From Institute of Environmental Medicine
Karolinska Institutet, Stockholm, Sweden

# MODIFIABLE RISK FACTORS, BLOOD PROTEINS, AND VENOUS THROMBOEMBOLISM

Shuai Yuan

袁 帅


Karolinska Institutet

Stockholm 2024

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2024

ISBN 978-91-8017-206-6
Cover illustration: The cover illustration was designed by Shuai Yuan using Midjourney (https://www.midjourney.com/app/). Shuai Yuan owns the copyright of the cover illustration under the Midjourney membership. The cover shows a tree of peripheral vessels and its roots from two lungs. The centered circle shows the formation of a thrombus in the micro perspective.

# Modifiable risk factors, blood proteins, and venous thromboembolism

Thesis for Doctoral Degree (Ph.D.)

By

**Shuai Yuan (袁 帅)**

The thesis will be defended in public at the Samuelsson Hall, Tomtebodavägen 6, Karolinska Institutet, Solna, March 8, 2024, at 9:30

**Principal Supervisor:**
Associate Professor Susanna C. Larsson
Karolinska Institutet
Institute of Environmental Medicine
Unit of Cardiovascular and Nutritional Epidemiology

**Co-supervisor(s):**
Professor Agneta Åkesson
Karolinska Institutet
Institute of Environmental Medicine
Unit of Cardiovascular and Nutritional Epidemiology

Associate Professor Maria Bruzelius
Karolinska Institutet
Department of Medicine

**Opponent:**
Professor Bengt Zöller
Lund University
Department of Clinical Sciences
Center for Primary Health Care Research

**Examination Board:**
Associate Professor Åsa Johansson
Uppsala University
Department of Immunology, Genetics and Pathology

Associate Professor Per Svensson
Karolinska Institutet
Department of Cardiology

Professor Lauren Lissner
University of Gothenburg
Institute of Medicine

# Popular science summary of the thesis

Venous thromboembolism (VTE) is a thrombotic disorder that typically originates in the deep veins, leading to deep vein thrombosis, and can potentially migrate to the lungs through circulation, resulting in pulmonary embolism. VTE has a high incidence of 1-2 ‰ as stroke and becomes particularly prevalent among the elderly. While VTE can be managed through anticoagulation therapies, it is essential to acknowledge that VTE *per se* or its treatments can result in significant health ramifications, such as elevated mortality rates, bleeding complications, pulmonary hypertension, physical impairment, and a decline in overall quality of life. Thus, it is crucial to identify modifiable risk factors to provide evidence support for effectively reducing the incidence of VTE from the outset. Although some VTE conditions caused by a certain gene mutation (e.g., Factor V Leiden thrombophilia) are unpreventable, fortunately, the risk of disease, particularly unprovoked VTE, is generally modifiable.

In this project, we conducted five studies aiming at addressing three primary research questions. Firstly, we investigated the modifiable risk factors for VTE, secondly, we explored the blood proteins that have a causal association with VTE, and finally, we delved into the protein pathways that underlie the connections between modifiable risk factors and VTE. These studies employed observational prospective cohort and Mendelian randomization methodologies, utilizing phenotypic and genetic data from extensive cohorts of European populations. Collectively, our findings demonstrated that various factors, including obesity, cigarette smoking, physical inactivity, diet, insomnia, and numerous blood proteins, were associated with the risk of VTE. Furthermore, we identified annexin II and coagulation factor XI as critical protein mediators in the links between obesity, smoking, insomnia, and an increased risk of VTE. In addition, low-density lipoprotein receptor-related protein 12 mediated most of the association between obesity and VTE.

These findings reveal important implications for VTE prevention and treatment. Initially, it appears advantageous to mitigate the risk of VTE by maintaining a healthy body weight and embracing a healthy lifestyle, aligning with the recommendations for overall cardiovascular disease prevention, despite the varying pathological mechanisms between VTE and other cardiovascular conditions. Additionally, the discovery of causal proteins associated with VTE provides valuable insights that can inform the development of potential drug targets, albeit within the realm of exploratory research. Such knowledge opens doors to novel therapeutic avenues. Lastly, the identification of protein pathways connecting modifiable risk factors and VTE greatly enhances our understanding of the underlying pathology driving VTE development. This newfound understanding can serve as a crucial guide for targeted VTE treatments, facilitating the precise modulation of protein mediators to effectively manage the condition.

# Abstract

Venous thromboembolism (VTE) refers to blood clots in the veins, which is an under-appreciated vascular disease that can cause disability and mortality. Although some triggers for VTE (e.g., surgery, fracture, infection, hospitalization, and cancer) have been established, the associations of modifiable risk factors and blood proteins with the risk of VTE remain uncertain. This PhD project aimed to 1) investigate the associations of obesity and lifestyle factors with VTE risk; 2) explore the associations of blood proteins with VTE risk; and 3) establish protein pathways linking modifiable risk factors to VTE development.

In **Paper I,** we explored the associations of overall and central obesity with the risk of VTE using both cohort and Mendelian randomization analyses. We found a potentially causal association between obesity and VTE risk. Waist circumference might be a preferable indicator linking obesity to VTE. Around 12.4% and 23.7% of VTE cases could be prevented if the population maintained a healthy body mass index and waist circumference, respectively. In **Paper II,** using the prospective cohort design, we investigated the associations of cigarette smoking, alcohol and coffee intake, physical activity, and diet with the risk of incident VTE. We found that high levels of physical activity and a healthy diet were associated with lower VTE risk in women and men. Cigarette smoking showed a positive association with VTE only in women. Alcohol and coffee intake was not associated with VTE. In **Paper III**, we explored the association between ultra-processed food intake and the risk of VTE using the prospective cohort design. A higher ultra-processed food intake was associated with a moderately increased risk of VTE. This association was not modified by age, sex, or body mass index. In **Paper IV,** we conducted a prospective cohort and Mendelian randomization study to estimate the associations of 257 blood proteins with VTE risk. The cohort analysis identified 21 blood proteins associated with incident VTE. Machine-learning analysis found that body mass index and von Willebrand factor shared an identical highest ranking concerning the contribution to the prediction model. Mendelian randomization analysis confirmed 7 protein-VTE associations. In **Paper V**, we performed a two-stage network Mendelian randomization analysis to decipher proteomic pathways underlying the associations of 15 modifiable risk factors with VTE. We found that several proteins, in particular annexin II and coagulation factor XI, mediated the associations of obesity, smoking, and insomnia with VTE. Proteome-wide Mendelian randomization analysis identified many VTE-associated proteins with druggable potentials.

In summary, the above five studies identified modifiable risk factors and blood proteins for VTE development and further revealed protein pathways underlying the associations between modifiable risk factors and VTE. These findings may deepen understanding of VTE pathogenesis and facilitate precision prevention and drug development for VTE.

# List of scientific papers

I. **Yuan S**, Bruzelius M, Xiong Y, Håkansson N, Åkesson A, Larsson SC. Overall and abdominal obesity in relation to venous thromboembolism. *Journal of Thrombosis and Haemostasis*. 2021;19(2):460-469.

II. **Yuan S**, Bruzelius M, Håkansson N, Åkesson A, Larsson SC. Lifestyle factors and venous thromboembolism in two cohort studies. *Thrombosis Research*. 2021;202:119-124.

III. **Yuan S**, Chen J, Fu T, Li X, Bruzelius M, Åkesson A, Larsson SC. Ultra-processed food intake and incident venous thromboembolism risk: prospective cohort study, *Clinical Nutrition*. 2023;42(8):1268-1275.

IV. **Yuan S**, Titova OE, Zhang K, Gou W, Schillemans T, Natarajan P, Chen J, Li X, Åkesson A, Bruzelius M, Klarin D, Damrauer SM, Larsson SC. Plasma protein and venous thromboembolism: prospective cohort and mendelian randomisation analyses. *British Journal of Haematology*. 2023;201(4):783-792.

V. **Yuan S**, Xu F, Zhang H, Chen J, Ruan X, Li Y, Burgess S, Åkesson A, Li X, Gill D, Larsson SC. Proteomic insights into modifiable risk of venous thromboembolism and cardiovascular comorbidities. *Journal of Thrombosis and Haemostasis*. 2023. doi: 10.1016/j.jtha.2023.11.013. Epub ahead of print.

# Other scientific papers related to the topic of this doctoral thesis

I. **Yuan S**, Bäck M, Bruzelius M, Mason AM, Burgess S, Larsson S. Plasma Phospholipid Fatty Acids, FADS1 and Risk of 15 Cardiovascular Diseases: A Mendelian Randomisation Study. *Nutrients*. 2019;11(12):3001.

II. **Yuan S**, Carter P, Bruzelius M, Vithayathil M, Kar S, Mason AM, Lin A, Burgess S, Larsson SC. Effects of tumour necrosis factor on cardiovascular disease and cancer: A two-sample Mendelian randomization study. *EBioMedicine*. 2020;59:102956.

III. **Yuan S**, Mason AM, Burgess S, Larsson SC. Genetic liability to insomnia in relation to cardiovascular diseases: a Mendelian randomisation study. *European Journal Epidemiology*. 2021;36(4):393-400.

IV. **Yuan S**, Burgess S, Laffan M, Mason AM, Dichgans M, Gill D, Larsson SC. Genetically Proxied Inhibition of Coagulation Factors and Risk of Cardiovascular Disease: A Mendelian Randomization Study. *Journal of the American Heart Association*. 2021;10(8):e019644.

V. **Yuan S**, Carter P, Mason AM, Burgess S, Larsson SC. Coffee Consumption and Cardiovascular Diseases: A Mendelian Randomization Study. *Nutrients*. 2021;13(7):2218.

VI. **Yuan S**, Bruzelius M, Damrauer SM, Håkansson N, Wolk A, Åkesson A, Larsson SC. Anti-inflammatory diet and venous thromboembolism: Two prospective cohort studies. *Nutrition, Metabolism and Cardiovascular Diseases*. 2021;31(10):2831-2838.

VII. **Yuan S**, Mason AM, Carter P, Burgess S, Larsson SC. Homocysteine, B vitamins, and cardiovascular disease: a Mendelian randomization study. *BMC Medicine*. 2021;19(1):97.

VIII. **Yuan S**, Zheng JS, Mason AM, Burgess S, Larsson SC. Genetically predicted circulating vitamin C in relation to cardiovascular disease. *European Journal of Preventive Cardiology*. 2022;28(16):1829-1837.

IX. **Yuan S**, Bruzelius M, Larsson SC. Causal effect of renal function on venous thromboembolism: a two-sample Mendelian randomization investigation. *Journal of Thrombosis and Thrombolysis*. 2022;53(1):43-50.

X. **Yuan S**, Li X, Morange PE, Bruzelius M, Larsson SC, On Behalf Of The Invent Consortium. Plasma Phospholipid Fatty Acids and Risk of Venous Thromboembolism: Mendelian Randomization Investigation. *Nutrients*. 2022;14(16):3354.

XI. **Yuan S**, Mason AM, Burgess S, Larsson SC. Differentiating Associations of Glycemic Traits With Atherosclerotic and Thrombotic Outcomes: Mendelian Randomization Investigation. *Diabetes*. 2022;71(10):2222-2232.

XII. Li H, Zhang Z, Qiu Y, Weng H, **Yuan S**, Zhang Y, Zhang Y, Xi L, Xu F, Ji X, Hao R, Yang P, Chen G, Zuo X, Zhai Z, Wang C. Proteome-wide mendelian randomization identifies causal plasma proteins in venous thromboembolism development. *Journal of Human Genetics*. 2023;68(12):805-812.

XIII. **Yuan S**, Sun Y, Chen J, Li X, Larsson SC. Long-term risk of venous thromboembolism among patients with gastrointestinal non-neoplastic and neoplastic diseases: A prospective cohort study of 484 211 individuals. *American Journal of Hematology*. 2023. doi: 10.1002/ajh.27106. Epub ahead of print.

# Contents

# List of abbreviations

| | |
|---|---|
| BMI | Body mass index |
| CFHR5 | Complement factor H related 5 protein |
| CI | Confidence interval |
| DVT | Deep vein thrombosis |
| EPHB4 | Ephrin type-B receptor 4 |
| FDR | False discovery rate |
| FXI | Coagulation factor XI |
| GWAS | Genome-wide association study |
| HR | Hazard ratio |
| ICD | International Classification of Diseases |
| IV | Instrumental variable |
| LD | Linkage disequilibrium |
| LightGBM | Light Gradient Boosting Machine |
| LRP12 | Low-density lipoprotein receptor-related protein 12 |
| MR | Mendelian randomization |
| MR-PRESSO | Mendelian Randomization Pleiotropy RESidual Sum and Outlier |
| MVP | Million Veteran Program |
| OR | Odds ratio |
| PE | Pulmonary embolism |
| PRS | Polygenic risk score |
| SERPINE1 (or PAI-1) | Plasminogen activator inhibitor 1 |
| SHAP | SHapley Additive exPlanations |
| SNPs | Single nucleotide polymorphsims |
| TNFRSF11A | Tumor necrosis factor receptor superfamily member 11A |
| UPFs | Ultra-processed foods |
| VTE | Venous thromboembolism |
| vWF | von Willebrand factor |
| WC | Waist circumference |
| WHR | Waist-to-hip ratio |

# 1 Introduction

Venous thromboembolism (VTE) is an under-recognized disorder of the circulatory system, impacting over 10 million people globally each year (1), and significantly escalating mortality risk (2). Despite evidence suggesting that the etiology of VTE distinctly deviates from that of atherosclerotic cardiovascular diseases (3), a comprehensive and robust understanding of the modifiable risk factors driving thrombus development has not yet been conclusively established, largely due to inconsistent and insufficient data (4). Thorough investigation into modifiable risk factors for VTE not only amplifies our comprehension of disease pathogenesis but also furnishes invaluable insights for disease prevention, risk assessment, and stratification as well as patient management, ultimately contributing to enhanced patient outcomes.

In clinical practice, anticoagulants and antiplatelet drugs are routinely administered to patients with VTE. However, these treatments often carry a heightened risk of excessive bleeding, and in severe cases, can be fatal (5). Even though several new antithrombotic drugs have been developed and are under testing (6), there remains a critical need for medications that deliver superior efficacy with fewer side effects. Blood proteins play a central role in maintaining molecular pathways, and many serve as therapeutic targets (7). Investigating the roles of circulating blood proteins in thrombosis could yield valuable insights into disease pathophysiology and unveil novel drug targets for treating blood clots. Moreover, there is a paucity of data elucidating the protein pathways linking modifiable risk factors to VTE development. Unraveling this complex interplay could substantiate evidence-based patient management and precision medicine strategies.

# 2 Background

## 2.1 Definition of VTE

VTE denotes a blood clotting disorder predominantly observed in the deep veins and the lungs. It is categorized into two subtypes: deep vein thrombosis (DVT) and pulmonary embolism (PE). DVT often begins with the formation of a blood clot in a deep vein, commonly located within the calf. This clot can subsequently extend to the more proximal veins. If a fragment of the clot dislodges, it may travel within the bloodstream to the lungs, resulting in PE. Nevertheless, it's important to note that not every patient with symptomatic PE will have imaging confirmed DVT in either the proximal or distal veins, and an even smaller number will present with clinically evident symptoms in the legs.

## 2.2 Pathophysiology

### 2.2.1 Molecular pathophysiology

The conventional model of thrombus or blood clot formation is encapsulated in Virchow's triad, a set of three main factors: hypercoagulability, blood flow stasis, and endothelial injury (8). However, emerging evidence suggests that inflammation intertwines with the coagulation process, introducing a more intricate interplay that broadens the underlying mechanism towards a concept known as immunothrombosis (1, 9).

In brief, immunothrombosis means activation of coagulation cascade triggers the immune system, and in turn, innate immune cells contribute to thrombus formation (**Figure 1**) (1). Venous thrombi are hypothesized to originate within the valve pockets of large veins, areas prone to blood stasis, especially during extended periods of immobility (10). Stasis and endothelial injury or dysfunction can give rise to hypoxia or inflammation (10). These processes subsequently

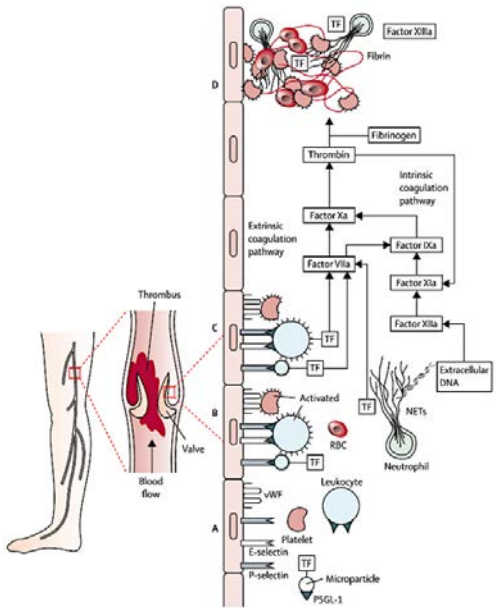



**Figure 1**. The interplay between coagulation and inflammation in the development of venous thrombosis. Abbreviations: NETs, neutrophil extracellular traps; PSGL-1, P-selectin glycoprotein ligand-1; RBC, red blood cell; TF, tissue factor; vWF, von Willebrand factor. (Source: Khan et al. 2021, *Lancet*. Permission obtained from Elsevier)

diminish the endothelium's natural anticoagulant properties, thereby ushering in a hypercoagulable state. The activation of procoagulant venous endothelial cells results in heightened expression of surface adhesion molecules. These molecules ease the subsequent attachment of circulating leukocytes, microparticles, and platelets. Once activated, leukocytes express tissue factor, a procoagulant thought to spark the extrinsic pathway of the coagulation cascade, with the intrinsic (contact) pathway activated via factors XII and XI. This cascade activation leads to the stimulation of factor X, the production of thrombin, and finally results in the creation of a thrombus composed of fibrin, red blood cells, and platelets.

### 2.2.2 Genetic pathophysiology

One of the prominent genes associated with VTE risk is the *F5* gene, which encodes factor V. A specific mutation in this gene, known as Factor V Leiden, is present in approximately 5% of Caucasians. This mutation doubles the risk of thrombus formation by hampering the protein's susceptibility to inactivation by the anticoagulant protease, activated protein C (11). Six other well-established genetic risk factors for VTE have been identified. These encompass mutations in genes encoding prothrombin, fibrinogen gamma, and non-O blood group, which, although relatively common in the general population, only mildly influence VTE risk. Conversely, mutations in genes encoding antithrombin, protein C, and protein S are found in less than 1% of the general population, yet they exert substantial effects, increasing the VTE risk by approximately tenfold (4).

Genome-wide association analysis is a powerful tool for identifying common genetic variants, such as single nucleotide polymorphisms (SNPs), that correlate with complex disorders, without the need for a preconceived hypothesis. The inaugural genome-wide association study on VTE was carried out in 2008 by Dutch researchers. They focused on three case-control studies of initial DVT, identifying three SNPs strongly linked to DVT in the genes *CYP4V2* (in linkage disequilibrium with *F11*), antithrombin (*SERPINC1*), and *GP6 (12)*. Since then, the genetic landscape of VTE has been progressively unveiled due to an increase in the sample size of GWASs, particularly because of the significant contributions from large-scale genetic consortia (13). In 2019, the International Network on VENous Thrombosis (INVENT) consortium amalgamated data from 18 studies involving 30,234 VTE cases and 172,122 controls, executing a meta-analysis of genome-wide association analyses (14). This resulted in the identification of 34 genetic signals implicated in VTE risk, which included both genes involved in traditional thrombosis pathways and genes with hitherto unknown functions (14). Fast forward to 2023, another expansive GWAS meta-analysis spotlighted 93 loci for VTE amongst 81,190 cases and 1,419,671 controls of European ancestry (15). This study underscored the crucial roles of the *F2* and *F5* genes in VTE development, but also underscored the significant impacts of other genes. This was discernible through the observation that both individuals in the top 0.1% of the polygenic risk score (PRS) distribution and carriers of homozygous or

compound heterozygous variants G20210A (c.*97 G > A) in *F2* and p.R534Q in *F5* exhibited similar VTE risk, while *F2* and *F5* mutation carriers in the bottom 10% of the PRS distribution demonstrated a VTE risk equivalent to that of the general population (15). Another GWAS meta-analysis involving 81,669 VTE cases of diverse ancestries pinpointed 135 independent loci for VTE (**Figure 2**) (16). Notably, this study expanded the genetic comprehension of VTE in non-European populations. Collectively, these studies shed light on novel pathophysiological mechanisms of VTE from the genetic perspective. Concurrently, the accessibility of genome-wide association analysis data forms a robust foundation for Mendelian randomization (MR) analysis.

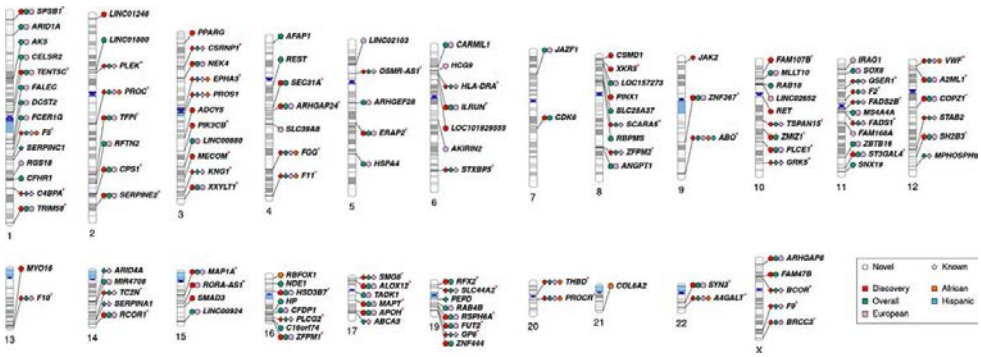



**Figure 2.** Genetic loci associated with VTE in a cross-ancestry GWAS meta-analysis. The study encompassed five meta-analyses: the discovery analysis (indicated in red), the comprehensive meta-analysis (denoted in green), along with analyses focusing on individuals of European ancestry (represented in purple), African ancestry (highlighted in orange), and Hispanic ancestry (portrayed in blue). Newly discovered loci are symbolized with circles, while known loci are represented with diamonds. Loci supported by replication evidence bear a red asterisk. (Source: Thibord et al. 2022, *Circulation.* Permission obtained from Wolters Kluwer Health, Inc.)

## 2.3 Epidemiology

VTE is a global health concern, characterized by substantial morbidity and mortality rates. Its prevalence is significant across various demographics, with the elderly being particularly susceptible. In the United States, the lifetime risk of VTE is approximately 8% overall among US citizens (17). The incidence of VTE is highest among African Americans, while it is lowest in Asian and Native American populations (18). The risk of VTE escalates exponentially with age (19) and approximately 30% of VTE cases reoccur within 5-10 years of the initial onset (20, 21). Around 20% of individuals succumb within a year of being diagnosed with VTE, often due to the VTE itself or underlying conditions that precipitated the event (22). Despite the considerable strain VTE places on healthcare systems (23, 24), public awareness about thrombosis in general, and VTE in particular, remains disappointingly low (25).

From a global perspective, VTE impacts approximately 10 million individuals each year, ranking as the third most common vascular disease, following coronary heart disease

and stroke (26). It has been estimated that the annual VTE incidence rate is about 1 to 2 cases per 1000 individuals, with a rate four times greater in high-income countries compared to low-income countries (1, 27). However, comprehensive global data on VTE incidence trends remain scarce. In Worcester, Massachusetts, USA, the VTE incidence rate either remained steady or increased from 1981 to 2009 (28). In Sweden, there are 2-3 patients with VTE (including both incident and recurrent cases) per 1000 inhabitants according to data from the patient register. The rate keeps stable from 2008 to 2022 and is slightly higher in men (**Figure 3**). The recent COVID-19 pandemic has led to an upsurge in the global VTE incidence rate, especially in cases of PE (29). The COVID-19 impact on PE can be observed in Sweden as well (an increase from 2020 to 2021).

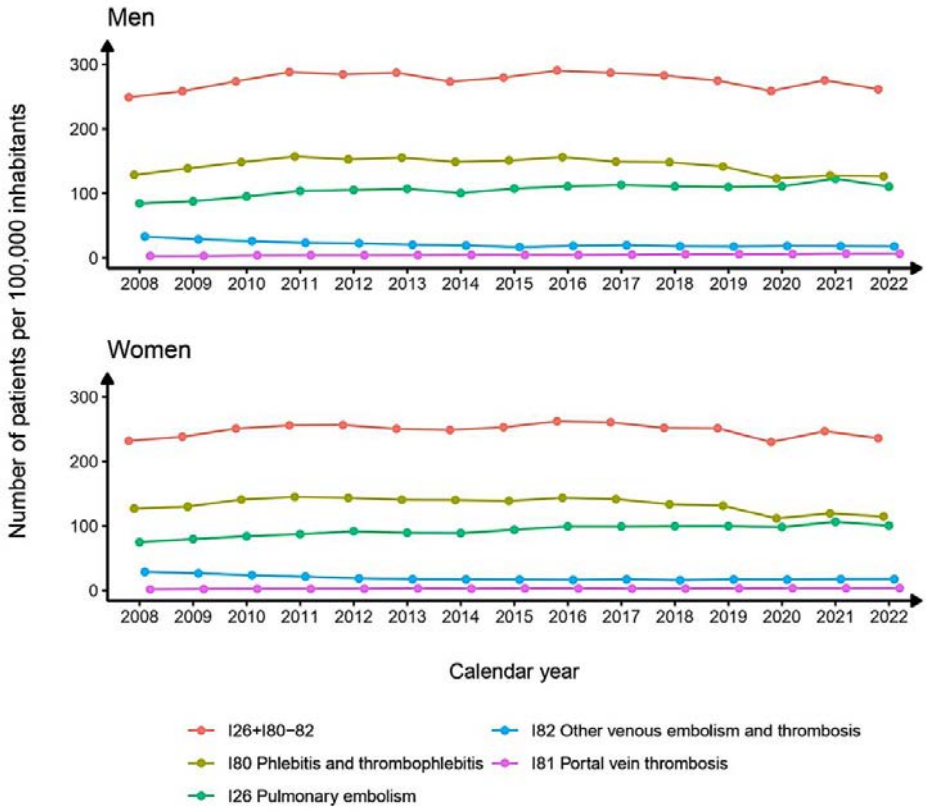



**Figure 3**. Number of patients with a VTE (ICD codes I26, I80, I81, or I82) diagnosis per 100,000 inhabitants aged 0-85+ in the Swedish patient register from 2008 to 2022. The number was age-standardized according to 2022 Swedish population structure. The data was extracted from https://www.socialstyrelsen.se/. (Source: Personal Collection)

While no significant overall disparity in VTE incidence has been observed between men and women (30, 31), the relationship between sex and VTE incidence is not straightforward (32). Among those under 45, women exhibit a slightly higher incidence

rate due to reproductive-related factors, whereas in adults aged over 45, men generally have a higher incidence rate than women (32). Beyond 85 years of age, women demonstrate a higher VTE incidence rate, attributable to their longer life expectancy (32).

In Sweden, the overall incidence rate is 36.2 per 100,000 years for women and 32.5 per 100,000 years for men with data from a nationwide epidemiological study based on hospitalizations (33). The age-specific incidence rates of VTE between women and men in Sweden align with the previously outlined pattern (33) (**Figure 4**).

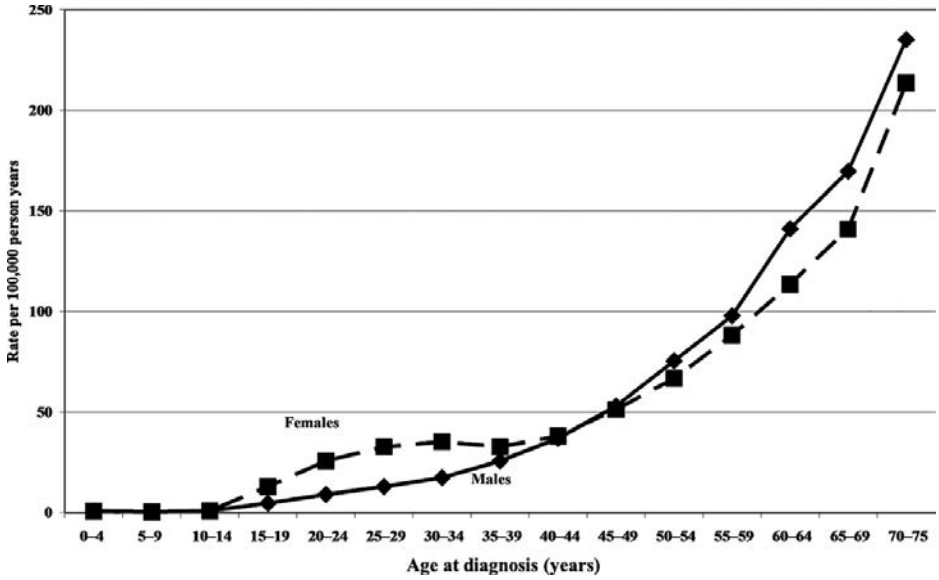



**Figure 4**. Age-specific incidence rates of VTE in siblings. (Source: Zöller et al. 2011, *Circulation*. Permission obtained from Wolters Kluwer Health, Inc.)

## 2.4 Risk factors for VTE

The development of VTE involves multifactorial factors and is influenced by a complex interplay of both genetic and environmental factors (34). In previous epidemiological and clinical studies, we usually classified VTE as provoked (occurring after triggering factors in the previous 3 months) or unprovoked (without clear triggering factors). However, the system has been challenged for being imprecise. For example, certain studies may categorize pregnancy and hormone related VTE as provoked incidents. Considering the multifactorial nature of VTE, it can be exceedingly difficult, if not impossible, to pinpoint a solitary attribute leading to VTE, or to label an accumulation of minor thrombotic challenges leading to a VTE event as unprovoked (35). The 2019 guidelines from the European Society of Cardiology and the European Respiratory Society for diagnosing and managing acute pulmonary embolism refrained from using terms like 'provoked' and 'unprovoked' (36). Instead, they centered their attention on

estimating the long-term risk of VTE recurrence (36). A recent review paper listed risk factors for VTE into four major categories, including acute trigger, subacute trigger, basal risk factor, and acquired risk factor according to the features of the risk factors and their associations with VTE (**Figure 5**) (35). In this thesis, nevertheless due to the widespread use of the terms 'provoked' and 'unprovoked' in existing literature, we employ this terminology. In addition, we majorly focus on the modifiable cluster of risk factors for VTE in this thesis.

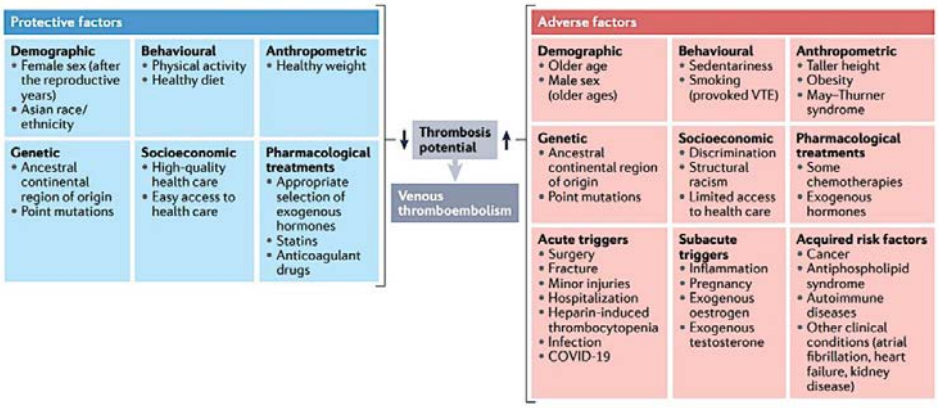



**Figure 5**. Factors with protective or adverse effects on thrombosis potential and risk of VTE. Demographic, behavioral, anthropometric, and genetic factors are labeled as basal risk factor. (Source: Lutsey et al. 2023, *Nat Rev Cardiol.* Permission obtained from Springer Nature)

### 2.4.1 Non-genetic, unmodifiable risk factors

Non-genetic, unmodifiable risk factors predominantly contribute to the risk of provoked VTE. For example, cancer and major surgery are two important acquired or acute albeit unmodifiable risk factors for VTE (4, 34). VTE risk differs across cancer types and stages (37). The overall risk of VTE has been estimated to be 13 (95% confidence interval [CI], 7-23) per 1000 person-years among average cancer risk patients and the risk is higher in patients with cancers of the pancreas, brain, and lung (37). The prothrombotic mechanism in cancer patients is complex and multifactorial, including effects from the host response to carcinoma, anticancer chemotherapy, and radiotherapy (38). However, the main pathway is led by prothrombotic properties expressed by malignant cells (38). These properties produce procoagulant, fibrinolytic, and pro-aggregating activities, release of proinflammatory and proangiogenic cytokines, and interact directly with host vascular and blood cells (38). At least 3-month low-molecular-weight heparin medication has been well demonstrated to be the treatment and management of VTE in cancer patients (39). Similarly, direct oral anticoagulants can be used to treat acute VTE in cancer patients; however, the bleeding risk should be carefully assessed (40). Trauma, surgery, and postoperative immobilization and hospitalization have been also identified as provoking risk factors for VTE (4). The VTE risk differs between surgical types,

duration, and clinical features of the patients (41-43). There are no consensus treatment strategies for patients undergoing different surgery or for those with same surgery albeit different clinical situations. Risk assessment, treatment, and management should set the basis at the individual model (41).

Use of certain medications, like exogenous hormones, corticosteroids, and statins, possibly alter VTE risk (4). For hormone replacement, oral therapy instead of transdermal treatment may increase VTE risk (44). Glucocorticoid use may increase VTE risk, and statin use (45), rosuvastatin use in particular, may decrease its risk (46). Some health issues have been associated with VTE risk, like pregnancy (47), HIV infection (48), kidney impairment (49), rheumatoid arthritis (50), inflammatory bowel disease (51), psoriasis (52), multiple Sclerosis (53), polycystic ovary syndrome (54), and Cushing's Syndrome (55).

### 2.4.2 Modifiable risk factors

In contrast to unmodifiable risk factors, modifiable risk factors, including obesity, cigarette smoking, an unhealthy diet, and physical inactivity, are attributable to the risk of unprovoked VTE, which makes up at least a half of VTE cases (56). Numerous studies have established links between these modifiable factors and the risk of VTE through observational research, yet the results vary significantly between studies. Furthermore, the causality of these associations remains uncertain due to potential limitations inherent in observational studies, such as confounding factors and reverse causation.

#### *2.4.2.1 Obesity and VTE*

A meta-analysis, which included three cohort and twelve case-control studies published prior to June 2006, reported a more than twofold increase in VTE risk for obese individuals with body mass index (BMI) >30 $kg/m^2$ compared to those with healthy BMI (57). This association was reaffirmed by an updated meta-analysis of twelve cohort studies, which also highlighted the linear nature of the relationship between BMI and VTE risk (58). The causality of this association was further substantiated by an MR study involving 7,507 VTE cases and 52,632 non-cases from the European population (59). Moreover, central obesity has been linked to VTE risk. A Danish cohort study of 641 incident VTE cases established a positive correlation between waist circumference (WC) and VTE risk (60), a finding later confirmed in a Norwegian cohort study involving both men and women (61). Both overall and abdominal obesity have been associated with VTE risk, but the relative impacts of these two measures of obesity on VTE are yet to be determined. Additionally, the question of whether the risk of VTE varies between individuals with different combinations of BMI and WC remains unresolved.

Several mechanisms have been suggested to elucidate the link between obesity and the heightened risk of VTE. Physically, obesity, particularly abdominal obesity, could

potentially impair venous return, increase intra-abdominal pressure, and decrease blood velocity in the femoral vein (62). From the molecular standpoint, there are possible pathways via influencing leptin levels, increasing activity of the coagulation cascade, and decreasing fibrinolysis (62). However, recent studies indicate that leptin may not play a mediating role in the connection between obesity and VTE (63, 64). Furthermore, obesity can trigger inflammation, oxidative stress, and endothelial dysfunction, which could subsequently elevate the risk of VTE (62). Despite these hypotheses, the definitive pathways underlying the obesity-VTE association largely remain to be determined, given the limited number of studies exploring these mechanistic pathways.

#### *2.4.2.2 Smoking and VTE*

The relationship between smoking and the risk of VTE was ambiguous in a meta-analysis of seven case-control studies and three cohort studies (57). However, an updated meta-analysis comprising 19 case-control studies and 13 cohort studies revealed a combined hazard ratio (HR) of 1.23 (95% CI, 1.14, 1.33) for current smokers and 1.10 (95% CI, 1.03, 1.17) for former smokers, as compared to those who had never smoked (65). This association was further confirmed in a subsequent study using data from 76 cohorts (66). The causal relationships of smoking with DVT and PE risk were reinforced by an MR study conducted using the UK Biobank data (67). The mechanisms underlying the association between smoking and VTE have been sparsely studied, but potential pathways might involve the effects of smoking on innate immunity, platelet function, and coagulation factor concentrations (68). Comorbid conditions, like cancer may mediate the association between cigarette smoking and VTE (69).

