekte die gekenmerkt wordt door hoge bloedsuikerwaardes voor langere tijd. Langdurige periodes van abnormaal hoge bloedsuikerwaardes kunnen leiden tot schade aan de bloedsomloop, het zenuwstelsel en het immuunsysteem, evenals aan meerdere organen zoals het hart, de nieren en de ogen. Deze schade maakt mensen met diabetes type 2 vatbaarder voor het ontwikkelen van hart- en vaatziekten (zoals atherosclerose), nieraandoeningen, oogziektes (zoals retinopathie) en huidinfecties. Voor patiënten met diabetes type 2 gaat deze ziekte ten koste van hun kwaliteit van leven, het verhoogt hun risico op overlijden en vermindert hun levensverwachting aanzienlijk. De prevalentie van diabetes type 2 is hoog, met aanzienlijke gezondheidsproblemen tot gevolg. Dit is een zware sociaaleconomische last voor de samenleving. Naar schatting waren er in 2021 wereldwijd ongeveer 480 miljoen mensen met diabetes type 2. Hoe diabetes type 2 te voorkomen – en de pandemie te beteugelen – is een van de meest cruciale en urgente kwesties voor de volksgezondheid en de samenleving.

Een ongezonde leefstijl is een van de belangrijkste oorzaken (vaak risicofactoren genoemd) van diabetes type 2. Een ongezonde leefstijl omvat onder andere slechte voeding, roken, drankmisbruik, gebrek aan lichaamsbeweging en een slechte slaapkwaliteit. Er is substantieel bewijs dat aantoont dat het verbeteren van leefstijl het risico op diabetes type 2 aanzienlijk kan verlagen. Leefstijlinterventies hebben dan ook prioriteit gekregen in de praktijk en in het gezondheidsbeleid, als een van de belangrijkste primaire preventiestrategieën voor diabetes type 2.

Met welke problemen en uitdagingen worden we geconfronteerd bij de preventie van diabetes?

In de afgelopen decennia hebben grootschalige leefstijlinterventieprogramma's gericht op de algemene bevolking echter beperkt succes geboekt. In de algemene bevolking worden leefstijlinterventies, ter preventie van diabetes type 2, voornamelijk geconfronteerd met de volgende problemen en uitdagingen:

(1) Diabetes type 2 wordt niet veroorzaakt door één enkele ongezonde leefstijlfactor, terwijl de huidige grootschalige leefstijlprogramma's zich vooral op één enkele factor richten. Er is onvoldoende aandacht voor de combinatie van meerdere ongezonde leefstijlfactoren (leefstijlpatronen). Bovendien is de relatie tussen leefstijlpatronen en het risico op het ontwikkelen van diabetes nog niet goed begrepen.

(2) De huidige leefstijlinterventieprogramma's passen voornamelijk een "one-size-fits-all"-benadering toe, waarbij iedereen dezelfde interventiemaatregelen krijgt. In de praktijk verschilt leefstijl echter tussen individuen. Een "one-size-fits-all"-strategie slaagt er niet in om effectief onderscheid te maken tussen de behoeften van verschillende groepen om de leefstijl binnen de bevolking te verbeteren, wat ten koste gaat van de effectiviteit van deze leefstijlinterventieprogramma's.

(3) Leefstijlprogramma's op populatieniveau houden zelden rekening met factoren die de leefstijl beïnvloeden (in dit proefschrift de onderliggende factoren van leefstijl genoemd), zoals sociaaleconomische status en contextuele factoren. Hoe deze factoren gezamenlijk het risico op diabetes type 2 beïnvloeden is ook grotendeels onbekend.

Welk onderzoek heb ik uitgevoerd in het licht van deze kwesties en uitdagingen?

Deel 1 van dit proefschrift beschrijft verschillende voedings- en leefstijlpatronen in de algemene bevolking. Onderzocht is wat de relaties zijn tussen deze voedings- en leefstijlpatronen en het risico op diabetes type 2. In **Deel 2** van dit proefschrift zijn de relaties van leefstijlfactoren met hun onderliggende factoren onderzocht, met speciale aandacht voor de individuele sociaaleconomische status en de sociaaleconomische status in de buurt. Vervolgens is onderzocht hoe deze factoren gezamenlijk het risico op diabetes type 2 beïnvloeden. Deze onderzoeken zijn gebaseerd op het Lifelines Cohort, een groot bevolkingscohort in Noord-Nederland.

Wat zijn de bevindingen van elk hoofdstuk?

Hoofdstuk 2 beschrijft een voedingspatroon. Een nauwere navolging van dit voedingspatroon hangt samen met een hoger risico op diabetes type 2. Dit voedingspatroon wordt geassocieerd met bloedlipiden, wat betekent dat wisselende

bloedlipideniveaus het risico op diabetes type 2 beïnvloeden. Het voedingspatroon wordt gekenmerkt door een hoge consumptie van toegevoegde suikers, suikerhoudende dranken en sap, en een lage consumptie van fruit, groenten, noten/zaden, volle granen en thee.

Hoofdstuk 3 laat zien dat het eten van meer ultrabewerkt voedsel gerelateerd is aan een hoger risico op diabetes type 2. Het volgen van een gezond dieet compenseert mogelijk niet voor de schadelijke effecten van ultrabewerkt voedsel op de gezondheid. Er moet echter worden opgemerkt dat niet alle soorten ultrabewerkt voedsel verband houden met het risico op diabetes type 2. Door consumptiepatronen van ultrabewerkt voedsel te bestuderen, ontdekten we dat een patroon met veel consumptie van koude hartige snacks (zoals kaas, vleeswaren en hartige spreads voor crackers en stokbrood) en een patroon met veel consumptie van warme hartige snacks (zoals gefrituurde snacks, frites en snacksaus) geassocieerd worden met een hoger risico op diabetes type 2. Een veelvoorkomend eetpatroon in de traditionele Nederlandse keuken (zoals gesneden brood, lunchvlees en jus) is echter niet geassocieerd met het risico op diabetes type 2, terwijl een patroon met veel zoete snacks (zoals koekjes, cake en chocolade) juist geassocieerd is met lager risico op diabetes type 2. Dit lagere risico van het zoete snackpatroon kan te wijten zijn aan de "reverse causation" in de epidemiologie. Om een specifieke verklaring te geven: mensen met een hoog risico op diabetes type 2 aan het begin van het onderzoek (zoals mensen met een familiegeschiedenis van de ziekte) kunnen proberen hun consumptie van zoet smakende voedingsmiddelen met een hoog suikergehalte in hun dagelijkse voeding te minimaliseren. We hebben ook bewijs gevonden dat het bestaan van deze omgekeerde oorzakelijkheid ondersteunt.

Hoofdstuk 4 beschrijft verschillende leefstijlpatronen in de bevolking en toont differentiële verbanden aan tussen deze leefstijlpatronen en het risico op diabetes type 2. Als we het "gezonde leefstijlpatroon" als referentie gebruiken, worden het "ongezonde leefstijlpatroon" en het "slechte voeding en weinig lichaamsbewegingspatroon" in verband gebracht met een hoger risico op diabetes type 2, terwijl het "couch potato-patroon" en het "risicodrinken-patroon" niet geassocieerd zijn met het risico op diabetes type 2. Interessant is dat mensen met het "couch potato-patroon", ondanks dat ze de langste tv-kijktijd van de hele bevolking hebben, het hoogste wekelijkse fysieke activiteitsniveau hebben en voornamelijk niet-rokers zijn. Mensen

A

uit het "risicodrinken-patroon" hebben de laagste tv-kijktijd onder de bevolking. Deze bevindingen suggereren dat leefstijlfactoren de neiging hebben om te clusteren in unieke gedragspatronen binnen de bevolking. Eerder onderzoek heeft aangetoond dat elke individuele ongezonde leefstijlfactor geassocieerd is met een hoger risico op diabetes type 2 wanneer deze afzonderlijk onderzocht worden. De combinatie van deze verschillende leefstijlfactoren - in de vorm van een leefstijlpatroon – correleert echter niet altijd met het risico op diabetes type 2. Daarom is het belangrijk om rekening te houden met de gecombineerde effecten van verschillende leefstijlfactoren op de gezondheid.