#### *2.4.2.3 Alcohol consumption and VTE*

The association between alcohol consumption and VTE risk remains inconsistent across previous studies, potentially varying depending on the types of alcoholic beverages consumed and the sex of the consumer. As for overall alcohol consumption, null (70), inverse (66), and positive (71) associations with VTE risk have been reported. A study in northern Sweden discovered that high alcohol consumption and alcohol dependency were linked with increased VTE risk in men, but not in women (71). Similarly, another Swedish study found that severe alcohol abuse was associated with a high risk of VTE (72). Conversely, a study utilizing data from the Emerging Risk Factors Collaboration and the UK Biobank study consistently found an inverse association between current alcohol consumption and VTE risk, which also applied to the quantity of alcohol consumed when the analysis was confined to current drinkers (66). However, a recent meta-analysis of 10 studies did not corroborate any associations with high alcohol consumption (70).

In terms of specific alcoholic beverages, a Norwegian study involving 26,662 participants found that liquor consumption and binge drinking were linked with an increased VTE risk,

while wine consumption was potentially associated with a reduced VTE risk after a median follow-up of 12.5 years (73).

#### *2.4.2.4 Coffee consumption and VTE*

Coffee, a globally popular beverage, particularly in Western countries, has been extensively studied for its association with cardiovascular disease risk. This association, as assessed in numerous observational studies, has been found to be nonlinear (74). However, the link between coffee consumption and VTW risk remains limited and yields inconsistent results (75). In the Tromsø study, which included 26,755 individuals followed for a median duration of 12.5 years, a U-shaped association between coffee consumption and VTE was discovered (76). This suggests that moderate coffee consumption might mitigate the risk of VTE. In contrast, an analysis conducted on the Iowa Women's Health Study, involving 37,393 women, indicated no clear association between coffee consumption and VTE risk (77). Consequently, future research is warranted to conclusively unravel the relationship between coffee intake and VTE risk.

#### *2.4.2.5 Diet and VTE*

Diet significantly impacts human health, but while many studies have investigated the relationship between dietary patterns, food components, nutrients, and VTE risk, the associations remain inconclusive. For instance, a healthy diet evaluated by the modified SmartDiet score showed no association with VTE risk in the Tromsø study (78). Similarly, adherence to the Dietary Approaches to Stop Hypertension (DASH) diet demonstrated no correlation with VTE risk in a study involving 34,827 women followed for an average of 14.6 years (79). However, in a recent cohort study involving 14,818 middle-aged adults, strong adherence to a prudent dietary pattern was linked with a lower VTE risk, while high adherence to a Western dietary pattern associated with a higher risk after an average follow-up period of 22 years (80). A positive association between a Western dietary pattern and VTE risk was observed in men but not in women in the Nurses' Health Study and Health Professionals Follow-up Study (81). Possible associations between VTE risk and intake of fruits and vegetables (82), red and processed meat (82), fish (83), fiber (81), vitamin B6 (81), vitamin E (81), marine omega-3 polyunsaturated fatty acids (84) have been suggested. However, current evidence remains insufficient to validate any of these proposed associations (85, 86).

#### *2.4.2.6 Physical activity and VTE*

Physical activity is a well-established factor for arterial thrombosis, but its association with venous thrombosis remains inconsistent. (4, 87). Several studies provide no evidence supporting an association between physical activity and VTE risk (3, 88). However, some research suggests a U-shaped correlation between physical activity and VTE risk, with strenuous daily physical activity seeming to increase the risk of VTE (89,

90). This elevated risk might be attributed to the increased likelihood of injuries resulting from strenuous physical activity (4). Despite the inconsistencies, current evidence leans towards a slight beneficial effect of moderate physical activity on the risk of incident VTE (91). A meta-analysis of 14 studies found a consistent inverse association between regular physical activity and VTE risk across various subgroups defined by geographical location, sex, or age (92). Nevertheless, the available evidence falls short of being conclusive in assessing the relationship between physical activity and VTE risk due to the inherent heterogeneity in the intensity, frequency, and duration of various forms of physical activity.

#### *2.4.2.7 Other modifiable risk factors and VTE*

Some traditional cardiovascular risk factors, such as diabetes, hyperlipidemia, and hypertension, have been unassociated with an increased risk of VTE (3, 69, 93, 94), which indicates the etiological differences between atherosclerosis and thrombosis. Despite limited data on sleep-related traits in relation to VTE, clinically diagnosed sleep disorders have been associated with a 1.79-fold higher risk of VTE among individuals without sleep apnea (95). In addition, insomnia may be causally associated with VTE risk in an MR study (96).

## 2.5 Blood proteins and VTE

Circulating proteins play a critical role in human health, including in thrombosis and hemostasis, as they primarily regulate molecular pathways. Owing to their fundamental roles, these blood proteins hold immense potential as diagnostic and predictive markers (97), as well as therapeutic targets (7). These protein biomarkers can be also used to illuminate the mechanistic pathways from the risk factor (e.g., certain treatment [tofacitinib versus tumor necrosis factor inhibitors] (98)) to the development of VTE.

Many important proteins have been identified to be associated with VTE risk. Increased levels of coagulation factor VIII (99) and XII (100) have been linked to a high risk of VTE. Likewise, high levels of blood von Willebrand factor (vWF) have been associated with an increases risk of incident VTE (101) and this association appears to be stronger for unprovoked events (102). In addition, high levels of plasma P-selectin (103) and lipopolysaccharide-binding protein (104) and lower levels of C1-inhibitor (105) have been associated with an increased risk of VTE. Elevated levels of complement factor H related 5 protein (CFHR5) are linked with an increased potential for thrombin generation, and in vitro studies have shown that recombinant CFHR5 enhances platelet activation; however, MR analysis generates no support of a causal association between CFHR5 and VTE (106). A cross-species study has identified heat shock protein 47 associated with VTE potential via attenuation of immune cell activation and neutrophil extracellular trap formation among immobile animals and patients (107).

**Table 1**. Publications of relevance for proteomics-based thrombosis research.

| | |
|---|---|
| Bruzelius et al., *Blood.* 2016 | This report describes the first large-scale affinity proteomics study in the venous thrombosis field. A total of 408 proteins are targeted in a discovery case/control study (VEBIOS, n = 88 cases and 85 controls) with validation in an independent case/control study (FARIVE, n = 580 cases, 589 controls). It describes the application of immuno-capture mass spectrometry to validate assay target specificity. Plasma level of platelet-derived growth factor β is identified as associated with venous thromboembolism risk. |
| Ten Cate et al., *Blood.* 2021 (109) | This report describes the application of machine learning techniques to analyze affinity proteomics data. A total of 444 proteins are targeted in a discovery cohort (Genotyping and Molecular Phenotyping of Venous ThomboEmbolism, n = 532 cases) with validation in an independent population cohort (Gutenberg Health Study, n = 5778). It describes the application of cis pQTL analysis to validate assay target specificity. Plasma levels of interferon-γ, glial cell-line derived neurotrophic factor, and interluekin-15Rα proteins are identified as associated with isolated pulmonary embolism. |
| Razzaq et al., *Sci Rep.* 2021 | This report describes an original integrated affinity proteomics and genetics strategy using a neural network approach, based on proteomics and genome-wide association studies data in the MARTHA study (n = 1388 cases) with replication in the EOVT study (n = 339 cases). PLXNA4 is identified as a new susceptibility gene for pulmonary embolism. |
| Iglesias et al., *Arterioscler Thromb Vasc Biol.* 2021 | This report describes a novel endothelial cell centric affinity proteomics strategy targeting 216 proteins with endothelial enriched expression in a population-based cohort (SCAPIS n = 1008). Plasma levels of 38 endothelial-derived proteins are identified as associated with cardiovascular disease risk. |
| Deutsch et al, *J Proteome Res.* 2021 | This publication provides a comprehensive overview of technological developments and applications of mass spectrometry- and affinity-based plasma proteomics methods, summarizing recent advances and challenges for translating plasma proteomics into clinical utility for precision medicine. It presents the Human Plasma PeptideAtlas build 2021-07 and the Human Extracellular Vesicle PeptideAtlas 2021-06. |

Source: Edfors et al. 2022, *Res Pract Thromb Haemost.* (Source: The paper was published under a CC BY NC ND license. Permission is not required for this non-commercial use.)

The advent of high-throughput methods for proteomic profiling, such as the SomaScan and Olink platforms, has opened opportunities to comprehensively study the associations between blood proteins and VTE risk. Several investigations have utilized data generated by these proteomic platforms and identified a multitude of proteins crucial for VTE (**Table 1**) (108). However, these studies, often limited by a small number of cases, may be subject to potential biases such as residual confounding and reverse causality, inherent in their observational design (108).

## 2.6 Mendelian randomization analysis

MR analysis is an epidemiological approach that can strengthen causal inference by using genetic variants as the instrumental variable (IV) to mimic the lifetime effects of the exposure of interest (109). It has been widely used in medical research nowadays due to its two major merits that are minimizing confounding and reverse causality (110). In brief, MR can diminish confounding since the used genetic variants as the IV are randomly assorted at conception and thus usually uncorrelated with confounders. This process resembles the randomization step in randomized controlled trials (**Figure 6**) (111). Additionally, the technique can minimize reverse causation because the germline genotypes are fixed and unmodifiable by the onset of progression of disease. However, a valid and robust MR association is heavily depending on three important assumptions: 1) the selected genetic instrument should be strongly associated with the exposure (relevance assumption); 2) the used genetic variants should not be associated with any confounders (independence assumption); and 3) the employed genetic variants should not impact the outcome directly or via alternative pathways (exclusion restriction assumption). Detailed elaborations on these assumptions have been discussed in the following sections.

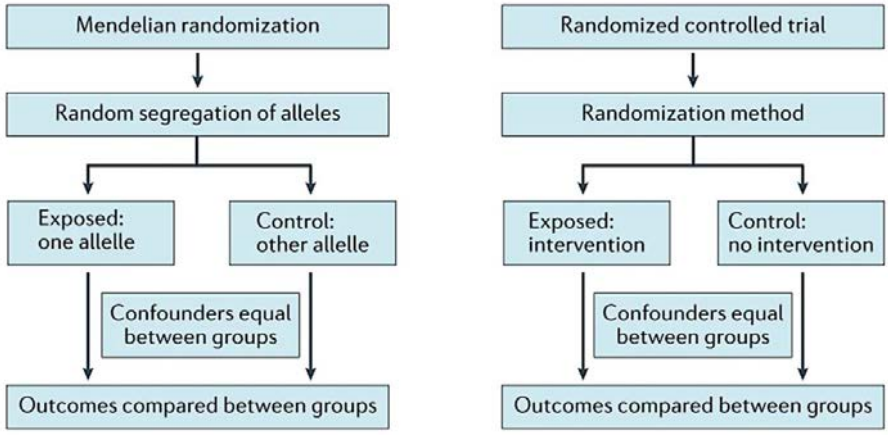



**Figure 6**. Comparison of the design of an MR study and a randomized controlled trial. In a randomized, controlled clinical trial, participants are randomly assigned to either the intervention (exposure) or control (non-exposure) groups, thereby equalizing any confounders that could impact the outcome. Similarly, Mendelian randomization involves the random allocation of alleles during meiosis, with one representing exposure and the other control. This natural allele randomization at the population level ensures that both groups are equally exposed to potential confounders. (Source: Robinson et al. 2016, *Nat Rev Rheumatol*. Permission obtained from Springer Nature)

### 2.6.1 Two-sample MR

MR was initially conceived in the format of one-sample MR, a method that requires access to individual-level data. However, a limitation of this method is that most

individual cohorts do not simultaneously measure multiple traits, which thus confines the application of this method. To address these limitations, two-sample MR was later introduced, utilizing summary-level data sources for the exposure and outcome from two different populations (**Figure 7**) (112). This approach allows for the simultaneous investigation of numerous exposures and outcomes. Compared to one-sample MR, two-sample MR has several additional advantages: 1) increased statistical power: Two-sample MR typically utilizes data from larger GWAS consortia with a large number of cases, which leads to greater statistical power and more reliable results; 2) lower risk of population stratification: Given that the genetic and outcome data are sourced from separate studies, the risk of population stratification (a type of confounding bias) is reduced in two-sample MR; 3) greater trait availability: In many cases, individual cohorts may not have all the necessary traits measured. Two-sample MR allows for the combination of results from different studies where different traits are measured, expanding the number of testable hypotheses; 4) less likely to violate the 'no measurement error' assumption: This is because the genetic associations used in two-sample MR often come from large-scale GWASs and are therefore more accurately estimated; 5) reduced participant overlap: In a one-sample MR, overlap between the IV and outcome samples may lead to "winner's curse" bias. This overlap is less likely in a two-sample MR; and 6) practicality and cost-effectiveness: Two-sample MR often uses summary data from publicly available GWAS studies, making it a more practical and cost-effective approach. The approach can be performed without needing access to individual-level data.

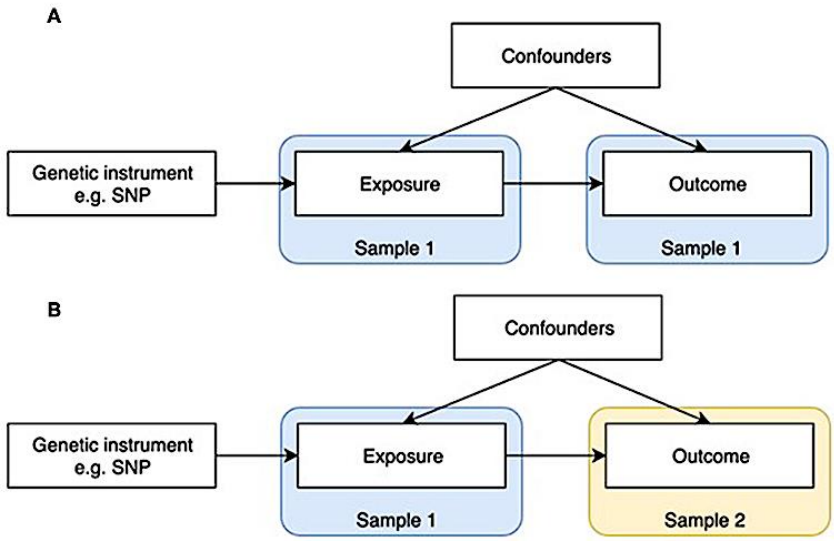



**Figure 7.** The schematic illustration of one-sample (A) and two-sample (B) MR design. SNP, single nucleotide polymorphisms. (Source: Zheng et al. 2019, *Front Endocrinol (Lausanne)*)

Two-sample MR also has its limitations. For instance, it is incapable of assessing the nonlinear nature of an association without individual-level data. Additionally, two-sample MR can struggle to examine associations within subgroups delineated by specific characteristics such as sex, age strata, and so on, particularly when no subgroup-specific summary-level data are available.

There are four principal steps for conducting two-sample MR analysis:

1) Identification of genetic instruments for the exposure.

Identify genetic variants (i.e., SNPs) that are robustly associated with the exposure of interest (e.g., usually at the genome-wide significance threshold, $P < 5\times10^{-8}$). These associations are typically identified from large GWASs. The pruning step is needed to minimize the influence of genetic correlations between these SNPs (also known as high linkage equilibrium [LD]) to lower the rate of type 1 error caused by collinearity. To assess the strength of these genetic instruments, *F* statistics are usually calculated (113). While there is not a universally agreed-upon cutoff, an *F*-statistic greater than 10 is commonly used as a rule of thumb to indicate that the instruments are sufficiently strong to avoid weak instrument bias. The variants left are used as IVs in the MR analysis.

2) Extraction data from the outcome GWAS.

Collect summary statistics (i.e., beta and standard error coefficients) for these genetic instruments related to the outcome from another GWAS. Ensure that the genetic instruments meet the other assumptions necessary for MR analysis. In this process, we can remove SNPs directly associated with the outcome or SNPs with pleiotropic effects (having effects on multiple traits, or at the first step) to minimize the potential horizontal pleiotropy.

3) Data harmonization.

Align the effect estimates for exposure and outcome GWASs to the same effect allele. This ensures that the direction of effects is consistent.

4) MR statistical analysis.

Apply an appropriate MR method to estimate the causal effect of the exposure on the outcome. There are different methods to perform MR analysis (e.g., the inverse-variance weighted, weighted median, MR-Egger, MR-PRESSO [Mendelian Randomization Pleiotropy RESidual Sum and Outlier], etc.) and the choice depends on the nature of data and assumptions met. We usually used the

inverse-variance weighted method as the primary statistical method since this method can provide the most accurate estimate. For the exposure with genetic IVs < 3 SNPs (e.g., the analysis for blood proteins; one or two cis-SNPs are used as IVs), the inverse-variance weighted method under the fixed-effects is usually used; otherwise, the method under the random-effects will be used to take potential heterogeneity in SNPs' estimates into consideration (e.g., the analysis for modifiable risk factors with many SNPs as IVs). Other methods are usually treated as sensitivity analyses with varying assumptions, strengths, and limitations.

### 2.6.2 Two-stage network MR and mediation estimation

Two-stage network MR and mediation estimation are sophisticated techniques in the MR analytical framework. They allow for a deeper exploration of intricate relationships among exposures, intermediary variables, and outcomes. These advanced methods provide a nuanced understanding of biological pathways and the mechanistic processes underlying observed associations.

Two-stage network MR is an advanced derivative of the standard two-sample MR approach, enabling the determination of indirect causal influences through one or more intermediary phenotypes. There are two steps (**Figure 8**). The first stage uses MR to establish the causal impact of the exposure on the intermediate phenotype(s). The second stage employs MR once again to deduce the causal influence of the intermediate phenotype(s) on the outcome. The indirect causal impact of the exposure on the outcome through the intermediate phenotype(s) is then estimated by multiplying these two results.

Subsequently, we can calculate mediation effects to quantify the portion of the overall effect of an exposure on an outcome that occurs via an intermediate phenotype. This is typically of interest when seeking to decipher the biological pathways that underlie observed associations. The method partitions the total effect of the exposure on the outcome into a direct effect (bypassing the intermediate) and an indirect effect (through the intermediate). While mediation estimation can be performed within the MR framework, it demands stringent assumptions, including no interaction between the exposure and the mediator, as well as no confounding between the mediator and the outcome, conditional on the exposure.

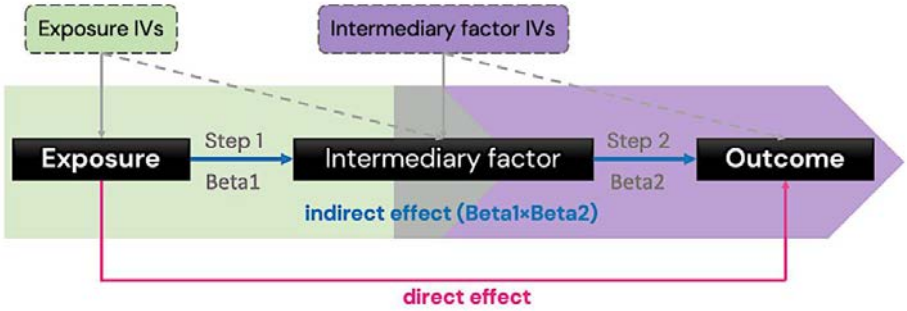



**Figure 8**. Study design of two-stage network Mendelian randomization analysis. IV, instrumental variables. (Source: Personal Collection)

### 2.6.3 Limitations

Despite their valuable insights into potential causal relationships, MR studies do come with several inherent limitations as listed in **Table 2**.

**Table 2.** Limitations of MR analysis.

| Potential limitations | Description |
|---|---|
| Pleiotropy | There are three types of pleiotropy including vertical, and balanced and unbalanced horizontal pleiotropy (details see figure 8). This is when genetic variants used as instruments have additional effects on the outcome independent of the exposure (the violation of the third assumption of MR). Usually, only unbalanced horizontal pleiotropy biases MR estimation. |
| Weak instrument bias | If the genetic instruments explain only a small proportion of the variance in the exposure, the study may suffer from weak instrument bias, which can bias the MR estimate and reduce its precision. This is a concern particularly in two-sample MR with sample overlap between the exposure and outcome datasets. |
| Population stratification | This occurs when frequency of genetic variants and the outcome both vary with population subgroups, which can confound the association between the genetic instrument and the outcome. |
| Nonlinear associations | Standard MR techniques usually assume linear effects, but real relationships may be non-linear. |

| | |
|---|---|
| Canalization | Genetic effects may be compensated by other physiological changes during development, leading to weaker observed associations than the true causal effect. |
| Generalizability | The findings from MR studies usually represent population-averaged effects and may not apply to individuals. Further, if the genetic variants have different effects in different populations, results may not generalize across populations. |
| Reverse causation | Although MR is typically robust to reverse causation, in some cases genetic variants can be influenced by the disease status, such as when the disease leads to death or other changes that affect the genotype frequencies in cases compared to controls. |

There are three types of pleiotropy (**Figure 9**). However, the major risk caused by pleiotropy in MR analysis is from the unbalanced horizontal pleiotropy, which is caused by the violation of the third assumption of MR (the genetic variant influences the outcome through pathways other than the one involving the exposure of interest). This type of pleiotropy can bias the estimates of the causal effect. To handle this, several robust MR methods have been developed, such as the MR-Egger regression (114), weighted median method (115), MR-PRESSO (116), and multivariable MR (117). These methods make different assumptions and offer different trade-offs between bias and precision.

MR-Egger regression allows for an unbalanced pleiotropy, but it requires the "Instrument Strength Independent of Direct Effect" (InSIDE) assumption and usually has lower power (114). Weighted median MR can provide a valid causal estimate as long as at least 50% of the weight comes from valid instruments (115). MR-PRESSO can detect outlying SNPs that may have pleiotropic effects and generate estimate after the removal of these identified SNP outliers (116). Multivariable MR allows for inclusion of multiple exposures in the same model, which can help when we suspect that the instruments for the exposure of interest also influence other traits that could confound the exposure-outcome relationship (117). While these methods help, they cannot fully eliminate the risk of bias due to pleiotropy, and the potential for this bias should always be considered when interpreting results from MR studies.

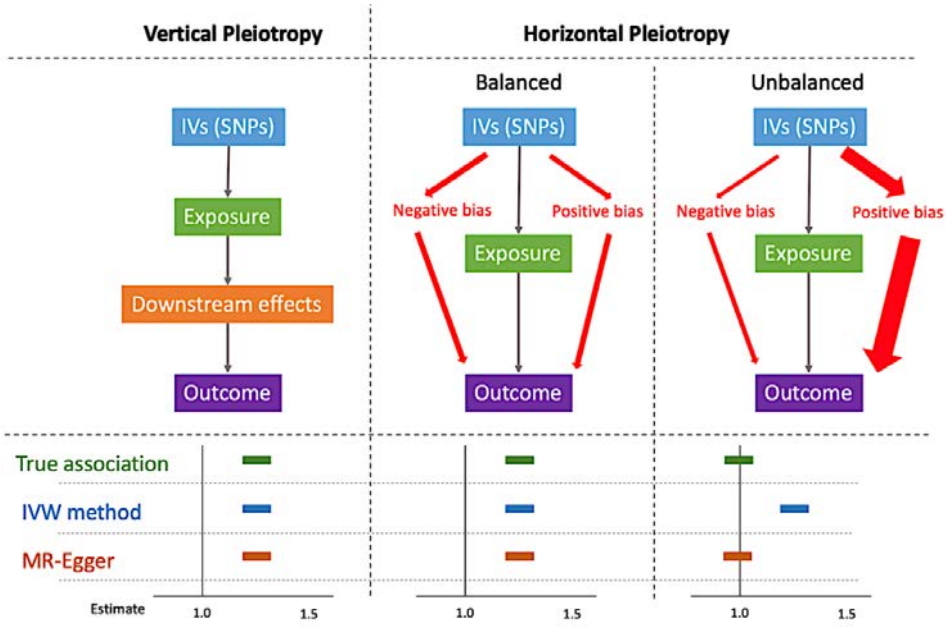



**Figure 9**. The schematic presentation of pleiotropy in Mendelian randomization analysis. We used the examples of MR estimates from the IVW and MR-Egger methods in different situations of pleiotropy to illustrate impact on the risk estimate. IV, instrumental variables; IVW, inverse variance weighted; SNPs, single nucleotide polymorphisms. (Source: Personal Collection)

## 2.7 Knowledge gaps

In a nutshell, the links between modifiable risk factors, specifically lifestyle habits, and the risk of VTE remain inconclusive. Most of these associations are derived from cross-sectional studies or those with short follow-up periods, with very few investigating sex-specific associations. Even though obesity is recognized as a basal risk factor for VTE, the associations between body shapes - determined by a combination of BMI and WC - and VTE risk, are yet to be clarified. Furthermore, no studies have been conducted on emerging modifiable risk factors such as the intake of ultra-processed foods (UPFs) in relation to VTE. Prospective studies assessing the associations of blood proteome with VTE in large samples are also lacking. The intricate interplay between modifiable risk factors and blood proteins in the context of VTE is an uncharted territory awaiting exploration.

## 2.8 Research hypotheses

Drawing upon existing knowledge, we proposed the following research hypotheses that form the foundation of investigations included in this thesis:

1) Elevated BMI and WC may independently contribute to a heightened risk of VTE.

2) Unhealthy lifestyle behaviors such as smoking, excessive alcohol and coffee consumption, physical inactivity, and poor diet may increase the risk of incident VTE.
3) Some of the associations between modifiable risk factors and VTE may exhibit sex-specific differences.
4) Numerous blood proteins, particularly those involved in the coagulation cascade, may be linked to VTE risk, with some serving as crucial intermediaries in the progression from unfavorable factors to the onset of VTE.
5) The modifiable risk factors may influence the risk of VTE via alternation of levels of certain blood proteins.

# 3 Research aims

The overall aim of this thesis is to explore the associations of modifiable risk factors and blood proteins with the development of VTE and to reveal protein pathways underlying the associations between modifiable risk factors and VTE. To achieve this aim, we planned five studies with specific aims:

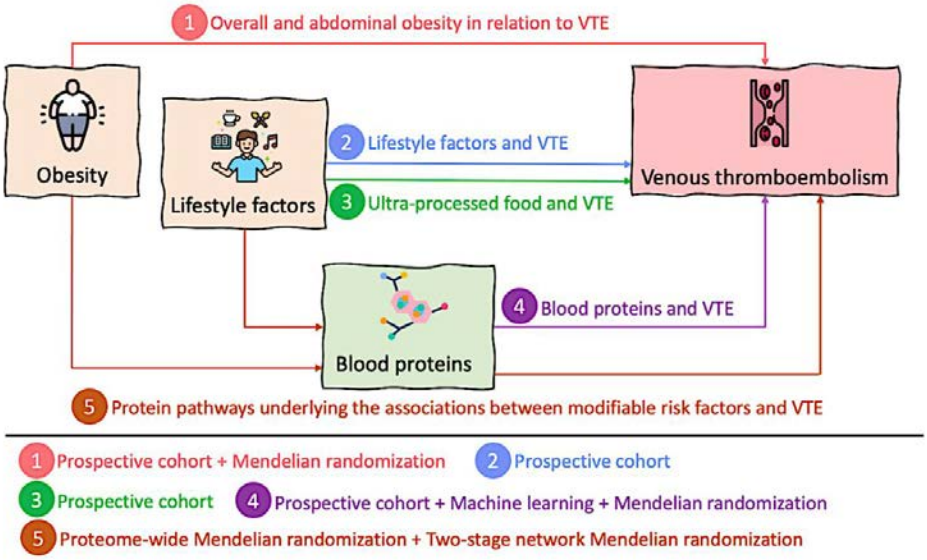
**Paper I** To explore the causal associations of overall and abdominal obesity with the risk of VTE using cohort and Mendelian randomization (MR) designs and to estimate the population attributable fraction for obesity for VTE in a Swedish population.

**Paper II** To assess the associations of modifiable lifestyle factors, including cigarette smoking, alcohol and coffee consumption, physical activity, and diet with the risk of incident VTE in two cohorts of Swedish middle-aged and old adults.

**Paper III** To estimate the associations of ultra-processed food intake and the risk of incident VTE in the UK Biobank study.

**Paper IV** To investigate the associations of blood proteins related to metabolism and cardiovascular diseases with the risk of incident VTE in two Swedish cohorts and to confirm the causality of identified associations using MR analysis.

**Paper V** To identify causal blood proteins in relation to VTE using proteome-wide MR analysis and to reveal protein pathways underlying the associations between modifiable risk factors and VTE using a two-stage network MR design.



# 4 Materials and methods

## 4.1 Prospective cohort study

### 4.1.1 Study participants

#### 4.1.1.1 Swedish Infrastructure for Medical Population-based Life-course and Environmental Research (SIMPLER)

SIMPLER (**Figure 10**) consists of two cohorts that are the Swedish Mammography Cohort (SMC) and the Cohort of Swedish Men (COSM). The **study I, II and IV** were based on data from this national infrastructure.

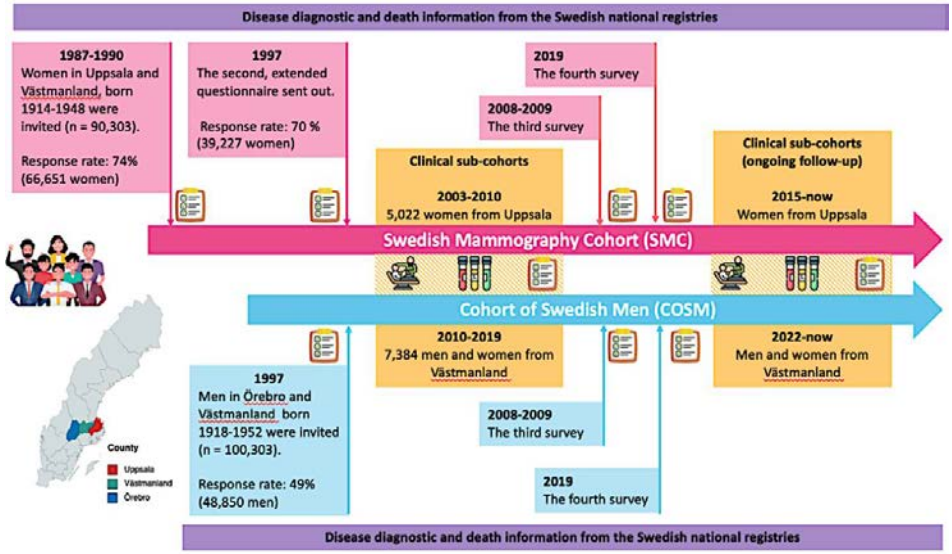



**Figure 10**. Schematic description of the SIMPLER. (Source: Personal Collection)

The SMC comprises all women (N = 90,303) who were born between the years 1914 and 1948 in Västmanland and Uppsala counties. These women were invited to participate in the study in conjunction with the initial mammography screening program conducted between 1987 and 1990. A total of 66,651 women (74% of the source population) returned a completed questionnaire. During the fall of 1997, an expanded questionnaire was sent via mail to women who were still alive and residing in the study area. This comprehensive questionnaire elicited information on various lifestyle factors and medical history. Out of the total participants, 39,227 women (70%) completed and returned the questionnaire. Similarly, during the autumn of 1997, the COSM enrolled 48,850 men born between 1918 and 1952, residing in central Sweden (Västmanland and Örebro counties). These men were enrolled by responding to a mailed questionnaire that was identical to the 1997 SMC questionnaire, except for a few sex-specific

questions. Individuals from the SMC and COSM were further followed up and invited to complete updated, albeit similar, questionnaires in 2008-2009, and 2019.

Clinical sub-cohorts were established to include clinical sampling and measurement of body composition in 2003 for women in Uppsala and in 2010 for women and men in Västerås. The first round of data collection in Uppsala was completed in 2009 (N = 5,022) and in Västerås was completed in 2019 (N = 7,384). Samples of blood, urine, feces, and fat tissue were collected and meanwhile participants were requested to filled in questionnaires elicited information on lifestyle factors and medical history.

The **study I and II** established the baseline using data from the 1997 wave (39,227 women from the SMC and 48,850 men from the COSM). The **study I** consisted of 35,691 women and 38,626 men after excluding individuals who met the following criteria: (a) those with incorrect or missing personal identity numbers, (b) individuals who had passed away before January 1, 1998, (c) participants with a history of cancer, VTE, or cardiovascular disease (heart failure and stroke), (d) those with missing or implausible BMI information, or (e) individuals classified as underweight with a BMI below 18.5 kg/m$^2$. The **study II** included 30,137 women and 36,193 men in the analysis after exclusion of individuals who had an incorrect or a missing personal identification number, died before January 1, 1998, had history of cancer or VTE, had missing information on any of lifestyle factors, or had extreme energy intakes (3 standard deviations from the $\log_e$-transformed mean energy intake in men and women separately). The **study IV** was based on two clinical sub-cohorts including 12,314 participants with available blood protein data. The analysis included 11,803 participants after removing 511 individuals with baseline VTE diagnosis. See **Figure 11** for inclusion and exclusion criteria.

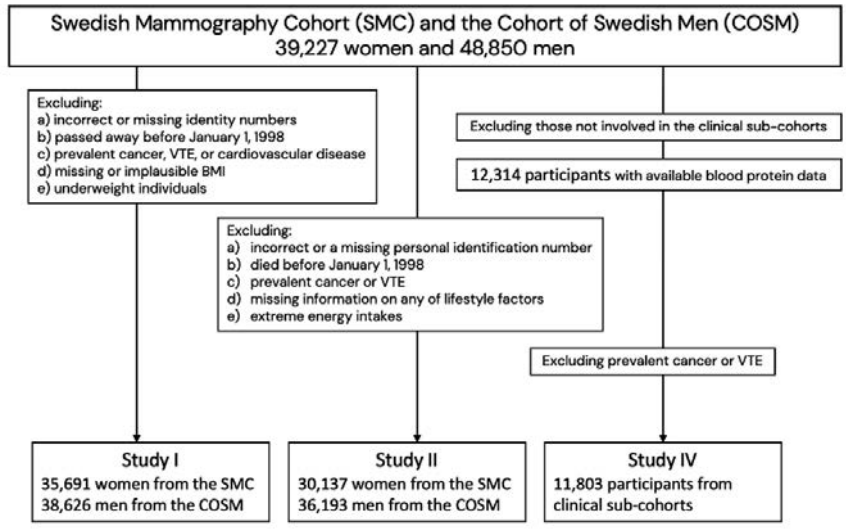



**Figure 11**. The study population for study I, II, and IV. (Source: Personal Collection)

#### 4.1.1.2 The UK Biobank study

The UK Biobank study is a large-scale ongoing cohort that recruited over 500,000 individuals, aged between 40 and 69 years, in 2007–2010 from 22 health centers in the United Kingdom (**Figure 12**). The following data were collected at the baseline assessment visit: a self-completed touch-screen questionnaire; brief computer-assisted interview; physical and functional measures; and collection of blood, urine, and saliva. More detailed information on the UK Biobank study has been described elsewhere (118). In **study III**, we enrolled a total of 191,897 participants who completed the 24-hour dietary recall questionnaire at least once and reported a reasonable energy intake (>0 and <18 MJ for females, >0 and <20 MJ for males) (119) in the UK Biobank. After excluding 5,574 individuals with baseline VTE, the analysis included 186,323 participants.


**Figure 12**. The UK Biobank cohort recruitment centers (Source: UK Biobank)

### 4.1.2 Modifiable risk factor measurement

In **study I and II**, information on body height and weight and lifestyle factors was obtained from the self-administrated questionnaires (**Appendix Figure 1 and 2**) completed by participants at baseline in 1997. The habitual dietary intake over the past year was evaluated using a validated 96-item food-frequency questionnaires (120).

Several lifestyle and dietary factors have been validated in the SIMPLER cohorts. The data obtained from the physical activity questionnaire underwent validation through comparison with results from both 7-day activity records and accelerometers (121). In this validation study involving 116 female participants, deattenuated concordance correlations were computed to assess the agreement between total daily activity levels obtained from a questionnaire, accelerometers, and records. The deattenuated concordance correlations for questionnaire-based measurements and accelerometer-derived measurements were found to be 0.38 (95% CI: 0.22–0.54) and 0.64 (95% CI: 0.45–0.83), respectively. The reproducibility and validity of the used food-frequency questionnaires have been assessed in comparison with multiple 24-hour recall interviews and/or diet records concerning foods, nutrients, dietary supplements, glycemic index, and glycemic load (120, 122-124). For example, the Spearman rank correlation coefficient was 0.81 when comparing alcohol consumption assessed by food-frequency questionnaires and that assessed by 14 interviews (each month during 1 year on randomly chosen days) using 24-hour recall of intake among 248 study participants (120). Nutrient intake has been further validated against circulating or

urinary biomarkers, like plasma total antioxidant capacity (125), fatty acids in subcutaneous adipose tissue (126, 127), and cadmium urine concentrations (128).

#### *4.1.2.1 Obesity indicators*

Body Mass Index (BMI) was derived by dividing body weight (in kilograms) by the square of body height (in meters). We used BMI and waist circumference (WC) as indicators for overall and abdominal obesity, respectively. Participants were grouped based on their BMI values:

Detailed categorization of BMI:

- 18.5–22.4 kg/m$^2$ (serving as the reference group)
- 22.5–24.9 kg/m$^2$
- 25.0–27.4 kg/m$^2$
- 27.5–29.9 kg/m$^2$
- ≥30.0 kg/m$^2$

Following the World Health Organization (WHO) criteria for European populations:

- Normal BMI: 18.5–24.9 kg/m$^2$ (reference group)
- Overweight: 25.0–29.9 kg/m$^2$
- Obesity: ≥30 kg/m$^2$

Waist Circumference (WC) in cm categorizations, in line with WHO classifications, were as follows:

- For men:
    - Normal: <94 cm (reference group)
    - Increased: 94–101 cm
    - Substantially increased: ≥102 cm
- For women:
    - Normal: <80 cm (reference group)
    - Increased: 80–87 cm
    - Substantially increased: ≥88 cm

#### *4.1.2.2 Cigarette smoking, alcohol and coffee consumption, and physical activity*

Participants reported their cigarette smoking habits, detailing both status and dosage, as well as their daily coffee consumption. Alcohol intake was gauged using six types of alcoholic beverages: light beer (alcohol by volume < 2.25%), regular beer (2.8–3.5%), strong beer (4.4–5.6%), wine (12–13.5%), fortified wine (15–22%), and liquor (40%). The questionnaire captured the quantity and frequency of beer, wine, and liquor consumed on individual occasions. To estimate alcohol (ethanol) intake, we multiplied the consumption frequency of each beverage type by the amount consumed, considering

the varying ethanol contents. Physical activity over the past year was evaluated using a validated questionnaire (121), which delved into the time participants dedicated to different activities. We combined the weekly time allocated to walking/bicycling and other exercises to gauge overall physical activity levels. For analysis, we classified participants based on:

- Cigarette Smoking:
    - Never smoker (reference group)
    - Past smoker
    - Current smoker with varying intensities: 1-5, 6-10, 11-20, or >20 cigarettes/day
- Alcohol Consumption:
    - Never drinker (reference group)
    - Past drinker
    - Current drinkers segmented by frequency: <1 drink/week, 1-7 drinks/week, 8-14 drinks/week, 15-21 drinks/week, and >22 drinks/week
- Coffee Intake:
    - Daily consumption categories: <1 (reference group), 1 to ≤2, 2 to ≤4, and >4 cups
- Physical Activity:
    - Time spent daily: <10 (reference group), 10-30, 31-60, and >60 minutes.

#### *4.1.2.3 Dietary intake in SIMPLER*

To measure the dietary quality, we formulated a modified Dietary Approaches to Stop Hypertension (mDASH) diet score based on the intake of seven food groups (129). This encompassed fruits, vegetables, nuts, legumes, whole grains, and low-fat dairy products as positive dietary choices, while red and processed meats, along with sweetened beverages, were considered negative components. Depending on their consumption quintiles for each food group, stratified by gender, participants earned a score between 1 and 5. The combined scores for all the food groups produced the mDASH diet score, ranging between 7 and 35, with a higher score signifying greater adherence to the mDASH diet. For analysis, we divided the mDASH diet scores into four categories, corresponding to quartiles 1 (reference group) through 4.

#### *4.1.2.4 Ultra-processed food intake in UK Biobank*

Dietary consumption of UK Biobank participants, as recalled over 24 hours, was evaluated using the Oxford WebQ questionnaire. This tool includes 206 food items and 32 drink and alcohol entries detailing consumption from the previous day. The 24-hour recall method, which the Oxford WebQ uses, was sent online five times to participants in 2011-2012. For those with multiple responses, an average from the 24-hour dietary recalls was computed to represent their consumption. Participants, once informed

about standardized portion sizes, indicated the number of portions they consumed for each item. The intake in grams was deduced by multiplying portion count with the grams in a standard portion size. Using this data, we then determined energy and nutrient intake by multiplying the gram intake of each food with its energy and nutrient composition as per The Composition of Foods 6th edition (2002) (130). A validation has been performed through repeated measures (131). The validation study encompassed 160 participants in London recruited between 2014 and 2016, assessing their biomarker levels at three nonconsecutive time points. Employing objective biomarkers as the reference standard, the Oxford WebQ demonstrated robust performance across essential nutrients when compared to more administratively burdensome interviewer-based 24-hour recalls (131). The starting point or "baseline" was marked by a participant's first completion of this online dietary survey.

UPFs were categorized as per the NOVA system, which classifies foods based on the extent and purpose of their industrial processing (132). In our study, UPFs covered foods and drinks that fall under the NOVA4 category and additionally ham and bacon (133). Estimation of UPF consumption was drawn from data on portion sizes and the nutrient and energy composition of each food. For our analysis, we utilized five quintile measures of UPF consumption to comprehensively capture the association between UPF intake and VTE risk by taking strengths and limitations of these UPF intake measures into consideration.

- Daily UPF intake in servings (a straightforward metric albeit no information on quantity)
- Daily UPF intake in grams (more precise about the quantity albeit easy to be driven by certain heavy food groups, like beverages)
- Daily proportion of grams consumed from UPFs (providing insights into the dietary compositions albeit missing the overall quality of the diet)
- Daily energy consumption from UPFs (taking energy intake and metabolism into consideration and more comparable between dietary patterns albeit providing indirect interpretation)
- Daily proportion of energy derived from UPFs (normalized values allowing for comparison between individuals albeit providing indirect interpretation)

### 4.1.3 Proteomic profiling

Proteomic data were measured using the Olink platform in SIMPLER clinical sub-cohorts and used in the **study IV.** After a 12-hour overnight fast, venous blood samples were collected, which were promptly centrifuged and stored at -80°C for subsequent analysis. A set of 276 protein biomarkers were examined using three high-capacity multiplex immunoassays: Olink CVD II, CVD III, and Metabolism panels, each targeting 92 specific proteins related to cardiovascular disease or metabolism. The analysis platform

delivered normalized protein expression data, standardized per analysis plate, on a log2 scale. Manufacturer-provided values below the limit of detection (LOD) guided our protein selection. The SciLifeLab at Uppsala University, Sweden conducted these analyses. In our study, we left out 19 proteins that had over 50% of samples falling below the LOD. Some samples were omitted due to not meeting the manufacturer's quality benchmarks, accounting for 3.6%, 0.8%, and 0.7% for the CVD II, CVD III, and metabolism panels respectively. A comprehensive list of the 257 included proteins is available in **Appendix Table 1**.

#### 4.1.4 Covariate measurement

In SIMPLER, we defined covariates (except for the aforementioned variables) as following: date of birth (years), sex (women and men), highest education (with or without postsecondary education), ever hormone therapy use (yes or no, only for women), history of diabetes and fractures (yes or no), regular aspirin use (≥7 or <7 tablets per week), total energy intake (continuous in log-transformed), incident cancer (yes or no), baseline diagnosis of cardiovascular disease including coronary artery disease, heart failure, stroke, and atrial fibrillation (yes or no), estimated glomerular filtration rate (eGFR, continuous in log-transformed), low- and high-density lipoprotein cholesterol (continuous in log-transformed), triglycerides (continuous in log-transformed), blood pressure (continuous), and blood glucose levels (continuous). Calculated total energy consumption was derived from the reported food intake, factoring in the energy content relevant to the age- and gender-specific serving sizes. Data on incident cancer and baseline cardiovascular disease diagnoses were sourced from the Swedish Cancer Register and the National Patient Register, respectively. During the health examination, eGFR, blood lipids and glucose concentrations, and blood pressure were measured. We presented detailed measurement information in **Appendix Table 2**.

In UK Biobank, we gathered data on participants' age (years), sex (women and men), ethnicity (white and other), highest education attainment (below college or college and above), smoking (current, never, or past), alcohol consumption (current, never, or past), physical activity (high, moderate, low, or missing) from the baseline questionnaires. During physical examinations conducted by trained nurses, participants' weight and height were measured. From these measurements, we calculated the BMI. We utilized the Townsend Deprivation Index as an indicator of socioeconomic status. Total energy intake, determined through the method previously outlined, was considered as a covariate in our analysis. Additional data points of interest included consumption patterns of tea and coffee, the duration of sedentary behavior, exposure to air pollution, instances of self-reported fractures and major surgeries, and the use of medications and treatments such as aspirin, oral contraceptive, and hormone-replacement therapies among women. We also incorporated health conditions represented by the Charlson

Comorbidity Index. A comprehensive breakdown and definitions of these covariates are provided in **Appendix Table 3**.

Three studies included different sets of covariates considered to be potential confounders. Confounding control was guided using the directed acyclic graphs (DAGs, **Figure 13**) framework.

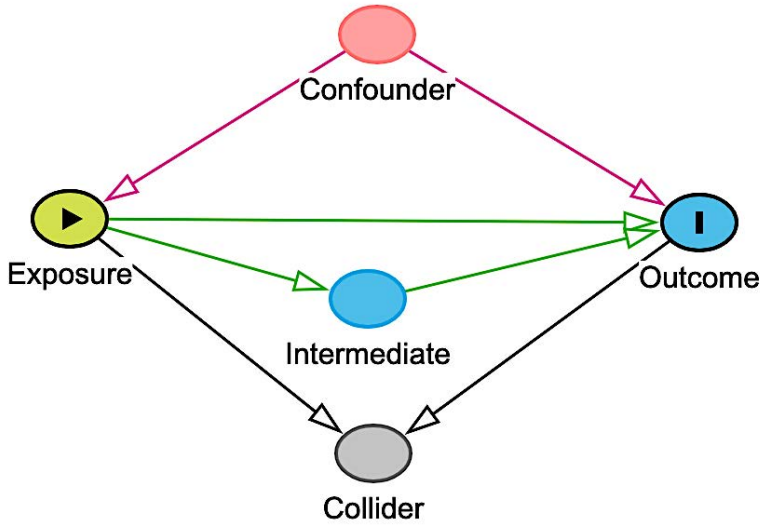



**Figure 13**. A DAG illustrating the relationship between an exposure, a confounder, an intermediate, a collider, and an outcome. For total effect estimation, we should only adjust for confounders that is variable shown in the literature to be causally associated with the outcome (i.e., VTE in this case) and associated with the exposure but not an intermediate variable in the causal pathway between the exposure and the outcome. Being different from the confounder, a collider is a variable that is affected by the exposure and outcome and, when conditioned upon (controlled for), induces an association between those two variables. This can lead to spurious associations and biased results.

By reading related materials and modelling these variables in DAGs, the detailed adjustments in each study refer below:

- **Study I** included age (the underlying time scale), sex, highest education attainment, alcohol consumption, cigarette smoking, mDASH score and total energy intake, physical activity, and ever hormone therapy use (only for women).
- **Study II** included age, sex, BMI, highest education attainment, history of diabetes and fracture, hormone therapy for women, regular aspirin use, incident cancer, cigarette smoking, alcohol and coffee consumption, physical activity, energy intake and mDASH Diet score.

- **Study III** included age, sex, ethnicity, BMI, Townsend Deprivation Index, highest education attainment, cigarette smoking, alcohol consumption, physical activity, and energy intake.
- **Study IV** included age (the underlying time scale), sex, plate, BMI, highest education attainment, baseline cardiovascular disease, alcohol consumption, cigarette smoking, physical activity, mDASH score, eGFR, low- and high-density lipoprotein cholesterol, triglycerides, blood pressure, and blood glucose levels.

### 4.1.5 VTE diagnosis and follow-up

In SIMPLER, VTE was determined by a clinical diagnosis attributed to VTE or its two subtypes: PE and DVT. This diagnosis was either the primary reason or a contributing factor. We obtained diagnostic information by connecting the cohort with the Swedish National Patient Register, which encompasses almost the entirety of hospital-based inpatient care. Since 2001, this register has grown to incorporate outpatient visits from specialist healthcare of both public and private providers (134). The criteria for diagnosing VTE, PE, and DVT adhered to the standards set by the International Classification of Diseases 9th and 10th Revisions (**Table 3**). Mortality information was obtained from the Swedish Cause of Death Register. The follow-up period for participants commenced on 01 January 1998 (**study I and II**) or the clinical visit date (**study IV**) and concluded either on the date of VTE diagnosis, the date of death, or the end of the follow-up period (i.e., 31 December 2017 for **study I** and 31 December 2019 for **study II and IV**, depending on which event occurred first.

In the UK Biobank, ICD-9 and -10 codes were also used to defined VTE cases. The data were sourced from the national inpatient database, primary care records, and the death registry. The monitoring of individuals persisted until they were diagnosed, passed away, were no longer traceable, or until the end of the follow-up period, whichever occurred earliest.

**Table 3**. ICD codes for VTE in SIMPLER.

| Outcome | 9th ICD code | 10th ICD code |
|---|---|---|
| VTE | 415.1, 451.1, 452, 453.0, 453.4 and 453.9 | I26, I80.1, I80.2, I80.3, I81, and I82.0 |
| PE | 415.1 | I26.0 and 126.9 |
| DVT | 451.1 | I80.1, I80.2 and I80.3 |

### 4.1.6 Statistical analysis

Missing values were labeled as missing (study I and II) or imputed (study III and IV) in these studies. For association assessments, we used Cox proportional hazards regression model with age as the underlying time scale to estimate the HRs and corresponding 95% CIs. The Cox model is a widely used statistical technique for

investigating the effect of several variables upon the time a specified event takes to happen. The Cox model is a semi-parametric with an assumption that the baseline hazard is proportionally distributed, which was tested using Schoenfeld residuals. Various models, adjusting for distinct covariates across the studies where suitable, were employed to mitigate confounding effects, evaluate the stability of results, and delineate a hierarchy of evidence. These tests were two-sided and performed in Stata/SE (version 15.0; StataCorp, College Station, TX, USA) or R software, version 4.2.1. The association with the $P < 0.05$ was considered statistically significant.

**Study I**. We used the Cox proportional hazard regression to estimate the HR of incident VTE across various categories of BMI, WC and combined. Initially, analyses were conducted separately for women and men, later combining the data with sex treated as a stratified variable. The multivariable model adjusted for educational attainment, alcohol consumption, smoking habits, total energy intake, mDASH score, time spent walking/bicycling, and the use of hormone therapy (for women only). To test trends across categories, we modeled median values of each anthropometric measure as continuous variables. Recognizing cancer as a potential risk amplifier for VTE, we conducted a stratified analysis to discern any variations in the obesity-VTE relationship between individuals diagnosed with cancer during the follow-up and those who were not, up until the VTE event date.

Assuming a causal relationship between obesity and VTE, we aimed to gauge the potential percentage reduction in VTE cases if all individuals sustained a healthy BMI within the 18.5 to 24.9 kg/m$^2$ or retained a normal WC. We employed the following equation to determine the population attributable risk: $p(HR-1)/(1+p[HR-1])$. In this equation, '$p$' signifies the exposure prevalence (overweight and obesity combined and increased and substantially increased WC combined) in the population. '$HR$' represents the hazard ratio contrasting exposed and unexposed participants. In a similar vein, we calculated the population attributable risk considering obesity (BMI ≥ 30 kg/m$^2$) and substantially enlarged WC.

**Study II**. We employed the Cox proportional hazard regression to determine the HRs and their 95% CIs for VTE based on lifestyle categories. Analyses were executed separately for women and men, combined, and for VTE subtypes, namely PE and DVT. All models considered age, sex (for combined analyses), BMI, education levels, histories of diabetes and fracture, hormone therapy usage among women, regular aspirin intake, newly diagnosed cancer cases, smoking habits, consumption of alcohol and coffee, physical activity levels, total energy intake, and the mDASH Diet score.

**Study III**. We assessed the associations between UPF consumption and the onset risk of VTE and its subtypes using the Cox proportional hazard regression. Two analytical models were employed: Model 1 accounted for age, sex, and ethnicity, while Model 2

extended adjustments to include BMI, Townsend Deprivation Index, education levels, smoking habits, alcohol consumption, physical activity, and total energy intake (not for daily energy consumption from UPFs or daily proportion of energy derived from UPFs). To probe for any nonlinear associations between UPF consumption and VTE risk, we utilized the Cox model with the UPF intake score input as a restricted cubic spline, assigning three knots at the 25th, 50th, and 75th percentiles. We further explored potential interactions on the multiplicative scale between UPF intake and factors like age (< and ≥60 years), sex, and BMI (≤ and >30 kg/m$^2$) in relation to VTE risk. Subsequent stratified analyses were executed to obtain these associations in specific subgroups. In a bid to distinguish between the impact of UPF food and beverage consumption on VTE, we ran separate analyses examining the associations of both UPF beverages and UPFs excluding beverages with VTE risk.

**Study IV**. We used the Cox proportional hazard regression to compute the HRs and their 95% CIs for the associations between circulating proteins and the onset risk of VTE. Our analysis incorporated two models: Model 1, which adjusted for sex and plate (of protein analysis), and Model 2, which considered sex, plate, BMI, educational levels, baseline cardiovascular disease diagnosis, smoking habits, alcohol intake, physical activity, mDASH score, eGFR, cholesterol levels (both low- and high-density lipoprotein), triglycerides, blood pressure, and blood glucose levels. To ensure the integrity of the associations, we conducted a sensitivity analysis, leaving out individuals with baseline cardiovascular conditions (totaling 2,548). We employed the false discovery rate (FDR) to address concerns regarding multiple testing of our data, considering results with an FDR < 0.1 as statistically significant.

Machine learning, a subset of artificial intelligence, has been increasingly applied in medical research to unravel complex associations within vast datasets. Unlike traditional statistical methods that might examine one protein at a time or rely heavily on pre-defined hypotheses, machine learning can simultaneously analyze multiple proteins, account for interactions, and uncover non-linear relationships. Furthermore, machine learning methods stand out in their ability to compare the relative importance of risk factors, even when those factors are measured in different units or scales. This comparative capability aids in prioritizing risk factors based on their impact. In **study IV**, we used machine learning to compare the importance between identified protein biomarkers and traditional lifestyle factors using a model with relaxed assumptions. To maintain equilibrium between the numbers of cases and controls within a machine-learning framework, we employed the MatchIt package (135) to resample each incident VTE case, pairing it with five controls matched for age and sex. Using the Light Gradient Boosting Machine (LightGBM) (136), a gradient-boosting framework, we probed the relative importance of pinpointed proteins in contrast to traditional VTE risk factors. The predictive efficacy of these factors was assessed through tenfold cross-validation.

Furthermore, the SHapley Additive exPlanations (SHAP) method (137) was harnessed to determine the average influence of each predictor on the composite model predictions.

## 4.2 Mendelian randomization analysis

### 4.2.1 GWAS data sources for VTE

In the **study IV**, we used the GWAS meta-analysis of the Million Veteran Program (MVP) and the UK Biobank study including a total of 26,066 cases and 624,053 controls as the discovery data (138) and the FinnGen study including 14,454 cases and 294,700 controls (the R7 release) as the replication data (139).

- The MVP is a groundbreaking national research initiative focused on exploring the intricate connections between genes, lifestyle choices, military experiences, and environmental exposures, aiming to gain invaluable insights into the impact of these factors on the health and well-being of United States Veterans. In MVP VTE GWAS, VTE cases were defined by International Classification of Diseases 9/10 codes (138). The VTE GWAS meta-analysis used data from MVP v.2.1 data including 11,844 VTE cases and 251,951 controls and the genetic associations were adjusted for age, sex, and top principal components (indicators of population structure) (138). Information on sample and variant quality control has been described in detail previously (140).
- The UK Biobank GWAS in the meta-analysis included individuals of European ancestry to avoid potential population stratification. VTE cases were defined by self-reported diagnosis reviewed by trained nurses, International Classification of Diseases 10 codes, and corresponding procedures codes (141). The analysis included 14,222 VTE cases and 372,102 controls and was adjusted for age, sex, and principal components (138). Detailed information on quality control can be found in the original GWAS (138).
- The FinnGen study is one of the most extensive public-private collaborations in the field of genomics (139). The FinnGen combines genome information with digital health record data from up to 500,000 Finnish biobank participants, nearly 10% of the country's population. VTE cases in the FinnGen study were defined by codes of ICD-8, -9 and -10, resulting in 14 454 cases and 294 700 controls in the R7 release after quality control. Association tests were adjusted for age, sex, 10 genetic principal components, and genotyping batch.

In **study V**, summary-level data on DNA-VTE associations were obtained from a genome-wide meta-analysis of six datasets including 81,190 cases and 1,419,671 controls of European ancestry (15). VTE cases were identified through hospital or registry records using the ICD-9 or ICD-10 codes. Comprehensive details on genotyping, imputation, and quality control both at the participant and gene levels are elaborated in another source

(15). The relationship between DNA variants and the likelihood of VTE was analyzed using logistic regression, incorporating at a minimum, age (or year of birth), sex, and principal components as factors. The IVs to proxy genetic liability to VTE were also selected based on this study for reverse MR analysis to examine the genetic liability to VTE on VTE-associated factor and proteins.

### 4.2.2 Genetic instrumental variable selection

#### *4.2.2.1 Modifiable factors*

We included 15 modifiable factors for VTE in the two-sample MR analysis, encompassing adiposity indicators (BMI, waist-to-hip ratio [WHR], and visceral adiposity), lifestyle variables (smoking initiation, lifetime smoking index, consumption of alcohol, coffee, and caffeine, moderate-to-vigorous physical activity, and leisure screen time), and sleep-related traits (sleep duration, short or long sleep, daytime napping, and insomnia). We selected SNPs for each trait with a significance of $P < 5\times10^{-8}$ from their respective GWAS (**Table 4**). SNPs exhibiting high linkage disequilibrium ($r^2 > 0.01$) were filtered out, retaining the SNP in linkage disequilibrium with the most significant *P* value, leading to the residual SNPs as IVs.

#### *4.2.2.2 Blood proteins*

We selected index cis-SNPs (the most significant SNP in the protein-encoding gene region) associated with protein levels at the genome-wide significance level ($P < 5\times10^{-8}$) as the IVs for proteins. Any SNP that was missing was substituted with a proxy SNP with linkage disequilibrium ($r^2 \geq 0.8$). Proteins lacking IVs were excluded from the analysis.

In **study IV**, IVs were obtained from six GWASs on blood proteins that were measured by SomaScan assay in four studies and Olink in two studies (**Table 5**). Genetic variants for the same protein from different studies were considered distinct IVs and thus were individually applied in the MR analysis to cross-verify each other.

In **study V**, IVs for 4,907 proteins were selected from a GWAS in 35,559 Icelanders (mean age of 55 years and 50% of women) (142). In this study, plasma proteins were assessed by the SomaScan version 4 assay and underwent rank-inverse normal transformation based on age, sex, and sample age, and were then standardized. For identified proteins, we utilized IVs sourced from the UK Biobank Pharma Proteomics Project's GWASs (N = 54,219) (143) and Fenland study (N = 10,708) (144) to validate the association when available. For the two-replication studies, the Olink and SomaScan platform conducted the proteomic profiling, respectively. The IVs were chosen based on the same criteria.

**Table 4.** Data sources for modifiable risk factors included in the two-sample MR analysis for studies

| Modifiable factors | Unit | SNPs | Sample size | Adjustment |
|---|---|---|---|---|
| Body mass index | SD | 308 | 806,834 | Sex, age at assessment, age at assessment squared and assessment center |
| Waist-to-hip ratio | SD | 559 | 697,734 | |
| Visceral adiposity | Kilogram | 288 | 397,170 | Age, first 15 genetic principal components, and a batch-effect variable to adjust for genotyping |
| Smoking initiation | SD in log odds | 308 | 1,232,091 | Age, sex, and the first 10 genetic principal components |
| Lifetime smoking index | SD | 123 | 462,690 | Genotyping chip and sex |
| Alcohol consumption | SD increase of log alcoholic drinks/week | 82 | 941,280 | Age, sex, and the first 10 genetic principal components |
| Caffeine consumption | SD | 24 | 362,316 | Age, sex, genotyping array, and the first 30 genetic principal components |
| Coffee consumption | 50% change | 12 | 375,833 | Age, sex, BMI, total energy, proportion of typical food intake, and 20 genetic principal components |
| Moderate-to-vigorous physical activity | Active vs. Inactive | 15 | 298,506 | Age, age-squared, principal components reflecting population structure and additional study-specific covariates |
| Leisure screen time | SD | 127 | 526,725 | |
| Sleep duration | Hour | 71 | 446,118 | Age, sex, 10 principal components of ancestry, genotyping array, and genetic correlation matrix |
| Short sleep | < 7 vs. 7-8 hours | 26 | 411,934 | |
| Long sleep | > 9 vs. 7-8 hours | 7 | 339,926 | |
| Daytime napping | Factor (never/rarely, sometimes, usually) | 34 | 993,966 | Age, sex, 10 principal components of ancestry, genotyping array, and genetic correlation matrix |
| Insomnia | Log odds ratio | 203 | 1,331,010 | Age, sex, genotype array, and 10 genetic principal components in the UK Biobank, and age, sex and the top five principal components in 23andMe. |

SD, standard deviation. SNPs, single nucleotide polymorphisms.

**Table 5.** Data sources for blood protein GWAS.

| First author | Sample size | Platform | Adjustment |
|---|---|---|---|
| Emilsson | 5,457 | SomaScan | Age |
| Folkersen | > 30,000 | Olink | Age, sex, population structure, and study-specific parameters |
| Yao | 6,861 | SomaScan | Age, sex, and study-specific parameters |
| Suhre | 1,000 | SomaScan | Age, sex, and body mass index |
| Sun | 3,301 | SomaScan | Age, sex, duration between blood draw and processing, and the top three principal components |
| Sun | 54,219 | Olink | Age, age$^2$, sex, age*sex, age$^2$*sex, batch, center, genetic array, time between blood sampling and measurement, and the first 20 genetic principal components |
| Ferkingstad | 35,559 | SomaScan | The data were rank-inverse normal transformed by age, sex, and sample age. |

### 4.2.3 Statistical analysis

For each protein, the *F* statistic was calculated to evaluate the strength of the applied IV. In MR analyses employing a single SNP for exposure (blood proteins), the association's odds ratio (OR) and its CI were estimated using the Wald ratio test and the delta method, respectively (145). In **study IV**, MR estimates for each protein from the FinnGen and the GWAS meta-analysis of MVP and UK Biobank studies were combined using the fixed-effect meta-analysis method.

Regarding MR analysis of traits with more than 3 SNPs (modifiable factors and reverse MR where the exposure is genetic liability to VTE), we used the inverse variance weighted method under multiplicative random effects to obtain the association estimates. Additionally, we undertook three sensitivity analyses including the weighted median (146), MR-Egger (114), and MR-PRESSO (116) methods to validate the results' stability and identify potential horizontal pleiotropy. In brief, the weighted median method can yield precise estimates when a majority of the IVs are valid (corresponding weight > 50%). The MR-Egger offers adjusted estimates by accounting for potential horizontal pleiotropy, as identified by its inherent intercept test, though such estimates often lack power. The MR-PRESSO identifies outlier SNPs that introduce pleiotropy and provides estimates after excluding these outliers. We used Cochran's Q test to assess the heterogeneity of SNP estimates.

When estimating mediation of protein in the association between modifiable risk factors and VTE, the proportion influenced by a protein was deduced by multiplying the estimated effect of the modifiable factor on protein levels with the estimated effect of

protein levels on VTE. The standard error for mediation was computed using the error propagation method.

For protein-VTE associations, using a Bayesian model, we conducted a colocalization analysis (147) to ascertain whether the protein and VTE are influenced by the same causal variant within the encoding gene region. This analysis aims to minimize the error caused by LD. For each locus, the Bayesian method assessed the support for the following five exclusive hypotheses: 1) no association with either trait; 2) association with trait 1 only; 3) association with trait 2 only; 4) both traits are associated, but distinct causal variants were for two traits; and 5) both traits are associated, and the same shares causal variant for both traits. The analysis provides posterior probabilities for each hypothesis testing (H0, H1, H2, H3, and H4). We set prior probabilities of the SNP being associated with trait 1 only (p1) at $1 \times 10^{-4}$; the probability of the SNP being associated with trait 2 only (p2) at $1 \times 10^{-4}$; and the probability of the SNP being associated with both traits (p12) at $1 \times 10^{-5}$. Two traits were considered to have strong evidence of colocalization if the posterior probability for shared causal variants (PH4) was ≥0.8.

## 4.3 Druggability assessment of VTE-associated proteins

To determine the druggability of the pinpointed VTE-associated proteins, we explored databases including DrugBank, Dependency Map, the Connectivity Map, and ChEMBL. We collated details on drug names and their developmental phases, subsequently categorizing these protein targets into five tiers: 1) approved (where at least one drug aimed at the protein has secured approval); 2) in clinical trials (with at least one drug targeting the protein undergoing clinical examination); 3) preclinical (featuring proteins with at least one associated drug in the preclinical development phase); 4) druggable (proteins not identifiable in drug databases but recognized as potential drug targets); and 5) not currently designated as druggable.

# 5 Results

## 5.1 Overall and central obesity and VTE risk (Paper I)

### 5.1.1 Cohort analysis

The cohort analysis included 74,317 Swedish women and men from SIMPLER. During a mean 16.7-year follow-up, we identified 4,332 incident VTE cases (2,175 in men and 2,157 in women). The crude incidence rate was 334.2 per 100,000 person-years in men and 350.9 per 100,000 person-years in women.

Elevated BMI and WC were associated with an increased VTE risk, with no discernible variation between sexes (**Table 6**). Both PE and DVT were positively associated with BMI and WC, but the strength of these associations was more pronounced for DVT than for PE. The association between WC and VTE had a slight reduction after accounting for BMI. In contrast, the association of BMI with VTE was considerably weakened when adjusted for WC. Obesity showed an association with VTE in individuals both with and without a prior cancer diagnosis before identifying VTE.

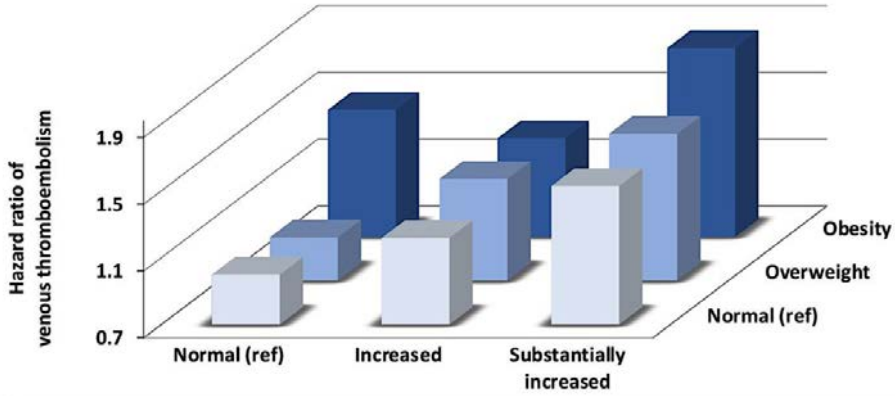
**Table 6**. Risk of incident VTE by BMI and WC in 74,317 Swedish adults.

| Obesity indicator | N | Cases (%) | Model 1 | | Model 2 | |
|---|---|---|---|---|---|---|
| | | | HR | 95% CI | HR | 95% CI |
| **BMI (kg/m$^2$)** | | | | | | |
| 18.5-22.4 (reference) | 14,594 | 4.4 | 1.00 | - | 1.00 | - |
| 22.5-24.9 | 22,319 | 5.6 | 1.24 | 1.12-1.36 | 1.15 | 1.04-1.27 |
| 25.0-27.4 | 19,754 | 5.8 | 1.26 | 1.14-1.39 | 1.08 | 0.97-1.20 |
| 27.5-29.9 | 10,089 | 7.2 | 1.59 | 1.43-1.78 | 1.27 | 1.12-1.44 |
| ≥30.0 | 7,561 | 7.7 | 1.76 | 1.57-1.97 | 1.36 | 1.19-1.55 |
| *P* for trend | | | <0.001 | | <0.001 | |
| **WC** | | | | | | |
| Normal | 24,295 | 4.4 | 1.00 | - | 1.00 | - |
| Increased | 19,938 | 5.9 | 1.27 | 1.17-1.38 | 1.22 | 1.12-1.33 |
| Substantially increased | 17,555 | 7.7 | 1.65 | 1.52-1.79 | 1.46 | 1.31-1.61 |
| *P* for trend | | | <0.001 | | <0.001 | |

Model 1 adjusted for age, sex, education levels, alcohol drinking, cigarette smoking, total energy intake, mDASH score, walking/bicycling time, and ever hormone therapy use (for women). Model 2 additionally adjusted for BMI in analyses of WC and for WC in the analyses of BMI.

In the joint categorization of individuals by BMI and WC (**Figure 14**), participants with a normal BMI but significantly elevated WC exhibited a 53% increased risk (HR 1.53; 95% CI, 1.28, 1.81) of VTE compared to those with normal WC. Within subgroups having identical BMI status, VTE risk exhibited a dose-response pattern across varying WC

statuses. Nonetheless, when considering individuals with the same WC but diverse BMI statuses, the VTE risk appeared consistent.




| Hazard ratio of venous thromboembolism | | | |
|---|---|---|---|
| Body mass index | Waist circumference | | |
| | Normal | Increased | Substantially increased |
| Normal | 1 (reference) | 1.22 (1.10-1.36) | 1.53 (1.28-1.81) |
| Overweight | 0.96 (0.82-1.13) | 1.31 (1.18-1.45) | 1.58 (1.43-1.74) |
| Obesity | 1.46 (0.73-2.94) | 1.29 (0.90-1.83) | 1.83 (1.63-2.05) |

| Number of events/individuals | | | |
|---|---|---|---|
| Body mass index | Waist circumference | | |
| | Normal | Increased | Substantially increased |
| Normal | 797/20 062 | 466/9042 | 126/2048 |
| Overweight | 151/4117 | 553/10 351 | 677/10 103 |
| Obesity | 7/116 | 29/545 | 390/5404 |

**Figure 14**. VTE risk by joint categorization for BMI and WC. The total population included in this analysis is 61,788 owing to missing information on WC for 12,529 participants. The analysis was adjusted for age, sex, education levels, history of hypertension, hypercholesterolemia, diabetes, and fracture, incident cancer, alcohol drinking; cigarette smoking, total energy intake; mDASH; walking/bicycling time, and ever aspirin use.

### 5.1.2 MR findings

Genetically proxied elevations in BMI, WC, and WC adjusted for BMI correlated with VTE risk in both the FinnGen consortium and UK Biobank (**Figure 15**). For each SD increment in BMI, WC, and WC adjusted for BMI, the combined OR for VTE was 1.51 (95% CI, 1.35, 1.70), 1.61 (95% CI, 1.31, 1.98), and 1.40 (95% CI, 1.19, 1.65), respectively. Sensitivity analyses upheld these findings, noting neither heterogeneity nor pleiotropy in the evaluations.

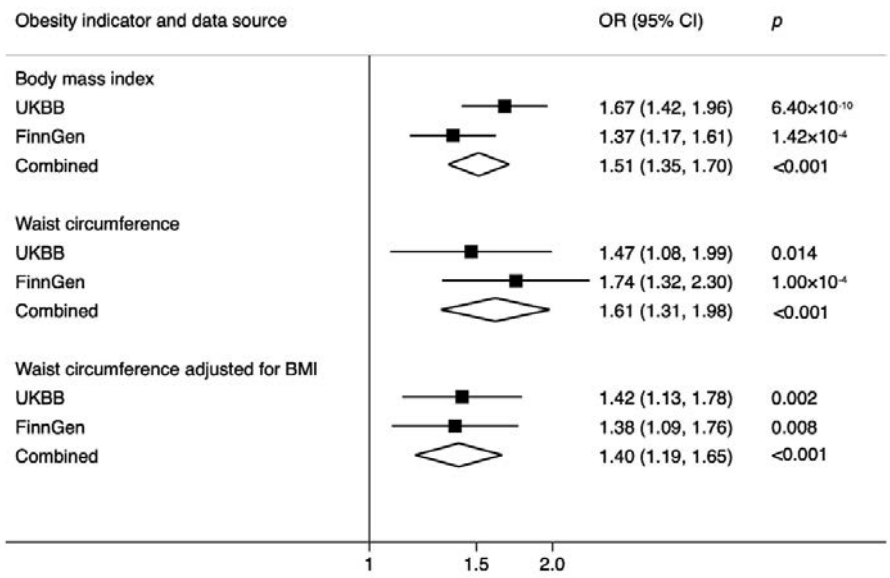



**Figure 15**. MR associations of genetically predicted BMI and WC with VTE in the UK Biobank and FinnGen studies. The associations were scaled to per SD increased in BMI and WC.