Hoofdstuk 5 onderzoekt een conceptueel model dat de belangrijkste risicofactoren voor diabetes type 2 omvat. Dit model omvat niet alleen risicofactoren uit verschillende domeinen, zoals sociaaleconomische status, bloedlipiden, obesitas en leefstijlfactoren, maar ook hun onderlinge relaties. De bevindingen van deze studie geven aan dat een grotere tailleomtrek in plaats van een hogere BMI en slechte bloedlipiden de grootste invloed hebben op het verhogen van het risico op diabetes type 2. Onder leefstijlfactoren kunnen het verminderen van de tv-kijktijd en het verhogen van het fysieke activiteitsniveau aanzienlijke voordelen opleveren bij het verbeteren van de bloedlipiden en het verminderen van de middelomtrek. Het is vermeldenswaardig dat een hoger opleidingsniveau, in tegenstelling tot een hoger inkomen, wordt geassocieerd met een betere leefstijl. Een lager inkomen is echter direct gekoppeld aan een hoger risico op diabetes type 2, wat niet kan worden verklaard door de bovengenoemde risicofactoren. Door de analyse van dit conceptuele model hebben we de onderlinge relaties tussen de belangrijkste risicofactoren voor diabetes type 2 binnen de bevolking gekwantificeerd. Door te focussen op de meest invloedrijke risicofactoren die binnen elk domein zijn geïdentificeerd als prioritaire preventiedoelen, kan de effectiviteit van diabetes type 2 preventie op populatieniveau worden verbeterd.

Hoofdstuk 6 onderzoekt de relaties tussen twee sociaaleconomische indicatoren, namelijk opleiding en inkomen, en de risico's op diabetes type 2 en hart- en vaatziekten. De resultaten laten zien dat een lagere opleiding en een lager inkomen beide geassocieerd zijn met hogere ziekterisico's, en deze associaties zijn onafhankelijk van elkaar. Risicofactoren zoals bloedlipiden, bloeddruk, BMI, leefstijlfactoren en geslacht kunnen deze associaties slechts gedeeltelijk verklaren. Bovendien valt op dat personen

met een lagere opleiding niet noodzakelijkerwijs degenen met een lager inkomen zijn, en vice versa. Dit onderzoek bevestigt verder de substantiële sociaaleconomische ongelijkheden op het gebied van gezondheid in de algemene bevolking. Het alleen verbeteren van de persoonlijke leefstijl, bloeddruk en bloedlipiden kan deze gezondheidsongelijkheid mogelijk niet effectief aanpakken. Bovendien is het bij het uitvoeren van gezondheidsonderzoek en het formuleren van gezondheidsinterventies en -beleid van cruciaal belang om rekening te houden met de afzonderlijke effecten van onderwijs en inkomen op de gezondheid, in plaats van ze samen te voegen.

Hoofdstuk 7 onderzoekt de invloed van de individuele sociaaleconomische status en de sociaaleconomische status van de buurt op leefstijl. Uit de studie is gebleken dat een lagere individuele sociaaleconomische status en een lagere sociaaleconomische status in de buurt beide onafhankelijk van elkaar samenhangen met een slechtere leefstijl. Bovendien is er statistisch gezien een interactie tussen de effecten van de individuele sociaaleconomische status en de sociaaleconomische status van de buurt op de leefstijl. Deze interactie suggereert dat bewoners die in buurten met een lagere sociaaleconomische status wonen en die ook zelf een lagere individuele sociaaleconomische status hebben, vatbaarder zijn voor de nadelige invloed van de achtergestelde buurt op hun leefstijl, in vergelijking met hun meer welvarende buren. Deze bevindingen benadrukken de noodzaak om niet alleen rekening te houden met de individuele sociaaleconomische status, maar ook om extra gezondheidsbronnen en sociale kansen te bieden aan sociaaleconomisch achtergestelde buurten bij het formuleren van gezondheidsprogramma's en -beleid.

A

Samengevat, hoe helpen deze onderzoeksresultaten bij de ontwikkeling van betere strategieën voor diabetes type 2 preventie voor de algemene bevolking?

(1) Focus op voedings- en leefstijlpatronen als interventiedoelen, in plaats van de generieke "one-size-fits-all"-benadering

Deel 1 van dit proefschrift beschrijft verschillende voedings- en leefstijlpatronen in de algemene bevolking, waarbij elk patroon individuen vertegenwoordigt met vergelijkbare voedings- of leefstijlgewoonten. Deze leefstijlpatronen zijn op verschillende manieren geassocieerd met het risico op diabetes type 2, wat elk hun klinische relevantie

aantoont. De unieke kenmerken van deze voedings- en leefstijlpatronen maken het mogelijk om voor elke voedings- en leefstijlpatroongroep op maat gemaakte voedings- en leefstijlinterventies te ontwerpen, wat een "gepersonaliseerde" leefstijlinterventie mogelijk maakt op een gemiddeld niveau van de bevolking. In tegenstelling tot de generieke "one-size-fits-all"-benadering, maakt deze benadering een effectievere differentiatie mogelijk van leefstijlinterventiebehoeften tussen verschillende groepen en vergemakkelijkt het gecombineerde interventies gericht op meerdere ongezonde leefstijlfactoren. Gezien deze voordelen biedt deze op patronen gebaseerde strategie een groot potentieel voor het bevorderen van een gezonde leefstijl en het voorkomen van diabetes type 2 bij de algemene bevolking.

(2) Het bevorderen van een gezonde leefstijl moet gericht zijn op onderliggende factoren van leefstijl

Het bevorderen van een gezonde leefstijl voor de algemene bevolking kan niet eenvoudig worden bereikt door middel van gezondheidsvoorlichting en op individuen gerichte interventies. De bevindingen in **Deel 2** van dit proefschrift benadrukken de noodzaak om de onderliggende factoren van leefstijlfactoren in de volksgezondheidspraktijk en beleidsvorming te verbeteren. Deze onderliggende factoren omvatten niet alleen de individuele sociaaleconomische status (zoals opleiding en inkomen), maar ook contextuele factoren (zoals de sociaaleconomische status in de buurt). Met betrekking tot de individuele sociaaleconomische status geeft **Hoofdstuk 5** aan dat opleiding een belangrijkere onderliggende factor is van leefstijlfactoren dan inkomen. Volksgezondheidsbeleid moet niet alleen ondersteuning bieden aan mensen met een lage opleiding, maar ook prioriteit geven aan het verbeteren van het opleidingsniveau van de algemene bevolking als een maatschappelijk langetermijndoel. Wat betreft omgevingsfactoren geven de resultaten van **Hoofdstuk 7** aan dat wonen in sociaaleconomisch achtergestelde buurten ongunstig is voor bewoners in termen van leefstijl. Dit benadrukt het belang van het verbeteren van de algehele buurtomgeving om een gezonde leefstijl te bevorderen. Praktische maatregelen kunnen zijn: het versterken van de natuurlijke omgeving, het veiliger maken van buurten, het toegankelijker maken van gezond en vers voedsel, het toevoegen van openbare beweeg- en fitnessfaciliteiten en het verminderen van geluidsoverlast. Vanuit een breder perspectief zijn sociale normen, gezondheidswaarden van de overheid en

maatregelen zoals het belasten van ongezond voedsel ook potentiële contextuele factoren die de leefstijl kunnen beïnvloeden. Het is duidelijk dat het verbeteren van deze onderliggende factoren verder gaat dan individuele verantwoordelijkheden en capaciteiten. Het bevorderen van een gezonde leefstijl vereist robuuste collectieve maatregelen en beleidsmaatregelen op populatieniveau om de realisatie ervan te stimuleren.

(3) Om ongelijkheden op gezondheidsgebied te verminderen, moeten fundamentele sociaaleconomische ongelijkheden worden aangepakt

Er bestaan hardnekkige ongelijkheden op het gebied van gezondheid tussen sociaaleconomisch achtergestelde groepen en meer welvarende groepen. Sociaaleconomisch achtergestelde groepen hebben in vergelijking met de welvarende bevolking een hoger risico om diabetes type 2 te ontwikkelen, evenals een hoger risico op sterfte en een kortere levensverwachting als ze de ziekte ontwikkelen. Traditioneel wordt de oorzaak van deze gezondheidsongelijkheid toegeschreven aan verschillen in individuele risicofactoren, waarbij kansarme groepen vaker een ongezonde leefstijl hebben. Traditioneel waren beleid en praktijk op het gebied van volksgezondheid vaak gericht op het verbeteren van individuele risicofactoren om de gezondheidstoestand van kansarme groepen te verbeteren, met als doel ongelijkheden op gezondheidsgebied te verminderen. Helaas ziet deze traditionele benadering over het hoofd dat ongelijkheden op gezondheidsgebied niet alleen worden veroorzaakt door verschillen in individuele risicofactoren. De wortels van ongelijkheden op gezondheidsgebied liggen in fundamentele ongelijkheden tussen individuen, inclusief de sociaaleconomische ongelijkheden die in dit proefschrift onderzocht zijn. In **Deel 2** van dit proefschrift tonen onze bevindingen duidelijk aan dat mensen met een lage sociaaleconomische status nog steeds een hoger risico lopen op het ontwikkelen van diabetes type 2, in vergelijking met mensen met een hoge sociaaleconomische status, zelfs wanneer rekening wordt gehouden met verschillen in een breed spectrum van individuele risicofactoren. Empirisch bewijs uit andere onderzoeken suggereert dat mensen met een hogere sociaaleconomische status grotere gezondheidsvoordelen kunnen behalen door verbeteringen in leefstijl dan mensen met een lagere sociaaleconomische status, wat mogelijk kan leiden tot grotere ongelijkheden op gezondheidsgebied. Mijn promotieonderzoek benadrukt,

A

samen met deze andere studies, de onevenredige gezondheidsschade die wordt ervaren door sociaaleconomisch achtergestelde mensen. Kansarme mensen kiezen er niet voor om ongezond te zijn. Op basis van deze onderzoeksresultaten benadrukt dit proefschrift de noodzaak van twee fundamentele strategieën om ongelijkheden op gezondheidsgebied te verminderen: verbetering van het algehele opleidingsniveau en vermindering van inkomensongelijkheid in de bevolking. Die strategieën adresseren de onderliggende sociaaleconomische ongelijkheden binnen de bevolking. Het verminderen van ongelijkheden op gezondheidsgebied kan niet alleen worden bereikt door individuele risicofactoren te verbeteren; het vereist alomvattende en collectieve sociale maatregelen en strategieën. Bovendien benadrukken de bevindingen van **Hoofdstuk 6** het enorme potentieel van het verbeteren van inkomensongelijkheid bij het verminderen van ongelijkheden op gezondheidsgebied, vooral gezien het feit dat de meerderheid van de algemene bevolking, wiens hoogste opleidingsniveau al is bereikt, moeilijk te veranderen is.

中文摘要

什么是2型糖尿病？对我们健康和社会有什么危害？

2型糖尿病是一种慢性病，其主要特征为长期高血糖。长期不正常的高血糖状态会损害身体的循环、神经和免疫系统，以及心脏、肾脏、眼睛等多个器官。这些损害使得2型糖尿病患者更容易患有心脑血管疾病（比如动脉粥样硬化）、肾脏病、眼部疾病（比如视网膜病变）和皮肤感染等疾病。这些疾病使2型糖尿病患者生活质量下降、死亡风险升高，并大大降低了2型糖尿病患者的预期寿命。2型糖尿病的患病率很高，这在人群中造成了巨大的健康问题和沉重的社会经济负担。据估计，2021年全球共有约4.8亿2型糖尿病患者。如何预防糖尿病——遏制糖尿病在人群中的"大流行"是十分重要且迫切的公共卫生和社会课题。

不良生活方式是导致2型糖尿病最重要的因素之一（或称之为风险因素）。不良生活方式包括不良饮食习惯、吸烟、过量饮酒、缺乏体育锻炼、久坐和睡眠质量差等。确凿的研究证据表明改变不良生活方式可以使患2型糖尿病的机会大大降低。公共卫生策略也将改善生活方式作为预防糖尿病最重要的措施之一。

对于预防2型糖尿病，我们面临什么问题和挑战？

然而，在过去几十年中，针对一般人群的大型生活方式干预项目鲜少取得成功。在一般人群中，针对2型糖尿病的生活方式干预主要面临以下问题和挑战：

（1）2型糖尿病的并不仅是由一种不良生活方式导致的，而目前的大型干预项目主要针对单个不良生活方式，缺乏针对多种不良生活方式的联合干预措施。此外， 生活方式模式（不同生活方式的组合）与糖尿病发病风险的关系尚不十分清楚。

（2）目前的生活方式干预项目主要采取"一刀切"的通用策略，即每个人都接受相同的干预措施。但由于现实中每个人的生活方式不同，对生活方式干预也有不同的需求，这种传统的"一刀切"策略因此无法有效区分人群中不同生活方式亚群体的需求，从而削弱了生活方式干预项目的有效性。

（3）人群水平的生活方式干预项目极少考虑到影响生活方式的潜在因素，比如社会经济地位和社区背景因素。这些因素与生活方式之间的相互作用对2型糖尿病发病风

险的影响也很少被考虑。

针对这些问题和挑战，我进行了什么研究？

在博士论文的第一部分（Part1），我描述了几种人群水平的膳食和生活方式模式，并探究了他们与2型糖尿病发病风险的关系。在博士论文的第二部分（Part2），我研究了生活方式与个人社会经济地位（教育和收入）和社区背景因素（社区社会经济地位）的关系，以及他们如何共同影响2型糖尿病的发病风险。这些研究基于Lifelines队列，这是一个荷兰北部的大型人群队列。

每一章有什么发现？

第二章（Chapter2）描述了一种膳食模式，饮食越接近这种膳食模式则与较高的2型糖尿病的风险相关。这种膳食模式与血脂相关，即其可能通过改变血脂从而影响2型糖尿病的风险。这种膳食模式的特征是糖、含糖饮料和果汁的摄入量较高，但水果、蔬菜、坚果/种子、全谷物和茶的摄入量较低。