### 5.1.3 Population-attributable risk

The estimated fraction of VTE cases due to combined overweight and obesity (with BMI ≥ 25 kg/m$^2$, encompassing 50.3% of the study participants) stood at 12.4% (95% CI 8.4%-16.5%). For obesity alone (with BMI ≥ 30 kg/m$^2$, making up 10.1% of the study participants), this was 5.1% (95% CI 3.8%-6.5%). The estimated percentage of VTE cases due to a combined increased and substantially increased WC (60.7% of the study participants) was 23.7% (95% CI 18.1%-29.4%). For only a substantially increased WC (28.4% of the study population), the proportion was 15.6% (95% CI 12.9%-18.3%).

## 5.2 Lifestyle factors and VTE risk (Paper II)

This cohort study included 66,330 individuals (30,137 women and 36,193 men) from SIMPLER. During follow-up (mean follow-up 17.3 years for women and 16.5 years for men), 1784 women and 2043 men were diagnosed with VTE.

In the combined analysis for both women and men, incident VTE was associated with cigarette smoking, physical activity, and mDASH diet score. The associations with physical activity and mDASH remained evident when analyzing women and men independently. While cigarette smoking was associated with an increased VTE risk in women, no such association was observed in men. Among female current smokers (N= 6,340), there was a 7% rise in VTE risk (95% CI, 0%, 15%) for every additional 5 cigarettes

consumed daily. Neither alcohol nor coffee consumption showed an associated with VTE in women or men.

When evaluating VTE subtypes, associations of physical activity and mDASH with PE remained consistent across women, men, and the combined group. Cigarette smoking was associated to both PE and DVT risks in women. Moreover, physical activity was associated with DVT in both genders. Other lifestyle factors did not exhibit a significant association with either PE or DVT.

**Table 7. Risk of incident VTE by lifestyle factors in SIMPLER**

| Lifestyle factor | Multivariable HR (95% CI) | | |
|---|---|---|---|
| | Combined | Women | Men |
| **Cigarette smoking** | | | |
| Never smoker | Reference | Reference | Reference |
| Past smoker | 1.02 (0.95, 1.10) | 1.16 (1.03, 1.29) | 0.93 (0.84, 1.02) |
| Current smoker, 1-5 cigs/day | 1.01 (0.87, 1.19) | 1.09 (0.87, 1.36) | 0.93 (0.74, 1.16) |
| Current smoker, 6-10 cigs/day | 1.11 (0.97, 1.26) | 1.24 (1.04, 1.48) | 0.96 (0.79, 1.17) |
| Current smoker, 11-20 cigs/day | 1.23 (1.09, 1.38) | 1.47 (1.24, 1.75) | 1.04 (0.88, 1.22) |
| Current smoker, > 20 cigs/day | 1.14 (0.86, 1.52) | 1.39 (0.80, 2.41) | 1.03 (0.73, 1.44) |
| **Alcohol consumption** | | | |
| Never drinker | Reference | Reference | Reference |
| Past drinker | 1.03 (0.86, 1.24) | 1.22 (0.96, 1.54) | 0.80 (0.59, 1.09) |
| Current drinker, <1 drink/wk | 1.03 (0.91, 1.17) | 1.05 (0.91, 1.22) | 1.00 (0.80, 1.24) |
| Current drinker, 1-7 drinks/wk | 1.02 (0.91, 1.14) | 1.09 (0.94, 1.25) | 0.90 (0.75, 1.08) |
| Current drinker, 8-14 drinks/wk | 1.00 (0.87, 1.14) | 0.97 (0.76, 1.22) | 0.93 (0.77, 1.12) |
| Current drinker, 15-21 drinks/wk | 1.00 (0.82, 1.21) | 0.91 (0.51, 1.62) | 0.93 (0.73, 1.18) |
| Current drinker, > 21 drinks/wk | 1.01 (0.76, 1.33) | 1.90 (0.70, 5.14) | 0.91 (0.66, 1.24) |
| **Coffee consumption** | | | |
| <1 cup/day | Reference | Reference | Reference |
| 1-2 cups/day | 0.96 (0.85, 1.09) | 1.02 (0.87, 1.20) | 0.89 (0.74, 1.07) |
| 2-4 cups/day | 0.94 (0.83, 1.06) | 1.00 (0.85, 1.17) | 0.88 (0.73, 1.06) |
| >4 cups/day | 1.00 (0.88, 1.14) | 1.15 (0.96, 1.38) | 0.90 (0.74, 1.08) |
| **Physical activity** | | | |
| <10 mins/day | Reference | Reference | Reference |
| 10-30 mins/day | 0.88 (0.78, 0.99) | 0.74 (0.62, 0.88) | 1.01 (0.86, 1.20) |
| 30-60 mins/day | 0.84 (0.75, 0.93) | 0.72 (0.62, 0.84) | 0.95 (0.82, 1.11) |
| >60 mins/day | 0.73 (0.65, 0.82) | 0.67 (0.58, 0.79) | 0.78 (0.67, 0.92) |
| **Quartile of mDASH diet score** | | | |
| Quartile 1 | Reference | Reference | Reference |
| Quartile 2 | 0.96 (0.87, 1.04) | 1.00 (0.88, 1.15) | 0.92 (0.82, 1.04) |
| Quartile 3 | 0.94 (0.86, 1.03) | 0.93 (0.81, 1.06) | 0.95 (0.84, 1.07) |
| Quartile 4 | 0.87 (0.80, 0.96) | 0.87 (0.75, 0.99) | 0.88 (0.80, 1.00) |

Models were adjusted for age, sex, BMI, education levels, energy intake, history of diabetes, history of fracture, regular aspirin use, hormone therapy (for women), incident cancer, cigarette smoking, alcohol consumption, coffee consumption, physical activity, and mDASH diet score where applicable.

## 5.3 UPF consumption and VTE risk (Paper III)

This cohort study included 186,323 individuals from the UK Biobank. Over a median follow-up duration of 10.5 years (with an interquartile range of 1.4 years), 4,235 new VTE cases were identified, comprising 1,855 DVT and 2,380 PE cases.

A higher consumption of UPF was associated with a heightened risk of VTE across various measures of UPF intake. These associations were largely consistent, regardless of whether UPF was considered as a categorical or continuous variable. While these associations were generally consistent for both VTE subtypes, the associations were more evident for PE, albeit with broader CIs. Differentiating between UPF food and beverage intake revealed that the association's strength was diminished for UPF beverages compared to UPF foods concerning VTE risk.

**Table 8**. Risk of incident VTE by UPF intake

| UPF intake | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | HR (95% CI) | *P* | HR (95% CI) | *P* |
| **UPF servings** | | | | |
| Per SD | 1.09 (1.06, 1.13) | <0.001 | 1.01 (0.98, 1.05) | 0.482 |
| Q1 | ref | | ref | |
| Q2 | 1.03 (0.93, 1.14) | 0.545 | 1.01 (0.91, 1.11) | 0.922 |
| Q3 | 1.04 (0.94, 1.15) | 0.466 | 0.98 (0.88, 1.09) | 0.700 |
| Q4 | 1.24 (1.12, 1.37) | <0.001 | 1.12 (1.01, 1.24) | 0.038 |
| Q5 | 1.29 (1.17, 1.42) | <0.001 | 1.05 (0.94, 1.17) | 0.365 |
| *P* for trend | | <0.001 | | 0.095 |
| **UPF grams** | | | | |
| Per SD | 1.13 (1.09, 1.16) | <0.001 | 1.04 (1.01, 1.08) | 0.008 |
| Q1 | ref | | ref | |
| Q2 | 1.02 (0.92, 1.12) | 0.771 | 1.00 (0.91, 1.11) | 0.957 |
| Q3 | 1.11 (1.00, 1.22) | 0.042 | 1.06 (0.95, 1.17) | 0.291 |
| Q4 | 1.21 (1.10, 1.33) | <0.001 | 1.11 (1.00, 1.22) | 0.049 |
| Q5 | 1.37 (1.24, 1.50) | <0.001 | 1.12 (1.01, 1.24) | 0.026 |
| *P* for trend | | <0.001 | | 0.004 |
| **UPF grams %** | | | | |
| Per SD | 1.13 (1.10, 1.17) | <0.001 | 1.06 (1.03, 1.09) | <0.001 |
| Q1 | ref | | ref | |
| Q2 | 0.96 (0.87, 1.06) | 0.421 | 0.94 (0.85, 1.04) | 0.221 |
| Q3 | 1.07 (0.97, 1.18) | 0.196 | 1.01 (0.92, 1.12) | 0.783 |
| Q4 | 1.18 (1.07, 1.30) | 0.001 | 1.08 (0.98, 1.19) | 0.122 |
| Q5 | 1.31 (1.19, 1.44) | <0.001 | 1.10 (1.00, 1.22) | 0.050 |
| *P* for trend | | <0.001 | | 0.002 |
| **UPF energy** | | | | |
| Per SD | 1.09 (1.05, 1.12) | <0.001 | 1.06 (1.03, 1.10) | <0.001 |
| Q1 | ref | | ref | |
| Q2 | 1.05 (0.95, 1.16) | 0.321 | 1.07 (0.97, 1.18) | 0.207 |
| Q3 | 1.04 (0.94, 1.15) | 0.400 | 1.06 (0.96, 1.17) | 0.262 |
| Q4 | 1.11 (1.00, 1.22) | 0.042 | 1.11 (1.01, 1.22) | 0.038 |
| Q5 | 1.26 (1.14, 1.38) | <0.001 | 1.21 (1.10, 1.33) | <0.001 |
| *P* for trend | | <0.001 | | <0.001 |

| | | | | |
|---|---|---|---|---|
| **UPF energy %** | | | | |
| Per SD | 1.08 (1.05, 1.12) | <0.001 | 1.05 (1.02, 1.08) | 0.002 |
| Q1 | ref | | ref | |
| Q2 | 1.00 (0.91, 1.11) | 0.941 | 1.02 (0.92, 1.12) | 0.741 |
| Q3 | 0.99 (0.89, 1.09) | 0.787 | 0.99 (0.90, 1.09) | 0.871 |
| Q4 | 1.11 (1.00, 1.22) | 0.041 | 1.09 (0.99, 1.20) | 0.070 |
| Q5 | 1.24 (1.13, 1.36) | <0.001 | 1.15 (1.05, 1.27) | 0.004 |
| *P* for trend | | <0.001 | | 0.001 |

Model 1 adjusted for age, sex, and ethnicity; Model 2 adjusted for age, sex, ethnicity, BMI, Townsend's deprivation index, educational levels, smoking and alcohol drinking status, physical activity, and energy intake (energy intake not for UPF energy or UPF energy %).

The association between UPF intake and VTE risk appeared nonlinear when evaluating UPF intake in terms of servings (*P* = 0.007) and grams (*P* = 0.047) (**Figure 16**). The risk of VTE seemed to surge notably at a UPF consumption range of 400-800 grams (approximately 6-14 servings) daily. However, for UPF intakes exceeding 800 grams or 14 servings per day, the association with VTE risk followed a positive linear trend.

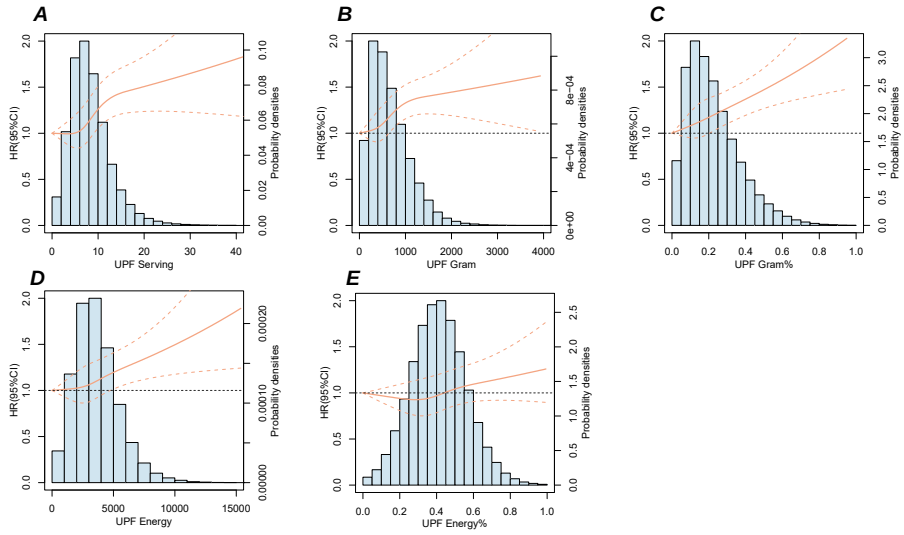



**Figure 16**. Associations of UPF intake with incident VTE risk in restricted cubic spline analyses. $P_{Nonlinearity}$ was 0.007 for A, 0.047 for B, 0.970 for C, 0.685 for D, and 0.164 for E.

## 5.4 257 cardiovascular blood proteins and VTE risk (Paper IV)

### 5.4.1 Cohort analysis

This cohort analysis included 11,803 individuals with baseline blood protein data from SIMPLER sub-clinical cohorts. Over a median observation period of 6.6 years, we identified 352 incident VTE cases.

Upon accounting for multiple testing (*FDR* < 0.1), elevated levels of 14 proteins were associated with a heightened risk of VTE, while raised levels of 7 proteins were associated with a reduced VTE risk (**Figure 17**). These associations persisted in the analysis adjusted for multiple variables and in a sensitivity review that omitted participants with pre-existing cardiovascular disease.

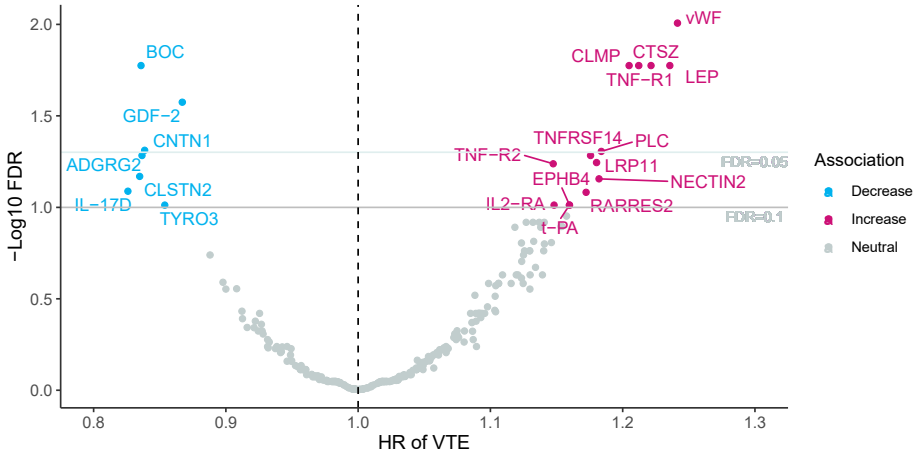



**Figure 17**. Volcano plot illustrating the findings from a protein-wide Cox regression, exploring the associations between 257 circulating proteins and incident VTE.
Abbreviations: ADGRG2, adhesion G-protein coupled receptor G2; BOC, brother of CDO; CLMP, CXADR-like membrane protein; CLSTN2, calsyntenin-2; CNTN1, contactin-1; CTSZ, cathepsin Z; EPHB4, ephrin type-B receptor 4; GDF-2, growth-differentiation factor 2; IL-17D, interleukin-17D; IL2-RA, interleukin-2 receptor subunit alpha; LEP, leptin; LRP11, low-density lipoprotein receptor-related protein 11; NECTIN2, nectin-2; PLC, perlecan; RARRES2, retinoic acid receptor responder protein 2; TNF-R1, tumor necrosis factor receptor 1; TNF-R2, tumor necrosis factor receptor 2; TNFRSF14, tumor necrosis factor receptor superfamily member 14; t-PA, tissue-type plasminogen activator; TYRO3, tyrosine-protein kinase receptor TYRO3; vWF, von Willebrand factor.

In a comparison of identified 21 proteins and 10 traditional risk factors using LightGBM, BMI emerged as the most important risk, evidenced by the highest SHAP value, with vWF trailing closely behind. BOC, CLSTN2, IL2-RA, GDF-2, and CNTN1 were found to have a higher importance than eGFR (**Figure 18**).

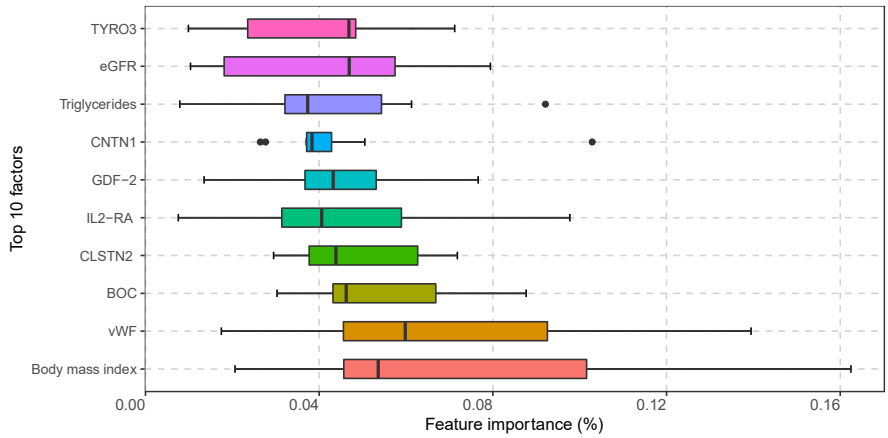
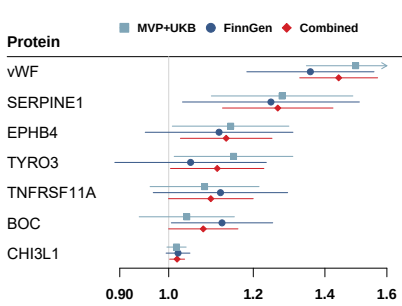



**Figure 18**. Top 10 VTE predictors among 21 identified proteins, 5 lifestyle factors, and 5 clinical features.

### 5.4.2 MR replication

The MR replication included 44 proteins associated with VTE at the nominal significance level in cohort analysis and with available genetic IVs. In the meta-analysis of MVP-UKB and FinnGen, the OR of VTE, per standard deviation increase in genetically predicted protein levels, was as follows: 1.44 (95% CI 1.33–1.57) for vWF; 1.26 (95% CI 1.12–1.42) for plasminogen activator inhibitor 1 (SERPINE1, also known as PAI-1), 1.13 (95% CI 1.03–1.25) for ephrin type-B receptor 4 (EPHB4); 1.11 (95% CI 1.00–1.23) for tyrosine-protein kinase receptor TYRO3; 1.10 (95% CI 1.00–1.20) for tumor necrosis factor receptor superfamily member 11A (TNFRSF11A); 1.08 (95% CI 1.00–1.16) for brother of CDO; and 1.02 (95% CI 1.00–1.03) for chitinase-3-like protein 1, as depicted in **Figure 19**.




| Protein | OR (95% CI) MVP+UKB | FinnGen | Combined | *P_combined* |
|---|---|---|---|---|
| vWF | 1.50 (1.35–1.66) | 1.36 (1.18–1.56) | 1.44 (1.33–1.57) | 1.17E−17 |
| SERPINE1 | 1.28 (1.10–1.49) | 1.25 (1.03–1.51) | 1.26 (1.12–1.42) | 1.09E−4 |
| EPHB4 | 1.14 (1.01–1.30) | 1.11 (0.95–1.31) | 1.13 (1.03–1.25) | 0.014 |
| TYRO3 | 1.15 (1.01–1.31) | 1.05 (0.89–1.23) | 1.11 (1.00–1.23) | 0.042 |
| TNFRSF11A | 1.08 (0.96–1.21) | 1.12 (0.97–1.29) | 1.10 (1.00–1.20) | 0.050 |
| BOC | 1.04 (0.94–1.15) | 1.12 (1.01–1.25) | 1.08 (1.00–1.16) | 0.050 |
| CHI3L1 | 1.02 (1.00–1.04) | 1.02 (0.99–1.05) | 1.02 (1.00–1.03) | 0.028 |

**Figure 19**. MR associations between genetically predicted blood proteins and VTE risk.

### 5.4.3 Druggability assessment

Of the seven proteins identified in the MR meta-analysis as potentially having a causal link to VTE, five are targeted in drug development efforts. There are approved drugs

aimed at vWF for the treatment of bleeding and clotting disorders. Although there were thrombolytic drugs targeting SERPINE1, they were taken off the market due to either inefficacy or adverse effects, including Drotrecogin alfa and Urokinase. Fostamatinib, which targets tyrosine-protein kinase receptor TYRO3, was approved by the US Food and Drug Administration in April 2018 for treating thrombocytopenia. Drugs aimed at EPHB4 and TNFRSF11A are in development for certain cancer treatments. As for brother of CDO and chitinase-3-like protein 1, there is currently no drug development information available.

## 5.5 Protein mediation in modifiable risk factor-VTE links (Paper V)

### 5.5.1 Proteome-wide MR for VTE

After filtering out proteins either lacking SNPs in the outcome dataset or possessing weak IVs (*F* statistic <10), our proteome-wide MR analysis incorporated 1151 plasma proteins. Post-Bonferroni multiple testing correction, we identified 34 proteins associated with VTE ($P < 0.05/1151$; **Figure 20A**). The OR of VTE for every SD increase in genetically predicted protein levels fluctuated between 0.55 (95% CI 0.45-0.68) for Protein S (PROS1) and 3.27 (95% CI 2.66-4.01) for low-density lipoprotein receptor-related protein 12 (LRP12) (**Figure 20B**). Out of these 34 proteins, 10 had IVs sourced from a separate GWAS. Utilizing these IVs for MR, nine associations were confirmed. From the 34 identified protein-VTE associations, 23 demonstrated strong colocalization evidence (PH4 >0.8), while one displayed moderate evidence (0.8> PH4 >0.5; **Figure 20C**). These VTE-associated proteins shared networks of co-expression, physical interactions, and pathway. Among these 34 proteins, 34 and 10 proteins had IVs from the Fenland study and UK biobank, respectively. We replicated 31 associations using IVs for protein from the Fenland study and 9 associations using IVs for proteins from the UK biobank (**Figure 20D**). In the reverse MR analysis, we found no evidence of associations of genetic liability to VTE with the levels of identified blood proteins after multiple testing correction.

We sourced clinical trial data from four drug databases, targeting the VTE-associated proteins identified in our proteome-wide MR analysis to ascertain their druggability. We discovered that 12 of these proteins had approved drugs targeting them, with nine (coagulation factor XI [FXI], prekallikrein, prothrombin, protein C, protein S, ATS13, annexin II, MARK3, and fibrinogen g-chain dimer) specifically used to treat thrombotic conditions. Notably, most of these proteins were deemed druggable, although they varied in their stages of druggability assessment.

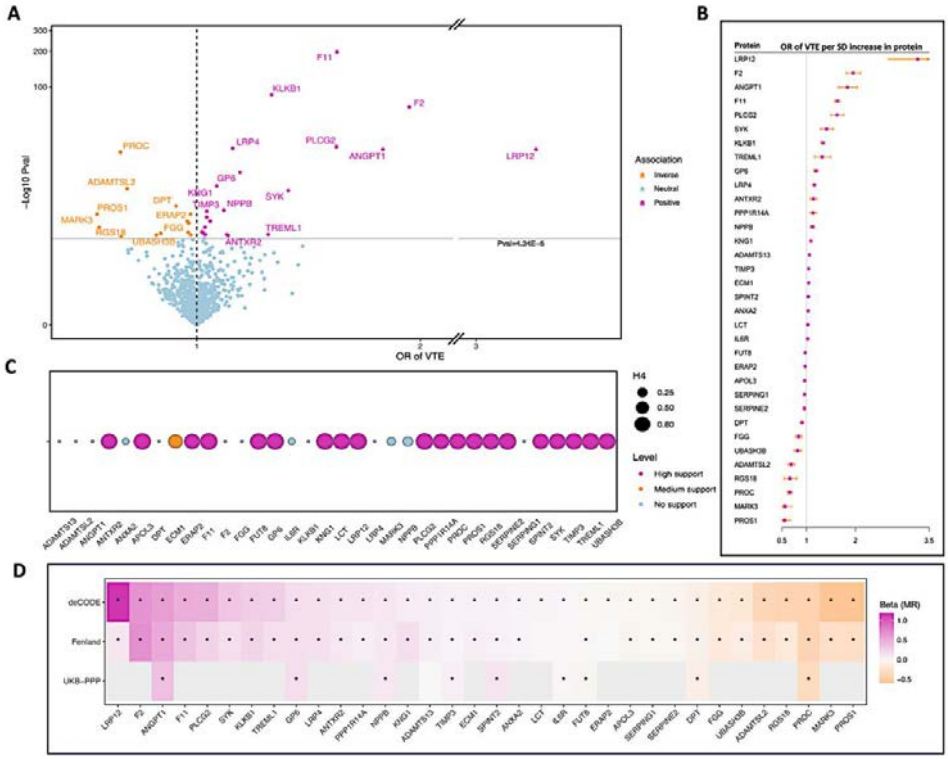



**Figure 20**. Proteome-wide MR and colocalization analyses on the associations between blood proteins and VTE risk. A. the volcano plot of results of the proteome-wide MR analysis of VTE using the discovery deCODE protein data. B. forest plot of identified MR associations between blood proteins and VTE risk using the discovery deCODE protein data. C. results of colocalization analysis based on deCODE protein data. D. comparison of associations in the discovery analysis based on deCODE protein data and replication analyses based on Fenland and UKB protein data. Abbreviations: ADAMTS13, A disintegrin and metalloproteinase with thrombospondin motifs 13; ADAMTSL2, ADAMTS-like protein 2; ANGPT1, Angiopoietin-1; ANTXR2, Anthrax toxin receptor 2; ANXA2, Annexin A2; APOL3, Apolipoprotein L3; DPT, Dermatopontin; ECM1, Extracellular matrix protein 1; ERAP2, Endoplasmic reticulum aminopeptidase 2; F11, Coagulation factor XI; F2, Prothrombin; FGG, Fibrinogen gamma chain; FUT8, Alpha-(1,6)-fucosyltransferase; GP6, Platelet glycoprotein VI; IL6R, Interleukin-6 receptor subunit alpha; KLKB1, Plasma kallikrein; KNG1, Kininogen-1; LCT, Lactase/phlorizin hydrolase; LRP4, Low-density lipoprotein receptor-related protein 4; LRP12, Low-density lipoprotein receptor-related protein 12; MARK3, MAP/microtubule affinity-regulating kinase 3; NPPB, Natriuretic peptides B; PLCG2, 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2; PPP1R14A, Protein phosphatase 1 regulatory subunit 14A; PROC, Vitamin K-dependent protein C; PROS1, Vitamin K-dependent protein S; RGS18, Regulator of G-protein signaling 18; SERPINE2, Glia-derived nexin; SERPING1, Plasma protease C1 inhibitor; SPINT2, Kunitz-type protease inhibitor 2; SYK, Tyrosine-protein kinase SYK; TIMP3, Metalloproteinase inhibitor 3; TREML1, Trem-like transcript 1 protein; UBASH3B, Ubiquitin-associated and SH3 domain-containing protein B.

### 5.5.2 Causal modifiable risk factor for VTE

Ten out of the fifteen genetically proxied modifiable factors displayed an association with VTE after multiple testing correction (**Figure 21**). Genetic predispositions to obesity, cigarette smoking, a sedentary lifestyle, short sleep durations, daytime napping, and insomnia were associated with an elevated VTE risk. These associations remained in sensitivity analyses. While heterogeneity was evident in most associations, there was scant evidence of horizontal pleiotropy as indicated by the MR-Egger intercept test (*P* >0.05).

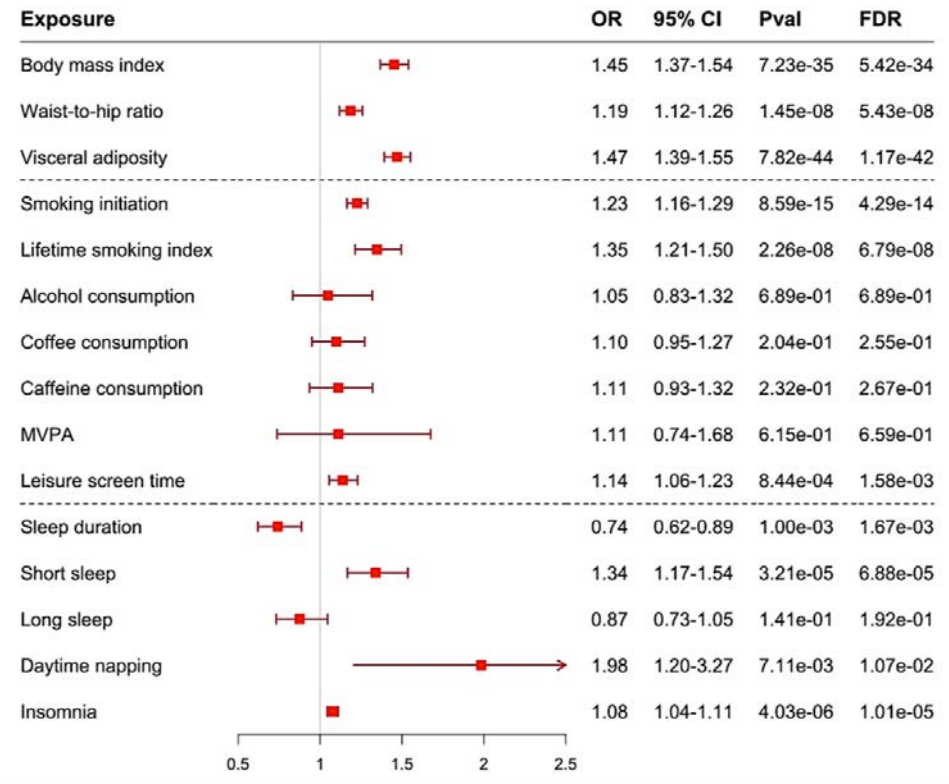



**Figure 21**. Associations of genetically proxied modifiable risk factors with VTE. MVPA, moderate-to-vigorous physical activity.

### 5.5.3 Mediation of proteins in the associations between modifiable risk factors and VTE

In the MR analysis examining the links between VTE-related modifiable factors and VTE-associated proteins, we set the threshold for significance at the nominal level to uncover as many potential mediation signals as possible. This approach identified a total of 86 pairs of associations. Among the VTE-related modifiable factors, indicators of obesity were associated with over 18 VTE-associated proteins. When considering VTE-

associated proteins, seven modifiable factors were associated with ADAMTS-like protein 2, five with anthrax toxin receptor 2 (ANTR2), and five with FXI.

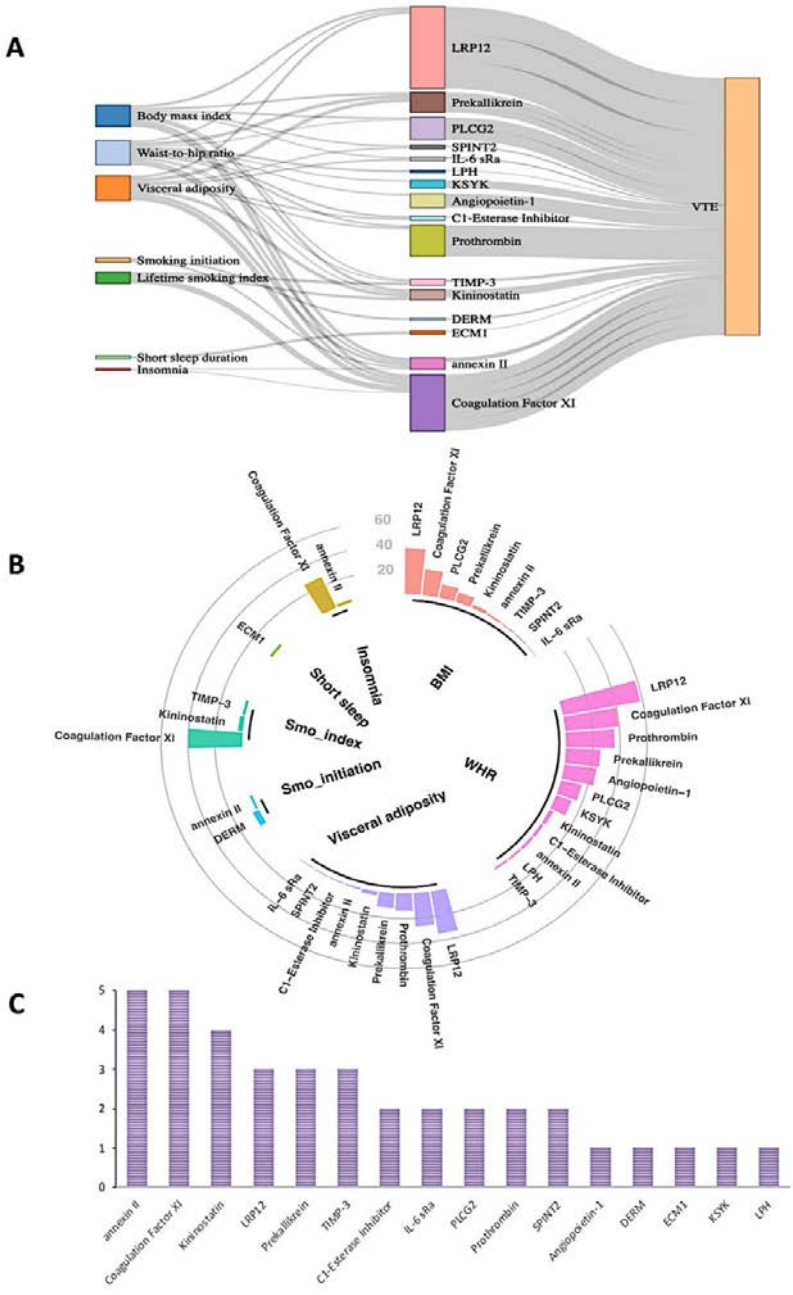



**Figure 22**. Mediation effects of proteins in the associations between modifiable risk factors and VTE. A. protein pathways linking obesity, smoking, and sleep-related traits to VTE. B.

the proportion of association between the modifiable factor and VTE mediated by a protein. C. the count of protein mediators among all identified pathways. Abbreviations: DERM, Dermatopontin; ECM1, Extracellular matrix protein 1; IL-6 sRa, Interleukin-6 receptor subunit alpha; KSYK, Tyrosine-protein kinase SYK; LPH, Lactase/phlorizin hydrolase; LRP12, Low-density lipoprotein receptor-related protein 12; PLCG2, 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2; SPINT2, Kunitz-type protease inhibitor 2; PLCG2, 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2; TIMP3, Metalloproteinase inhibitor 3.

Out of the identified 86 pairs, we assessed the mediation of 38 combinations in which the modifiable factor-protein-VTE direction aligned with the influence through the mediator. This was determined by the directionality of the total effect (beta of the modifiable factor-VTE relation) being consistent with the mediating effect (beta of the modifiable factor-protein relation multiplied by the beta of the protein-VTE relation). Among 38 combinations, 30 were related to obesity indicators, and five proteins - annexin II, FXI, Kininostatin-1, LRP12, and prekallikrein - consistently mediated the associations of three obesity indicators with VTE risk (**Figure 22A**).

In terms of the mediation magnitude, LRP12, FXI, and prothrombin were the most influential in the obesity-VTE association (**Figure 22B**). FXI played a pivotal role, mediating 40% of the cigarette smoking-VTE association (95% CI 20-60%). Similarly, FXI mediated 27% of the association between sleep-related traits and VTE (95% CI 5-49%). Overall, both annexin II and FXI emerged as key mediators for many associations between the examined modifiable factors and VTE (**Figure 22C**).

# 6 Discussion

## 6.1 Summary of main findings

This PhD thesis encompasses five studies, highlighting the intricate relationship between modifiable risk factors, blood proteins, and the susceptibility to VTE. Key findings include:

**Paper I)** Obesity was causally associated with a heightened risk of VTE. When assessing obesity in relation to VTE risk, WC appeared to be a more discerning metric compared to BMI. By ensuring a population-wide adherence to recommended BMI and WC guidelines, more than 20% of VTE cases could be prevented.

**Paper II)** Higher physical activity and a healthy diet were associated with a lower risk of incident VTE. Cigarette smoking was associated with a higher risk of VTE in women. Conversely, no significant associations were observed between VTE risk and alcohol or coffee consumption.

**Paper III)** Elevated UPF consumption was associated with a heightened risk of developing VTE.

**Paper IV)** Numerous protein biomarkers were associated with the onset of VTE. Notably, proteins such as vWF, SERPINE1, EPHB4, and TNFRSF11A showed similar associations in the MR analysis, suggesting a likely causal relationship. Circulating vWF levels might play an equally important role in VTE development comparable to that of BMI.

**Paper V)** Multiple proteins, especially annexin II and FXI, played a mediating role in the relationship between obesity, smoking, insomnia, and VTE risk. A significant number of VTE-associated proteins demonstrated strong druggability potential, especially in relation to coagulation-related disorders.

## 6.2 Interpretation of the findings

### 6.2.1 Comparison to previous studies

#### *6.2.1.1 Modifiable risk factors and VTE*

**Obesity**: The association of obesity (66, 148-150) with VTE risk was consistently revealed in many population-based observational studies. The causality of the association between higher BMI and a heightened risk of VTE was reinforced in MR analysis (151, 152). Based on Swedish cohorts and MR analysis with a larger sample size, our findings further strengthened the association, which emphasizes the role of obesity in thrombus formation. Additionally, our work suggested that central obesity indicator, like WC appeared to be a better index compared to overall obesity indicator, like BMI to discriminate VTE risk across individuals with different obesity status. This finding

highlights the pronounced influence of adipose tissue distribution on VTE risk and consequently underscores the importance of a more precise characterization of obesity beyond just BMI for a deeper understanding of VTE's etiology.

**Physical activity/inactivity**: Likewise, the inverse association between physical activity and VTE risk has been identified in previous studies (92), which is in line with our findings. Interestingly, in our study, the protective influence of physical activity seemed more pronounced in women than in men, and more evident for DVT compared to PE. However, the inverse association between moderate-to-vigorous physical activity and VTE was not clearly observed in our MR analysis. Instead, we identified a positive link between prolonged leisure screen time (representing a sedentary lifestyle) and VTE risk.

**Cigarette smoking**: While the relationship between cigarette smoking and VTE has been ambiguous (57), many studies lean towards a positive correlation (65, 66, 153). Some research suggests that the risk may be not directly from smoking, but rather from smoking-associated diseases (57). After adjusting for such diseases, like cancer and fracture, our cohort study identified a positive correlation between smoking and VTE risk exclusively in women. This absence of association in men aligns with certain findings (3) and is further supported by a recent MR study (67). This study indicated that a genetic predisposition to start smoking increased the VTE risk in the UK Biobank (comprising roughly 50% men), but not in the MVP (with 92% men) (67). One potential explanation for this gender-specific link might be a substantially higher tobacco dose threshold to trigger risk in men than in women (154). Further research is needed to delve deeper into the gender-based nuances of the association between smoking and VTE.

**Diet**: The results from a handful of observational studies examining diet quality in relation to VTE have been inconsistent. One prospective cohort study involving 129,430 US participants identified a link between the Western dietary pattern and a 16% increased VTE risk in men, with no such correlation observed in women (81). Conversely, some studies have found connections between either prudent or Western dietary patterns and VTE risk (77, 78). In a separate cohort study tracking 675 VTE cases, following the DASH diet did not correlate with VTE risk in women over an average follow-up of 14.6 years (79). Yet, our research suggests a potential protective relationship between stricter adherence to the DASH diet and VTE, particularly PE. Along with potential inverse associations with other major cardiovascular diseases (129, 155), DASH diet can be recommended as a primary overall cardiovascular disease prevention strategy.

**Coffee consumption**: Research regarding the relationship between coffee consumption and VTE is limited. One study hinted at a U-shaped correlation, suggesting that consuming 5–6 cups daily might reduce VTE risk compared to non-drinkers (76). However, another study found a link between coffee and VTE; though, this association

diminished after factoring in body mass index and diabetes (77). In our research, there was no association between coffee intake and VTE risk in either women or men.

**Alcohol consumption**: Concerning alcohol intake, our null findings align with a meta-analysis involving roughly 400,000 participants and over 10,000 VTE cases from 10 studies (70), which found no significant relationship between alcohol and VTE development. This is further supported by a recent MR study showing no evident link between alcohol and VTE risk (156).

**UPF consumption**: This project identified a novel risk factor for VTE, which is excessive UPF intake. The UPF intake has been associated with the risk of coronary artery disease and cerebrovascular disease in some studies (157, 158); however, its association with VTE has been established for the first time in our analysis using data from the UK Biobank. This finding may not only imply a negative influence of UPF on thrombus formation but also further proves the importance of diet in VTE prevention.

**Sleep-related traits**: Obstructive sleep apnea (159, 160) and sleep disorders (95) have been associated with an increased risk of VTE. Our MR analysis found three sleep-related issues, including short sleep duration, daytime napping, and insomnia, positively associated with the risk of VTE. Due to few studies on these topics, these findings are considered novel and need verification.

#### *6.2.1.2 Blood protein and VTE*

**vWF**: vWF's association with VTE risk has been consistently noted across numerous cohort studies and genetic analyses (101, 102, 161, 162). Our research reaffirms the causal significance of heightened vWF levels in VTE development. Intriguingly, our machine-learning results position vWF levels and body mass index on an equal footing regarding VTE onset. This thus suggests monitoring vWF might be as crucial as managing a healthy BMI to mitigate VTE risk. This revelation could shape future clinical interventions and VTE management strategies. Moreover, the therapeutic potential of vWF for VTE has gained traction, with certain nonspecific (like heparin and aurintricarboxylic acid) and specific (like caplacizumab) vWF antagonists already approved for thrombotic disorders treatment (163). Additionally, Egaptivon pegol, a novel drug targeting vWF, is currently under scrutiny.

From both biological and genetic standpoints, vWF and factor VIII exhibit a strong correlation (161, 164). The question of whether vWF's association with VTE risk is independently of factor VIII remains. A prior MR study, drawing from extensive consortium data, could not establish an independent causal connection between vWF levels and VTE that was devoid of factor VIII's influence, primarily due to the absence of genetic loci exclusively linked to vWF, unlike with factor VIII (161). Given lack of data on factor VIII, we could not dissect the independent roles of vWF and factor VIII concerning

VTE risk. Future research should aim to elucidate whether vWF's association with VTE functions independently of factor VIII.

**SERPINE1 or PAI-1**: In a case-control study involving 770 patients and 743 controls, elevated SERPINE1 levels, commonly referred to as PAI-1, were found to underlie the observed positive link between prolonged plasma clot lysis time (hypofibrinolysis) and VTE risk (165). A similar association was found in a case-control study from Norway, examining 383 VTE cases alongside 782 age- and sex-matched controls (166). While the association of *PAI-1* gene with VTE risk was inconsistent (167, 168), both our prospective cohort and MR analyses suggest that high PAI-1 levels could have a potentially causative negative impact on VTE development. It is worth noting that while certain drugs affecting PAI-1 levels have received approval for treating thrombotic disorders, multiple PAI-1 targeting drugs were discontinued due to their limited efficacy (169).

**Other established blood proteins with previous evidence support**: A comprehensive MR study involving 81,669 VTE cases across diverse ancestries pinpointed 23 proteins linked to VTE (170). Many of these, including FXI, prekallikrein, prothrombin, and protein S, which have clear roles in thrombosis, were validated in our research. Echoing another protein-centric study (171), we found kininogen 1 and protein C to be associated with VTE. For other proteins in thrombosis, our observations are bolstered by earlier research, specifically on phospholipase C gamma 2 (172), angiopoietin-1 (173, 174), glycoprotein VI (175), tyrosine kinase Syk (176), N-terminal pro-BNP (177), metalloproteinase inhibitor 3 (178), extracellular matrix protein 1 (179), ADAMTS13 (A disintegrin and metalloproteinase with thrombospondin motifs 13) (180), protease nexin-1 (181), annexin II (182), interleukin-6 receptor subunit alpha (183), plasma protease C1 inhibitor (184), fibrinogen g-chain dimer (185), Trem-like transcript 1 protein (186), ubiquitin-associated and SH3 domain-containing protein B (also known as T-cell ubiquitin ligand 2) (187), and regulator of G-protein signaling 18 (188).

**Other identified proteins with limited previous data**: While we discerned potential causal relationships of EPHB4, tyrosine-protein kinase receptor TYRO3, TNFRSF11A, brother of CDO, and chitinase-3-like protein 1 with VTE, these novel observations require further validation. Of note, inconsistencies exist in the associations of tyrosine-protein kinase receptor TYRO3 and brother of CDO with VTE between cohort and MR analyses, which might be caused by variations in proteomic profiling assays used in the two analyses. For the identified proteins associated with VTE in the proteome-wide MR analysis, while direct evidence connecting LRP12 to VTE is sparse, this lipid metabolism-centric protein is linked to platelet internalization (189) and vascular endothelial functionality (190), potentially influencing thrombus development indirectly. Similarly, the relationship between ADAMTS-like protein 2 and VTE is not well-documented. Yet, it is relevant that ADAMTS-like protein 2 shares a protein superfamily with ADAMTS13 (A disintegrin and metalloproteinase with thrombospondin motifs 13) (191), which has an

established role in thrombosis. Associations of VTE with proteins like low-density lipoprotein receptor-related protein 4, dermatopontin, endoplasmic reticulum aminopeptidase 2, apolipoprotein L3, microtubule affinity-regulating kinase 3, lactase-phlorizin hydrolase, protein phosphatase 1 regulatory subunit 14A, alpha-(1,6)-fucosyltransferase, and anthrax toxin receptor 2are minimally explored in existing literature. These proteins have connections to lipid metabolism, oncology, or immunological responses. Further investigations are imperative to validate these potential associations.

### 6.2.2 Underlying mechanisms

Regarding obesity, in particular central obesity, in relation to VTE, systemic inflammation may be a potential underlying pathway. Abdominal obesity has been shown to exhibit a distinct effect on inflammation (192). Results of our previous MR studies showed causal effects of inflammation-related fatty acids, especially arachidonic acid (193), and tumor necrosis factor (194) on VTE. Our cohort analysis also found a link between antiinflammation diet and VTE risk (195). In addition, obesity-driven chronic inflammation and impaired fibrinolysis have been postulated to be major effector mechanisms of thrombosis (196). Furthermore, a review article suggested that central obesity could amplify ectopic fat accumulation, potentially accelerating venous thrombosis (197).

The mechanism underlying the association between physical activity and VTE can be explained by the Virchow's triad. High levels of physical activity largely reduce stasis, which thus lower the risk of VTE. Whether physical activity can influence the risk of VTE via altering coagulation status or endothelial injury/dysfunction is unclear. In addition, the stronger association between physical activity and VTE observed in women may be linked to differences in sex hormones. A meta-analysis from randomized controlled trials revealed that physical activity led to reduced circulating sex hormone levels in healthy women, an effect not solely attributable to weight loss (198).

For the positive association between UPF and VTE, several hypotheses have been proposed. Firstly, high UPF consumption correlates with obesity, especially abdominal fat (199, 200), a recognized VTE risk factor as we established. Even after accounting for BMI in our study, the UPF-VTE link remained, suggesting other potential pathways. Secondly, UPF consumption has been tied to inflammatory markers such as elevated interleukin-6 (201) and high-sensitivity C-reactive protein (202), both of which have been linked to heightened VTE risk (203). Thirdly, excessive UPF intake might contribute to kidney dysfunction (204), with reduced glomerular filtration rates (indicative of renal issues) recognized as a VTE risk factor (49). Fourthly, UPF consumption may suggest an overall poor dietary quality, which has been associated with increased VTE risk. Yet, the Moli-sani Study showed that the UPF-mortality link was not merely due to diet quality

(205). Similarly, our study, even after adjusting for diet quality, still observed a significant UPF-VTE relationship, suggesting factors beyond just diet quality.

For the associations of obesity, smoking, short sleep duration, and insomnia with VTE, our study V revealed some protein underlying pathways. For example, annexin II and FXI serve as mediators between various modifiable factors (such as obesity, smoking, and insomnia) and VTE. Furthermore, LPR12, FXI, and prothrombin consistently underpin the association between diverse obesity markers and VTE. Pinpointing these pathways could enhance our understanding of the pathogenesis underlying VTE development, especially unprovoked VTE. Moreover, considering the druggable nature of some of these protein mediators, this knowledge could also inform VTE therapeutic strategies.

## 6.3 Methodological considerations

This PhD project employed both prospective cohort and MR methodologies across **studies I** to **V**. While each design has its inherent limitations that merit consideration when evaluating the results, our comprehensive assessment of potential biases suggests a minimal likelihood that our conclusions are significantly skewed.

### 6.3.1 Prospective cohort design

In cohort studies, the accuracy and reliability of research findings are influenced by the presence of two types of errors: random errors (chance) and systematic errors (bias) (206). Random errors are inherent fluctuations in measurements that occur by chance. On the other hand, systematic errors, or bias, introduce a consistent and directional distortion in study outcomes. Bias can result from flaws in study design, data collection methods, or participant selection, leading to a systematic deviation of the study findings from the true values.

The prospective cohort design utilizing data from Swedish and UK Biobank cohorts was applied in **studies I, II, III,** and **IV**. Intending to probe the relationships between obesity, lifestyle factors, UPF intake, and blood proteins with the risk of VTE, these studies aimed to establish precise connections indicative of causality. While randomized controlled trials are often hailed as the gold standard for causal inference, they demand significant resources, ethical considerations, and extended timelines, which might not be feasible for every research question, particularly for most VTE (unprovoked type) that needs a comparatively long time of follow-up. Cohort studies offer a practical alternative, capturing real-world data over extended periods. However, limitations inherent to cohort designs including both random errors and systematic errors can challenge the strict requirements of causal inference. Thus, while cohorts can provide robust associations, drawing definitive causal conclusions requires careful consideration of these limitations. We have discussed each limitation in detail in the following paragraphs.

#### *6.3.1.1 Random error (chance)*

Random error in prospective cohort studies is usually caused by sampling error and measurement error (207). However, measurement error can manifest as both random error and bias, which is determined by whether the error occurs by chance or in a systematic deviation of the observed measurements from the true values.

Sampling error means the characteristics of the studied sample do not perfectly reflect the characteristics of the entire population. It is a natural consequence of using a subset of the population to make inferences about the entire population. It influences the precision of estimates; however, it does not systematically distort the estimates in one direction.

Measurement error indicates inaccuracies or imprecision in the measurement of key variables, such as exposures, outcomes, or covariates. Measurement error can contribute to random error when it leads to imprecision or variability in the observed measurements. Random error is characterized by fluctuations in data that occur by chance and can affect the precision of estimates. This error can be less or more for different traits and measurements. It does not have a systematic direction and tends to balance out with repeated measurements.

To minimize random error, several approaches can be used from different perspectives like using high-quality instruments, training staff on how to take measurements, designing good questionnaires, increasing sample size, repeating measurements, using advanced statistical methods, etc. However, it cannot be eliminated. We usually use *p*-values and confidence intervals to assess random error in epidemiology (208). In this thesis, random error in study I to IV should be minimized due to 1) large sample sizes; 2) well-validated questionnaires; 3) well-trained staff for data collection; and 4) repeated 24-h dietary recall measurements in UK Biobank.

#### *6.3.1.2 Systematic errors (bias)*

##### 6.3.1.2.1 Confounding

A confounder is a variable that influences both the dependent variable and independent variable, causing a spurious association. We usually make corresponding adjustments in the statistical models to minimize confounding. Except measured confounders, unmeasured confounders influence causal inference. In contrast to retrospective cohort studies and other forms of observational research, prospective cohort analyses are anticipated to exhibit fewer challenges arising from unmeasured confounding factors due to their inherent study design. Nevertheless, exceptions exist, particularly when certain outcomes of interest are identified after the collection of baseline data. In such instances, the comprehensive inclusion of all pertinent risk factors for these unforeseen outcomes may not have been achieved during the initial data collection phase. Except

from unmeasured factors, residual confounding can come from inadequate measurement of confounders.

In our studies, despite accounting for many critical covariates, as outlined using DAGs (209), and adjusting for them in our analyses, potential unmeasured or unknown confounders or measurement errors in included confounders may still exist. This can introduce what is termed "residual confounding." It is crucial to evaluate this bias in relation to the specific exposure and outcome under consideration, as it can influence the likelihood of Type I or II errors, depending on the effects of the unaccounted confounder on both the exposure and outcome. Take, for instance, the potential influence of anti-inflammatory drugs on the relationship between obesity and VTE risk. The usage of nonsteroidal anti-inflammatory drugs has been shown to be inversely associated with obesity (210) and associated with an increased VTE risk (211). Our cohort's absence of data on anti-inflammatory drug usage might lead to its residual confounding. However, given the divergent effects of anti-inflammatory drugs on obesity and VTE, our hypothesis is that any residual confounding would likely moderate the observed relationship between obesity and VTE, making our results more conservative. Taking another scenario into account, consider the potential for a healthier lifestyle with possibly higher levels of physical activity among users of anti-inflammatory drugs. This assumption could introduce residual confounding to the observed inverse relationship between physical activity and VTE risk. In such a situation, the bias would likely amplify the risk of a Type I error, casting doubt on the causative nature of the association between physical activity and VTE.

Similarly, in the relationship between smoking and VTE, there is a prevailing hypothesis suggesting that VTE risk is more closely associated with smoking-related comorbidities rather than direct smoking itself (57). While our analysis did account for certain smoking-related conditions, like fractures and cancer, the association between smoking and VTE risk in women could still be influenced by residual confounding from other unaccounted-for diseases. In addition, there is significant uncertainty in pinpointing the factors that influence blood protein levels. This ambiguity complicates the identification of confounders in the cohort analysis exploring the relationship between blood proteins and VTE, potentially leading to further residual confounding. However, most observed associations in the cohort analyses have been replicated using MR design, which partly reinforced the causality of these links by minimizing the bias.

#### 6.3.1.2.2 Selection bias

Selection bias in prospective cohorts means the distortion of associations between exposures and outcomes due to systematic differences in the characteristics between those who are included in the study and those who are excluded or lost to follow-up. It mainly jeopardizes external validity. Selection bias includes non-response bias, healthy

worker effect, and attrition bias. Of note, it is not clear to determine whether a selection of participant affects only external validity/generalizability or additionally impair the internal validity. The former does not question the results obtained within the study, but question whether it can be transferred to external populations. In the later scenario, selection bias can be introduced during recruitment in case-control studies and in tracing of participants to ascertain their outcome status in prospective cohorts. According to Rothman K.J., non-response bias and missing information bias resulting from differential selection at recruitment are viewed as confounding bias, since they are not conditioned on an outcome that has not yet occurred.

**Non-response bias** It occurs when individuals who choose to participate (or respond to certain questions or follow-up efforts) differ systematically in their exposures or outcomes from those who do not. Furthermore, when participants choose not to respond to specific questions in a survey or questionnaire, especially if their reasons correlate with the exposure or outcome, it introduces another layer of bias. In the context of SIMPLER, this is likely not a major concern given the minimal rate of missing data for most of our independent variables. Specifically concerning dietary intake, when participants provided partial responses, we interpreted this as infrequent or no consumption. Our rationale for this assumption stems from a Swedish study on the "zero approach" for handling partial non-responses, which found that about 74.1% of the missing responses aligned with genuine non-consumption (212, 213). In the UK Biobank, there were systematic patterns observed in item nonresponse to survey questionnaires (214). While the association between UPF and VTE in this dataset held firm across various imputation strategies for covariates, it is still essential to approach the generalizability of this association with caution.

**Healthy volunteer effect** This occurs when the study population consists of individuals who are healthier than the general population due to social stigma etc. Participants of SMC and COMS were generally comparable to the general Swedish population concerning at least age distribution, education level, and BMI (155, 215). However, the UK Biobank may have the “healthy volunteer” selection bias; however, it does not necessarily compromise the accuracy of the identified associations (216).

**Attrition bias** This results from differential loss to follow-up, where participants who drop out of the study have different characteristics or outcomes compared to those who remain (missing not at random) (217). Given that both SIMPLER (218, 219) and the UK Biobank (216) obtain VTE diagnosis from national registers and have a low loss to follow-up rates, this bias may be minimal.

6.3.1.2.3 Information bias

Information bias refers to systematic errors in the collection, measurement, or recording of information related to exposure, outcome, or confounding variables. This includes misclassification bias, response bias, and detection bias.

**Misclassification bias** This bias occurs when participants are incorrectly classified with respect to their exposure or outcome status. This misclassification can be non-differential (misclassification is random and unrelated to true exposure or outcome status) or differential. For exposure misclassification in prospective cohorts, it is usually non-differential given that the exposure information is collocated at the baseline before the onset the outcome (220). Regarding blood proteins, potential sources of measurement error include variations in sample handling and techniques (e.g., batch effects), and inherent detection limits, even when using high-sensitivity and specificity platforms like Olink (221). This non-differential exposure misclassification tends to bias associations towards the null and therefore result in reduced statistical power and an increased likelihood of Type II errors. Exposure misclassification can also be a consequence of temporal variability that denotes the changes in an individual's exposure level over a period, potentially leading to exposure misclassification. This misclassification is likely to be non-differential, given that exposures were ascertained before VTE diagnosis, possibly attenuating the observed strength of the association. Of note, whether the risk estimates are biased away from the null sometimes is different for the exposure defined as continuous, dichotomous, or polytomous. For example, the risk estimates for intermediate levels can be biased away from the null; however, dichotomous exposure categorization is often to the null.

The outcome misclassification is more related to the sensitivity and specificity of the diagnostic definition. If the sensitivity is low, which means individuals with the outcome will be incorrectly classified as non-cases (false negative), this outcome misclassification will bias the association to the null. If the specificity is low, which means individuals without the outcome will be incorrectly classified as cases (false positive), this outcome misclassification can bias the association away from the null. For studies in this thesis, the exposure misclassification might attenuate the associations in a conservative way due to the prospective design, which alarms whether the null association is truly null. Regarding the outcome misclassification, the sensitivity and specificity of cardiovascular disease was found to be satisfying in the Swedish patient register (both >90%) (134) used in SIMPLER. For the UK Biobank (222), the sensitivity and specificity appear to vary largely between outcomes (223, 224) and are potentially influenced by datasets used (225).

**Response bias** This can occur if a group of participants reports their data systematically different from their counterparts since they know their healthy behaviors

may influence health. Due to the prospective nature, this bias may be minimal in SIMPLER and the UK Biobank.

**Detection bias** This occurs when there are systematic differences in the identification, diagnosis, or ascertainment of outcomes between comparison groups. For example, smokers may have a higher chance of being diagnosed with lung cancer compared with non-smokers due to more frequent check-ups for respiratory symptoms. However, this bias is less likely to apply to VTE since early DVT is usually asymptomatic and PE is urgent for hospitalization.

### 6.3.2 MR design

#### *6.3.2.1 Unbalanced horizontal pleiotropy*

Horizontal pleiotropy, when unbalanced, can bias MR estimates by contravening the third assumption. To mitigate this, rigorous examination of genetic IVs for exposure is crucial. In our MR evaluation, we utilized index cis-SNPs as the IVs for blood proteins, which typically adhere to this assumption (226, 227). For complex traits, we carried out a range of sensitivity tests designed to identify and adjust for horizontal pleiotropy. The consistency across these results and the overall minimal evidence of horizontal pleiotropy underscores the reliability and precision of the MR associations we observed.

#### *6.3.2.2 Inadequate statistical power*

Given that a quite large sample of VTE GWAS used in our MR analysis, the statistical power can be inadequate when using genetic instruments that explain a low proportion of phenotypic variance for the exposure. The lack of association between moderate-to-vigorous physical activity and VTE and between alcohol intake and VTE may be related to lack of power.

#### *6.3.2.3 Bidirectional causality*

While the MR analysis minimizes the risk of reverse causality due to the fixed nature of genetic variants, which are not influenced by disease onset or progression, we still cannot entirely dismiss the potential for bidirectional causality (228). Moreover, conducting the MR analysis with exposure and outcome data from separate populations can introduce bias if a considerable fraction of VTE cases exists in the exposure data. To address this in **Study V**, we carried out the reverse MR analyses for proteins associated with VTE and lifestyle factors to statistically assess this potential bias. Our findings revealed limited associations between genetic predisposition to VTE and both VTE-associated proteins and lifestyle factors, suggesting a minimal impact from reverse causation.

#### *6.3.2.4 Measurement error*

Measurement error can bias MR analysis, primarily stemming from the metrics and definitions of traits prior to gene-trait association evaluations. Such errors can diminish the analytical power and elevate the likelihood of type II errors (229). Although we cannot completely exclude the possibility of measurement error in our MR study, given that it relies on summary-level data, any significant bias is unlikely. This is reinforced by our utilization of data from well-conducted, high-quality published GWAS as well as consistent results obtained from a series of sensitivity analyses.

#### *6.3.2.5 Sample overlap*

In a two-sample MR analysis, sample overlap refers to the exposure and outcome GWAS studies sharing a complete or partial set of individuals. Such overlap can skew the causal estimate in the direction of the observational association, especially when the genetic instruments have weak strength, leading to model overfitting (113). In our MR study, given the minimal sample overlap and the good strength of the selected instruments, the bias introduced due to sample overlap is likely negligible.

#### *6.3.2.6 Population structure bias and generalizability*

Leveraging data from diverse population ancestries in MR studies can bring about confounding effects, often termed as population structure bias (230). In our MR analyses, we solely sourced genetic data from European populations, thereby curtailing potential bias from mixed population structures. Nonetheless, this singular ancestry focus might constrain the broader applicability of our findings to populations of different ethnic backgrounds.

#### *6.3.2.7 Nonlinearity*

The two-sample MR analysis operates under the assumption of a linear relationship, limiting its ability to evaluate non-linear associations. This poses a constraint, especially when considering certain blood proteins that might exhibit a non-linear relationship with VTE risk.

### 6.3.3 Generalizability

Generalizability, also known as external validity, refers to the extent to which research findings from one group (or cohort) can be reliably extended or applied to other groups or settings. Our associations derived from Swedish cohorts might be particularly indicative for the generation starting in 1997, as this study population mirrors the general Swedish demographic in terms of age distribution, education, and body mass index (155, 215). Additionally, only a minimal set of individuals was excluded in our analyses. Yet, given this cohort's prevalent healthy lifestyle, it is uncertain if the observed associations hold true for other demographic groups, such as high-risk populations. Similarly, while

the UK Biobank offers a treasure trove of data for biomedical research, its participants, primarily healthier volunteers of White British origin, might not be entirely representative of the wider UK population. This is compounded by a noted nonresponse pattern tied to socioeconomic factors. Hence, extrapolating its findings to more diverse or broader populations necessitates prudence. As for our MR analysis, the findings have a restricted scope, mainly limited to European ancestry.

# 7 Conclusions

In summary, this thesis explored the interplay between modifiable risk factors, blood proteins, and the onset of VTE. Through cohort studies, we identified that obesity, especially abdominal obesity, physical inactivity, poor diet, and a high consumption of UPF were associated with a higher risk of developing VTE. Cigarette smoking had a positive association with VTE risk specifically in women. The MR analysis reinforced the causal links of obesity, lack of physical activity, and smoking with VTE. It further unveiled associations of short sleep durations, daytime napping, and insomnia with VTE. A variety of circulating proteins were found to have a potentially causal relationship with VTE risk, with vWF being as identically important as BMI as disease risk predictor. Proteins like annexin II and FXI were pivotal in bridging the associations between obesity, smoking, insomnia, and VTE.

# 8 Points of perspective

The findings from this PhD project offer significant clinical implications. Recognizing obesity, particularly abdominal obesity, physical inactivity, poor diet, and higher consumption of UPF as high-risk factors for VTE, can guide clinicians in early detection and prevention strategies. The sex-specific association of cigarette smoking with VTE in women may necessitate a gender-specific approach in health advisories. Furthermore, the discovery that vWF play an equivalently pivotal role in predicting VTE risk as BMI means that clinical tests assessing the protein levels could become routine in evaluating VTE susceptibility. These proteins, such as annexin II and FXI, can be potential therapeutic targets or diagnostic markers, especially in patients with the identified modifiable risk factors.

This thesis also has opened new avenues for future exploration in understanding VTE risk. For example, future studies should delve deeper into the mechanisms by which obesity, smoking, and sleep patterns impact the levels of key proteins like annexin II and FXI. The gender-specific association of smoking in women with VTE risk may suggest there might be underlying hormonal or genetic factors at play, meriting further investigation. Additionally, intervention studies are crucial to ascertain if modifying these risk factors can indeed change protein levels and, consequently, VTE risk. As our understanding of the relationship between modifiable risk factors, blood proteins, and VTE grows, there is an opportunity to develop targeted therapeutic interventions and predictive tools that can better address and prevent this condition.

# 9 Acknowledgements

I would like to take this opportunity to express my deepest gratitude to those who have supported and guided me throughout the journey of my 4-year PhD training.

First and foremost, I extend my heartfelt appreciation to my main supervisor, **Dr. Susanna C. Larsson**, whose unwavering support, expertise, and invaluable insights have been instrumental in shaping this research. Your dedication to excellence and patience in mentoring me have been truly remarkable. Thank you for guiding and accompanying me through almost all my wonderful six years at Karolinska Institutet and being a role model for my academic pursue. I also want to thank you for providing a constant source of encouragement, instilling in me the confidence to navigate the intricacies of this research with conviction and determination.

I would like to express my sincere gratitude to my co-supervisor, **Dr. Agneta Åkesson**, who has been a pillar of mental and encyclopedic support throughout this journey. I can always find the answer and direction after coming back from your office. Your empathetic approach, understanding, and consistent encouragement were more than just academic guidance; they were lifelines during the most challenging times. The personal interest you showed in my well-being and progress made a monumental difference in my ability to persevere and thrive. The cups of coffee, shared laughter, and your hugs will always be cherished memories from this journey.

Equally, my sincere thanks go to **Dr. Maria Bruzelius** for the unparalleled academic and professional insight you brought to this research. Your insightful understanding of VTE, meticulous attention to detail, and dedication to scholarly excellence have been nothing short of inspiring.

I am deeply grateful to my mentor, **Dr. Jonas F. Ludvigsson**, for your keen interest in the Chinese language and culture, which enriched our discussions and instilled pride in my heritage. Beyond cultural appreciation, your unparalleled guidance in my additional academic interest in gastrointestinal disease exemplifies a mentorship that seamlessly combines cultural respect and academic brilliance.

I would be remiss not to extend my heartfelt thanks to my incredible colleagues (**Mr. Fredrik Söderlund, Dr. Tessa Schillemans, Dr. Niclas Håkansson, Miss. Emilie Helte, Dr. Federica Laguzzi, Miss. Stephanie Pitt, Dr. Karin Leander, Dr. Alicja Wolk, Dr. Max Vikström,** and **Dr. Helen Håkansson)**, whose companionship has greatly enriched my academic journey. From lunches and casual chats to insightful discussions, your camaraderie has been a comforting constant. I deeply appreciate the countless times you lent a listening ear and the wonderful atmosphere of mutual respect and warmth you have cultivated, turning our workspace into more than just a place of research, but a haven of collaboration and friendship.

I wish to express my profound gratitude to my collaborators, **Dr. Xue Li**, **Dr. Jie Chen, Dr. Fengzhe Xu**, **Dr. Stephen Burgess, Dr. Amy Mason, Dr. Dipender Gill**, and **Dr. Scott M. Damrauer**. Our partnership on many studies has been both enlightening and enriching. Your expertise, insight, and dedication to our shared goals have greatly elevated the quality of our work. It has been an honor to work alongside someone of your caliber, and I deeply appreciate the synergy we have shared in our academic endeavors.

I also want to thank my close friends (**Ying Xiong, Bowen Tang, Jing Wu,** and **Yuying Li**) who provided camaraderie, encouragement, and a listening ear during the highs and lows of this academic endeavor. Your friendship has been a source of motivation and relief throughout this process.

I owe a special debt of gratitude to my roommate, **Sichao Li**, who has been so much more than just a person sharing a living space with me. Through late-night conversations, shared meals, and countless moments of laughter and understanding, you have transformed the shared space into a true home. Additionally, I want to thank my cherished cat, **Deng Deng**, for being my steadfast companion through the long nights and early mornings.

To my parents, **Zili Yuan** and **Sanyuan Wu**, words cannot adequately capture the depth of my gratitude. You have always instilled in me the values of hard work, perseverance, and the pursuit of knowledge. Your unwavering belief in my abilities, even when I doubted myself, has been the driving force behind every milestone I have achieved. The love you have shown, and the wisdom you have imparted have been my guiding lights throughout this journey. This accomplishment, while my own, is also a testament to the strong foundation you have built for me. Thank you for being my rock and my biggest cheerleaders.

Last but not least, I extend my heartfelt gratitude to **Karolinska Institutet** for providing an environment that fosters academic growth and innovation. In addition, I thank Sweden for embracing me with its rich culture and traditions, and for offering opportunities that have significantly shaped my academic journey.

感谢遇见的所有和所有的遇见。

# 10 References


1. Khan F, Tritschler T, Kahn SR, Rodger MA. Venous thromboembolism. Lancet. 2021;398(10294):64-77.

2. Tagalakis V, Patenaude V, Kahn SR, Suissa S. Incidence of and mortality from venous thromboembolism in a real-world population: the Q-VTE Study Cohort. Am J Med. 2013;126(9):832.e13-21.

3. Tsai AW, Cushman M, Rosamond WD, Heckbert SR, Polak JF, Folsom AR. Cardiovascular risk factors and venous thromboembolism incidence: the longitudinal investigation of thromboembolism etiology. Arch Intern Med. 2002;162(10):1182-9.

4. Crous-Bou M, Harrington LB, Kabrhel C. Environmental and Genetic Risk Factors Associated with Venous Thromboembolism. Semin Thromb Hemost. 2016;42(8):808-20.

5. Castellucci LA, Cameron C, Le Gal G, Rodger MA, Coyle D, Wells PS, et al. Efficacy and safety outcomes of oral anticoagulants and antiplatelet drugs in the secondary prevention of venous thromboembolism: systematic review and network meta-analysis. Bmj. 2013;347:f5133.

6. Mackman N, Bergmeier W, Stouffer GA, Weitz JI. Therapeutic strategies for thrombosis: new targets and approaches. Nat Rev Drug Discov. 2020;19(5):333-52.

7. Meissner F, Geddes-McAlister J, Mann M, Bantscheff M. The emerging role of mass spectrometry-based proteomics in drug discovery. Nat Rev Drug Discov. 2022;21(9):637-54.

8. Virchow R. Gesammelte abhandlungen zur wissenschaftlichen medizin: Meidinger; 1856.

9. Engelmann B, Massberg S. Thrombosis as an intravascular effector of innate immunity. Nat Rev Immunol. 2013;13(1):34-45.

10. Mackman N. New insights into the mechanisms of venous thrombosis. J Clin Invest. 2012;122(7):2331-6.

11. Segers K, Dahlbäck B, Nicolaes GA. Coagulation factor V and thrombophilia: background and mechanisms. Thromb Haemost. 2007;98(3):530-42.

12. Bezemer ID, Bare LA, Doggen CJ, Arellano AR, Tong C, Rowland CM, et al. Gene variants associated with deep vein thrombosis. Jama. 2008;299(11):1306-14.

13. Baylis RA, Smith NL, Klarin D, Fukaya E. Epidemiology and Genetics of Venous Thromboembolism and Chronic Venous Disease. Circ Res. 2021;128(12):1988-2002.

14. Lindström S, Wang L, Smith EN, Gordon W, van Hylckama Vlieg A, de Andrade M, et al. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. Blood. 2019;134(19):1645-57.

15. Ghouse J, Tragante V, Ahlberg G, Rand SA, Jespersen JB, Leinøe EB, et al. Genome-wide meta-analysis identifies 93 risk loci and enables risk prediction equivalent to monogenic forms of venous thromboembolism. Nat Genet. 2023;55(3):399-409.

16. Thibord F, Klarin D, Brody JA, Chen M-H, Levin MG, Chasman DI, et al. Cross-Ancestry Investigation of Venous Thromboembolism Genomic Predictors. medRxiv. 2022.

17. Bell EJ, Lutsey PL, Basu S, Cushman M, Heckbert SR, Lloyd-Jones DM, et al. Lifetime Risk of Venous Thromboembolism in Two Cohort Studies. Am J Med. 2016;129(3):339.e19-26.

18. Heit JA. Epidemiology of venous thromboembolism. Nature Reviews Cardiology. 2015;12(8):464-74.

19. Heit JA, Petterson TM, Farmer SA, Bailey KR, Melton LJ, III. Trends in the Incidence of Deep Vein Thrombosis and Pulmonary Embolism: A 35-Year Population-Based Study. Blood. 2006;108(11):1488-.

20. Kyrle PA, Rosendaal FR, Eichinger S. Risk assessment for recurrent venous thrombosis. Lancet. 2010;376(9757):2032-9.

21. Verso M, Agnelli G, Ageno W, Imberti D, Moia M, Palareti G, et al. Long-term death and recurrence in patients with acute venous thromboembolism: the MASTER registry. Thromb Res. 2012;130(3):369-73.

22. Søgaard KK, Schmidt M, Pedersen L, Horváth-Puhó E, Sørensen HT. 30-year mortality after venous thromboembolism: a population-based cohort study. Circulation. 2014;130(10):829-36.

23. Barco S, Woersching AL, Spyropoulos AC, Piovella F, Mahan CE. European Union-28: An annualised cost-of-illness model for venous thromboembolism. Thromb Haemost. 2016;115(4):800-8.