第三章（Chapter3）发现吃越多的超加工食物可能与更高的2型糖尿病风险相关，并且健康饮食并不能完全弥补超加工食物对健康带来的不良影响。然而，值得注意的是，可能并不是所有的超加工食物都与2型糖尿病风险相关。通过研究超加工食物的膳食模式，我们发现以热咸味超加工小吃（比如油炸小吃、薯条和用于咸味小吃的酱料）和以冷咸味超加工小吃（比如奶酪块、肉肠片和用于涂抹饼干和切片法棍的咸味酱料）摄入量高为特征的超加工食物膳食模式与较高的2型糖尿病风险相关；而以面包、肉肠片（用于面包）和肉汁摄入量较高的传统荷兰菜肴超加工食物膳食模式则与2型糖尿病风险无关；以甜味小吃（比如饼干、蛋糕和巧克力）摄入量较高的超加工食物膳食模式则与较低的2型糖尿病风险相关。甜味小吃与较低的2型糖尿病风险相关可能是由于流行病学中的"反向因果"现象。具体解释来说，在研究开始时有较高糖尿病风险的人（比如有家族遗传史），他们可能在日常饮食中尽量避开甜味的含糖量较高的食物。我们也在研究中也进一步发现了这种反向因果关联存在的证据。

第四章（Chapter4）描述了人群中几种不同的生活方式模式，并且发现这些生活方式模式与2型糖尿病风险有不同的关联。以健康生活方式模式为参考，多种不健康生活

方式模式和低膳食质量低体育锻炼生活方式模式与较高的2型糖尿病风险相关；而"沙发土豆"（couch potato，意指看电视时间长）生活方式模式和过量饮酒生活方式模式则与2型糖尿病风险无关。有趣的是，"沙发土豆"生活方式模式人群虽然在全人群中看电视时间最长，但他们每周体育锻炼时间也最长，并且他们主要是戒烟者。而过量饮酒生活方式模式人群看电视的时间是全人群中最低的。这些发现表明不同的生活方式可能在人群中形成独特的生活方式模式集群。并且，即使先前的研究发现每一个不健康的生活方式在被单一研究时都与较高的2型糖尿病风险相关，但当这些不同生活方式以生活方式模式的形式组合在一起的时候，它们并不总是和2型糖尿病风险相关，我们也需要考虑它们彼此可能的相互作用对健康的影响。

第五章（Chapter5）分析了一个包含2型糖尿病主要风险因素的概念模型。这个概念模型不仅包含了不同类别的风险因素（比如社会经济地位、血脂、肥胖指标和生活方式），也涵盖了这些不同类别风险因素之间的复杂关系。这项研究发现，较大的腰围（而不是较高的体重指数BMI）和较差的血脂状态对2型糖尿病风险增加有最大的影响；减少看电视时间和增加体育锻炼对改善血脂和减小腰围有最大的益处；较高的教育水平（而不是较高的收入）与更好的生活方式相关。然而，较低的收入水平对2型糖尿病风险增高有直接的关联，并且这种关联并不能被不良生活方式、肥胖和血脂状态所解释。通过分析这个概念模型，我们量化了人群中2型糖尿病不同类别风险因素之间的关系。通过比较它们对2型糖尿病风险相对贡献的大小，我们可以在不同类别风险因素之中找到影响最大的优先干预目标，从而提升2型糖尿病在人群水平预防的有效性。

第六章（Chapter6）研究了两个社会经济地位指标，即教育水平和收入水平，与2型糖尿病风险和心脑血管疾病风险的关系。研究发现，较低的教育水平和较低的收入水平均与较高的疾病风险相关，并且这种关联彼此相互独立。血脂、血压、体重指数、生活方式和性别等风险因素仅能部分解释教育和收入与疾病风险间的关联。此外，我们也发现拥有较低教育水平的人群并不一定是有较低收入的人群，而拥有较低收入水平的人群也并不一定是有较低教育水平的人群。这项研究进一步证实了一般人群在健康方面存在着严重的社会经济地位不平等，并且这种健康不平等可能并不能单纯地通过改善个人生活方式、血压和血脂等措施来得到改善。此外，在进行健康研究和制定健康干预措施与政策时，我们需要同时考虑教育和收入两个不同的社会经济地位指标对健康的影响，而不是把它们混为一谈。

第七章（Chapter7）研究了个人社会经济地位和邻里社会经济地位对生活方式的影响。研究发现，较低的个人社会经济地位与较低的邻里社会经济地位均与较差的生活方式独立相关。此外，在统计学上，个人社会经济地位和邻里社会经济地位与生活方式的关联存在交互作用。这种交互作用表明，对居住在较低邻里社会经济地位社区的居民来说，如果他们个人社会经济地位也较低，那么相比他们具有更高个人社会经济地位的邻居来说，他们更容易受到较差邻里环境对其生活方式带来的不良影响。这些结果提示在制定促进个人生活方式的干预措施和健康政策时，我们不仅需要考虑个人社会经济地位，我们也要为社会经济弱势社区提供更多的健康资源和社会机会。

总结而言，这些研究结果为制定对一般人群的2型糖尿病预防策略有什么帮助？

（1）以膳食和生活方式模式为基础和目标进行生活方式干预，而不是采取"一刀切"的单一通用干预模式

论文的第一部分发现在人群中存在着不同的膳食与生活方式模式，并且它们与2型糖尿病风险有着不同的关联。这些不同的膳食和生活方式模式各自具有其独特的的特征，这使得我们可以根据每个不同膳食和生活方式模式群体的需求，为他们"量体裁衣"，制定针对每个群体的"群体个性化"膳食和生活方式干预措施。相比传统的"一刀切"式的全民统一干预方法，这种在人群水平基于膳食和生活方式模式的干预措施可以更有效地区分不同群体的生活方式干预需求，并且可以针对多种不良生活方式进行联合干预。基于这些优点，相比一般通用的人群水平干预策略而言，这种基于膳食与生活方式模式的干预策略可能更加有效，在促进人群良好生活方式和预防2型糖尿病的实践中有着巨大的潜力。

（2）促进健康生活方式需要重视和改善生活方式的潜在因素

促进全民实践健康生活方式，仅仅通过健康教育和对个人的干预是不够的。论文第二部分的研究结果强调了在公共卫生实践和政策制定中需要重视和改善影响生活方式的潜在因素。这些潜在因素不仅包括个人社会经济地位（比如教育和收入），还包括背景因素（比如邻里社会经济地位）。对于个人社会经济地位，第五章的研究表明教育是相比收入更重要的生活方式潜在因素。公共卫生政策不仅需要为低教育水平人群提供支持，更需要以提高全民教育水平作为长期的社会性目标。而对于背景因素，第七

章的研究结果表明居住在社会经济弱势社区不利于居民实践健康的生活方式，这强调了通过改善社区的整体环境可以促进居民的健康生活方式。具体的实践措施包括改善社区自然环境、提高社区安全性、使社区中健康新鲜食品更容易获得、增加公共运动健身场所和减少噪音污染等。除此之外，社会规范、政府健康宣传观念，以及对不健康食品征税等措施都是在社区层面之上对健康生活方式有影响的潜在背景因素。显而易见，改善这些潜在因素远远超出了个人的能力范围，健康生活方式也并不完全是我们的个人责任。促进健康生活方式需要强有力的公共集体措施和政策去推动实现。