24. Gussoni G, Foglia E, Frasson S, Casartelli L, Campanini M, Bonfanti M, et al. Real-world economic burden of venous thromboembolism and antithrombotic prophylaxis in medical inpatients. Thromb Res. 2013;131(1):17-23.

25. Wendelboe AM, McCumber M, Hylek EM, Buller H, Weitz JI, Raskob G. Global public awareness of venous thromboembolism. J Thromb Haemost. 2015;13(8):1365-71.

26. Raskob GE, Angchaisuksiri P, Blanco AN, Buller H, Gallus A, Hunt BJ, et al. Thrombosis: a major contributor to global disease burden. Arterioscler Thromb Vasc Biol. 2014;34(11):2363-71.

27. Siegal DM, Eikelboom JW, Lee SF, Rangarajan S, Bosch J, Zhu J, et al. Variations in incidence of venous thromboembolism in low-, middle-, and high-income countries. Cardiovasc Res. 2021;117(2):576-84.

28. Huang W, Goldberg RJ, Anderson FA, Kiefe CI, Spencer FA. Secular trends in occurrence of acute venous thromboembolism: the Worcester VTE study (1985-2009). Am J Med. 2014;127(9):829-39.e5.

29. Bikdeli B, Madhavan MV, Jimenez D, Chuich T, Dreyfus I, Driggin E, et al. COVID-19 and Thrombotic or Thromboembolic Disease: Implications for Prevention, Antithrombotic Therapy, and Follow-Up: JACC State-of-the-Art Review. J Am Coll Cardiol. 2020;75(23):2950-73.

30. Anderson FA, Jr., Wheeler HB, Goldberg RJ, Hosmer DW, Patwardhan NA, Jovanovic B, et al. A population-based perspective of the hospital incidence and case-

fatality rates of deep vein thrombosis and pulmonary embolism. The Worcester DVT Study. Arch Intern Med. 1991;151(5):933-8.

31. Silverstein MD, Heit JA, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ, 3rd. Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. Arch Intern Med. 1998;158(6):585-93.

32. Heit JA. Epidemiology of venous thromboembolism. Nat Rev Cardiol. 2015;12(8):464-74.

33. Zöller B, Li X, Sundquist J, Sundquist K. Age- and gender-specific familial risks for venous thromboembolism: a nationwide epidemiological study based on hospitalizations in Sweden. Circulation. 2011;124(9):1012-20.

34. Mackman N. Triggers, targets and treatments for thrombosis. Nature. 2008;451(7181):914-8.

35. Lutsey PL, Zakai NA. Epidemiology and prevention of venous thromboembolism. Nat Rev Cardiol. 2023;20(4):248-62.

36. Konstantinides SV, Meyer G, Becattini C, Bueno H, Geersing GJ, Harjola VP, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS): The Task Force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). Eur Respir J. 2019;54(3).

37. Horsted F, West J, Grainge MJ. Risk of Venous Thromboembolism in Patients with Cancer: A Systematic Review and Meta-Analysis. PLOS Medicine. 2012;9(7):e1001275.

38. Prandoni P, Falanga A, Piccioli A. Cancer and venous thromboembolism. Lancet Oncol. 2005;6(6):401-10.

39. Farge D, Bounameaux H, Brenner B, Cajfinger F, Debourdeau P, Khorana AA, et al. International clinical practice guidelines including guidance for direct oral anticoagulants in the treatment and prophylaxis of venous thromboembolism in patients with cancer. Lancet Oncol. 2016;17(10):e452-e66.

40. Mulder FI, Bosch FTM, Young AM, Marshall A, McBane RD, Zemla TJ, et al. Direct oral anticoagulants for cancer-associated venous thromboembolism: a systematic review and meta-analysis. Blood. 2020;136(12):1433-41.

41. Graham WC, Flanigan DC. Venous thromboembolism following arthroscopic knee surgery: a current concepts review of incidence, prophylaxis, and preoperative risk assessment. Sports Med. 2014;44(3):331-43.

42. Ho KM, Bham E, Pavey W. Incidence of Venous Thromboembolism and Benefits and Risks of Thromboprophylaxis After Cardiac Surgery: A Systematic Review and Meta-Analysis. J Am Heart Assoc. 2015;4(10):e002652.

43. Murphy PB, Vogt KN, Lau BD, Aboagye J, Parry NG, Streiff MB, et al. Venous Thromboembolism Prevention in Emergency General Surgery: A Review. JAMA Surg. 2018;153(5):479-86.

44. Vinogradova Y, Coupland C, Hippisley-Cox J. Use of hormone replacement therapy and risk of venous thromboembolism: nested case-control studies using the QResearch and CPRD databases. Bmj. 2019;364:k4810.

45. Ayodele OA, Cabral HJ, McManus DD, Jick SS. Glucocorticoids and Risk of Venous Thromboembolism in Asthma Patients Aged 20-59 Years in the United Kingdom's CPRD 1995-2015. Clin Epidemiol. 2022;14:83-93.

46. Kunutsor SK, Seidu S, Khunti K. Statins and primary prevention of venous thromboembolism: a systematic review and meta-analysis. Lancet Haematol. 2017;4(2):e83-e93.

47. Sultan AA, West J, Tata LJ, Fleming KM, Nelson-Piercy C, Grainge MJ. Risk of first venous thromboembolism in and around pregnancy: a population-based cohort study. Br J Haematol. 2012;156(3):366-73.

48. Rasmussen LD, Dybdal M, Gerstoft J, Kronborg G, Larsen CS, Pedersen C, et al. HIV and risk of venous thromboembolism: a Danish nationwide population-based cohort study. HIV Med. 2011;12(4):202-10.

49. Yuan S, Bruzelius M, Larsson SC. Causal effect of renal function on venous thromboembolism: a two-sample Mendelian randomization investigation. J Thromb Thrombolysis. 2022;53(1):43-50.

50. Molander V, Bower H, Frisell T, Askling J. Risk of venous thromboembolism in rheumatoid arthritis, and its association with disease activity: a nationwide cohort study from Sweden. Ann Rheum Dis. 2021;80(2):169-75.

51. Kuenzig ME, Bitton A, Carroll MW, Kaplan GG, Otley AR, Singh H, et al. Inflammatory Bowel Disease Increases the Risk of Venous Thromboembolism in Children: A Population-Based Matched Cohort Study. J Crohns Colitis. 2021;15(12):2031-40.

52. Chung WS, Lin CL. Increased risks of venous thromboembolism in patients with psoriasis. A Nationwide Cohort Study. Thromb Haemost. 2017;117(8):1637-43.

53. Christensen S, Farkas DK, Pedersen L, Miret M, Christiansen CF, Sørensen HT. Multiple sclerosis and risk of venous thromboembolism: a population-based cohort study. Neuroepidemiology. 2012;38(2):76-83.

54. Bird ST, Hartzema AG, Brophy JM, Etminan M, Delaney JA. Risk of venous thromboembolism in women with polycystic ovary syndrome: a population-based matched cohort analysis. Cmaj. 2013;185(2):E115-20.

55. Stuijver DJ, van Zaane B, Feelders RA, Debeij J, Cannegieter SC, Hermus AR, et al. Incidence of venous thromboembolism in patients with Cushing's syndrome: a multicenter cohort study. J Clin Endocrinol Metab. 2011;96(11):3525-32.

56. Ageno W, Farjat A, Haas S, Weitz JI, Goldhaber SZ, Turpie AGG, et al. Provoked versus unprovoked venous thromboembolism: Findings from GARFIELD-VTE. Res Pract Thromb Haemost. 2021;5(2):326-41.

57. Ageno W, Becattini C, Brighton T, Selby R, Kamphuisen PW. Cardiovascular risk factors and venous thromboembolism: a meta-analysis. Circulation. 2008;117(1):93-102.

58. Rahmani J, Haghighian Roudsari A, Bawadi H, Thompson J, Khalooei Fard R, Clark C, et al. Relationship between body mass index, risk of venous thromboembolism and pulmonary embolism: A systematic review and dose-response meta-analysis of cohort studies among four million participants. Thromb Res. 2020;192:64-72.

59. Lindström S, Germain M, Crous-Bou M, Smith EN, Morange PE, van Hylckama Vlieg A, et al. Assessing the causal relationship between obesity and venous thromboembolism through a Mendelian Randomization study. Hum Genet. 2017;136(7):897-902.

60. Severinsen MT, Kristensen SR, Johnsen SP, Dethlefsen C, Tjønneland A, Overvad K. Anthropometry, body fat, and venous thromboembolism: a Danish follow-up study. Circulation. 2009;120(19):1850-7.

61. Horvei LD, Brækkan SK, Mathiesen EB, Njølstad I, Wilsgaard T, Hansen JB. Obesity measures and risk of venous thromboembolism and myocardial infarction. Eur J Epidemiol. 2014;29(11):821-30.

62. Allman-Farinelli MA. Obesity and venous thrombosis: a review. Semin Thromb Hemost. 2011;37(8):903-7.

63. Frischmuth T, Hindberg K, Aukrust P, Ueland T, Brækkan SK, Hansen JB, et al. Plasma Levels of Leptin and Risk of Future Incident Venous Thromboembolism. Thromb Haemost. 2021.

64. Ten Cate V, Koeck T, Prochaska J, Schulz A, Panova-Noeva M, Rapp S, et al. A targeted proteomics investigation of the obesity paradox in venous thromboembolism. Blood Adv. 2021;5(14):2909-18.

65. Cheng YJ, Liu ZH, Yao FJ, Zeng WT, Zheng DD, Dong YG, et al. Current and former smoking and risk for venous thromboembolism: a systematic review and meta-analysis. PLoS Med. 2013;10(9):e1001515.

66. Gregson J, Kaptoge S, Bolton T, Pennells L, Willeit P, Burgess S, et al. Cardiovascular Risk Factors Associated With Venous Thromboembolism. JAMA Cardiol. 2019;4(2):163-73.

67. Larsson SC, Mason AM, Bäck M, Klarin D, Damrauer SM, Michaëlsson K, et al. Genetic predisposition to smoking in relation to 14 cardiovascular diseases. Eur Heart J. 2020;41(35):3304-10.

68. Reitsma PH, Versteeg HH, Middeldorp S. Mechanistic view of risk factors for venous thromboembolism. Arterioscler Thromb Vasc Biol. 2012;32(3):563-8.

69. BK M, M C, I AN, MA A, WJ B, SK B, et al. Association of Traditional Cardiovascular Risk Factors With Venous Thromboembolism: An Individual Participant Data Meta-Analysis of Prospective Studies. Circulation. 2017;135(1).

70. Chen M, Ji M, Chen T, Hong X, Jia Y. Alcohol Consumption and Risk for Venous Thromboembolism: A Meta-Analysis of Prospective Studies. Front Nutr. 2020;7:32.

71. Johansson M, Johansson L, Wennberg M, Lind M. Alcohol Consumption and Risk of First-Time Venous Thromboembolism in Men and Women. Thromb Haemost. 2019;119(6):962-70.

72. Zöller B, Ji J, Sundquist J, Sundquist K. Alcohol use disorders are associated with venous thromboembolism. J Thromb Thrombolysis. 2015;40(2):167-73.

73. Hansen-Krone IJ, Brækkan SK, Enga KF, Wilsgaard T, Hansen JB. Alcohol consumption, types of alcoholic beverages and risk of venous thromboembolism - the Tromsø Study. Thromb Haemost. 2011;106(2):272-8.

74. Ding M, Bhupathiraju SN, Satija A, van Dam RM, Hu FB. Long-term coffee consumption and risk of cardiovascular disease: a systematic review and a dose-response meta-analysis of prospective cohort studies. Circulation. 2014;129(6):643-59.

75. Lippi G, Mattiuzzi C, Franchini M. Venous thromboembolism and coffee: critical review and meta-analysis. Ann Transl Med. 2015;3(11):152.

76. Enga KF, Braekkan SK, Hansen-Krone IJ, Wilsgaard T, Hansen JB. Coffee consumption and the risk of venous thromboembolism: the Tromsø study. J Thromb Haemost. 2011;9(7):1334-9.

77. Lutsey PL, Steffen LM, Virnig BA, Folsom AR. Diet and incident venous thromboembolism: the Iowa Women's Health Study. Am Heart J. 2009;157(6):1081-7.

78. Hansen-Krone IJ, Enga KF, Njølstad I, Hansen JB, Braekkan SK. Heart healthy diet and risk of myocardial infarction and venous thromboembolism. The Tromsø Study. Thromb Haemost. 2012;108(3):554-60.

79. Fitzgerald KC, Chiuve SE, Buring JE, Ridker PM, Glynn RJ. Comparison of associations of adherence to a Dietary Approaches to Stop Hypertension (DASH)-style diet with risks of cardiovascular disease and venous thromboembolism. J Thromb Haemost. 2012;10(2):189-98.

80. Yi SY, Steffen LM, Lutsey PL, Cushman M, Folsom AR. Contrasting Associations of Prudent and Western Dietary Patterns with Risk of Developing Venous Thromboembolism. Am J Med. 2021;134(6):763-8.e3.

81. Varraso R, Kabrhel C, Goldhaber SZ, Rimm EB, Camargo CA, Jr. Prospective study of diet and venous thromboembolism in US women and men. Am J Epidemiol. 2012;175(2):114-26.

82. Steffen LM, Folsom AR, Cushman M, Jacobs DR, Jr., Rosamond WD. Greater fish, fruit, and vegetable intakes are related to lower incidence of venous thromboembolism: the Longitudinal Investigation of Thromboembolism Etiology. Circulation. 2007;115(2):188-95.

83. Hansen-Krone IJ, Enga KF, Südduth-Klinger JM, Mathiesen EB, Njølstad I, Wilsgaard T, et al. High fish plus fish oil intake is associated with slightly reduced risk of venous thromboembolism: the Tromsø Study. J Nutr. 2014;144(6):861-7.

84. Isaksen T, Evensen LH, Johnsen SH, Jacobsen BK, Hindberg K, Brækkan SK, et al. Dietary intake of marine n-3 polyunsaturated fatty acids and future risk of venous thromboembolism. Res Pract Thromb Haemost. 2019;3(1):59-69.

85. Zhang Y, Ding J, Guo H, Liang J, Li Y. Associations of Fish and Omega-3 Fatty Acids Consumption With the Risk of Venous Thromboembolism. A Meta-Analysis of Prospective Cohort Studies. Front Nutr. 2020;7:614784.

86. Lippi G, Cervellin G, Mattiuzzi C. Red meat, processed meat and the risk of venous thromboembolism: friend or foe? Thromb Res. 2015;136(2):208-11.

87. Danin-Mankowitz H, Ugarph-Morawski A, Braunschweig F, Wändell P. The risk of venous thromboembolism and physical activity level, especially high level: a systematic review. J Thromb Thrombolysis. 2021;52(2):508-16.

88. Borch KH, Hansen-Krone I, Braekkan SK, Mathiesen EB, Njolstad I, Wilsgaard T, et al. Physical activity and risk of venous thromboembolism. The Tromso study. Haematologica. 2010;95(12):2088-94.

89. van Stralen KJ, Doggen CJ, Lumley T, Cushman M, Folsom AR, Psaty BM, et al. The relationship between exercise and risk of venous thrombosis in elderly people. J Am Geriatr Soc. 2008;56(3):517-22.

90. Armstrong ME, Green J, Reeves GK, Beral V, Cairns BJ. Frequent physical activity may not reduce vascular disease risk as much as moderate activity: large prospective study of women in the United Kingdom. Circulation. 2015;131(8):721-9.

91. Evensen LH, Brækkan SK, Hansen JB. Regular Physical Activity and Risk of Venous Thromboembolism. Semin Thromb Hemost. 2018;44(8):765-79.

92. Kunutsor SK, Mäkikallio TH, Seidu S, de Araújo CGS, Dey RS, Blom AW, et al. Physical activity and risk of venous thromboembolism: systematic review and meta-analysis of prospective cohort studies. Eur J Epidemiol. 2020;35(5):431-42.

93. MacDonald CJ, Madika AL, Lajous M, Canonico M, Fournier A, Boutron-Ruault MC. Association between cardiovascular risk-factors and venous thromboembolism in a large longitudinal study of French women. Thromb J. 2021;19(1):58.

94. Yuan S, Mason AM, Burgess S, Larsson SC. Differentiating Associations of Glycemic Traits with Atherosclerotic and Thrombotic Outcomes: Mendelian Randomization Investigation. Diabetes. 2022.

95. Chung WS, Chen YF, Lin CL, Chang SN, Hsu WH, Kao CH. Sleep disorders increase the risk of venous thromboembolism in individuals without sleep apnea: a nationwide population-based cohort study in Taiwan. Sleep Med. 2015;16(1):168-72.

96. Yuan S, Mason AM, Burgess S, Larsson SC. Genetic liability to insomnia in relation to cardiovascular diseases: a Mendelian randomisation study. Eur J Epidemiol. 2021;36(4):393-400.

97. Mohammed Y, Touw CE, Nemeth B, van Adrichem RA, Borchers CH, Rosendaal FR, et al. Targeted proteomics for evaluating risk of venous thrombosis following traumatic lower-leg injury or knee arthroscopy. J Thromb Haemost. 2022;20(3):684-99.

98. Weitz JI, Szekanecz Z, Charles-Schoeman C, Vranic I, Sahin B, Paciga SA, et al. Biomarkers to predict risk of venous thromboembolism in patients with rheumatoid arthritis receiving tofacitinib or tumour necrosis factor inhibitors. RMD Open. 2022;8(2).

99. Tsai AW, Cushman M, Rosamond WD, Heckbert SR, Tracy RP, Aleksic N, et al. Coagulation factors, inflammation markers, and venous thromboembolism: the longitudinal investigation of thromboembolism etiology (LITE). Am J Med. 2002;113(8):636-42.

100. Cushman M, O'Meara ES, Folsom AR, Heckbert SR. Coagulation factors IX through XIII and the risk of future venous thrombosis: the Longitudinal Investigation of Thromboembolism Etiology. Blood. 2009;114(14):2878-83.

101. Bruzelius M, Iglesias MJ, Hong MG, Sanchez-Rivera L, Gyorgy B, Souto JC, et al. PDGFB, a new candidate plasma biomarker for venous thromboembolism: results from the VEREMA affinity proteomics study. Blood. 2016;128(23):e59-e66.

102. Edvardsen MS, Hindberg K, Hansen ES, Morelli VM, Ueland T, Aukrust P, et al. Plasma levels of von Willebrand factor and future risk of incident venous thromboembolism. Blood Adv. 2021;5(1):224-32.

103. Fenyves BG, Mehta A, Kays KR, Beakes C, Margolin J, Goldberg MB, et al. Plasma P-selectin is an early marker of thromboembolism in COVID-19. Am J Hematol. 2021;96(12):E468-e71.

104. Jensen SB, Latysheva N, Hindberg K, Ueland T. Plasma lipopolysaccharide-binding protein is a biomarker for future venous thromboembolism: Results from discovery and validation studies. J Intern Med. 2022;292(3):523-35.

105. Grover SP, Snir O, Hindberg K, Englebert TM, Braekkan SK, Morelli VM, et al. High plasma levels of C1-inhibitor are associated with lower risk of future venous thromboembolism. J Thromb Haemost. 2023;21(7):1849-60.

106. Iglesias MJ, Sanchez-Rivera L, Ibrahim-Kosta M, Naudin C, Munsch G, Goumidi L, et al. Elevated plasma complement factor H related 5 protein is associated with venous thromboembolism. Nat Commun. 2023;14(1):3280.

107. Thienel M, Müller-Reif JB, Zhang Z, Ehreiser V, Huth J, Shchurovska K, et al. Immobility-associated thromboprotection is conserved across mammalian species from bear to human. Science. 2023;380(6641):178-87.

108. Edfors F, Iglesias MJ, Butler LM, Odeberg J. Proteomics in thrombosis research. Res Pract Thromb Haemost. 2022;6(3):e12706.

109. Burgess S, Thompson SG. Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation. London, UK: Chapman and Hall/CRC; 2015.

110. Yuan S, Merino J, Larsson SC. Causal factors underlying diabetes risk informed by Mendelian randomisation analysis: evidence, opportunities and challenges. Diabetologia. 2023.

111. Robinson PC, Choi HK, Do R, Merriman TR. Insight into rheumatological cause and effect through the use of Mendelian randomization. Nat Rev Rheumatol. 2016;12(8):486-96.

112. Zheng J, Frysz M, Kemp JP, Evans DM, Davey Smith G, Tobias JH. Use of Mendelian Randomization to Examine Causal Inference in Osteoporosis. Front Endocrinol (Lausanne). 2019;10:807.

113. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40(7):597-608.

114. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32(5):377-89.

115. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.

116. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50(5):693-8.

117. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181(4):251-60.

118. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.

119. Chen J, Wellens J, Kalla R, Fu T, Deng M, Zhang H, et al. Intake of ultra-processed foods is associated with an increased risk of Crohn's disease: a cross-sectional and prospective analysis of 187,154 participants in the UK Biobank. J Crohns Colitis. 2022.

120. Messerer M, Johansson SE, Wolk A. The validity of questionnaire-based micronutrient intake estimates is increased by including dietary supplement use in Swedish men. J Nutr. 2004;134(7):1800-5.

121. Orsini N, Bellocco R, Bottai M, Hagströmer M, Sjöström M, Pagano M, et al. Validity of self-reported total physical activity questionnaire among older women. Eur J Epidemiol. 2008;23(10):661-7.

122. Messerer M, Wolk A. Sensitivity and specificity of self-reported use of dietary supplements. Eur J Clin Nutr. 2004;58(12):1669-71.

123. Wolk A, Larsson SC, Johansson JE, Ekman P. Long-term fatty fish consumption and renal cell carcinoma incidence in women. Jama. 2006;296(11):1371-6.

124. Levitan EB, Westgren CW, Liu S, Wolk A. Reproducibility and validity of dietary glycemic index, dietary glycemic load, and total carbohydrate intake in 141 Swedish men. Am J Clin Nutr. 2007;85(2):548-53.

125. Rautiainen S, Serafini M, Morgenstern R, Prior RL, Wolk A. The validity and reproducibility of food-frequency questionnaire-based total antioxidant capacity estimates in Swedish women. Am J Clin Nutr. 2008;87(5):1247-53.

126. Wolk A, Ljung H, Vessby B, Hunter D, Willett WC. Effect of additional questions about fat on the validity of fat estimates from a food frequency questionnaire. Study Group of MRS SWEA. Eur J Clin Nutr. 1998;52(3):186-92.

127. Wolk A, Vessby B, Ljung H, Barrefors P. Evaluation of a biological marker of dairy fat intake. Am J Clin Nutr. 1998;68(2):291-5.

128. Julin B, Wolk A, Bergkvist L, Bottai M, Akesson A. Dietary cadmium exposure and risk of postmenopausal breast cancer: a population-based prospective cohort study. Cancer Res. 2012;72(6):1459-66.

129. Larsson SC, Wallin A, Wolk A. Dietary Approaches to Stop Hypertension Diet and Incidence of Stroke: Results From 2 Prospective Cohorts. Stroke. 2016;47(4):986-90.

130. Perez-Cornago A, Pollard Z, Young H, van Uden M, Andrews C, Piernas C, et al. Description of the updated nutrition calculation of the Oxford WebQ questionnaire

and comparison with the previous version among 207,144 participants in UK Biobank. Eur J Nutr. 2021;60(7):4019-30.

131. Greenwood DC, Hardie LJ, Frost GS, Alwan NA, Bradbury KE, Carter M, et al. Validation of the Oxford WebQ Online 24-Hour Dietary Questionnaire Using Biomarkers. Am J Epidemiol. 2019;188(10):1858-67.

132. Monteiro CA, Cannon G, Moubarac JC, Levy RB, Louzada MLC, Jaime PC. The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. Public Health Nutr. 2018;21(1):5-17.

133. Rauber F, da Costa Louzada ML, Steele EM, Millett C, Monteiro CA, Levy RB. Ultra-Processed Food Consumption and Chronic Non-Communicable Diseases-Related Dietary Nutrient Profile in the UK (2008-2014). Nutrients. 2018;10(5).

134. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011;11:450.

135. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political analysis. 2007;15(3):199-236.

136. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems. 2017;30.

137. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Advances in neural information processing systems. 2017;30.

138. Klarin D, Busenkell E, Judy R, Lynch J, Levin M, Haessler J, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. Nat Genet. 2019;51(11):1574-9.

139. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613(7944):508-18.

140. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat Genet. 2018;50(11):1514-23.

141. Klarin D, Emdin CA, Natarajan P, Conrad MF, Kathiresan S. Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk Factor. Circ Cardiovasc Genet. 2017;10(2).

142. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet. 2021;53(12):1712-21.

143. Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, et al. Genetic regulation of the human plasma proteome in 54,306 UK Biobank participants. bioRxiv. 2022.

144. Pietzner M, Wheeler E, Carrasco-Zanini J, Cortes A, Koprulu M, Wörheide MA, et al. Mapping the proteo-genomic convergence of human diseases. Science. 2021;374(6569):eabj1541.

145. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res. 2017;26(5):2333-55.

146. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology. 2017;28(1):30-42.

147. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383.

148. Cushman M, O'Meara ES, Heckbert SR, Zakai NA, Rosamond W, Folsom AR. Body size measures, hemostatic and inflammatory markers and risk of venous thrombosis: The Longitudinal Investigation of Thromboembolism Etiology. Thromb Res. 2016;144:127-32.

149. Borch KH, Braekkan SK, Mathiesen EB, Njølstad I, Wilsgaard T, Størmer J, et al. Anthropometric measures of obesity and risk of venous thromboembolism: the Tromso study. Arterioscler Thromb Vasc Biol. 2010;30(1):121-7.

150. Glise Sandblad K, Jern S, Aberg M, Robertson J, Toren K, Lindgren M, et al. Obesity in adolescent men increases the risk of venous thromboembolism in adult life. J Intern Med. 2020.

151. Larsson SC, Bäck M, Rees JMB, Mason AM, Burgess S. Body mass index and body composition in relation to 14 cardiovascular conditions in UK Biobank: a Mendelian randomization study. Eur Heart J. 2020;41(2):221-6.

152. Lindstrom S, Germain M, Crous-Bou M, Smith EN, Morange PE, van Hylckama Vlieg A, et al. Assessing the causal relationship between obesity and venous thromboembolism through a Mendelian Randomization study. Hum Genet. 2017;136(7):897-902.

153. Mahmoodi BK, Cushman M, Anne Næss I, Allison MA, Bos WJ, Brækkan SK, et al. Association of Traditional Cardiovascular Risk Factors With Venous Thromboembolism: An Individual Participant Data Meta-Analysis of Prospective Studies. Circulation. 2017;135(1):7-16.

154. Severinsen MT, Kristensen SR, Johnsen SP, Dethlefsen C, Tjonneland A, Overvad K. Smoking and venous thromboembolism: a Danish follow-up study. J Thromb Haemost. 2009;7(8):1297-303.

155. Akesson A, Larsson SC, Discacciati A, Wolk A. Low-risk diet and lifestyle habits in the primary prevention of myocardial infarction in men: a population-based prospective cohort study. J Am Coll Cardiol. 2014;64(13):1299-306.

156. Larsson SC, Burgess S, Mason AM, Michaëlsson K. Alcohol Consumption and Cardiovascular Disease: A Mendelian Randomization Study. Circ Genom Precis Med. 2020;13(3):e002814.

157. Srour B, Fezeu LK, Kesse-Guyot E, Allès B, Méjean C, Andrianasolo RM, et al. Ultra-processed food intake and risk of cardiovascular disease: prospective cohort study (NutriNet-Santé). Bmj. 2019;365:l1451.

158. Juul F, Vaidean G, Lin Y, Deierlein AL, Parekh N. Ultra-Processed Foods and Incident Cardiovascular Disease in the Framingham Offspring Study. J Am Coll Cardiol. 2021;77(12):1520-31.

159. Peng YH, Liao WC, Chung WS, Muo CH, Chu CC, Liu CJ, et al. Association between obstructive sleep apnea and deep vein thrombosis / pulmonary embolism: a population-based retrospective cohort study. Thromb Res. 2014;134(2):340-5.

160. Trzepizur W, Gervès-Pinquié C, Heudes B, Blanchard M, Meslier N, Jouvenot M, et al. Sleep Apnea and Incident Unprovoked Venous Thromboembolism: Data from the Pays de la Loire Sleep Cohort. Thromb Haemost. 2023;123(4):393-401.

161. Sabater-Lleal M, Huffman JE, de Vries PS, Marten J, Mastrangelo MA, Song C, et al. Genome-Wide Association Transethnic Meta-Analyses Identifies Novel Associations Regulating Coagulation Factor VIII and von Willebrand Factor Plasma Levels. Circulation. 2019;139(5):620-35.

162. Edvardsen MS, Hansen ES, Hindberg K, Morelli VM, Ueland T, Aukrust P, et al. Combined effects of plasma von Willebrand factor and platelet measures on the risk of incident venous thromboembolism. Blood. 2021;138(22):2269-77.

163. Calabrò P, Gragnano F, Golia E, Grove EL. von Willebrand Factor and Venous Thromboembolism: Pathogenic Link and Therapeutic Implications. Semin Thromb Hemost. 2018;44(3):249-60.

164. Lacroix-Desmazes S, Repessé Y, Kaveri SV, Dasgupta S. The role of VWF in the immunogenicity of FVIII. Thromb Res. 2008;122 Suppl 2:S3-6.

165. Meltzer ME, Lisman T, de Groot PG, Meijers JC, le Cessie S, Doggen CJ, et al. Venous thrombosis risk associated with plasma hypofibrinolysis is explained by elevated plasma levels of TAFI and PAI-1. Blood. 2010;116(1):113-21.

166. Frischmuth T, Hindberg K, Aukrust P, Ueland T, Braekkan SK, Hansen JB, et al. Elevated plasma levels of plasminogen activator inhibitor-1 are associated with risk of future incident venous thromboembolism. J Thromb Haemost. 2022;20(7):1618-26.

167. Wang J, Wang C, Chen N, Shu C, Guo X, He Y, et al. Association between the plasminogen activator inhibitor-1 4G/5G polymorphism and risk of venous thromboembolism: a meta-analysis. Thromb Res. 2014;134(6):1241-8.

168. Ridker PM, Hennekens CH, Lindpaintner K, Stampfer MJ, Miletich JP. Arterial and venous thrombosis is not associated with the 4G/5G polymorphism in the promoter of the plasminogen activator inhibitor gene in a large cohort of US men. Circulation. 1997;95(1):59-62.

169. Angus DC. Drotrecogin alfa (activated)...a sad final fizzle to a roller-coaster party. Crit Care. 2012;16(1):107.

170. Thibord F, Klarin D, Brody JA, Chen MH, Levin MG, Chasman DI, et al. Cross-Ancestry Investigation of Venous Thromboembolism Genomic Predictors. Circulation. 2022;146(16):1225-42.

171. Li H, Zhang Z, Qiu Y, Weng H, Yuan S, Zhang Y, et al. Proteome-wide mendelian randomization identifies causal plasma proteins in venous thromboembolism development. J Hum Genet. 2023.

172. Elvers M, Pozgaj R, Pleines I, May F, Kuijpers MJ, Heemskerk JM, et al. Platelet hyperreactivity and a prothrombotic phenotype in mice with a gain-of-function mutation in phospholipase Cgamma2. J Thromb Haemost. 2010;8(6):1353-63.

173. Gangaraju R, Song J, Kim SJ, Tashi T, Reeves BN, Sundar KM, et al. Thrombotic, inflammatory, and HIF-regulated genes and thrombosis risk in polycythemia vera and essential thrombocythemia. Blood Adv. 2020;4(6):1115-30.

174. Kreft IC, Huisman EJ, Cnossen MH, van Alphen FPJ, van der Zwaan C, van Leeuwen K, et al. Proteomic landscapes of inherited platelet disorders with different etiologies. J Thromb Haemost. 2023;21(2):359-72.e3.

175. Perrella G, Nagy M, Watson SP, Heemskerk JWM. Platelet GPVI (Glycoprotein VI) and Thrombotic Complications in the Venous System. Arterioscler Thromb Vasc Biol. 2021;41(11):2681-92.

176. Reilly MP, Sinha U, André P, Taylor SM, Pak Y, Deguzman FR, et al. PRT-060318, a novel Syk inhibitor, prevents heparin-induced thrombocytopenia and thrombosis in a transgenic mouse model. Blood. 2011;117(7):2241-6.

177. Alonso-Martínez JL, Urbieta-Echezarreta M, Anniccherico-Sánchez FJ, Abínzano-Guillén ML, Garcia-Sanchotena JL. N-terminal pro-B-type natriuretic peptide predicts the burden of pulmonary embolism. Am J Med Sci. 2009;337(2):88-92.

178. Santos-Martínez MJ, Medina C, Jurasz P, Radomski MW. Role of metalloproteinases in platelet function. Thromb Res. 2008;121(4):535-42.

179. Bergmeier W, Hynes RO. Extracellular matrix proteins in hemostasis and thrombosis. Cold Spring Harb Perspect Biol. 2012;4(2).

180. Muia J, Zhu J, Gupta G, Haberichter SL, Friedman KD, Feys HB, et al. Allosteric activation of ADAMTS13 by von Willebrand factor. Proc Natl Acad Sci U S A. 2014;111(52):18584-9.

181. Bouton MC, Boulaftali Y, Richard B, Arocas V, Michel JB, Jandrot-Perrus M. Emerging role of serpinE2/protease nexin-1 in hemostasis and vascular biology. Blood. 2012;119(11):2452-7.

182. Cañas F, Simonin L, Couturaud F, Renaudineau Y. Annexin A2 autoantibodies in thrombosis and autoimmune diseases. Thromb Res. 2015;135(2):226-30.

183. Senchenkova EY, Komoto S, Russell J, Almeida-Paula LD, Yan LS, Zhang S, et al. Interleukin-6 mediates the platelet abnormalities and thrombogenesis associated with experimental colitis. Am J Pathol. 2013;183(1):173-81.

184. Grover SP, Kawano T, Wan J, Tanratana P, Polai Z, Shim YJ, et al. C1 inhibitor deficiency enhances contact pathway-mediated activation of coagulation and venous thrombosis. Blood. 2023;141(19):2390-401.

185. Grünbacher G, Weger W, Marx-Neuhold E, Pilger E, Köppel H, Wascher T, et al. The fibrinogen gamma (FGG) 10034C>T polymorphism is associated with venous thrombosis. Thromb Res. 2007;121(1):33-6.

186. Smith CW, Raslan Z, Parfitt L, Khan AO, Patel P, Senis YA, et al. TREM-like transcript 1: a more sensitive marker of platelet activation than P-selectin in humans and mice. Blood Adv. 2018;2(16):2072-8.

187. Zhou Y, Abraham S, Andre P, Edelstein LC, Shaw CA, Dangelmaier CA, et al. Anti-miR-148a regulates platelet Fc γ RIIA signaling and decreases thrombosis in vivo in mice. Blood. 2015;126(26):2871-81.

188. Alshbool FZ, Karim ZA, Vemana HP, Conlon C, Lin OA, Khasawneh FT. The regulator of G-protein signaling 18 regulates platelet aggregation, hemostasis and thrombosis. Biochem Biophys Res Commun. 2015;462(4):378-82.

189. Miao S, Zhang Q, Ding W, Hou B, Su Z, Li M, et al. Platelet Internalization Mediates Ferroptosis in Myocardial Infarction. Arterioscler Thromb Vasc Biol. 2023;43(2):218-30.

190. Debette S, Visvikis-Siest S, Chen MH, Ndiaye NC, Song C, Destefano A, et al. Identification of cis- and trans-acting genetic variants explaining up to half the variation in circulating vascular endothelial growth factor levels. Circ Res. 2011;109(5):554-63.

191. Le Goff C, Cormier-Daire V. The ADAMTS(L) family and human genetic disorders. Hum Mol Genet. 2011;20(R2):R163-7.

192. Nishida M, Moriyama T, Sugita Y, Yamauchi-Takihara K. Abdominal obesity exhibits distinct effect on inflammatory and anti-inflammatory proteins in apparently healthy Japanese men. Cardiovasc Diabetol. 2007;6:27.

193. Yuan S, Li X, Morange PE, Bruzelius M, Larsson SC, On Behalf Of The Invent C. Plasma Phospholipid Fatty Acids and Risk of Venous Thromboembolism: Mendelian Randomization Investigation. Nutrients. 2022;14(16).

194. Yuan S, Carter P, Bruzelius M, Vithayathil M, Kar S, Mason AM, et al. Effects of tumour necrosis factor on cardiovascular disease and cancer: A two-sample Mendelian randomization study. EBioMedicine. 2020;59:102956.

195. Yuan S, Bruzelius M, Damrauer SM, Håkansson N, Wolk A, Åkesson A, et al. Anti-inflammatory diet and venous thromboembolism: Two prospective cohort studies. Nutr Metab Cardiovasc Dis. 2021;31(10):2831-8.