（3）减少健康不平等需要解决根本的社会经济不平等

社会经济弱势群体与更富裕的群体之间长期存在着巨大的健康差异。对于2型糖尿病而言，社会经济弱势群体（相比富裕群体）有更高的患病风险，在患病后也有更高的死亡风险和更短的预期寿命。一直以来，这种健康差异的来源被归咎于个人风险因素的差异，即弱势群体更常有不健康的生活方式。传统公共卫生政策和实践也常通过改善个人风险因素的方式去提高弱势群体的健康状况，从而试图减少健康不平等。然而，遗憾的是，这种传统思维策略忽略了健康不平等并不仅仅是由于个体风险因素差异而造成的。健康不平等的根源来自于人与人之间的根本不平等，这包括了本论文着重研究的社会经济不平等。在论文的第二部分，我们的研究结果清晰地表明对于低社会经济地位群体，即使考虑了他们与高社会经济地位群体之间的个人风险因素差异，他们也仍旧有更高的2型糖尿病患病风险。一些来自其他研究的经验性证据也表明，相比社会经济弱势群体，高社会经济地位群体可能从生活方式改善中获得更多健康益处，这可能会导致更严重的健康不平等。我的研究和这些其他研究都强调了社会经济弱势群体在健康中受到的不成比例的伤害。弱势群体并不是自己选择了不健康。基于这些研究结果，我的论文强调了减少健康不平等需要采取两个根本的社会性策略，即提高全民教育水平和减少收入差距不平等，即减少人群中的根本社会经济不平等。减少健康不平等无法单纯通过改善个人风险因素来实现，需要采取广泛的社会性集体策略。此外，第六章的研究结果也进一步强调了即使对于大部分一般人群而言，他们的最高教育水平已经达到并且很难改变，改善收入水平差距也对减少健康不平等有巨大的潜力。

A

Acknowledgments
致谢 Dankwoord

I am about to graduate again!

As I wrote this, a subtle feeling of surprise came to me. In Chinese, a PhD is accompanied by "graduation", but not "promotion" in Dutch. And for this time the graduation, it is no longer about saying goodbye to classmates around, nor is it about leaving a classroom with things that remained habitually unchanged.

How I wish I could be as sharp as when I wrote papers. However, I know that thoughts and feelings cannot be categorized smoothly like a cluster analysis. Alongside these reflections are the memories of everything that has happened during my five-year-long PhD journey, starting from 2018.

Although time is continuous, we seem to be used to pause at important points to look back and summarize.

Five years may sound like a long time, but when compared to a lifetime, or to the vastness of the universe, it is not even a mere instant. It makes me ponder – within this vast universe, all the particles, and rocks, have persisted billions of years. While their forms constantly change, carbon remains carbon, and silicon remains silicon. Have the sands I walked on ever been touched by creatures from the Cambrian era?

Over the past five years, it feels as if time has frozen at a particular stage. I have spent such a long time alone, sitting behind the screen, reading, learning, writing, and having meetings online that has remained unchanged since then.

Over the past five years, I have witnessed a rapid condensation and acceleration of things in my life – an overwhelming experience I had never encountered before.

Over the past five years, there have also been numerous choices, discussions, resistance, as well as moments of anxiety and sadness.

However, despite all, I must emphasize, that over the past five years, I have conducted research that I am passionate about, acquired knowledge, and have been able to explore and travel the world. I so much enjoy my research and I am very proud of myself. To attain this, it has not been an easy feat.

What an extraordinary and joyful journey to celebrate!

A

Whether in the face of challenges or amidst idle hours, I always enjoy to take a walk in the nature. I cannot help but wonder, how many years it took for those trees to grow into their current form.

I bought this Schefflera two years ago. One day, as I was randomly browsing the photos, I was delighted to see how much it had grown. Look at its beautiful leaves! (Although I think some pruning is necessary... haha.)

Over the past 5 years, I have been fortunate to meet you, and along with the cherished companionship with those I have known for many years. I have learned from you, discussed with you. We drink coffee, eat, travel, and chat, creating countless joyful and happy moments together. And this, is not a goodbye.

Dear **Prof. Gerjan Navis** and **Dr. Louise Dekker**, I am so lucky to have you as my PhD supervisors. All of this is not only about your guidance for my research, but also your encouragement for me to explore. You welcome and support my research ideas. We discussed results and papers together. Your knowledge, experience, and wisdom have greatly inspired me! I have learned a lot. Besides research, I also greatly appreciate your care and support, especially during difficult times. You were always there by my side. Thank you!

Dear **Gerjan**, thank you for all your knowledge and wisdom. I greatly appreciate and have learned a lot from you, particularly your insights into healthy aging, primary prevention strategies, health inequalities, and so on. Your enthusiasm and passion for investigating better prevention strategies and strategies to tackle health inequalities have been a significant inspiration to me. I am also confident that, I will continue my research to make my contribution towards this field. Thank you!

Dear **Louise**, the simplest yet most powerful words I can find are: thank you for everything! Our weekly meetings have been enjoyable, with discussions about research where we have solved many big and small issues. We've also had pleasant conversations about our daily lives. You are so kind, caring, and supportive. When I mentioned my wish to do a research internship, you were very willing to help and helped arrange everything, which provided me with a valuable opportunity! Thank you!

During my PhD journey, I had the opportunity to work with **Dr. Eva Corpeleijn**, **Prof. Juan-Jesus Carrero**, **Prof. Jochen Mierau**, and **Prof. Stephan Bakker**. I appreciate your insightful comments, help, and encouragement in my research. Dear **Eva**, attending the Lifestyle Unit Meeting is always a delight! There were always many interesting topics and research. Your sharp insights into ultra-processed food add a fun yet important new knowledge to the paper and the field. I also want to thank you for sharing knowledge as well as information about conferences. Dear **Juan-Jesus**, in the final year of my PhD,

A

I finally had the chance to visit you in Stockholm! The coffee at KI was very nice! I am grateful for your guidance and suggestions that greatly helped improve my writing. I can clearly see the progress and differences in my work over the past four years!

For my PhD thesis, I would like to thank the assessment committee, **Prof. Edith Feskens**, **Prof. Erik Buskens**, and **Prof. Karien Stronks**, for your time and effort in reading and reviewing my thesis.

In addition, I would like to thank my supervisors for my research internship at the RIVM. Thank you, **Dr. Henk Hilderink** and **Dr. René Poos**, for providing me with the opportunity to explore and study health inequalities in mortality among diabetes patients, using the CBS and Nivel data (which I consider the best available data in the country). Thank you for your guidance and help in this. I would also like to thank **Marjanne**, **Ellen**, **Marianne**, and **Joost** for your willingness to help with many practical questions.

Dear **Qingqing**, **Petra**, **Yinjie**, **Maryse**, and other co-authors that I worked with. It's so nice that we had the chance to work together on research and we have achieved these beautiful paper outcomes! Some of the papers were quite challenging, like with more than 15 times of submission or rebuttal letters as long as 60 pages. But we made it! I am proud of our beautiful papers.

My PhD research is mainly based on the **Lifelines** data. I sincerely acknowledge the services of the Lifelines Cohort study, especially the **data management team**, that I worked with throughout my PhD research. You helped solved a lot of practical issues. I would also like to thank the contributing research centers delivering data to Lifelines, and all the study participants!

Dear **İdil**, **Vicente**, **Qiming**, and **Li**, thank you very much for being my paranymphs and helping arrange things for my promotion!

Dear **İdil**, **Vika**, **Arianna**, and **Yasmina**, I so much enjoy hanging out with all of you! It's so lucky to have you throughout my PhD journey. Expressing my affection in just a few sentences is impossible. I will definitely miss all the spontaneous meet-ups for coffee, drinks, and city trips. **Arianna**, you are the strong bond that keeps us connected! It was always happy to have our weekly coffee and food meet-ups somewhere in the city.

Vika, it is always pleasant to talk and share our love for plants, nature, coffee, and the little cute details of life. I am very happy that we did a great job for the final symposium. **Yasmina**, it was always so much fun to participate in various activities with you. Thank you that I got to know about the nice events and activities of Gopher.