196. Blokhin IO, Lentz SR. Mechanisms of thrombosis in obesity. Curr Opin Hematol. 2013;20(5):437-44.

197. Morange PE, Alessi MC. Thrombosis in central obesity and metabolic syndrome: mechanisms and epidemiology. Thromb Haemost. 2013;110(4):669-80.

198. Ennour-Idrissi K, Maunsell E, Diorio C. Effect of physical activity on sex hormones in women: a systematic review and meta-analysis of randomized controlled trials. Breast Cancer Res. 2015;17(1):139.

199. Beslay M, Srour B, Méjean C, Allès B, Fiolet T, Debras C, et al. Ultra-processed food intake in association with BMI change and risk of overweight and obesity: A prospective analysis of the French NutriNet-Santé cohort. PLoS Med. 2020;17(8):e1003256.

200. Rauber F, Chang K, Vamos EP, da Costa Louzada ML, Monteiro CA, Millett C, 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.

201. Silva Dos Santos F, Costa Mintem G, de Oliveira IO, Horta BL, Ramos E, Lopes C, et al. Consumption of ultra-processed foods and interleukin-6 in two cohorts from high- and middle-income countries. Br J Nutr. 2022:1-28.

202. Lane MM, Lotfaliany M, Forbes M, Loughman A, Rocks T, O'Neil A, et al. Higher Ultra-Processed Food Consumption Is Associated with Greater High-Sensitivity C-Reactive Protein Concentration in Adults: Cross-Sectional Results from the Melbourne Collaborative Cohort Study. Nutrients. 2022;14(16).

203. Evensen LH, Folsom AR, Pankow JS, Hansen JB, Allison MA, Cushman M, et al. Hemostatic factors, inflammatory markers, and risk of incident venous thromboembolism: The Multi-Ethnic Study of Atherosclerosis. J Thromb Haemost. 2021;19(7):1718-28.

204. Rey-García J, Donat-Vargas C, Sandoval-Insausti H, Bayan-Bravo A, Moreno-Franco B, Banegas JR, et al. Ultra-Processed Food Consumption is Associated with Renal Function Decline in Older Adults: A Prospective Cohort Study. Nutrients. 2021;13(2).

205. Bonaccio M, Di Castelnuovo A, Ruggiero E, Costanzo S, Grosso G, De Curtis A, et al. Joint association of food nutritional profile by Nutri-Score front-of-pack label and ultra-processed food intake with mortality: Moli-sani prospective cohort study. Bmj. 2022;378:e070688.

206. Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ. Modern Epidemiology. 4th ed. Philadelphia, PA: Wolters Kluwer / Lippincott Williams & Wilkins; 2021.

207. Kirkwood BR, Sterne JAC. Essential Medical Statistics. 2nd Edition ed. Hoboken: Wiley; 2003.

208. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31(4):337-50.

209. Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. Int J Epidemiol. 2021;50(2):620-32.

210. Hosoi T, Yamaguchi R, Noji K, Matsuo S, Baba S, Toyoda K, et al. Flurbiprofen ameliorated obesity by attenuating leptin resistance induced by endoplasmic reticulum stress. EMBO Mol Med. 2014;6(3):335-46.

211. Meaidi A, Mascolo A, Sessa M, Toft-Petersen AP, Skals R, Gerds TA, et al. Venous thromboembolism with use of hormonal contraception and non-steroidal anti-inflammatory drugs: nationwide cohort study. Bmj. 2023;382:e074450.

212. Hansson LM, Galanti MR. Diet-associated risks of disease and self-reported food consumption: how shall we treat partial nonresponse in a food frequency questionnaire? Nutr Cancer. 2000;36(1):1-6.

213. Di Giuseppe D, Alfredsson L, Bottai M, Askling J, Wolk A. Long term alcohol intake and risk of rheumatoid arthritis in women: a population based cohort study. Bmj. 2012;345:e4230.

214. Mignogna G, Carey CE, Wedow R, Baya N, Cordioli M, Pirastu N, et al. Patterns of item nonresponse behaviour to survey questionnaires are systematic and associated with genetic loci. Nat Hum Behav. 2023;7(8):1371-87.

215. Harris H, Håkansson N, Olofsson C, Stackelberg O, Julin B, Åkesson A, et al. The Swedish mammography cohort and the cohort of Swedish men: study design and

characteristics of 2 population-based longitudinal cohorts. OA Epidemiology. 2013;1(2):16.

216. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol. 2017;186(9):1026-34.

217. Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004;19(8):751-60.

218. Larsson SC, Bergkvist L, Wolk A. Processed meat consumption, dietary nitrosamines and stomach cancer risk in a cohort of Swedish women. Int J Cancer. 2006;119(4):915-9.

219. Byberg L, Bellavia A, Larsson SC, Orsini N, Wolk A, Michaëlsson K. Mediterranean Diet and Hip Fracture in Swedish Men and Women. J Bone Miner Res. 2016;31(12):2098-105.

220. Blair A, Stewart P, Lubin JH, Forastiere F. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. Am J Ind Med. 2007;50(3):199-207.

221. Cui M, Cheng C, Zhang L. High-throughput proteomics: a methodological mini-review. Lab Invest. 2022;102(11):1170-81.

222. Commission TA. Improving data quality in the NHS Annual report on the PbR assurance programme Health 2010.; 2010.

223. Woodfield R, Sudlow CL. Accuracy of Patient Self-Report of Stroke: A Systematic Review from the UK Biobank Stroke Outcomes Group. PLoS One. 2015;10(9):e0137538.

224. Wilkinson T, Schnier C, Bush K, Rannikmäe K, Henshall DE, Lerpiniere C, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. Eur J Epidemiol. 2019;34(6):557-65.

225. Bassett E, Broadbent J, Gill D, Burgess S, Mason AM. Inconsistency in UK Biobank Event Definitions From Different Data Sources and Its Impact on Bias and Generalizability: A Case Study of Venous Thromboembolism. Am J Epidemiol. 2023.

226. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet. 2020;52(10):1122-31.

227. Chen J, Xu F, Ruan X, Sun J, Zhang Y, Zhang H, et al. Therapeutic targets for inflammatory bowel disease: proteome-wide Mendelian randomization and colocalization analyses. EBioMedicine. 2023;89:104494.

228. Burgess S, Swanson SA, Labrecque JA. Are Mendelian randomization investigations immune from bias due to reverse causation? Eur J Epidemiol. 2021;36(3):253-7.

229. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. Epidemiology. 2014;25(3):427-35.

230. Brumpton B, Sanderson E, Heilbron K, Hartwig FP, Harrison S, Vie G, et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. Nat Commun. 2020;11(1):3519.

# 11 Appendices

**Appendix Table 1.** Included proteins in the study IV.

| Panel | Label Assay UniProt | Long name |
|---|---|---|
| CVD2 | BMP-6 (P22004) | Bone morphogenetic protein 6 |
| CVD2 | ANGPT1 (Q15389) | Angiopoietin-1 |
| CVD2 | ADM (P35318) | ADM |
| CVD2 | CD40-L (P29965) | CD40 ligand |
| CVD2 | SLAMF7 (Q9NQ25) | SLAM family member 7 |
| CVD2 | PGF (P49763) | Placenta growth factor |
| CVD2 | ADAM-TS13 (Q76LX8) | A disintegrin and metalloproteinase with thrombospondin motifs 13 |
| CVD2 | BOC (Q9BWV1) | Brother of CDO |
| CVD2 | IL-4RA (P24394) | Interleukin-4 receptor subunit alpha |
| CVD2 | SRC (P12931) | Proto-oncogene tyrosine-protein kinase Src |
| CVD2 | IL-1ra (P18510) | Interleukin-1 receptor antagonist protein |
| CVD2 | IL6 (P05231) | Interleukin-6 |
| CVD2 | TNFRSF10A (O00220) | Tumor necrosis factor receptor superfamily member 10A |
| CVD2 | STK4 (Q13043) | Serine/threonine-protein kinase 4 |
| CVD2 | IDUA (P35475) | Alpha-L-iduronidase |
| CVD2 | TNFRSF11A (Q9Y6Q6) | Tumor necrosis factor receptor superfamily member 11A |
| CVD2 | PAR-1 (P25116) | Proteinase-activated receptor 1 |
| CVD2 | TRAIL-R2 (O14763) | TNF-related apoptosis-inducing ligand receptor 2 |
| CVD2 | PRSS27 (Q9BQR3) | Serine protease 27 |
| CVD2 | TIE2 (Q02763) | Angiopoietin-1 receptor |
| CVD2 | TF (P13726) | Tissue factor |
| CVD2 | IL1RL2 (Q9HB29) | Interleukin-1 receptor-like 2 |
| CVD2 | PDGF subunit B (P01127) | Platelet-derived growth factor subunit B |
| CVD2 | IL-27 (Q8NEV9,Q14213) | Interleukin-27 |
| CVD2 | IL-17D (Q8TAD2) | Interleukin-17D |
| CVD2 | CXCL1 (P09341) | C-X-C motif chemokine 1 |
| CVD2 | LOX-1 (P78380) | Lectin-like oxidized LDL receptor 1 |
| CVD2 | Gal-9 (O00182) | Galectin-9 |
| CVD2 | GIF (P27352) | Gastric intrinsic factor |
| CVD2 | SCF (P21583) | Stem cell factor |
| CVD2 | IL18 (Q14116) | Interleukin-18 |
| CVD2 | FGF-21 (Q9NSA1) | Fibroblast growth factor 21 |
| CVD2 | PIgR (P01833) | Polymeric immunoglobulin receptor |
| CVD2 | RAGE (Q15109) | Receptor for advanced glycosylation end products |
| CVD2 | SOD2 (P04179) | Superoxide dismutase [Mn], mitochondrial |
| CVD2 | CTRC (Q99895) | Chymotrypsin C |
| CVD2 | FGF-23 (Q9GZV9) | Fibroblast growth factor 23 |
| CVD2 | SPON2 (Q9BUD6) | Spondin-2 |
| CVD2 | GH (P01241) | Growth hormone |
| CVD2 | FS (P19883) | Follistatin |
| CVD2 | GLO1 (Q04760) | Lactoylglutathione lyase |
| CVD2 | CD84 (Q9UIB8) | SLAM family member 5 |
| CVD2 | SERPINA12 (Q8IW75) | Serpin A12 |
| CVD2 | REN (P00797) | Renin |
| CVD2 | DECR1 (Q16698) | 2,4-dienoyl-CoA reductase, mitochondrial |
| CVD2 | MERTK (Q12866) | Tyrosine-protein kinase Mer |
| CVD2 | KIM1 (Q96D42) | Kidney Injury Molecule |
| CVD2 | THBS2 (P35442) | Thrombospondin-2 |
| CVD2 | TM (P07204) | Thrombomodulin |
| CVD2 | VSIG2 (Q96IQ7) | V-set and immunoglobulin domain-containing protein 2 |
| CVD2 | AMBP (P02760) | Protein AMBP |
| CVD2 | PRELP (P51888) | Prolargin |
| CVD2 | HO-1 (P09601) | Heme oxygenase 1 |
| CVD2 | XCL1 (P47992) | Lymphotactin |

| | | |
|---|---|---|
| CVD2 | IL16 (Q14005) | Pro-interleukin-16 |
| CVD2 | SORT1 (Q99523) | Sortilin |
| CVD2 | CEACAM8 (P31997) | Carcinoembryonic antigenrelated cell adhesion molecule 8 |
| CVD2 | PTX3 (P26022) | Pentraxin-related protein PTX3 |
| CVD2 | PSGL-1 (Q14242) | P-selectin glycoprotein ligand 1 |
| CVD2 | CCL17 (Q92583) | C-C motif chemokine 17 |
| CVD2 | CCL3 (P10147) | C-C motif chemokine 3 |
| CVD2 | MMP7 (P09237) | Matrix metalloproteinase-7 |
| CVD2 | IgG Fc receptor II-b (P31994) | Low affinity immunoglobulin gamma Fc region receptor II-b |
| CVD2 | ITGB1BP2 (Q9UKP3) | Melusin |
| CVD2 | DCN (P07585) | Decorin |
| CVD2 | Dkk-1 (O94907) | Dickkopf-related protein 1 |
| CVD2 | LPL (P06858) | Lipoprotein lipase |
| CVD2 | PRSS8 (Q16651) | Prostasin |
| CVD2 | AGRP (O00253) | Agouti-related protein |
| CVD2 | HB-EGF (Q99075) | Proheparin-binding EGF-like growth factor |
| CVD2 | GDF-2 (Q9UK05) | Growth/differentiation factor 2 |
| CVD2 | FABP2 (P12104) | Fatty acid-binding protein, intestinal |
| CVD2 | THPO (P40225) | Thrombopoietin |
| CVD2 | MARCO (Q9UEW3) | Macrophage receptor MARCO |
| CVD2 | GT (P51161) | Gastrotropin |
| CVD2 | BNP (P16860) | Natriuretic peptides B |
| CVD2 | MMP12 (P39900) | Matrix metalloproteinase-12 |
| CVD2 | ACE2 (Q9BYF1) | Angiotensin-converting enzyme 2 |
| CVD2 | PD-L2 (Q9BQ51) | Programmed cell death 1 ligand 2 |
| CVD2 | CTSL1 (P07711) | Cathepsin L1 |
| CVD2 | hOSCAR (Q8IYS5) | Osteoclast-associated immunoglobulin-like receptor |
| CVD2 | TNFRSF13B (O14836) | Tumor necrosis factor receptor superfamily member 13B |
| CVD2 | TGM2 (P21980) | Protein-glutamine gamma-glutamyltransferase 2 |
| CVD2 | LEP (P41159) | Leptin |
| CVD2 | CA5A (P35218) | Carbonic anhydrase 5A, mitochondrial |
| CVD2 | HSP 27 (P04792) | Heat shock 27 kDa protein |
| CVD2 | CD4 (P01730) | T-cell surface glycoprotein CD4 |
| CVD2 | NEMO (Q9Y6K9) | NF-kappa-B essential modulator |
| CVD2 | VEGFD (O43915) | Vascular endothelial growth factor D |
| CVD2 | HAOX1 (Q9UJM8) | Hydroxyacid oxidase 1 |
| CVD3 | MEPE (Q9NQ76) | Matrix extracellular phosphoglycoprotein |
| CVD3 | TNFRSF14 (Q92956) | Tumor necrosis factor receptor superfamily member 14 |
| CVD3 | LDL receptor (P01130) | Low-density lipoprotein receptor |
| CVD3 | ITGB2 (P05107) | Integrin beta-2 |
| CVD3 | IL-17RA (Q96F46) | Interleukin-17 receptor A |
| CVD3 | TNF-R2 (P20333) | Tumor necrosis factor receptor 2 |
| CVD3 | MMP-9 (P14780) | Matrix metalloproteinase-9 |
| CVD3 | EPHB4 (P54760) | Ephrin type-B receptor 4 |
| CVD3 | IL2-RA (P01589) | Interleukin-2 receptor subunit alpha |
| CVD3 | OPG (O00300) | Osteoprotegerin |
| CVD3 | ALCAM (Q13740) | CD166 antigen |
| CVD3 | TFF3 (Q07654) | Trefoil factor 3 |
| CVD3 | SELP (P16109) | P-selectin |
| CVD3 | CSTB (P04080) | Cystatin-B |
| CVD3 | MCP-1 (P13500) | Monocyte chemotactic protein 1 |
| CVD3 | CD163 (Q86VB7) | Scavenger receptor cysteine-rich type 1 protein M130 |
| CVD3 | Gal-3 (P17931) | Galectin-3 |
| CVD3 | GRN (P28799) | Granulins |
| CVD3 | BLM hydrolase (Q13867) | Bleomycin hydrolase |
| CVD3 | PLC (P98160) | Perlecan |
| CVD3 | LTBR (P36941) | Lymphotoxin-beta receptor |
| CVD3 | Notch 3 (Q9UM47) | Neurogenic locus notch homolog protein 3 |
| CVD3 | TIMP4 (Q99727) | Metalloproteinase inhibitor 4 |
| CVD3 | CNTN1 (Q12860) | Contactin-1 |
| CVD3 | CDH5 (P33151) | Cadherin-5 |

| | | |
|---|---|---|
| CVD3 | TLT-2 (Q5T2D2) | Trem-like transcript 2 protein |
| CVD3 | FABP4 (P15090) | Fatty acid-binding protein, adipocyte |
| CVD3 | TFPI (P10646) | Tissue factor pathway inhibitor |
| CVD3 | PAI (P05121) | Plasminogen activator inhibitor 1 |
| CVD3 | CCL24 (O00175) | C-C motif chemokine 24 |
| CVD3 | TR (P02786) | Transferrin receptor protein 1 |
| CVD3 | TNFRSF10C (O14798) | Tumor necrosis factor receptor superfamily member 10C |
| CVD3 | GDF-15 (Q99988) | Growth/differentiation factor 15 |
| CVD3 | SELE (P16581) | E-selectin |
| CVD3 | AZU1 (P20160) | Azurocidin |
| CVD3 | DLK-1 (P80370) | Protein delta homolog 1 |
| CVD3 | MPO (P05164) | Myeloperoxidase |
| CVD3 | CXCL16 (Q9H2A7) | C-X-C motif chemokine 16 |
| CVD3 | IL-6RA (P08887) | Interleukin-6 receptor subunit alpha |
| CVD3 | RETN (Q9HD89) | Resistin |
| CVD3 | IGFBP-1 (P08833) | Insulin-like growth factor-binding protein 1 |
| CVD3 | CHIT1 (Q13231) | Chitotriosidase-1 |
| CVD3 | TR-AP (P13686) | Tartrate-resistant acid phosphatase type 5 |
| CVD3 | PSP-D (P35247) | Pulmonary surfactant-associated protein D |
| CVD3 | PI3 (P19957) | Elafin |
| CVD3 | Ep-CAM (P16422) | Epithelial cell adhesion molecule |
| CVD3 | AP-N (P15144) | Aminopeptidase N |
| CVD3 | AXL (P30530) | Tyrosine-protein kinase receptor UFO |
| CVD3 | IL-1RT1 (P14778) | Interleukin-1 receptor type 1 |
| CVD3 | MMP-2 (P08253) | Matrix metalloproteinase-2 |
| CVD3 | FAS (P25445) | Tumor necrosis factor receptor superfamily member 6 |
| CVD3 | MB (P02144) | Myoglobin |
| CVD3 | TNFSF13B (Q9Y275) | Tumor necrosis factor ligand superfamily member 13B |
| CVD3 | PRTN3 (P24158) | Myeloblastin |
| CVD3 | PCSK9 (Q8NBP7) | Proprotein convertase subtilisin/kexin type 9 |
| CVD3 | U-PAR (Q03405) | Urokinase plasminogen activator surface receptor |
| CVD3 | OPN (P10451) | Osteopontin |
| CVD3 | CTSD (P07339) | Cathepsin D |
| CVD3 | PGLYRP1 (O75594) | Peptidoglycan recognition protein 1 |
| CVD3 | CPA1 (P15085) | Carboxypeptidase A1 |
| CVD3 | JAM-A (Q9Y624) | Junctional adhesion molecule A |
| CVD3 | Gal-4 (P56470) | Galectin-4 |
| CVD3 | IL-1RT2 (P27930) | Interleukin-1 receptor type 2 |
| CVD3 | SHPS-1 (P78324) | Tyrosine-protein phosphatase non-receptor type substrate 1 |
| CVD3 | CCL15 (Q16663) | C-C motif chemokine 15 |
| CVD3 | CASP-3 (P42574) | Caspase-3 |
| CVD3 | uPA (P00749) | Urokinase-type plasminogen activator |
| CVD3 | CPB1 (P15086) | Carboxypeptidase B |
| CVD3 | CHI3L1 (P36222) | Chitinase-3-like protein 1 |
| CVD3 | ST2 (Q01638) | ST2 protein |
| CVD3 | t-PA (P00750) | Tissue-type plasminogen activator |
| CVD3 | SCGB3A2 (Q96PL1) | Secretoglobin family 3A member 2 |
| CVD3 | EGFR (P00533) | Epidermal growth factor receptor |
| CVD3 | IGFBP-7 (Q16270) | Insulin-like growth factor-binding protein 7 |
| CVD3 | CD93 (Q9NPY3) | Complement component C1q receptor |
| CVD3 | IL-18BP (O95998) | Interleukin-18-binding protein |
| CVD3 | COL1A1 (P02452) | Collagen alpha-1 (I) chain |
| CVD3 | PON3 (Q15166) | Paraoxonase |
| CVD3 | CTSZ (Q9UBR2) | Cathepsin Z |
| CVD3 | MMP-3 (P08254) | Matrix metalloproteinase-3 |
| CVD3 | RARRES2 (Q99969) | Retinoic acid receptor responder protein 2 |
| CVD3 | ICAM-2 (P13598) | Intercellular adhesion molecule 2 |
| CVD3 | KLK6 (Q92876) | Kallikrein-6 |
| CVD3 | PDGF subunit A (P04085) | Platelet-derived growth factor subunit A |
| CVD3 | TNF-R1 (P19438) | Tumor necrosis factor receptor 1 |
| CVD3 | IGFBP-2 (P18065) | Insulin-like growth factor-binding protein 2 |
| CVD3 | vWF (P04275) | von Willebrand factor |

| | | |
|---|---|---|
| CVD3 | PECAM-1 (P16284) | Platelet endothelial cell adhesion molecule |
| CVD3 | CCL16 (O15467) | C-C motif chemokine 16 |
| Metab | CLMP (Q9H6B4) | CXADR-like membrane protein |
| Metab | LRIG1 (Q96JA1) | Leucine-rich repeats and immunoglobulin-like domains protein 1 |
| Metab | NPTXR (O95502) | Neuronal pentraxin receptor |
| Metab | THOP1 (P52888) | Thimet oligopeptidase |
| Metab | CTSO (P43234) | Cathepsin O |
| Metab | FCRL1 (Q96LA6) | Fc receptor-like protein 1 |
| Metab | CD164 (Q04900) | Sialomucin core protein 24 |
| Metab | DDC (P20711) | Aromatic-L-amino-acid decarboxylase |
| Metab | ACP6 (Q9NPH0) | Lysophosphatidic acid phosphatase type 6 |
| Metab | TFF2 (Q03403) | Trefoil factor 2 |
| Metab | ANGPT2 (O15123) | Angiopoietin-2 |
| Metab | CD2AP (Q9Y5K6) | CD2-associated protein |
| Metab | ANGPTL7 (O43827) | Angiopoietin-related protein 7 |
| Metab | CLEC5A (Q9NY25) | C-type lectin domain family 5 member A |
| Metab | TINAGL1 (Q9GZM7) | Tubulointerstitial nephritis antigen-like |
| Metab | ENO2 (P09104) | Gamma-enolase |
| Metab | NADK (O95544) | NAD kinase |
| Metab | GHRL (Q9UBU3) | Appetite-regulating hormone |
| Metab | SERPINB8 (P50452) | Serpin B8 |
| Metab | SERPINB6 (P35237) | Serpin B6 |
| Metab | CDHR5 (Q9HBB8) | Cadherin-related family member 5 |
| Metab | CCDC80 (Q76M96) | Coiled-coil domain-containing protein 80 |
| Metab | CA13 (Q8N1Q1) | Carbonic anhydrase 13 |
| Metab | SEMA3F (Q13275) | Semaphorin-3F |
| Metab | KLK10 (O43240) | Kallikrein-10 |
| Metab | PILRB (Q9UKJ0) | Paired immunoglobulin-like type 2 receptor beta |
| Metab | ANGPTL1 (O95841) | Angiopoietin-related protein 1 |
| Metab | APLP1 (P51693) | Amyloid-like protein 1 |
| Metab | ADGRG2 (Q8IZP9) | Adhesion G-protein coupled receptor G2 |
| Metab | TYMP (P19971) | Thymidine phosphorylase |
| Metab | GRAP2 (O75791) | GRB2-related adapter protein 2 |
| Metab | LILRA5 (A6NI73) | Leukocyte immunoglobulin-like receptor subfamily A member 5 |
| Metab | ALDH1A1 (P00352) | Retinal dehydrogenase 1 |
| Metab | CD79B (P40259) | B-cell antigen receptor complex-associated protein beta chain |
| Metab | SIGLEC7 (Q9Y286) | Sialic acid-binding Ig-like lectin 7 |
| Metab | QDPR (P09417) | Dihydropteridine reductase |
| Metab | SNAP23 (O00161) | Synaptosomal-associated protein 23 |
| Metab | APEX1 (P27695) | DNA-(apurinic or apyrimidinic site) lyase |
| Metab | ENTPD5 (O75356) | Ectonucleoside triphosphate diphosphohydrolase 5 |
| Metab | CLSTN2 (Q9H4D0) | Calsyntenin-2 |
| Metab | COMT (P21964) | Catechol O-methyltransferase |
| Metab | CLUL1 (Q15846) | Clusterin-like protein 1 |
| Metab | HDGF (P51858) | Hepatoma-derived growth factor |
| Metab | CHRDL2 (Q6WN34) | Chordin-like protein 2 |
| Metab | NOMO1 (Q15155) | Nodal modulator 1 |
| Metab | SOST (Q9BQB4) | Sclerostin |
| Metab | FAM3C (Q92520) | Protein FAM3C |
| Metab | TXNDC5 (Q8NBS9) | Thioredoxin domain-containing protein 5 |
| Metab | PPP1R2 (P41236) | Protein phosphatase inhibitor 2 |
| Metab | LRP11 (Q86VZ4) | Low-density lipoprotein receptor-related protein 11 |
| Metab | ADGRE2 (Q9UHX3) | Adhesion G protein-coupled receptor E2 |
| Metab | ENPP7 (Q6UWV6) | Ectonucleotide pyrophosphatase/phosphodiesterase family member 7 |
| Metab | SSC4D (Q8WTU2) | Scavenger receptor cysteine-rich domain-containing group B protein |
| Metab | MCFD2 (Q8NI22) | Multiple coagulation factor deficiency protein 2 |
| Metab | REG4 (Q9BYZ8) | Regenerating islet-derived protein 4 |

| | | |
|---|---|---|
| Metab | SUMF2 (Q8NBJ7) | Sulfatase-modifying factor 2 |
| Metab | CANT1 (Q8WVQ1) | Soluble calcium-activated nucleotidase 1 |
| Metab | CD1C (P29017) | T-cell surface glycoprotein CD1c |
| Metab | GAL (P22466) | Galanin peptides |
| Metab | CDH2 (P19022) | Cadherin-2 |
| Metab | TYRO3 (Q06418) | Tyrosine-protein kinase receptor TYRO3 |
| Metab | CRKL (P46109) | Crk-like protein |
| Metab | IGFBPL1 (Q8WX77) | Insulin-like growth factor-binding protein-like 1 |
| Metab | RTN4R (Q9BZR6) | Reticulon-4 receptor |
| Metab | VCAN (P13611) | Versican core protein |
| Metab | FBP1 (P09467) | Fructose-1,6-bisphosphatase 1 |
| Metab | TSHB (P01222) | Thyrotropin subunit beta |
| Metab | BAG6 (P46379) | Large proline-rich protein BAG6 |
| Metab | NECTIN2 (Q92692) | Nectin-2 |
| Metab | SDC4 (P31431) | Syndecan-4 |
| Metab | PAG1 (Q9NWQ8) | Phosphoprotein associated with glycosphingolipid-enriched microdomains 1 |
| Metab | KYAT1 (Q16773) | Kynurenine--oxoglutarate transaminase 1 |
| Metab | NPDC1 (Q9NQX5) | Neural proliferation differentiation and control protein 1 |
| Metab | METRNL (Q641Q3) | Meteorin-like protein |
| Metab | MEP1B (Q16820) | Meprin A subunit beta |
| Metab | ROR1 (Q01973) | Inactive tyrosine-protein kinase transmembrane receptor ROR1 |
| Metab | RNASE3 (P12724) | Eosinophil cationic protein |
| Metab | NT-proBNP (N/A) | N-terminal prohormone brain natriuretic peptide |

CVD, cardiovascular disease; Metab, metabolism.

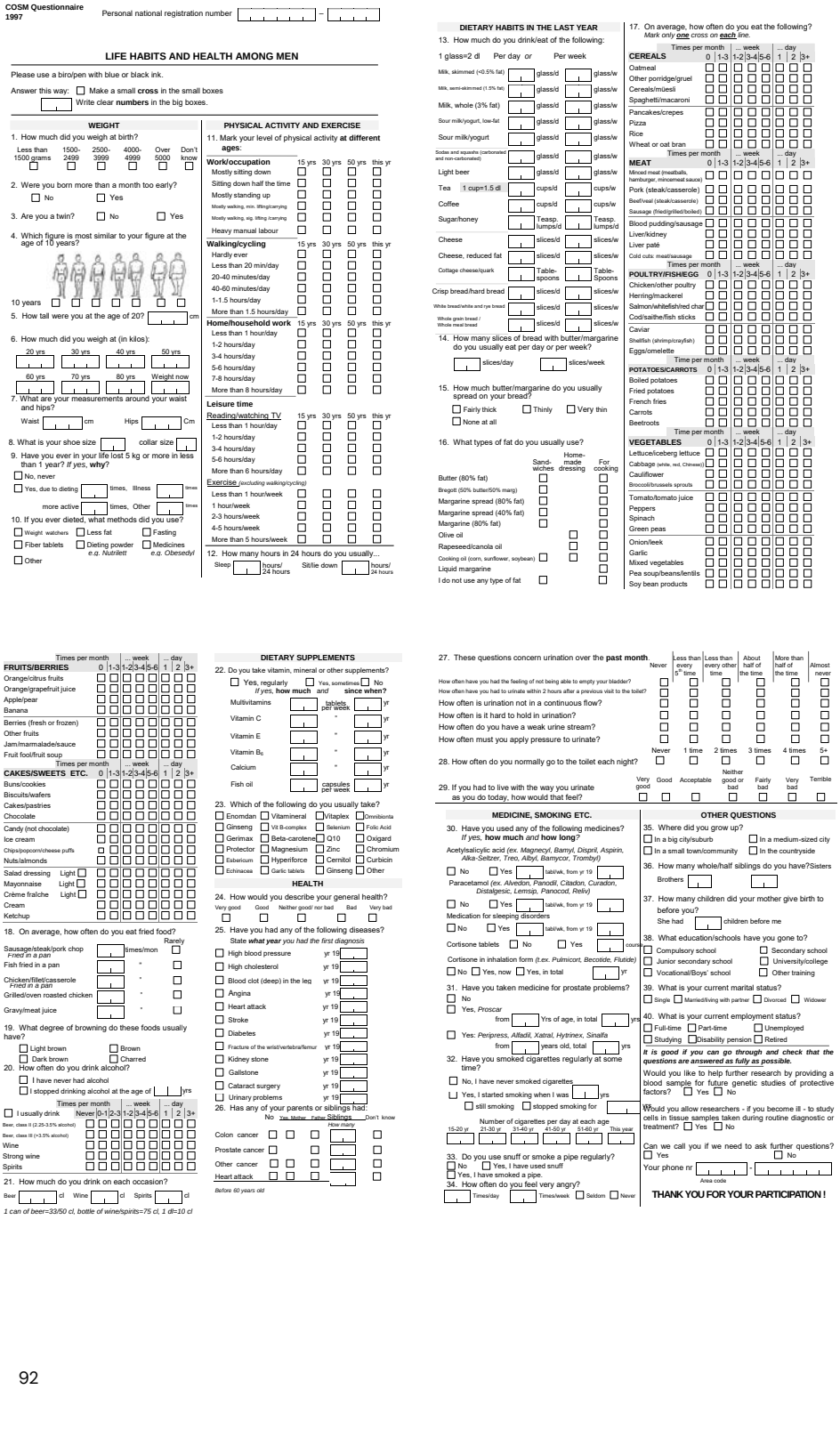
**Appendix Table 2.** Detailed information on covariates in SIMPLER

| Covariate | Type and category | Definition |
|---|---|---|
| Age | Continuous | Questionnaire date minus birth date |
| Sex | Men and women | Self-reported data |
| Body mass index | Continuous | Weight in kg divided by square of height in m |
| Education attainment | ≤9, 10–12, >12 years | Self-reported data |
| Smoking | Never smoker<br>Ever smokers | Never and ever smokers (including past and current smokers) |
| Alcohol consumption | Never<br>Moderate<br>Excessive | Derived from dietary intake questionnaire and measured by six alcoholic beverages including light beer, moderate beer, strong beer, wine, fortified wine, and liquor.<br>Never (0 drink/day)<br>Moderate (0-1 drink/day for women and 0-2 drinks/day for men)<br>Excessive (> 1 drink/day for women and >2 drinks/day for men) |
| Physical activity | <10 mins/day<br>10-30 mins/day<br>30-60 mins/day<br>>60 mins/day | Measured by sum of time spent walking or bicycling and leisure time exercise. |
| Diet quality | Continuous | mDSAH score ranging from 7 to 35 was based on 7 food groups (fruits, vegetables, nuts and legumes, whole grains, and low-fat dairy products as healthy components and red and processed meat and sweetened beverages as unhealthy components) with information from dietary intake questionnaire. The average intake frequency and serving size were reported in the questionnaire. Individuals received a score of 1 to 5 according to sex-specific quintiles of consumption of each food and the scores were summed to create a mDASH diet score. A higher score represented a higher adherence to the mDASH diet. |
| eGFR | Continuous in mL/min/1.73m 2 | Estiamted using formula: $144 \times (S_{cr}/61.9)^{-0.329} \times (0.993)^{age}$ for female and $141 \times (S_{cr}/79.6)^{-0.411} \times (0.993)^{age}$ for male. $S_{cr}$ means serum creatine in μmol/L, which was measured in fasting blood samples using the enzymatic method. |
| Lipids | Continuous in mmol/L | The levels were measured in fasting blood sample using the photometric method. |
| Blood pressure | Continuous in mm Hg | The levels were measured twice by nurses. We calculated mean of measurements. |
| Fasting glucose | Continuous in mmol/L | The levels were measured in fasting blood sample using glucose dehydrogenase reagent from Bergman & Beving, instrument Advia 1650. |
| Diagnosis of cardiovascular disease | Yes and no | Baseline diagnosis of coronary artery disease, heart failure, stroke, or atrial fibrillation |

**Appendix Table 3.** Detailed information on covariates in the UK Biobank

| Covariates | Definition |
|---|---|
| Physical activity | UK Biobank Touchscreen questionnaire on the reported type and duration of physical activity (including walking, DIY, moderate and vigorous physical activity, strenuous sports, etc.). It was measured by International Physical Activity Questionnaire (IPAQ). IPAQ was calculated by weighting each type of activity by its energy requirements defined in METs to yield a score in MET–minutes. |
| Tea and coffee consumption | Tea: sum of Intake of standard tea, rooibos tea, green tea, herbal tea, and other tea.<br>Coffee: sum of intake of instant coffee, filtered coffee, cappuccino intake, latte intake, espresso, and other coffee.<br>Data collected by 24-h dietary questionnaires. |
| Sedentary time | Sum of time spent on television watching, computer using, and car driving collected by baseline touchscreen questionnaires. |
| Air pollution | Air pollution of nitrogen dioxide, nitrogen oxides, particulate matter (pm10), particulate matter (pm2.5), and particulate matter 2.5-10um. |
| Self-reported fracture, major operation, and aspirin use | Self-reported major operation: any major operations<br>Self-reported fracture: fractured/broken bones in last 5 years<br>Self-reported aspirin use: self-reported medication for pain relief, constipation, heartburn (Aspirin).<br>Data collected by baseline touchscreen questionnaires. |
| Comorbidities index | The Charlson Comorbidity Index: calculated by 17 comorbidities including myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, mild liver disease, diabetes without chronic complication, diabetes with chronic complication, hemiplegia or paraplegia, renal disease, any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin, moderate or severe liver disease, metastatic solid tumor, acquired immunodeficiency syndrome according to International Classification of Disease (ICD) codes. |
| Oral contraceptive pill and hormone-replacement therapy for women | Ever taken oral contraceptive pill or ever used hormone-replacement therapy (HRT) for female only.<br>Data collected by baseline touchscreen questionnaires |

# **Appendix Figure 1.** COSM Questionnaire 1997



COSM Questionnaire 1997

Personal national registration number ____ – ____

## LIFE HABITS AND HEALTH AMONG MEN

Please use a biro/pen with blue or black ink.

Answer this way: ☐ Make a small **cross** in the small boxes

☐ Write clear **numbers** in the big boxes.

### WEIGHT

1. How much did you weigh at birth?

Less than 1500 grams ☐ | 1500-2499 ☐ | 2500-3999 ☐ | 4000-4999 ☐ | Over 5000 ☐ | Don't know ☐

2. Were you born more than a month too early?

☐ No ☐ Yes

3. Are you a twin? ☐ No ☐ Yes

4. Which figure is most similar to your figure at the age of 10 years?



10 years ☐ ☐ ☐ ☐ ☐ ☐ ☐

5. How tall were you at the age of 20? ____ cm

6. How much did you weigh at (in kilos):

20 yrs ____ | 30 yrs ____ | 40 yrs ____ | 50 yrs ____

60 yrs ____ | 70 yrs ____ | 80 yrs ____ | Weight now ____

7. What are your measurements around your waist and hips?

Waist ____ cm Hips ____ Cm

8. What is your shoe size ____ collar size ____

9. Have you ever in your life lost 5 kg or more in less than 1 year? *If yes,* **why?**

☐ No, never

☐ Yes, due to dieting ____ times, Illness ____ times

more active ____ times, Other ____ times

10. If you ever dieted, what methods did you use?

☐ Weight watchers ☐ Less fat ☐ Fasting

☐ Fiber tablets ☐ Dieting powder *e.g. Nutrilett* ☐ Medicines *e.g. Obesedyl*

☐ Other

### PHYSICAL ACTIVITY AND EXERCISE

11. Mark your level of physical activity **at different ages**:

| **Work/occupation** | 15 yrs | 30 yrs | 50 yrs | this yr |
|---|---|---|---|---|
| Mostly sitting down | ☐ | ☐ | ☐ | ☐ |
| Sitting down half the time | ☐ | ☐ | ☐ | ☐ |
| Mostly standing up | ☐ | ☐ | ☐ | ☐ |
| Mostly walking, min. lifting/carrying | ☐ | ☐ | ☐ | ☐ |
| Mostly walking, wlg. lifting /carrying | ☐ | ☐ | ☐ | ☐ |
| Heavy manual labour | ☐ | ☐ | ☐ | ☐ |
| **Walking/cycling** | 15 yrs | 30 yrs | 50 yrs | this yr |
| Hardly ever | ☐ | ☐ | ☐ | ☐ |
| Less than 20 min/day | ☐ | ☐ | ☐ | ☐ |
| 20-40 minutes/day | ☐ | ☐ | ☐ | ☐ |
| 40-60 minutes/day | ☐ | ☐ | ☐ | ☐ |
| 1-1.5 hours/day | ☐ | ☐ | ☐ | ☐ |
| More than 1.5 hours/day | ☐ | ☐ | ☐ | ☐ |
| **Home/household work** | 15 yrs | 30 yrs | 50 yrs | this yr |
| Less than 1 hour/day | ☐ | ☐ | ☐ | ☐ |
| 1-2 hours/day | ☐ | ☐ | ☐ | ☐ |
| 3-4 hours/day | ☐ | ☐ | ☐ | ☐ |
| 5-6 hours/day | ☐ | ☐ | ☐ | ☐ |
| 7-8 hours/day | ☐ | ☐ | ☐ | ☐ |
| More than 8 hours/day | ☐ | ☐ | ☐ | ☐ |
| **Leisure time** | | | | |
| Reading/watching TV | 15 yrs | 30 yrs | 50 yrs | this yr |
| Less than 1 hour/day | ☐ | ☐ | ☐ | ☐ |
| 1-2 hours/day | ☐ | ☐ | ☐ | ☐ |
| 3-4 hours/day | ☐ | ☐ | ☐ | ☐ |
| 5-6 hours/day | ☐ | ☐ | ☐ | ☐ |
| More than 6 hours/day | ☐ | ☐ | ☐ | ☐ |
| Exercise *(excluding walking/cycling)* | | | | |
| Less than 1 hour/week | ☐ | ☐ | ☐ | ☐ |
| 1 hour/week | ☐ | ☐ | ☐ | ☐ |
| 2-3 hours/week | ☐ | ☐ | ☐ | ☐ |
| 4-5 hours/week | ☐ | ☐ | ☐ | ☐ |
| More than 5 hours/week | ☐ | ☐ | ☐ | ☐ |

12. How many hours in 24 hours do you usually...

Sleep ____ hours/24 hours Sit/lie down ____ hours/24 hours

### DIETARY HABITS IN THE LAST YEAR

13. How much do you drink/eat of the following:

| 1 glass=2 dl | Per day *or* | Per week |
|---|---|---|
| Milk, skimmed (<0.5% fat) | ____ glass/d | ____ glass/w |
| Milk, semi-skimmed (1.5% fat) | ____ glass/d | ____ glass/w |
| Milk, whole (3% fat) | ____ glass/d | ____ glass/w |
| Sour milk/yogurt, low-fat | ____ glass/d | ____ glass/w |
| Sour milk/yogurt | ____ glass/d | ____ glass/w |
| Sodas and squashs (carbonated and non-carbonated) | ____ glass/d | ____ glass/w |
| Light beer | ____ glass/d | ____ glass/w |
| Tea 1 cup=1.5 dl | ____ cups/d | ____ cups/w |
| Coffee | ____ cups/d | ____ cups/w |
| Sugar/honey | ____ Teasp. lumps/d | ____ Teasp. lumps/d |
| Cheese | ____ slices/d | ____ slices/w |
| Cheese, reduced fat | ____ slices/d | ____ slices/w |
| Cottage cheese/quark | ____ Table-spoons | ____ Table-Spoons |
| Crisp bread/hard bread | ____ slices/d | ____ slices/w |
| White bread/white and rye bread | ____ slices/d | ____ slices/w |
| Whole grain bread / Whole meal bread | ____ slices/d | ____ slices/w |

14. How many slices of bread with butter/margarine do you usually eat per day or per week?

____ slices/day ____ slices/week

15. How much butter/margarine do you usually spread on your bread?

☐ Fairly thick ☐ Thinly ☐ Very thin

☐ None at all

16. What types of fat do you usually use?

| | Sand-wiches | Home-made dressing | For cooking |
|---|---|---|---|
| Butter (80% fat) | ☐ | | ☐ |
| Bregott (50% butter/50% marg) | ☐ | | ☐ |
| Margarine spread (80% fat) | ☐ | | ☐ |
| Margarine spread (40% fat) | ☐ | | ☐ |
| Margarine (80% fat) | ☐ | | ☐ |
| Olive oil | | ☐ | ☐ |
| Rapeseed/canola oil | | ☐ | ☐ |
| Cooking oil (corn, sunflower, soybean) | ☐ | ☐ | ☐ |
| Liquid margarine | | | ☐ |
| I do not use any type of fat | ☐ | | ☐ |

17. On average, how often do you eat the following?

*Mark only* ***one*** *cross on* ***each*** *line.*

| | Times per month | | ... week | | | ... day | | |
|---|---|---|---|---|---|---|---|---|
| **CEREALS** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Oatmeal | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other porridge/gruel | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cereals/müesli | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Spaghetti/macaroni | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Pancakes/crepes | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Pizza | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Rice | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Wheat or oat bran | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | Times per month | | ... week | | | ... day | | |
| **MEAT** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Minced meat (meatballs, hamburger, mincemeat sauce) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Pork (steak/casserole) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Beef/veal (steak/casserole) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Sausage (fried/grilled/boiled) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Blood pudding/sausage | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Liver/kidney | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Liver paté | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cold cuts: meat/sausage | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | Times per month | | ... week | | | ... day | | |
| **POULTRY/FISH/EGG** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Chicken/other poultry | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Herring/mackerel | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Salmon/whitefish/red char | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cod/saithe/fish sticks | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Caviar | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Shellfish (shrimp/crayfish) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Eggs/omelette | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | Time per month | | ... week | | | ... day | | |
| **POTATOES/CARROTS** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Boiled potatoes | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Fried potatoes | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| French fries | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Carrots | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Beetroots | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | Time per month | | ... week | | | ... day | | |
| **VEGETABLES** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Lettuce/iceberg lettuce | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cabbage (white, red, Chinese) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cauliflower | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Broccoli/brussels sprouts | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Tomato/tomato juice | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Peppers | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Spinach | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Green peas | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Onion/leek | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Garlic | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Mixed vegetables | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Pea soup/beans/lentils | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Soy bean products | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | Times per month | | ... week | | | ... day | | |
|---|---|---|---|---|---|---|---|---|
| **FRUITS/BERRIES** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Orange/citrus fruits | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Orange/grapefruit juice | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Apple/pear | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Banana | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Berries (fresh or frozen) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other fruits | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Jam/marmalade/sauce | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Fruit fool/fruit soup | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | Times per month | | ... week | | | ... day | | |
| **CAKES/SWEETS ETC.** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Buns/cookies | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Biscuits/wafers | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cakes/pastries | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Chocolate | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Candy (not chocolate) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Ice cream | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Chips/popcorn/cheese puffs | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Nuts/almonds | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Salad dressing Light ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Mayonnaise Light ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Crème fraîche Light ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cream | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Ketchup | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

18. On average, how often do you eat fried food?

| | | Rarely |
|---|---|---|
| Sausage/steak/pork chop *Fried in a pan* | ____ times/mon | ☐ |
| Fish fried in a pan | ____ " | ☐ |
| Chicken/fillet/casserole *Fried in a pan* | ____ " | ☐ |
| Grilled/oven roasted chicken | ____ " | ☐ |
| Gravy/meat juice | ____ " | ☐ |

19. What degree of browning do these foods usually have?

☐ Light brown ☐ Brown

☐ Dark brown ☐ Charred

20. How often do you drink alcohol?

☐ I have never had alcohol

☐ I stopped drinking alcohol at the age of ____ yrs

| | Times per month | | | ... week | | | ... day | | |
|---|---|---|---|---|---|---|---|---|---|
| ☐ I usually drink | Never | 0-1 | 2-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Beer, class II (2.25-3.5% alcohol) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Beer, class III (>3.5% alcohol) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Wine | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Strong wine | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Spirits | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

21. How much do you drink on each occasion?

Beer ____ cl Wine ____ cl Spirits ____ cl

*1 can of beer=33/50 cl, bottle of wine/spirits=75 cl, 1 dl=10 cl*

### DIETARY SUPPLEMENTS

22. Do you take vitamin, mineral or other supplements?

☐ Yes, regularly ☐ Yes, sometimes ☐ No

*If yes,* **how much** *and* **since when?**

| | | tablets per week | |
|---|---|---|---|
| Multivitamins | ____ | | ____ yr |
| Vitamin C | ____ | " | ____ yr |
| Vitamin E | ____ | " | ____ yr |
| Vitamin $B_6$ | ____ | " | ____ yr |
| Calcium | ____ | " | ____ yr |
| Fish oil | ____ | capsules per week | ____ yr |

23. Which of the following do you usually take?

☐ Enomdan ☐ Vitamineral ☐ Vitaplex ☐ Omnibionta

☐ Ginseng ☐ Vit B-complex ☐ Selenium ☐ Folic Acid

☐ Gerimax ☐ Beta-carotene ☐ Q10 ☐ Oxigard

☐ Protector ☐ Magnesium ☐ Zinc ☐ Chromium

☐ Esbericum ☐ Hyperiforce ☐ Cernitol ☐ Curbicin

☐ Echinacea ☐ Garlic tablets ☐ Ginseng ☐ Other

### HEALTH

24. How would you describe your general health?

Very good ☐ Good ☐ Neither good/ nor bad ☐ Bad ☐ Very bad ☐

25. Have you had any of the following diseases?

State ***what year*** *you had the first diagnosis*

| | | How many |
|---|---|---|
| ☐ High blood pressure | yr 19____ | |
| ☐ High cholesterol | yr 19____ | |
| ☐ Blood clot (deep) in the leg | yr 19____ | |
| ☐ Angina | yr 19____ | |
| ☐ Heart attack | yr 19____ | |
| ☐ Stroke | yr 19____ | |
| ☐ Diabetes | yr 19____ | |
| ☐ Fracture of the wrist/vertebra/femur | yr 19____ | |
| ☐ Kidney stone | yr 19____ | |
| ☐ Gallstone | yr 19____ | |
| ☐ Cataract surgery | yr 19____ | |
| ☐ Urinary problems | yr 19____ | |

26. Has any of your parents or siblings had:

| | No | Yes, Mother | Yes, Father | Yes, Siblings | Don't know |
|---|---|---|---|---|---|
| Colon cancer | ☐ | ☐ | ☐ | ____ | ☐ |
| Prostate cancer | ☐ | | ☐ | ____ | ☐ |
| Other cancer | ☐ | ☐ | ☐ | ____ | ☐ |
| Heart attack | ☐ | ☐ | ☐ | ____ | ☐ |

*Before 60 years old*

27. These questions concern urination over the **past month**.

| | Never | Less than 1 time every 5 time | Less than every other time | About half of the time | More than half of the time | Almost always |
|---|---|---|---|---|---|---|
| How often have you had the feeling of not being able to empty your bladder? | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| How often have you had to urinate within 2 hours after a previous visit to the toilet? | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| How often is urination not in a continuous flow? | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| How often is it hard to hold in urination? | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| How often do you have a weak urine stream? | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| How often must you apply pressure to urinate? | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | Never | 1 time | 2 times | 3 times | 4 times | 5+ |
|---|---|---|---|---|---|---|
| 28. How often do you normally go to the toilet each night? | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | Very good | Good | Acceptable | Neither good or bad | Fairly bad | Very bad | Terrible |
|---|---|---|---|---|---|---|---|
| 29. If you had to live with the way you urinate as you do today, how would that feel? | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

### MEDICINE, SMOKING ETC.

30. Have you used any of the following medicines?

*If yes,* **how much** *and* **how long**?

Acetylsalicylic acid *(ex. Magnecyl, Bamyl, Dispril, Aspirin, Alka-Seltzer, Treo, Albyl, Bamycor, Trombyl)*

☐ No ☐ Yes ____ tabl/wk, from yr 19____

Paracetamol *(ex. Alvedon, Panodil, Citadon, Curadon, Distalgesic, Lemsip, Panocod, Reliv)*

☐ No ☐ Yes ____ tabl/wk, from yr 19____

Medication for sleeping disorders

☐ No ☐ Yes ____ tabl/wk, from yr 19____

Cortisone tablets ☐ No ☐ Yes ____ courses

Cortisone in inhalation form *(t.ex. Pulmicort, Becotide, Flutide)*

☐ No ☐ Yes, now ☐ Yes, in total ____ yr

31. Have you taken medicine for prostate problems?

☐ No

☐ Yes, *Proscar*

from ____ Yrs of age, in total ____ yrs

☐ Yes: *Peripress, Alfadil, Xatral, Hytrinex, Sinalfa*

from ____ years old, total ____ yrs

32. Have you smoked cigarettes regularly at some time?

☐ No, I have never smoked cigarettes

☐ Yes, I started smoking when I was ____ yrs

☐ still smoking ☐ stopped smoking for ____ yrs

Number of cigarettes per day at each age

| 15-20 yr | 21-30 yr | 31-40 yr | 41-50 yr | 51-60 yr | This year |
|---|---|---|---|---|---|
| | | | | | |

33. Do you use snuff or smoke a pipe regularly?

☐ No ☐ Yes, I have used snuff

☐ Yes, I have smoked a pipe.

34. How often do you feel very angry?

____ Times/day ____ Times/week ☐ Seldom ☐ Never

### OTHER QUESTIONS

35. Where did you grow up?

☐ In a big city/suburb ☐ In a medium-sized city

☐ In a small town/community ☐ In the countryside

36. How many whole/half siblings do you have?Sisters ____

Brothers ____

37. How many children did your mother give birth to before you?

She had ____ children before me

38. What education/schools have you gone to?

☐ Compulsory school ☐ Secondary school

☐ Junior secondary school ☐ University/college

☐ Vocational/Boys' school ☐ Other training

39. What is your current marital status?

☐ Single ☐ Married/living with partner ☐ Divorced ☐ Widower

40. What is your current employment status?

☐ Full-time ☐ Part-time ☐ Unemployed

☐ Studying ☐ Disability pension ☐ Retired

***It is good if you can go through and check that the questions are answered as fully as possible.***

Would you like to help further research by providing a blood sample for future genetic studies of protective factors? ☐ Yes ☐ No

Would you allow researchers - if you become ill - to study cells in tissue samples taken during routine diagnostic or treatment? ☐ Yes ☐ No

Can we call you if we need to ask further questions?

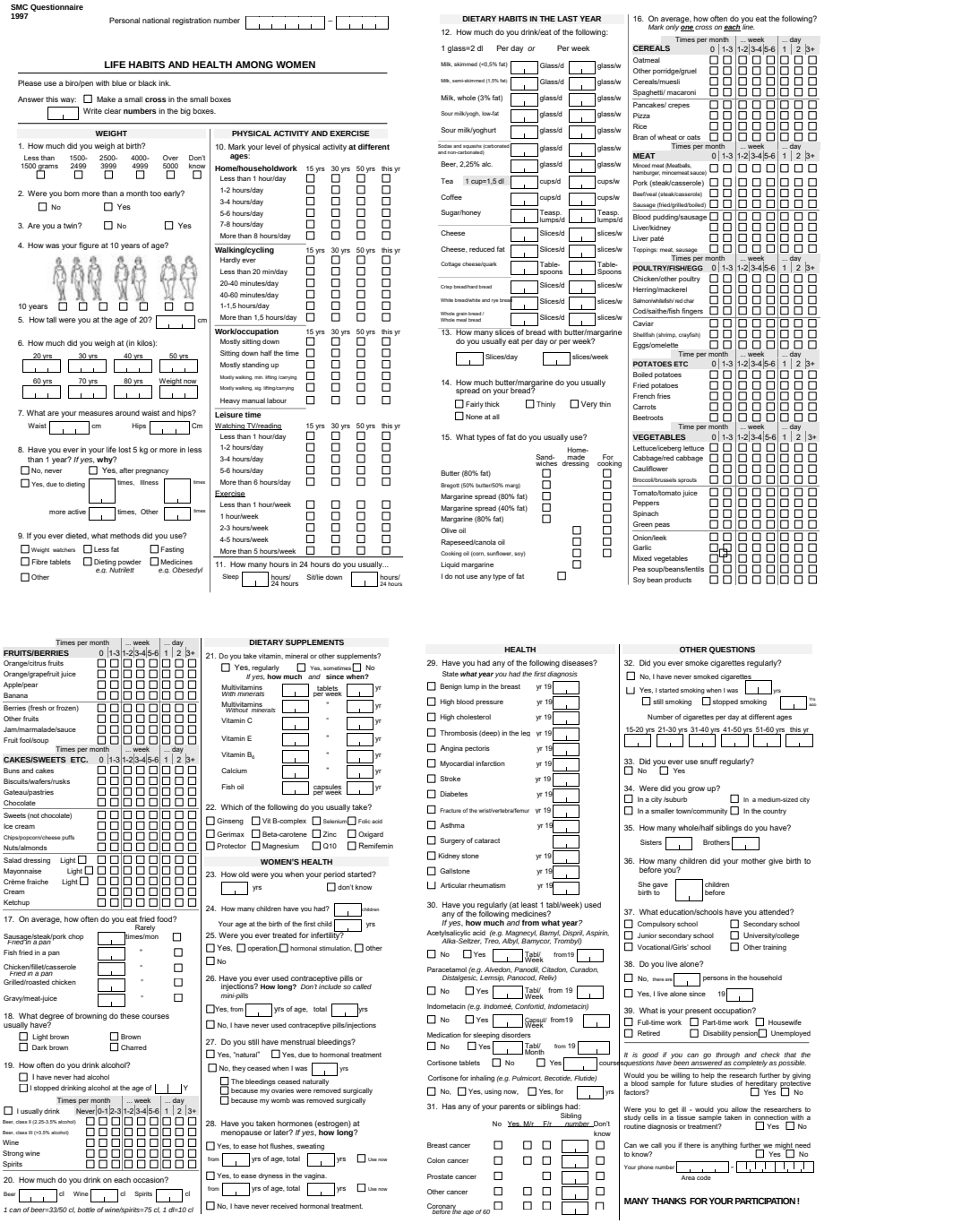
☐ Yes ☐ No

Your phone nr ____ – ____

Area code

**THANK YOU FOR YOUR PARTICIPATION !**

# Appendix Figure 2. SMC Questionnaire 1997



**SMC Questionnaire 1997**

Personal national registration number ____ – ____

## LIFE HABITS AND HEALTH AMONG WOMEN

Please use a biro/pen with blue or black ink.

Answer this way: ☐ Make a small **cross** in the small boxes
☐ Write clear **numbers** in the big boxes.

### WEIGHT

1. How much did you weigh at birth?
   Less than 1500 grams ☐ | 1500-2499 ☐ | 2500-3999 ☐ | 4000-4999 ☐ | Over 5000 ☐ | Don't know ☐
2. Were you born more than a month too early? ☐ No ☐ Yes
3. Are you a twin? ☐ No ☐ Yes
4. How was your figure at 10 years of age?
   10 years ☐ ☐ ☐ ☐ ☐
5. How tall were you at the age of 20? ____ cm
6. How much did you weigh at (in kilos):
   20 yrs ____ | 30 yrs ____ | 40 yrs ____ | 50 yrs ____
   60 yrs ____ | 70 yrs ____ | 80 yrs ____ | Weight now ____
7. What are your measures around waist and hips?
   Waist ____ cm Hips ____ Cm
8. Have you ever in your life lost 5 kg or more in less than 1 year? *If yes*, **why**?
   ☐ No, never ☐ Yes, after pregnancy
   ☐ Yes, due to dieting ____ times, Illness ____ times
   more active ____ times, Other ____ times
9. If you ever dieted, what methods did you use?
   ☐ Weight watchers ☐ Less fat ☐ Fasting
   ☐ Fibre tablets ☐ Dieting powder *e.g. Nutrilett* ☐ Medicines *e.g. Obesedyl*
   ☐ Other

### PHYSICAL ACTIVITY AND EXERCISE

10. Mark your level of physical activity **at different ages**:

| **Home/householdwork** | 15 yrs | 30 yrs | 50 yrs | this yr |
|---|---|---|---|---|
| Less than 1 hour/day | ☐ | ☐ | ☐ | ☐ |
| 1-2 hours/day | ☐ | ☐ | ☐ | ☐ |
| 3-4 hours/day | ☐ | ☐ | ☐ | ☐ |
| 5-6 hours/day | ☐ | ☐ | ☐ | ☐ |
| 7-8 hours/day | ☐ | ☐ | ☐ | ☐ |
| More than 8 hours/day | ☐ | ☐ | ☐ | ☐ |
| **Walking/cycling** | 15 yrs | 30 yrs | 50 yrs | this yr |
| Hardly ever | ☐ | ☐ | ☐ | ☐ |
| Less than 20 min/day | ☐ | ☐ | ☐ | ☐ |
| 20-40 minutes/day | ☐ | ☐ | ☐ | ☐ |
| 40-60 minutes/day | ☐ | ☐ | ☐ | ☐ |
| 1-1,5 hours/day | ☐ | ☐ | ☐ | ☐ |
| More than 1,5 hours/day | ☐ | ☐ | ☐ | ☐ |
| **Work/occupation** | 15 yrs | 30 yrs | 50 yrs | this yr |
| Mostly sitting down | ☐ | ☐ | ☐ | ☐ |
| Sitting down half the time | ☐ | ☐ | ☐ | ☐ |
| Mostly standing up | ☐ | ☐ | ☐ | ☐ |
| Mostly walking, min. lifting /carrying | ☐ | ☐ | ☐ | ☐ |
| Mostly walking, eg. lifting/carrying | ☐ | ☐ | ☐ | ☐ |
| Heavy manual labour | ☐ | ☐ | ☐ | ☐ |
| **Leisure time** | | | | |
| Watching TV/reading | 15 yrs | 30 yrs | 50 yrs | this yr |
| Less than 1 hour/day | ☐ | ☐ | ☐ | ☐ |
| 1-2 hours/day | ☐ | ☐ | ☐ | ☐ |
| 3-4 hours/day | ☐ | ☐ | ☐ | ☐ |
| 5-6 hours/day | ☐ | ☐ | ☐ | ☐ |
| More than 6 hours/day | ☐ | ☐ | ☐ | ☐ |
| Exercise | | | | |
| Less than 1 hour/week | ☐ | ☐ | ☐ | ☐ |
| 1 hour/week | ☐ | ☐ | ☐ | ☐ |
| 2-3 hours/week | ☐ | ☐ | ☐ | ☐ |
| 4-5 hours/week | ☐ | ☐ | ☐ | ☐ |
| More than 5 hours/week | ☐ | ☐ | ☐ | ☐ |

11. How many hours in 24 hours do you usually...
    Sleep ____ hours/24 hours | Sit/lie down ____ hours/24 hours

### DIETARY HABITS IN THE LAST YEAR

12. How much do you drink/eat of the following:

| 1 glass=2 dl | Per day *or* | | Per week | |
|---|---|---|---|---|
| Milk, skimmed (<0,5% fat) | ____ | Glass/d | ____ | glass/w |
| Milk, semi-skimmed (1,5% fat) | ____ | Glass/d | ____ | glass/w |
| Milk, whole (3% fat) | ____ | glass/d | ____ | glass/w |
| Sour milk/yogh, low-fat | ____ | glass/d | ____ | glass/w |
| Sour milk/yoghurt | ____ | glass/d | ____ | glass/w |
| Sodas and squashs (carbonated and non-carbonated) | ____ | glass/d | ____ | glass/w |
| Beer, 2,25% alc. | ____ | glass/d | ____ | glass/w |
| Tea 1 cup=1,5 dl | ____ | cups/d | ____ | cups/w |
| Coffee | ____ | cups/d | ____ | cups/w |
| Sugar/honey | ____ | Teasp, lumps/d | ____ | Teasp, lumps/d |
| Cheese | ____ | Slices/d | ____ | slices/w |
| Cheese, reduced fat | ____ | Slices/d | ____ | slices/w |
| Cottage cheese/quark | ____ | Table-spoons | ____ | Table-Spoons |
| Crisp bread/hard bread | ____ | Slices/d | ____ | slices/w |
| White bread/white and rye bread | ____ | Slices/d | ____ | slices/w |
| Whole grain bread / Whole meal bread | ____ | Slices/d | ____ | slices/w |

13. How many slices of bread with butter/margarine do you usually eat per day or per week?
    ____ Slices/day ____ slices/week
14. How much butter/margarine do you usually spread on your bread?
    ☐ Fairly thick ☐ Thinly ☐ Very thin ☐ None at all
15. What types of fat do you usually use?

| | Sand-wiches | Home-made dressing | For cooking |
|---|---|---|---|
| Butter (80% fat) | ☐ | | ☐ |
| Bregott (50% butter/50% marg) | ☐ | | ☐ |
| Margarine spread (80% fat) | ☐ | | ☐ |
| Margarine spread (40% fat) | ☐ | | ☐ |
| Margarine (80% fat) | ☐ | | ☐ |
| Olive oil | | ☐ | ☐ |
| Rapeseed/canola oil | | ☐ | ☐ |
| Cooking oil (corn, sunflower, soy) | | ☐ | ☐ |
| Liquid margarine | | ☐ | |
| I do not use any type of fat | ☐ | | |

16. On average, how often do you eat the following?
    *Mark only **one** cross on **each** line.*

| | Times per month | | | ... week | | | ... day | | |
|---|---|---|---|---|---|---|---|---|---|
| **CEREALS** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Oatmeal | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other porridge/gruel | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cereals/muesli | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Spaghetti/ macaroni | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Pancakes/ crepes | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Pizza | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Rice | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Bran of wheat or oats | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | Times per month | | | ... week | | | ... day | | |
| **MEAT** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Minced meat (Meatballs, hamburger, mincemeat-sauce) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Pork (steak/casserole) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Beef/veal (steak/casserole) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Sausage (fried/grilled/boiled) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Blood pudding/sausage | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Liver/kidney | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Liver paté | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Toppings: meat, sausage | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | Times per month | | | ... week | | | ... day | | |
| **POULTRY/FISH/EGG** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Chicken/other poultry | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Herring/mackerel | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Salmon/whitefish/ red char | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cod/saithe/fish fingers | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Caviar | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Shellfish (shrimp, crayfish) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Eggs/omelette | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | Time per month | | | ... week | | | ... day | | |
| **POTATOES ETC** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Boiled potatoes | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Fried potatoes | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| French fries | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Carrots | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Beetroots | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | Time per month | | | ... week | | | ... day | | |
| **VEGETABLES** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Lettuce/iceberg lettuce | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cabbage/red cabbage | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cauliflower | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Broccoli/brussels sprouts | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Tomato/tomato juice | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Peppers | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Spinach | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Green peas | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Onion/leek | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Garlic | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Mixed vegetables | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Pea soup/beans/lentils | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Soy bean products | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | Times per month | | | ... week | | | ... day | | |
|---|---|---|---|---|---|---|---|---|---|
| **FRUITS/BERRIES** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Orange/citrus fruits | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Orange/grapefruit juice | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Apple/pear | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Banana | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Berries (fresh or frozen) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other fruits | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Jam/marmalade/sauce | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Fruit tool/soup | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | Times per month | | | ... week | | | ... day | | |
| **CAKES/SWEETS ETC.** | 0 | 1-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Buns and cakes | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Biscuits/wafers/rusks | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Gateau/pastries | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Chocolate | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Sweets (not chocolate) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Ice cream | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Chips/popcorn/cheese puffs | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Nuts/almonds | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Salad dressing Light ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Mayonnaise Light ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Crème fraiche Light ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cream | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Ketchup | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

17. On average, how often do you eat fried food?

| | | Rarely |
|---|---|---|
| Sausage/steak/pork chop *Fried in a pan* | ____ | times/mon ☐ |
| Fish fried in a pan | ____ | " ☐ |
| Chicken/fillet/casserole *Fried in a pan* | ____ | " ☐ |
| Grilled/roasted chicken | ____ | " ☐ |
| Gravy/meat-juice | ____ | " ☐ |

18. What degree of browning do these courses usually have?
    ☐ Light brown ☐ Brown ☐ Dark brown ☐ Charred
19. How often do you drink alcohol?
    ☐ I have never had alcohol
    ☐ I stopped drinking alcohol at the age of ____ Y

| | Times per month | | | ... week | | | ... day | | |
|---|---|---|---|---|---|---|---|---|---|
| ☐ I usually drink | Never | 0-1 | 2-3 | 1-2 | 3-4 | 5-6 | 1 | 2 | 3+ |
| Beer, class II (2.25-3.5% alcohol) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Beer, class III (>3.5% alcohol) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Wine | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Strong wine | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Spirits | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

20. How much do you drink on each occasion?
    Beer ____ cl Wine ____ cl Spirits ____ cl
    *1 can of beer=33/50 cl, bottle of wine/spirits=75 cl, 1 dl=10 cl*

### DIETARY SUPPLEMENTS

21. Do you take vitamin, mineral or other supplements?
    ☐ Yes, regularly ☐ Yes, sometimes ☐ No
    *If yes,* **how much** *and* **since when?**

| | | | | |
|---|---|---|---|---|
| Multivitamins *With minerals* | ____ | tablets per week | ____ | yr |
| Multivitamins *Without minerals* | ____ | " | ____ | yr |
| Vitamin C | ____ | " | ____ | yr |
| Vitamin E | ____ | " | ____ | yr |
| Vitamin $B_6$ | ____ | " | ____ | yr |
| Calcium | ____ | " | ____ | yr |
| Fish oil | ____ | capsules per week | ____ | yr |

22. Which of the following do you usually take?
    ☐ Ginseng ☐ Vit B-complex ☐ Selenium ☐ Folic acid
    ☐ Gerimax ☐ Beta-carotene ☐ Zinc ☐ Oxigard
    ☐ Protector ☐ Magnesium ☐ Q10 ☐ Remifemin

### WOMEN'S HEALTH

23. How old were you when your period started?
    ____ yrs ☐ don't know
24. How many children have you had? ____ children
    Your age at the birth of the first child ____ yrs
25. Were you ever treated for infertility?
    ☐ Yes, ☐ operation, ☐ hormonal stimulation, ☐ other
    ☐ No
26. Have you ever used contraceptive pills or injections? **How long?** *Don't include so called mini-pills*
    ☐ Yes, from ____ yrs of age, total ____ yrs
    ☐ No, I have never used contraceptive pills/injections
27. Do you still have menstrual bleedings?
    ☐ Yes, "natural" ☐ Yes, due to hormonal treatment
    ☐ No, they ceased when I was ____ yrs
    ☐ The bleedings ceased naturally
    ☐ because my ovaries were removed surgically
    ☐ because my womb was removed surgically
28. Have you taken hormones (estrogen) at menopause or later? *If yes*, **how long**?
    ☐ Yes, to ease hot flushes, sweating
    from ____ yrs of age, total ____ yrs ☐ Use now
    ☐ Yes, to ease dryness in the vagina.
    from ____ yrs of age, total ____ yrs ☐ Use now
    ☐ No, I have never received hormonal treatment.

### HEALTH

29. Have you had any of the following diseases?
    State ***what year*** *you had the first diagnosis*

| | | | |
|---|---|---|---|
| ☐ Benign lump in the breast | yr 19 | ____ | |
| ☐ High blood pressure | yr 19 | ____ | |
| ☐ High cholesterol | yr 19 | ____ | |
| ☐ Thrombosis (deep) in the leq | yr 19 | ____ | |
| ☐ Angina pectoris | yr 19 | ____ | |
| ☐ Myocardial infarction | yr 19 | ____ | |
| ☐ Stroke | yr 19 | ____ | |
| ☐ Diabetes | yr 19 | ____ | |
| ☐ Fracture of the wrist/vertebra/femur | yr 19 | ____ | |
| ☐ Asthma | yr 19 | ____ | |
| ☐ Surgery of cataract | | ____ | |
| ☐ Kidney stone | yr 19 | ____ | |
| ☐ Gallstone | yr 19 | ____ | |
| ☐ Articular rheumatism | yr 19 | ____ | |

30. Have you regularly (at least 1 tabl/week) used any of the following medicines?
    *If yes,* **how much** *and* **from what year***?*
    Acetylsalicylic acid *(e.g. Magnecyl, Bamyl, Dispril, Aspirin, Alka-Seltzer, Treo, Albyl, Bamycor, Trombyl)*
    ☐ No ☐ Yes ____ Tabl/Week from19 ____
    Paracetamol *(e.g. Alvedon, Panodil, Citadon, Curadon, Distalgesic, Lemsip, Panocod, Reliv)*
    ☐ No ☐ Yes ____ Tabl/Week from 19 ____
    Indometacin *(e.g. Indomeé, Confortid, Indometacin)*
    ☐ No ☐ Yes ____ Caesul/Week from19 ____
    Medication for sleeping disorders
    ☐ No ☐ Yes ____ Tabl/Month from 19 ____
    Cortisone tablets ☐ No ☐ Yes ____ courses
    Cortisone for inhaling *(e.g. Pulmicort, Becotide, Fluticle)*
    ☐ No, ☐ Yes, using now, ☐ Yes, for ____ yrs
31. Has any of your parents or siblings had:

| | No | Yes, M/r | F/r | Sibling number | Don't know |
|---|---|---|---|---|---|
| Breast cancer | ☐ | ☐ | ☐ | ____ | ☐ |
| Colon cancer | ☐ | ☐ | ☐ | ____ | ☐ |
| Prostate cancer | ☐ | | ☐ | ____ | ☐ |
| Other cancer | ☐ | ☐ | ☐ | ____ | ☐ |
| Coronary *before the age of 60* | ☐ | ☐ | ☐ | ____ | ☐ |

### OTHER QUESTIONS

32. Did you ever smoke cigarettes regularly?
    ☐ No, I have never smoked cigarettes
    ☐ Yes, I started smoking when I was ____ yrs
    ☐ still smoking ☐ stopped smoking ____ this year
    Number of cigarettes per day at different ages
    15-20 yrs ____ | 21-30 yrs ____ | 31-40 yrs ____ | 41-50 yrs ____ | 51-60 yrs ____ | this yr ____
33. Did you ever use snuff regularly?
    ☐ No ☐ Yes
34. Were did you grow up?
    ☐ In a city /suburb ☐ In a medium-sized city
    ☐ In a smaller town/community ☐ In the country
35. How many whole/half siblings do you have?
    Sisters ____ Brothers ____
36. How many children did your mother give birth to before you?
    She gave birth to ____ children before
37. What education/schools have you attended?
    ☐ Compulsory school ☐ Secondary school
    ☐ Junior secondary school ☐ University/college
    ☐ Vocational/Girls' school ☐ Other training
38. Do you live alone?
    ☐ No, there are ____ persons in the household
    ☐ Yes, I live alone since 19____
39. What is your present occupation?
    ☐ Full-time work ☐ Part-time work ☐ Housewife
    ☐ Retired ☐ Disability pension ☐ Unemployed

*It is good if you can go through and check that the questions have been answered as completely as possible.*

Would you be willing to help the research further by giving a blood sample for future studies of hereditary protective factors? ☐ Yes ☐ No

Were you to get ill - would you allow the researchers to study cells in a tissue sample taken in connection with a routine diagnosis or treatment? ☐ Yes ☐ No

Can we call you if there is anything further we might need to know? ☐ Yes ☐ No

Your phone number ____ – ____ Area code

**MANY THANKS FOR YOUR PARTICIPATION!**