Dear **İdil**, I can have so many personal, honorable titles for you, appreciating all the beautiful, funny, lovely, happy, or occasionally sad moments during our PhD journey. You are my best corona time dinner buddy, my best walking buddy, my best temperature-feeling buddy, saying 10 degrees feels like summer, and so much more. You make my PhD journey beautiful!

And also, thank you to all my **PROMINENT** fellow PhDs: **Angelina**, **Anna**, **Faizan**, **İdil**, **Karla**, **Martina**, **Noura**, **Sajad**, **Sara**, **Sara**, **Sok Cin**, **Sonia**, **Vicente**, **Vika**, and **Zaid**. PROMINENT is a very unique project that brought 16 of us from different countries together, starting our PhD at more or less the same time. We studied diabetes together, but from diverse perspectives, sometimes struggling to understand lectures from the field of some of us, but unfamiliar to the others. And of course, we also shared our complaints about something haha. We attended many interesting courses and successfully organized two symposiums! What an experience filled with so much fun!

Dear **Vicente**, it is very happy to know you! Although we didn't meet each other often, I enjoy attending courses and activities with you. You are a fun and cheerful person, and I always enjoy talking with you! I admire your dedication to your work and your perseverance. With effort from you, **İdil**, and other PROMINENT colleagues, we were able to have such a beautiful video. Also, thank you for inviting me to happy celebration events!

Dear **Sok Cin**, no need to say more, but 能一起说中文就是超级开心的事！And for a period of time, we were neighbors. You are so caring and kind. We also shared many interesting things that we experienced in our daily lives. And I also very much appreciate your care and support, especially when I was sick. Thank you!

Dear **Sara**, it was nice to be colleague with you in Nephrology. I enjoyed our chats, and I was happy that we attended the EASD conference together with **Sok Cin** in Stockholm after covid.

And also, very importantly, thank you, dear **Sieta**. I'd like to say that without you, PROMINENT wouldn't run as smoothly as it does. I sincerely appreciate all the work

A

and care you have put into supporting us for the past four years, especially with the administrative matters behind the scenes. Additionally, thank you for your efforts in organizing all the courses and events!

In addition, I would like to thank **Prof. Hiddo Lambers Heerspink**, and all other professors within the PROMINENT. Thank you for creating this wonderful project!

And also, thank you to my colleagues from Nephrology: **Arno**, **Amber**, **Anna-Sophie**, **Camilo**, **José**, **Li**, **Lyanne**, **Maryse**, **Manuela**, **Qingqing**, **Sara**, **Stanley**, **Suzanne**, **Shuqi**, **Yinjie**, and many. Thank you for the nice moments we shared, like Sinterklaas and the trip to somewhere in Noord Brabant.

I would also like to thank **Winnie** and **Agnes**. Thank you for helping solving practical issues. I am deeply sorry to hear about Winnie's passing, and I hope she may rest in peace.

Dear **Qingqing**, **Yinjie**, **Shuqi**, and **Li**, I'm happy that we celebrated many Chinese festivals together. We made 粽子 and shared many interesting things of life.

Dear **Rumei**, I remember what you wrote in your thesis. Indeed, it is hard to believe how fast time flies! So many years have passed by! We started chatting before we went to Groningen. We not only shared information about practical matters like accommodations, but we also discussed academic topics. It was very happy to get to know you.

Dear **Guanzhi**, I am very happy that you came to Groningen to study! You are beautiful, smart, and kind. You have also been the best neighbor at Paddelpoel, and you helped take very good care of my plants. Additionally, thank you for the delicious 牛肉粉.

Dear **Iris** and **Qingqing**, it was such a joyful trip to Dublin in 2019 to attend the FENS conference together!

And also, thank you **Eva** and colleagues from the **Lifestyle Unit meeting**: **Elise**, **Lu**, **Mian**, **Petra**, **Qihua**, **Tian**, and many. Thank you for all the ideas, discussions, and for sharing your research findings. I have learned a lot.

Also, thank you **大姐Julie**, for nice coffee meet-ups and city trips.

Dear **Marco**, **Arend**, **Pat**, **Yang**, and friends of **Ewoud**, it is very happy to get to know all of you. Although we did not know each other for long, every meet-up was happy and

fun! Thank you, **Pat**, for always being cheerful and sharing your love for art and life. Thank you, **Lieve Marco**, for you being loving and caring. It is always happy to chat with you, whether about research or just anything fun. And also, I am sincerely happy for you for lovely **Isa**. And also, very importantly, please stay together with **Ewoud**! Thank you, **Arend**, for always being cheerful and kind. It is always a pleasure to talk to you! And also, I am sincerely happy for you for joyful **Tamar**, as well as the little one who is eagerly expecting to greet the world! And thank you, **Yang**, 用中文聊一些有的没的的就是让人开心的事。

Ik wil mijn **Nederlandse familie** bedanken! Bedankt dat jullie zo goed voor me hebben gezorgd, bedankt voor de gezellige tijd, jullie hulp bij de verhuizing, het bowlen, de koffie en het eten. Jullie hebben me hier erg thuis laten voelen. En ook, bedankt **Suzan** voor alle mooie tekeningen~

绝对不能忘记的是，感谢这么多年来与我一直相伴的各路爱人们。我们的各大聊天群貌似都被我在某一个时间点开始带领大家冠以了爱的群名：爱的团聚，爱的见证人，爱的情感大揭秘，爱人们相约，爱鸭美，爱的还愿…… 感谢你们，我们有这么多开心快乐的回忆。尤其在疫情期间，相聚是千金难换的奢侈。而幸运的是，我们始终在朝着美好与团聚前进。

致我们彼此真挚的友谊！

蕾姐姐Zhilei，我知道我们都相信爱是注定的，从我还在根特的时候，或者更早某一天还在排练的时候，可能就注定了我们终会在荷兰团聚。从根特，到秋日Roosendaal，到"在格村要坚强"，到基本可以随意串门的现在，冥冥中都是爱的吸引力。但我们最终还是不约而同的需要去瓦村。尤其是在我读博士期间，我们不需要太多煽情的话，但能去HoogCatharijne逛逛名创优品，吃个大顺，聊聊日常，就是开心快乐的！

腻腻Yini，和孩每一天度过的都是开心快乐的时光。感谢博士前两年孩北上300多公里坐5个多小时火车屁痛痛探望长。博士第三年孩也来了北边，然后我们每天去超市遛弯，饭后算塔罗，饭后家庭剔牙，刷b站也算塔罗，画画并欣赏彼此的艺术造诣。在漫长的疫情期间却和孩度过了最闲适的时光！

孙启明Qiming，对你的爱是不变的。从我们刚认识的时候，一路上做出的每一个小的选择，都精妙的组合在一起，最终让我们可以一起走过这么长一段亲密的时光！我们一起旅行，一起分享开心快乐的事，在艰难的时刻给予彼此最坚定的支持。写不出你文采飞扬的诗，就用简单的话来感谢你曾经的付出和陪伴！

还有**梦梦Meng**、**大厨Penghan**、**溪溪Xiaoxi**，太开心我们这博士小组，和大家一起，这么多年来一直保持开心畅谈！**梦梦**，我们一起聊开心快乐，还有那些难过的事情。你对科学前沿的不懈追求是本群的学术之光！感谢**大厨**和**腻腻**，用极高的标准保证了我们在旅行中的珍馐美馔。感谢**大厨**在意大利的富贵美食美酒招待！

溪溪、**何瑶Yao**、**黄子娴Zia**、**宣宣（爱的见证人）Xuan**、**何直橙EricChan**，回国隔离21天怎能阻挡我们美丽的爱的相聚！感谢你们一直以来线上线下的陪伴。2015年我们住进了根特的HomeGroningen，在2018年我真的去了Groningen！即使过去几年我们都在不同的地方，但我从未觉得我们离得很远。无论是在什么时候，这种陪伴是帮我度过艰难时期最好的力量！以及答应我，之后不论在哪，一些爱的固定项目不会改变，我们还会一起去雍和宫虔诚，去潘家园配眼镜！

美宋和**瑶瑶**，感谢爱的姐妹对我博士4年多的支持和陪伴！神奇的是，我们在过去4年都见证了彼此走向甚至有些离奇，但却又最终合理的人生轨迹。在我第一次去格罗宁根的时候，还是和**美宋**一起！爱的姐妹有太多有趣的故事，光靠挖坟就可以爱的开心一辈子。要和**瑶瑶**一直唱k唱下去，每次回国第一顿也都要吃渝信。愿**美宋**早日脱困田纳西，让我们一起在北京唱大鱼吃云南菜。

冰冰、**丽丽**、**田异星**，感谢姐妹从大学以来的陪伴！本群姐妹最能上山下海，彼此鼓励不断探索美丽的世界。相信我，让我们先来定个小目标，四人游会凑齐的，南极会去的！

青青Qingqing、**爽姐Senshuang**、**麦乐迪Yinjie**、**邻居Zheng**，多开心在格罗宁根有你们相伴！我们一起庆祝了很多节日，包了爱的月饼，爱的粽子，送了彼此爱的平安果，还有**爽姐**送给**青姐**的我还年轻，还有我们一起送给**青姐**的给我再来一杯mojito。我们还一起在疫情期间成功出游去了**青姐**总念不对的哥廷根。回想起来，还发生了很多搞笑无厘头的事情。这些大大小小的事情，使我在格罗的时光丰富多彩起来！尤其感谢**青姐**，在很多艰难的时候的鼓励和支持！

最后，感谢我的家人们。你们是我坚强的后盾和温暖的避风港，让我可以放心大胆地的去自由选择。你们所有的爱都凝聚在一点一滴的生活细节中。

爸爸妈妈，感谢你们的爱，感谢你们支持鼓励我去探索。家是我丰富多彩的神奇小宇宙，就比如虽然我和我妈总嫌弃我爸做菜咸，但神奇的是我也在不知不觉中熟练继承了比如扁豆焖鸡腿、凉拌彩椒等老段招牌家常菜（不咸版）。那叫一个地道！这些再平常不过的饭菜不知不觉已经陪我走过了从2015年到现在这么多年的留学时光，而在这每一顿再平常不过的饭菜中，总有最安心的家的味道。

妈妈，你是最美丽的冒险家！爬过北极的山，看过绚烂的极光，也曾在热带雨林中探索。我无法体会和想象你在2020年经历的那场狂风暴雨。但我知道我们又准备好了，期待和你再次一起出发探索世界！

姥姥姥爷，在万寿路长大是多开心的事！感谢你们为我创造了舒适安逸的成长环境。虽然过去几年回家次数不多，但每次回去都是熟悉安稳的感觉，让我感到放松、开心和快乐。在此，也致敬我已故的**姥姥**，有绿水青山围绕，愿你开心快乐，一路平安！

感谢**爷爷奶奶**，**二爸二妈**，和我所有的家人们，我们都彼此惦记着！比心！

Dear **Ewoud**, my bun. Thank you! Thank you for being there with me all the time. I don't have any more fancy words to express my love and gratitude for you. In Chinese, there is a somewhat clichéd but genuinely sincere sentence that says, "Being with each other together, is the most profound expression of love". I am so happy for all the joyful moments we have spent together. Whether big decisions or small details, together we plan, and we move forward. We make our home beautiful and classy. We do our best to make each other feel relaxed after a long day of work. We play and travel together. Together, we do house chores, we complain some ridiculous things, we cook delicious food and drink teatea. We will get old together, but we will still be each other's bunbun. Thank you for taking me on many local trips across the three northern provinces. Thank you for taking me to the forest for walks when I was stressed. Thank you for making bun decaf coffeeffee every night before sleep. Thank you for your gentle and tender care and support during many difficult times. Thank you for your strong and unwavering support in every aspect of my life, always being by my side. I count myself so lucky to have you in my life. **Ewoud**, my bun. Thank you!

A

Description of the Artworks

It was just a simple thought at the beginning to design the thesis cover using artificial intelligence (AI). I was very curious to see what this 'much more advanced data-driven' technique would bring to me. And indeed, this thesis is the beautiful outcome. Although there were some struggles (including the currently very vague legal and copyright regulations for AI-generated contents), it is possible and good to use them in this thesis.

When designing the artworks, I did some experiments. There are some painters that I like, and I was curious to see what would happen if I used the title of my paper as an instruction for the painting and requested a painting in the style of those painters.

The table on the right shows the examples of the instructions that led to the artworks in this thesis.

It is not difficult to notice that, although these paintings are beautiful and unique, they do not necessarily adhere to the given instructions or follow the style of the painter. In fact, they often deviated substantially from the intended style. Nevertheless, it is incredibly impressive to see how AI interpreted instructions in its own unique way and generated paintings that align with the rather abstract title of the paper.

Some disclaimers:

All the artworks in this thesis, including the cover but excluding the figures for scientific results, were created using DALL·E 2 (openai.com/dall-e-2) under my instruction. The use of the images adheres to the Content Policies and Terms of DALL·E.

It should also be noted that DALL·E only generates square images, and there is an 80% chance that the generated contents may not be entirely satisfactory, which therefore require additional editing. Furthermore, it is important to note that the paintings of the referenced artists currently belong to the public domain.

Place	Reference style (of the painter)	Typed instructions: to illustrate a painting with
Cover	Robert Delaunay	bikes, books, people with different lifestyle, some fruits stands
Chapter 1	Abstractionism	a sharp open door in the midnight, next to a big tall tree
Part 1	Edvard Munch	a mass chess board, lively colors
Chapter 2	Claude Monet	blood lipids-related dietary patterns and diabetes
Chapter 3	Grant Wood	ultra-processed food and incident type 2 diabetes studying the underlying consumption patterns to unravel the health effects of this heterogeneous food category
Chapter 4	Édouard Manet	lifestyle patterns
Part 2	Pieter Bruegel the Elder	a tip of an iceberg in a big ocean
Chapter 5	Sandro Botticelli	using structural equation modeling to untangle pathways of risk factors associated with incident type 2 diabetes
Chapter 6	Camille Pissarro	education, income, and health
Chapter 7	Robert Delaunay	effects of individual and neighborhood socio-economic disadvantage on health-related lifestyle risk factors
Chapter 8	Joaquín Sorolla	pine trees growing on black soil full of pine needles; iris flower fields in a open ground and blue sky; a Dutch tulip field, peaceful weather; poplars in China; tropical jungle forests full of exotic plants; and cactus growing on red deserts. (Each was a separate request.)
Appendices	Frédéric Bazille	an art studio with books
PhD Portfolio and List of Publications	Frédéric Bazille	books in a delicate botanical garden

A

PhD Portfolio

Courses and Workshops

Statistics

Mixed models for clustered data

Longitudinal and incomplete data (FLAMES summer school statistics, Gent)

Advanced (non)linear regression techniques in R

Public Health

Food environments and public health (KNAW, Amsterdam)

Research Skills

Publishing in English

Science writing for clinical and epidemiological research

Critical appraisal of literature

Transferrable Skills

Data visualization

Teacher training workshop

PhD Essential Courses

Ethics of research and scientific integrity

Research data awareness workshop

Managing your PhD

My future career

PROMINENT Training Schools and Activities in Personalized Medicine

Training Schools and Courses

- Personalized medicine in diabetes practice: tools to identify patient response

- Regulatory aspects of personalized medicine: drug registration experiences in oncology

- Personalized medicine and industry: models of drug development, cost-effectiveness

- Mechanisms for response variations in personalized medicine
- Personalized lifestyle management and drug treatment in type 2 diabetes
- Personalized medicine and payers in diabetes care
- Future biomarker discovery for drug response variation

PROMINENT Symposiums and Meetings

Midterm symposium: Barriers and future directions of personalized medicine: From bench to the patients, Online, 2020

Final symposium: Translating personalized medicine research from bench to bedside: opportunities and challenges for diabetes, Forum, Groningen, 2022.
Member of the organization team.

PROMINENT monthly research meeting (2 presentations), 2018-2022

Research Internship
Socioeconomic inequalities in mortality among people with and without diabetes in the Netherlands

In this study, I investigated the mortality rates and life expectancy among people with and without diabetes across different income groups. This analysis is based on the data from the Nivel Primary Care registry data in conjunction with the Dutch population data from the CBS.
Supervision: Dr. HBM Hilderink, Dr. R Poos, Dr. LH Dekker, National Institute for Public Health and the Environment (RIVM), 2022.

Research Meetings
Lifestyle unit meeting (2 presentations), 2020-2022
Kidney center meeting (1 presentation), 2019-2020

Other Activities
Attending PhD Day 2021
Supervision of 1 master research internship project

Conferences and Presentations

European Nutrition Conference FENS	Oral presentation	Dublin, Ireland	2019
World Congress on Public Health / European Public Health Conference (joint conference)	Oral presentation × 2	Online	2020
Annual Dutch Diabetes Research Meeting	Oral presentation	Online	2020
European Diabetes Epidemiology Group Conference	Oral presentation × 2	Online	2021
Nutrition Live Online American Society for Nutrition	Poster presentation	Online	2021
European Association for the Study of Diabetes Annual Meeting	Short oral presentation	Stockholm, Sweden	2022
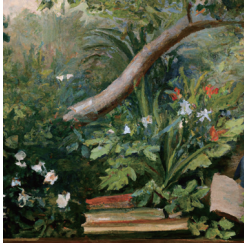
List of Publications

Publications included in this thesis

Duan MJ, Dekker LH, Carrero JJ, Navis G. Blood lipids-related dietary patterns derived from reduced rank regression are associated with incident type 2 diabetes. *Clinical Nutrition*. 2021; 40(7): 4712-4719. doi: 10.1016/j.clnu.2021.04.046.

Zhu Y*, **Duan MJ***, Riphagen IJ, Minovic I, Mierau JO, Carrero JJ, Bakker SJL, Navis GJ, Dekker LH. Separate and combined effects of individual and neighbourhood socio-economic disadvantage on health-related lifestyle risk factors: a multilevel analysis. *International Journal of Epidemiology*. 2021; 50(6): 1959-1969. doi: 10.1093/ije/dyab079.

Duan MJ*, Vinke PC*, Navis G, Corpeleijn E, Dekker LH. Ultra-processed food and incident type 2 diabetes: studying the underlying consumption patterns to unravel the health effects of this heterogeneous food category in the prospective Lifelines cohort. *BMC Medicine*. 2022; 20(1): 7. doi: 10.1186/s12916-021-02200-4.

Duan MJ, Dekker LH, Carrero JJ, Navis G. Using Structural Equation Modeling to Untangle Pathways of Risk Factors Associated with Incident Type 2 Diabetes: the Lifelines Cohort Study. *Prevention Science*. 2022; 23(7): 1090-1100. doi: 10.1007/s11121-022-01357-5.

Duan MJ, Dekker LH, Carrero JJ, Navis G. Lifestyle patterns and incident type 2 diabetes in the Dutch Lifelines cohort study. *Preventive Medicine Reports*. 2022; 30:102012. doi: 10.1016/j. pmedr.2022.102012.

Duan MJF*, Zhu Y*, Dekker LH, Mierau JO, Corpeleijn E, Bakker SJL, Navis G. Effects of Education and Income on Incident Type 2 Diabetes and Cardiovascular Diseases: a Dutch Prospective Study. *Journal of General Internal Medicine*. 2022; 37(15): 3907-3916. doi: 10.1007/s11606-022-07548-8.

A

Other publications

Zhu Y, **Duan MJ**, Dijk HH, Freriks RD, Dekker LH, Mierau JO; Lifelines Corona Research Initiative. Association between socioeconomic status and self-reported, tested and diagnosed COVID-19 status during the first wave in the Northern Netherlands: a general population-based cohort from 49 474 adults. *BMJ Open*. 2021; 11(3):e048020. doi: 10.1136/bmjopen-2020-048020.

Osté MCJ, **Duan MJ**, Gomes-Neto AW, Vinke PC, Carrero JJ, Avesani C, Cai Q, Dekker LH, Navis GJ, Bakker SJL, Corpeleijn E. Ultra-processed foods and risk of all-cause mortality in renal transplant recipients. *The American Journal of Clinical Nutrition*. 2022; 115(6):1646-1657. doi: 10.1093/ajcn/nqac053.

Cai Q, **Duan MJ**, Dekker LH, Carrero JJ, Avesani CM, Bakker SJL, de Borst MH, Navis GJ. Ultraprocessed food consumption and kidney function decline in a population-based cohort in the Netherlands. *The American Journal of Clinical Nutrition*. 2022; 116(1): 263-273. doi: 10.1093/ajcn/nqac073.

*Indicating joint first authorship

About the Author

Ming-Jie (Frederick) Duan was born on May 12th, 1992, in Beijing Haidian, China. After completing his secondary education at the High School Affiliated to Beijing Institute of Technology, he began his undergraduate studies, majoring in Food Science and Engineering at China Agricultural University. He conducted experimental research for his bachelor thesis on the topic "Research and development of lupin-based functional foods".

In 2015, Frederick was admitted to the master program in Human Nutrition and Rural Development at Ghent University in Belgium and was awarded a full scholarship from the Flemish government. The multidisciplinary master courses broadened his perspectives and sparked his scientific interest in public health nutrition. For his master thesis, he analyzed the effects of unconditional cash transfers on improving household WaSH (water, sanitation, and hygiene) conditions in Burkina Faso. This research was conducted as part of the project MAM'OUT, which aimed to evaluate the effects of unconditional cash transfers for the prevention of childhood moderate and acute malnutrition. Frederick graduated with cum laude in 2017.

A

His strong interest in public health nutrition made him move further and in 2018 he was selected as one of the PhD candidates for the PROMINENT project at the University Medical Center Groningen. Frederick focused his studies on analyzing lifestyle patterns and the underlying factors for better type 2 diabetes prevention, under the supervision of Prof. GJ Navis and Dr. LH Dekker. This thesis is the beautiful outcome. During his PhD studies, he presented his research at various scientific conferences and also participated in multidisciplinary training programs. In 2022, he did his research internship at the RIVM, focusing on socioeconomic inequalities in mortality among people with and without diabetes in the Netherlands.

Frederick is currently working as a postdoctoral researcher at Wageningen University & Research in the Department of Human Nutrition and Health. His research primarily focuses on assessing the population non-communicable diseases burden due to suboptimal diets, as well as the socioeconomic inequalities in sustainable diet transitions.